Genetic variation associated with α-synuclein pathology in Caenorhabditis elegans

Yiru A. Wang

Propositions

- 1. The expression and aggregation of α -synuclein and the associated phenotypic traits are substantially modified by the genetic background in *C. elegans*. (this thesis)
- 2. Transgenic *C. elegans* recombinant inbred lines harboring α -synuclein provide a powerful platform to investigate the gene transcriptional architecture of α -synuclein toxicity. (this thesis)
- 3. Scientists should not conclude that two studies conflict because one had a statistically significant result and the other did not.
- 4. The requirements to obtain a PhD degree should be equal among different universities in The Netherlands.
- 5. Being independent is not the same as doing everything alone.
- 6. Unconventional ideas are as essential for scientific advances as scientific reasoning.

Propositions belonging to the thesis, entitled Genetic variation associated with α-synuclein pathology in *Caenorhabditis elegans*

> Yiru Wang Wageningen, 13 November 2019

Genetic variation associated with α -synuclein pathology in Caenorhabditis elegans

Yiru A. Wang

Thesis committee

Promotor

Prof. Dr J.E.Kammenga Personal chair at the Laboratory of Nematology Wageningen University & Research

Co-promotor

Dr M.G. Sterken Assistant professor, Laboratory of Nematology Wageningen University & Research

Prof. Dr S.C. Harvey Professor of Evolutionary Biology Canterbury Christ Church University, UK

Other members

Prof. Dr B.J. Zwaan, Wageningen University & Research Prof. Dr H.C. Korswagen, Utrecht University Dr M. Boxem, Utrecht University Dr E.C. Andersen, Northwestern University, Evanston (IL), USA

This research was conducted under the auspices of the Graduate School Production Ecology and Research Conservation.

Genetic variation associated with α -synuclein pathology in Caenorhabditis elegans

Yiru A. Wang

Thesis

submitted in fulfillment of the requirements for the degree of doctor at Wageningen University by the authority of the Rector Magnificus, Prof. Dr A.P.J. Mol, in the presence of the Thesis Committee appointed by the Academic Board to be defended in public on Wednesday 13 November 2019 at 1:30 p.m. in the Aula.

Yiru Anning Wang

Genetic variation associated with α -synuclein pathology in *Caenorhabditis elegans*,

145 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)

With references, with summary in English

DOI-link: https://doi.org/10.18174/501981

ISBN: 978-94-6395-151-7

Contents

Chapter I	General Introduction	7
Chapter 2	Genetic variation in NGD and its accessibility in the model organism <i>C. elegans</i>	21
Chapter 3	Genetic background modifies phenotypic and transcriptional responses to α -synuclein	43
Chapter 4	lpha-synuclein aggregates and associated toxicities	71
Chapter 5	eQTL mapping in newly constructed CB4856/N2 RILs carrying α -synuclein gene	89
Chapter 6	General Discussion	117
	Summary	134
	Acknowledgments	136
	About the author	142
	PE&RC Training and Education Statement	143

Chapter I

General Introduction

Yiru A. Wang

Parkinson's disease, α -synuclein, and neurodegeneration

An increasing number of elderly people are affected by neurodegenerative diseases (NGD) as life expectancy generally increases over time [1]. Parkinson's disease (PD), as the second common NGDs after Alzheimer's disease (AD), is characterized by the formation of aberrant fibrillar aggregates in nerve cells of the brain, *i.e.* the Lewy bodies. These abundant cytoplasmic Lewy body inclusions consist of α -synuclein protein aggregates [2]. Within aggregates, α -synuclein is present as a misfolded protein and its oligomers, as well as fibrils, induce cytotoxicity by disrupting different cellular functions. A series of studies have pointed out that the α -synuclein misfolding and aggregation elicits metabolic changes associated with autophagy [3], lipid membrane binding [4], cellular ionic homeostasis [5], and mitochondrial homeostasis [6]. Many of these processes are closely linked to organelle biogenesis, of which dynamic decline and even structural/functional deterioration accompany the aging process. Moreover, genetic evidence also demonstrates that the gene SNCA (α -synuclein) mutants (A30P, E46K, H50Q, G51D, and A53T) relate to the onset and progression of PD by genetically interacting with cellular pathways [7, 8], such as the ubiquitin proteasome system (UPS) [9, 10] and the autophagy-lysosomal pathway [11, 12]. Furthermore, the progression of α -synuclein aggregation with advancing age has been identified to cause the loss of dopaminergic neuron functioning and synaptic abnormalities leading to mobility impairment (ataxias) and/or cognitive problems (dementias) [13]. In conclusion, α -synuclein plays a central role in the aetiology and progression of PD.

Research into understanding the molecular pathological mechanisms of NGDs has rapidly expanded. Besides PD, dementia due to progressive neurodegeneration has also been investigated by studying AD as well as Huntington's disease (HD) [14–16]. Although AD and HD each have their distinct causal genes and factors, these neurodegenerative dementias share neuropathological similarities (reviewed in [17], including overlapping cognitive and behavioural symptoms, protein misfolding and aggregation, and the associated neuroanatomical and biochemical changes. For example, next to AD, the pronounced tau pathology has also been described in familial cases of α -synuclein (*SNCA*) mutations inducing PD [18]. Interestingly, α -synuclein has been originally identified as the precursor of the non-A β component of AD amyloid, acting as a presynaptic nerve terminal protein [19]. Further investigations indicated that the intraneuronal filamentous inclusions of AD, tau, co-exist with α -synuclein aggregates in PD as well as other NGDs [20]. Furthermore, α -synuclein aggregates can directly bind to

tau and also interact with hyperphosphorylated tau, which can contribute to the polymerization into fibrils [21, 22]. However, the molecular mechanisms of α -synuclein assembly are complex. Taking α -synuclein phosphorylation for instance - it involves different associated kinases [23], metal ions [24, 25], and oxidative modifications [26]. Currently, we have limited understanding on how natural genetic variation affects the phenotypic traits associated with neurodegeneration.

Genetic variation affects onset and progression of NGD

Genetic variation strongly contributes to the phenotypic variation observed in almost all complex traits, including health span [27], the immune system [28], and many common diseases from NGDs to cancers [29, 30]. In contrast to essentially monogenic disorders (*e.g.* cystic fibrosis), most common diseases arise from the effect of multiple genes, the environment, and the interaction between the two. These diseases can be classified as multifactorial disorders, *e.g.*: obesity [31, 32], diabetes [33], and chromosomal diseases (*e.g.* down syndrome). Genetic variation affects many of these diseases by either multiple variants in the casual genes or genetic variants in background, robustly conferring risks for the onset of disorder [34]. Due to abnormal variants, the disease causal gene(s) may genetically interact with other genes (epistasis) thereby modifying the susceptibility of developing or worsening a disease. These interacting genes, in turn, can harbour genetic variation as well, hence resulting in differential functioning.

Human genetic variation is described as the naturally occurring genetic differences between individual people. Genetic variants include single nucleotide variants and structural variations (*i.e.* small insertion/deletions, induced copy number variants, and chromosomal rearrangements). Genetic factors can strongly influence phenotypic traits, by affecting regulatory elements in different genetic backgrounds. In human disease studies, the causal relationships between disease associated genetic variant(s) and phenotypic diversity is a major challenge [35, 36]. An important question in understanding the mechanisms that give rise to neurodegenerative disorders, is how genetic variation affects the (cellular-level) pathology and disease formation (reviewed in [37, 38]). Answering these questions requires detailed molecular insights of the underlying mechanisms of NGD formation. Investigation on NGDs in small families as well as among groups suggests the involvement of differentially distributed alleles in variable disease associated genes with moderate to severe effects. How the genetic background interaction generates the phenotypic

differences in NGDs remains unclear [39]. These insights are difficult to obtain from humans due to practical and ethical reasons. Therefore, model species like the nematode worm *Caenorhabditis elegans* offer a means to gain this insight at the molecular genetic level.

The nematode C. elegans and its career as a human disease model

A number of advantages of the free-living nematode *C. elegans* support its career in the laboratory. It has a relatively short life-cycle which includes embryonic development, further post-embryonic L1-L4 larval stages, pre-reproductive and reproductive adulthood while feeding sufficiently on the bacterium *Escherichia coli* [40]. Generally, it takes 3.5 days at 20°C from egg to sexually mature adult. The major mode of reproduction of *C. elegans* is self-fertilizing hermaphroditism (XX) that produces up to 300 progeny per animal, while the naturally rare males can mate with hermaphrodites, resulting in an increase of progeny to over 1000 [41, 42]. The fertility of *C. elegans* allows for abundant progeny by natural sexual/self-crossings or via experimental manipulation resulting in selective reproduction. Furthermore, this transparent worm offers experimental tractability, *e.g.* fluorescence visualization in live worms. Comprehensive information on biological aspects has been collected and shared on WormBook (www.wormbook.org) and WormAtlas (www.wormatlas.org), including on available tools, protocols (WormMethods). Also genomic databases are freely-accessible via WormBase (www.wormbase.org). These major technological advances and its relatively simple biology, in combination with a relatively small genome (100 Mb) do contribute for studying complex traits in this versatile model organism.

C. elegans has been instrumental for the development of forward and reverse genetic screens. It has been used to characterize gene expression, function, and regulation in the worms biological processes, including processes involved in many disease mechanisms [43]. For example, by generating transgenic worms expressing human disease-related genes under the control of a *C. elegans* promoter [44, 45]. Furthermore, RNA interference (RNAi) contributes to silence individual genes or even conduct genome-scale screens leading to the identification of various genes in pathways of interest and/or their functions in the nematode genome. More recently, the popular genome editing CRISPR-based systems have expanded the possibilities even further [46].

Notably, *C. elegans* was the first multicellular organism to have its genome sequenced in 1998. Many cellular processes and neuronal signaling pathways are conserved between *C. elegans* and humans (Wormbook, [47]), such as

the insulin/IGF-1 signaling (IIS) pathway and the target-of-rapamycin (TOR) pathways [48–50]. Biological and genetic evidence revealed these pathways contribute to the aging phenotype. For example, the insulin signaling (IIS) pathway is involved in aging as well as age-related diseases by affecting energy homeostasis through both metabolic and mitogenic responses [51]. Mammalian TOR (mTOR) is implicated in several disorders, including cancer, obesity, and diabetes [52]. Hence, *C. elegans* is a suitable model for understanding the mechanisms of aging, mitochondrial dysfunction, and other complex metabolism pathways possibly involved in disorders; especially the ones relating to aging. The benefits of this model organism allow researchers to track biological and genetic events underlying cancers, characteristic NGDs as well as other human diseases [53–57]. Identification of the human orthologs that are involved in biological pathways of *C. elegans* illustrate the worm's value for genetic studies relevant to human biology and disease [58].

Exploring natural genetic variation in C. elegans associated with disease pathways

Investigating disease-related pathways in *C. elegans* has mainly relied on studies being carried out in a single genetic background, the canonical strain Bristol N2. This has revealed invaluable insight into the molecular process of a range of complex traits and disease related phenotypes [59–63]. However, when the goal is to investigate the underlying variation in disease pathways, multiple genetic backgrounds serve as a basis to study the effects of genetic diversity on the phenotypic outcome of a disease-causing mutation.

Wild isolates of *C. elegans* offer an opportunity to explore natural genetic variation. *C. elegans* wild strains are widely distributed across the globe and display both worldwide and local genetic diversity that can be tapped as a reservoir of natural genetic variation [64]. More strains that are divergent from the canonical laboratory strain N2 have been taken along in genetic studies, especially the Hawaiian strain CB4856 [65, 66]. Additionally, there are wild isolates from two sites in France: Orsay and Santeuil [67] that have been used to develop a multi-parental mapping population [68]. Those wild isolates are both geographically and genetically divergent compared to Bristol N2. Most natural variation is attributed to founder effects and genetic drift among *C. elegans* wild isolates [69]. However, also genetic variation linked to local sites has been reported [70]. Considering the occurrence of genetic variation, polymorphic regions are mostly located on the distal ends of the chromosomes and include single nucleotide polymorphisms (SNPs), insertions, and deletions [69–71].

Further investigation of the genetic polymorphisms and phenotypic traits on divergent wild type strains has revealed the importance of natural variation in studying genes regulating complex traits. The genetic diversity among strains has been shown to impact feeding-related behaviour [70], pathogen resistance [72], longevity [73], stress and behavioural responses [60, 74], and disease-associated progression [75, 76]. More detailed information, such as isolation site, genomic, and identified polymorphisms of isolated wild strains are updated and freely shared on CeNDR (www.elegansvariation.org).

Complex traits are only rarely the result of single genes in humans, but also in *C. elegans*. Recombinant inbred lines (RILs) and near isogenic lines (NILs) have been constructed for capturing natural genetic variation and facilitating mapping of genomic loci (quantitative trait loci, QTL) associated with phenotypic traits [77, 78]. Mapping gene expression QTL (eQTL) is used to detect polymorphic loci underlying variation in gene transcript abundances [79]. eQTL contribute to the detection of functional genes which are involved and/or affect transcript levels of many genes and transcriptional pathways underlying the genotype–phenotype relationship.

Genetic background interacts with α -synuclein expression and aggregation in humans and worms

The processes of α -synuclein expression and aggregation can be influenced by transcription factors (TF) and/or genetic background modifiers, which contribute to phenotypic disease variation. Hence, in addition to the major causative genes, the individual disease variability in PD emphasizes the role of the genetic background. It is challenging to detect the relative contribution of genetic background factors interacting with disease-causing genes and to identify how those factors modify it in general and specific ways. For α -synuclein, GATA2 and ZSCAN21 have been identified as its transcription factors (TFs). Variation in GATA2 in humans has been identified in families that suffer from haematological, immunological, lymphatic, and/or other different chronic disorders [80]. Although they are extreme examples of GATA-deficiency related diseases, it suggests that the transcriptional regulation of genes can be affected by natural genetic variation which is associated with an elevated risk of developing PD.

For PD, *SNCA* gene mutants, including allelic mutants or duplicates/triplicates, cause PD [81], which was identified in the early-onset familial and sporadic PD. Moreover, many other genes also cause and influence PD progression, such as LRRK2, MAPT, RAB7L1, etc. [82–85].

At the moment, relatively little is known about the general mechanisms underlying this effect of the genetic background on misfolded-protein-related deficits. Phenotypic consequences can be differently affected by genetic variation, which is also presented in the α -synuclein pathological progression (e.g. [86]). It supposes that there are genes interacting with disease related SNCA gene and/or this interaction is responsible for disease resistance/susceptibility, possibly at variable levels. However, the genetic architecture of diseaseassociated phenotypic traits is often complex due to the underlying genetic basis of different backgrounds and their regulatory properties (e.g. [87]). Like non-coding RNA, its deregulation has also been included as part of cellular processes response to neurodegeneration [88, 89]. Numerous variant alleles, genes, and loci are involved in the molecular pathways associated with α -synuclein-pathology. More complex still, even loci and variants that function indirect and are not annotated to functions in said pathway. These variants could be identified by further studying natural variation in different genetic backgrounds.

To address this issue, investigations based on the polymorphisms between wild strains and Bristol N2 provides the opportunity to understand the impact of natural genetic variation on biological and molecular responses to the α -synuclein protein. Genetic variation in *C. elegans* offers the opportunity to identify genetic variants involved in the regulation of complex traits. For example, natural variation in the nath-10 gene was found to affect the vulval phenotype when sensitized with different mutations in N2, but not in wild-type strains (*e.g.* Australian isolate AB1, [90]). Also, the polymorphic Ras/MAPK modifier amx-2 was found by measuring vulval index and gene expression QTL mapping in a *C. elegans* N2xCB4856 RIL population containing a gain-of-function let-60 mutation [91, 92].

Scope of this thesis

The goal of this thesis is to understand the role of genetic regulatory variation in response to α -synuclein. Together, the results provide important insights into the role of the *C. elegans* genetic background in α -synuclein pathological alterations, and highlights the added value of this model organism for human disease research.

The present thesis addresses the effect of genetic variation on α -synucleinassociated phenotypic deficits by using a newly constructed model system of *C. elegans*. First, I review the role of natural genetic variation involved in onset and disease progression of NGDs and their related characteristics to genetic background variants in Chapter 2. Within this chapter, natural variation of the model organism *C. elegans* and genetic analysis of α -synuclein pathological mechanisms are discussed. On this basis, six different genetic backgrounds have been used to study phenotypic traits associated with α -synuclein expression and aggregation in Chapter 3. With the introduction of α -synuclein into different genetic backgrounds, a series of life-history traits was measured. Chapter 3 investigates how phenotypic and molecular characterizations contribute to reveal the existence of core and/or important genetic diversity in response to α -synuclein in the different genetic backgrounds were then tracked in Chapter 4. The potential influences of its aggregates on molecular biology, *i.e.* lipid formation, was studied as well.

To map genetic background loci being affected by α -synuclein expression, a new panel of RILs was constructed, forming a population with recombination events between the divergent genotypes N2 and CB4856. The whole genome except the α -synuclein transgene introgression on chromosome IV was studied this way (Chapter 5). The RIL panel provides the genetic basis for mapping eQTL associated with the expression and aggregation of α -synuclein. Chapter 6 discusses the most important findings of this thesis related to the relationship of natural genetic variants in model organism *C. elegans* with α -synuclein associated pathological mechanisms. It highlights the importance of the contribution and utilization of natural variation in studies of α -synuclein pathology.

References

- 1. Licher, S. *et al.*, Lifetime risk and multimorbidity of non-communicable diseases and disease-free life expectancy in the general population: A population-based cohort study. *PLoS Med*, 2019. 16(2): e1002741.
- Wakabayashi, K. et al., The Lewy body in Parkinson's disease and related neurodegenerative disorders. Mol Neurobiol, 2013. 47(2): p. 495–508.
- 3. Koch, J. C. *et al.*, Alpha-Synuclein affects neurite morphology, autophagy, vesicle transport and axonal degeneration in CNS neurons. *Cell Death & Disease*, 2015. 6: e1811.
- Galvagnion, C. *et al.*, Chemical properties of lipids strongly affect the kinetics of the membrane-induced aggregation of α-synuclein. *Proc Natl Acad Sci U S A*, 2016. 113(26): p. 7065–7070.
- Büttner, S. et al., The Ca²⁺/Mn²⁺ ion-pump PMR1 links elevation of cytosolic Ca²⁺ levels to α-synuclein toxicity in Parkinson's disease models. *Cell Death Differ*, 2013. 20(3): p. 465–477.
- 6. **Guardia-Laguarta**, **C.** *et al.*, α-Synuclein is localized to mitochondria-associated ER membranes. *J Neurosci*, 2014. 34(1): p. 249–259.
- 7. Lázaro, D. F. *et al.*, Systematic comparison of the effects of alpha-synuclein mutations on its oligomerization and aggregation. *PLos Genet*, 2014. 10(11): e1004741.
- 8. Schulz-Schaeffer, W. J., The synaptic pathology of alpha-synuclein aggregation in dementia with Lewy bodies, Parkinson's disease and Parkinson's disease dementia. *Acta Neuropathol*, 2010. 120(2): p. 131–143.
- Liu, Y. et al., The UCH-L1 gene encodes two opposing enzymatic activities that affect alpha-synuclein degradation and Parkinson's disease susceptibility. *Cell*, 2002. 111(2): p. 209–218.
- 10. Zondler, L. et al., Proteasome impairment by α-synuclein. PLoS One, 2017. 12(9): e0184040.
- 11. Gan-Or, Z., P.A. Dion, and G.A. Rouleau, Genetic perspective on the role of the autophagy-lysosome pathway in Parkinson disease. *Autophagy*, 2015. 11(9): p. 1443–1457.
- 12. Sala, G. et al., Role of Chaperone-Mediated Autophagy Dysfunctions in the Pathogenesis of Parkinson's Disease. *Front Mol Neurosci*, 2016. 9: p. 157.
- 13. Klein, J. C. *et al.*, Neurotransmitter changes in dementia with Lewy bodies and Parkinson disease dementia in vivo. *Neurology*, 2010. 74(11): p. 885–892.
- 14. Aarsland, D. and M. W. Kurz, The epidemiology of dementia associated with Parkinson disease. *J Neurol Sci*, 2010. 289: p. 18–22.
- 15. **Aybek**, **S**. *et al.*, Hippocampal atrophy predicts conversion to dementia after STN-DBS in Parkinson's disease. *Parkinsonism & Related Disorders*, 2009. 15(7): p. 521–524.
- 16. **Kivipelto, M.** *et al.*, Obesity and vascular risk factors at midlife and the risk of dementia and Alzheimer disease. *Arch Neurol*, 2005. 62(10): p. 1556–1560.
- 17. **Inzelberg, R.** *et al.*, Association between amantadine and the onset of dementia in Parkinson's disease. *Mov Disord*, 2006. 21(9): p. 1375–1379.
- 18. **Moussaud, S.** *et al.*, Alpha-synuclein and tau: teammates in neurodegeneration? *Mol Neurodegener*, 2014. 9: p. 43.
- 19. **Iwai**, A. *et al.*, The precursor protein of non-A beta component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system. *Neuron*, 1995. 14(2): p. 467–475.
- 20. **Erro Aguirre**, **M. E.** *et al.*, Midbrain catecholaminergic neurons co-express α-synuclein and tau in progressive supranuclear palsy. *Front Neuroanat*, 2015. 9: p. 25.
- 21. **Duka**, **T.** *et al.*, Alpha-Synuclein contributes to GSK-3beta-catalyzed Tau phosphorylation in Parkinson's disease models. *FASEB J*, 2009. 23(9): p. 2820–2830.

- 22. **Oikawa, T.** *et al.*, α-Synuclein Fibrils Exhibit Gain of Toxic Function, Promoting Tau Aggregation and Inhibiting Microtubule Assembly. *J Biol Chem*, 2016. 291(29): p. 15046–15056.
- 23. **Zhang, S.** *et al.*, LK6/Mnk2a is a new kinase of alpha synuclein phosphorylation mediating neurodegeneration. *Sci Rep*, 2015. 5: p. 12564.
- 24. **Nübling, G.** *et al.*, Synergistic influence of phosphorylation and metal ions on tau oligomer formation and coaggregation with α-synuclein at the single molecule level. *Mol Neurodegener*, 2012. 7: p. 35.
- Oueslati, A., Implication of Alpha-Synuclein Phosphorylation at S129 in Synucleinopathies: What Have We Learned in the Last Decade? *Journal of Parkinson's Disease*, 2016. 6(1): p. 39–51.
- 26. **Binolfi, A.** *et al.*, Intracellular repair of oxidation-damaged α-synuclein fails to target C-terminal modification sites. *Nat Commun*, 2016. 7: p. 10251.
- 27. Jackson, F. L. C., Human genetic variation and health: new assessment approaches based on ethnogenetic layering. *Br Med Bull*, 2004. 69: p. 215–235.
- 28. Liston, A., E. J. Carr, and M. A. Linterman, Shaping Variation in the Human Immune System. *Trends Immunol*, 2016. 37(10): p. 637–646.
- 29. Mocellin, S. *et al.*, Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. *Gut*, 2015. 64(8): p. 1209–1219.
- Parikshak, N.N., M.J. Gandal, and D.H. Geschwind, Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet*, 2015. 16(8): p. 441–458.
- Clément, K. and P. Ferré, Genetics and the pathophysiology of obesity. *Pediatr Res*, 2003. 53(5): p. 721–725.
- 32. Hofker, M. and C. Wijmenga, A supersized list of obesity genes. *Nat Genet*, 2009. 41(2): p. 139–140.
- 33. **Chambers, J. C.** *et al.*, Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes & Endocrinology*, 2015. 3(7): p. 526–534.
- 34. McClellan, J. and M.-C. King, Genetic heterogeneity in human disease. *Cell*, 2010. 141(2): p. 210–217.
- Deplancke, B., D. Alpern, and V. Gardeux, The Genetics of Transcription Factor DNA Binding Variation. *Cell*, 2016. 166(3): p. 538–554.
- 36. Gajbhiye, R., J. N. Fung, and G. W. Montgomery, Complex genetics of female fertility. *npj Genomic Med*, 2018. 3: p. 29.
- Seelaar, H. et al., Distinct genetic forms of frontotemporal dementia. *Neurology*, 2008. 71(16): p. 1220–1226.
- 38. Gratwicke, J., M. Jahanshahi, and T. Foltynie, Parkinson's disease dementia: a neural networks perspective. *Brain*, 2015. 138: p. 1454–1476.
- 39. **Krüger, S.** *et al.*, Rare Variants in Neurodegeneration Associated Genes Revealed by Targeted Panel Sequencing in a German ALS Cohort. *Front Mol Neurosci*, 2016. 9: p. 92.
- 40. Corsi, A. K., B. Wightman, and M. Chalfie, A Transparent Window into Biology: A Primer on Caenorhabditis elegans. *Genetics*, 2015. 200(2): p. 387–407.
- 41. Hodgkin, J., The Nematode Caenorhabditis elegans. In: ed. by W. B. Wood. Cold Spring Harbor Monographs, 1988. Chap. Sexual Dimorphism and Sex Determination, p. 243–279.
- 42. **Frézal, L.** *et al.*, Natural Genetic Variation in a Multigenerational Phenotype in C. elegans. *Curr Biol*, 2018. 28(16): 2588–2596.e8.
- Kutscher, L.M. and S. Shaham, Forward and reverse mutagenesis in C. elegans. WormBook, 2014: p. 1–26.
- 44. **Cooper, J.F. and J.M. Van Raamsdonk**, Modeling Parkinson's Disease in C. elegans. *Journal of Parkinson's Disease*, 2018. 8(1): p. 17–32.

- 45. Gaeta, A. L., K. A. Caldwell, and G. A. Caldwell, Found in Translation: The Utility of *C. elegans* Alpha-Synuclein Models of Parkinson's Disease. *Brain Sciences*, 2019. 9(4).
- 46. **Unniyampurath, U., R. Pilankatta, and M. N. Krishnan**, RNA Interference in the Age of CRISPR: Will CRISPR Interfere with RNAi? *Int J Mol Sci*, 2016. 17(3): p. 291.
- Hua, Q.-X. et al., A divergent INS protein in *Caenorhabditis elegans* structurally resembles human insulin and activates the human insulin receptor. *Genes Dev*, 2003. 17(7): p. 826–831.
- 48. Alvarez-Ponce, D., M. Aguadé, and J. Rozas, Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 Drosophila genomes. *Genome Res*, 2009. 19(2): p. 234–242.
- 49. Kaletsky, R. and C. T. Murphy, The role of insulin/IGF-like signaling in *C. elegans* longevity and aging. *Disease Models & Mechanisms*, 2010. 3: p. 415–419.
- Schieber, M. and N. S. Chandel, TOR signaling couples oxygen sensing to lifespan in C. elegans. *Cell reports*, 2014. 9(1): p. 9–15.
- 51. **Templeman, N.M. and C.T. Murphy**, Regulation of reproduction and longevity by nutrient-sensing pathways. *J Cell Biol*, 2018. 217(1): p. 93–106.
- 52. **González**, **N**. *et al.*, 2017 update on the relationship between diabetes and colorectal cancer: epidemiology, potential molecular mechanisms and therapeutic implications. *Oncotarget*, 2017. 8(11): p. 18456–18485.
- Faber, P. W. et al., Glutamine/proline-rich PQE-1 proteins protect *Caenorhabditis elegans* neurons from huntingtin polyglutamine neurotoxicity. *Proc Natl Acad Sci U S A*, 2002. 99(26): p. 17131–17136.
- 54. **Harrington, A. J.** *et al., C. elegans* as a model organism to investigate molecular pathways involved with Parkinson's disease. *Dev Dyn,* 2010. 239(5): p. 1282–1295.
- 55. **Hassan, W. M.** *et al.*, Identifying Aβ-specific pathogenic mechanisms using a nematode model of Alzheimer's disease. *Neurobiol Aging*, 2015. 36(2): p. 857–866.
- 56. Link, C. D., C. *elegans* models of age-associated neurodegenerative diseases: lessons from transgenic worm models of Alzheimer's disease. *Exp Gerontol*, 2006. 41(10): p. 1007–1013.
- 57. **Settivari, R.** *et al.*, The Nrf2/SKN-1-dependent glutathione S-transferase π homologue GST-1 inhibits dopamine neuron degeneration in a *Caenorhabditis elegans* model of manganism. *Neurotoxicology*, 2013. 38: p. 51–60.
- 58. Shaye, D. D. and I. Greenwald, OrthoList: a compendium of *C. elegans* genes with human orthologs. *PLoS One*, 2011. 6(5): e20085.
- Calahorro, F. and M. Ruiz-Rubio, *Caenorhabditis elegans* as an experimental tool for the study of complex neurological diseases: Parkinson's disease, Alzheimer's disease and autism spectrum disorder. *Invert Neurosci*, 2011. 11(2): p. 73–83.
- 60. **Rodriguez, M.** *et al.*, Genetic variation for stress-response hormesis in *C. elegans* lifespan. *Exp Gerontol*, 2012. 47(8): p. 581–587.
- 61. Sin, O., H. Michels, and E. A. A. Nollen, Genetic screens in *Caenorhabditis elegans* models for neurodegenerative diseases. *Biochim Biophys Acta*, 2014. 1842(10): p. 1951–1959.
- 62. **Teschendorf, D. and C. D. Link**, What have worm models told us about the mechanisms of neuronal dysfunction in human neurodegenerative diseases? *Mol Neurodegener*, 2009. 4: p. 38.
- 63. Wang, Y. A., J. E. Kammenga, and S. C. Harvey, Genetic variation in neurodegenerative diseases and its accessibility in the model organism *Caenorhabditis elegans*. *Hum Genomics*, 2017. 11(1): p. 12.
- 64. **Cook, D. E.** *et al.,* The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics,* 2016. 204(1): p. 371–383.
- Gaertner, B.E. and P.C. Phillips, *Caenorhabditis elegans* as a platform for molecular quantitative genetics and the systems biology of natural variation. *Genetics Research*, 2010. 92: p. 331–348.

- 66. **Thompson, O. A.** *et al.*, Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. *Genetics*, 2015. 200(3): p. 975–989.
- 67. Félix, M.-A. and C. Braendle, The natural history of Caenorhabditis elegans. *Curr Biol*, 2010. 20(22): R965–R969.
- 68. **Snoek, B. L.** *et al.*, A multi-parent recombinant inbred line population of *C. elegans* allows identification of novel QTLs for complex life history traits. *BMC Biol*, 2019. 17(1): p. 24.
- 69. Andersen, E. C. *et al.*, Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet*, 2012. 44(3): p. 285–290.
- 70. **Volkers, R. J. M.** *et al.*, Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations. *BMC Biol*, 2013. 11: p. 93.
- 71. Rockman, M. V. and L. Kruglyak, Recombinational landscape and population genomics of Caenorhabditis elegans. *PLos Genet*, 2009. 5(3): e1000419.
- 72. **Martin, N., J. Singh, and A. Aballay**, Natural Genetic Variation in the *Caenorhabditis elegans* Response to *Pseudomonas aeruginosa*. *G3*, 2017. 7(4): p. 1137–1147.
- 73. **Reynolds, R. M. and P. C. Phillips**, Natural variation for lifespan and stress response in the nematode Caenorhabditis remanei. *PLoS One*, 2013. 8(4): e58212.
- 74. **Sterken, M.G.** *et al.*, The laboratory domestication of Caenorhabditis elegans. *Trends Genet*, 2015. 31(5): p. 224–231.
- 75. **Gidalevitz, T.** *et al.*, Natural genetic variation determines susceptibility to aggregation or toxicity in a *C. elegans* model for polyglutamine disease. *BMC Biol*, 2013. 11: p. 100.
- Kamkina, P. et al., Natural Genetic Variation Differentially Affects the Proteome and Transcriptome in Caenorhabditis elegans. *Molecular & Cellular Proteomics*, 2016. 15(5): p. 1670–1680.
- 77. **Gao, A. W.** *et al.*, Natural genetic variation in *C. elegans* identified genomic loci controlling metabolite levels. *Genome Res*, 2018. 28(9): p. 1296–1308.
- 78. Kammenga, J. E. *et al.*, Beyond induced mutants: using worms to study natural variation in genetic pathways. *Trends Genet*, 2008. 24(4): p. 178–185.
- 79. Jansen, R. C. and J. P. Nap, Genetical genomics: the added value from segregation. *Trends Genet*, 2001. 17(7): p. 388–391.
- 80. **Hahn, C. N.** *et al.*, Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*, 2011. 43(10): p. 1012–1017.
- Brockmann, K. et al., SNCA: major genetic modifier of age at onset of Parkinson's disease. Mov Disord, 2013. 28(9): p. 1217–1221.
- 82. Gilks, W.P. et al., A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet*, 2005. 365(9457): p. 415–416.
- 83. **He, Q.** *et al.*, DMT1 polymorphism and risk of Parkinson's disease. *Neurosci Lett*, 2011. 501(3): p. 128–131.
- 84. MacLeod, D. A. *et al.*, RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron*, 2013. 77(3): p. 425–439.
- 85. Nalls, M. A. *et al.*, Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*, 2014. 46(9): p. 989–993.
- 86. **Huang, Y.** *et al.*, SNCA Gene, but Not MAPT, Influences Onset Age of Parkinson's Disease in Chinese and Australians. *Biomed Res Int*, 2015. 2015: p. 135674.
- 87. **Simon-Sanchez, J.** *et al.*, Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet*, 2007. 16(1): p. 1–14.
- 88. Ciccodicola, A. et al., Non-coding RNA in Neurodegeneration. *Current Geriatrics Reports*, 2012. 1(4): p. 219.

- 89. Cipolla, G. A. *et al.*, Long Non-Coding RNAs in Multifactorial Diseases: Another Layer of Complexity. *Non-coding RNA*, 2018. 4(2).
- 90. **Duveau, F. and M.-A. Félix**, Role of pleiotropy in the evolution of a cryptic developmental variation in Caenorhabditis elegans. *PLoS Biol*, 2012. 10(1): e1001230.
- 91. Schmid, T. et al., Systemic Regulation of RAS/MAPK Signaling by the Serotonin Metabolite 5-HIAA. *PLos Genet*, 2015. 11(5): e1005236.
- 92. Sterken, M. G. *et al.*, Ras/MAPK Modifier Loci Revealed by eQTL in *C. elegans*. *G*3, 2017. 7(9): p. 3185–3193.

Chapter 2

Genetic variation in neurodegenerative diseases and its accessibility in the model organism Caenorhabditis elegans

Yiru A. Wang Jan E. Kammenga Simon C. Harvey

This chapter has been published as:

Y. A. Wang, J. E. Kammenga, and S. C. Harvey, Genetic variation in neurodegenerative diseases and its accessibility in the model organism *Caenorhabditis elegans. Hum Genomics*, 2017. 11(1): p. 12

Abstract

Neurodegenerative diseases (NGDs) such as Alzheimer's and Parkinson's are debilitating and largely untreatable conditions strongly linked to age. The clinical, neuropathological, and genetic components of NGDs indicate that neurodegeneration is a complex trait determined by multiple genes and by the environment. The symptoms of NGDs differ among individuals due their genetic background and this variation affects the onset and progression of NGD and NGD-like states. Such genetic variation affects the molecular and cellular processes underlying NGDs, leading to differential clinical phenotypes. So far we have a limited understanding of the mechanisms of individual background variation. Here, we consider how variation between genetic backgrounds of *C. elegans* affects the mechanisms of ageing and proteostasis in NGD phenotypes. Additionally, we discuss advances in *C. elegans* methods that can facilitate the identification of NGD regulators and/or networks.

Key words: neurodegenerative diseases, natural variation, Quantitative genetics, *C. elegans*

Introduction

Neurodegenerative diseases (NGDs) cause disability and premature death, primarily among older people. These chronic and fatal illnesses include Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's diseases (HD), spinocerebellar ataxia, prion diseases (PrD) (*i.e.* transmissible spongiform encephalopathies), frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) [2]. Most NGDs are age-dependent with their incidence increasing with advancing age. This makes understanding NGDs increasingly important given the recent increase in human lifespan seen in many countries. Many disease causing mutations underlying different NGDs have phenotypic effects that result from the misfolding of proteins and/or mitochondrial dysfunctions leading to widespread damage in different parts of the nervous system (Figure 1). This damage leads to a range of symptoms and the overlap between symptoms in different NGDs can make it difficult to precisely diagnose patients. Diagnosis is further complicated by the individual variability apparent from the onset of disease [3]. For example, PD is characterized by motor issues, including tremor, slowing of movement, and an unstable gait, and by cognitive symptoms, but only some patients suffer from cognitive impairment and develop dementia [4].

The rate of disease progression (*i.e.*, the duration of a given neuropathological stage) and clinical presentation also vary from one patient to another. Young onset PD patients, for example, often have a more frequent family history of PD and a more variable survival rate relative to those without the familial history [11]. Several studies of the amyloid- β protein aggregates, which cause AD, also indicate that the existence of distinct shapes in A β aggregates, 40 residue A β (A β 40) and 42 residue A β fibril structures [12], and identify the distinct strain-specific traits (defined as "strainness") of the forms of AD by the different conformation of the aggregates [13, 14].

There is however limited knowledge about the mechanisms that determine individual variation. Individuals that carry the same mutation in the same disease causing gene may display a range of different clinical symptoms. For example, assessing 6-year change in verbal memory, processing speed and executive function in AD identified effects of MS4A4E, CLU, and NME8 in whites and of ZCWPW1 and CDS33 variants in African-Americans. For MS4A4E and CLU this association was only significant in individuals bearing at least one APOE $\epsilon 4$ allele (AD risk gene) [15]. Apart from lifestyle and environment, these individual differences are caused by the unique genetic background of each person. The genetic background, thus, could be defined as

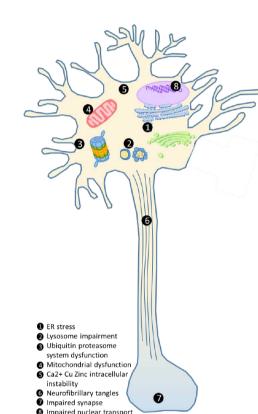


Figure 1: Schematic diagram of the pathological mechanisms associated with neurodegeneration in neurons. Eight main events exist in the cellular biological processes related to neurodegeneration: 1) The protein misfolding process inhibits ER to Golgi trafficking and alters ER-associated degradation, inducing the ER stress [5] 2) Damaged lysosomes disrupt processes that maintains lysosomal homeostasis [6]; 3) Accumulation of misfolded proteins generates positive feedback exacerbating other effects [7]; 4) Dysfunctional mitochondrial maintenance directly affects mitochondrial biogenesis and induction of autophagy [8], *e.g.* the production of ROS and ATP in the cell; 5) Altered homeostasis generates excessive influx of calcium, copper and zinc, due to ER and mitochondrial stresses [6]; 6) Plaques and neurofibrillary tangles, due to Tau phosphorylation and aggregates, reduce neurotransmitter release, which also 7) weaken synaptic strength [7]; 8) Signalling pathways in the stressed organelle or the cytoplasm induce the transductions of the signals to the nucleus, which provoke DNA damage [9, 10].

the genetic makeup of all alleles that interact with the disease related to the "disease causing" mutation [16].

Detailed mechanistic studies into background modifiers of NGDs are however difficult to conduct at the individual level in humans for ethical and technical reasons. Model organisms such as yeast, insects, worms, fish and rodents are therefore critical to furthering our understanding of differences between individual in NGDs. This use is facilitated by the range of methods and approaches available to construct transgenic models of human diseases in model species with different genetic backgrounds [17–20]. They are likewise an important resource for investigating the genetic underpinnings of quantitative traits, including complex disease phenotypes. Here we review the onset and progression of NGDs in the context of genetic background and illustrate how work on the model species *Caenorhabditis elegans* can illuminate the underlying mechanisms of individual variation.

NGD phenotypes depend on the genetic background

Individual genetic backgrounds differ from by thousands to millions of genetic variants, that will range from single nucleotide polymorphisms (SNPs) to, potentially very large, copy number variants (CNVs) [21, 22]. This genetic variation is a major determinate of differences in predisposition to disease [17] where risk-increasing variants are numerous, display intricate patterns of interaction with each other as well as with non-genetic variables, and unlike classical Mendelian ("monogenic") disorders will often exhibit no single mode of inheritance. Differential phenotypes for most NGDs arise from multiple genetic variants and their interaction with each other, as well as environmental factors. Hence, the genetics of these diseases is considered "complex" based on the heterogeneity in pathology and the disease polygenicity [23].

Genetic background can affect an important feature shared by different NGDs, specifically the formation of cerebral deposits of misfolded protein aggregates (also called prion-like proteins). For example, AD is characterised by the presence of neurofibrillary tangles and beta amyloid peptide (A β) in neural plaques, which are abnormal accumulations of microtubule-associated protein tau in a hyperphosphorylated state [24]. Similarly, the accumulation of proteins with polyglutamine-rich extensions is characteristic of HD and associated polyglutamine diseases [25], while PD involves the loss of dopaminergic neurons and the presence of Lewy bodies and Lewy neurites that are the aggregates of the synaptic protein alpha-synuclein. However, exactly how genetic variation modifies and affects specific parts of the NGD pathways is

mostly unclear. We suggest that the genetic variation associated with NGDs can be grouped into two classes. Firstly, variation present directly in disease genes. Secondly, variation in regulatory factors that modulate onset and progression of the NGDs. Critically this distinction separates those variations that are causative of disease from those that modify the disease, but that on their own, cannot cause NGD.

In an example of the first such class of variants, multiple rare mutations in amyloid precursor protein (APP), or in the presenilin-1 and 2 genes (PSEN1 and PSEN2) can cause early onset AD (Figure 3). Similarly, multiple mutations in LRRK2, a gene associated with PD, are known to be related to the sporadic late-onset form of the disease. Here, the G2385R and R1628P LRRK2 variants are validated risk factors for PD in Asian populations, while the G2019S variant has been identified in different populations worldwide [26]. Examples of the second class of variants are those known to act in the IIS/mTOR pathway (Figures 2 and 3). This pathway modulates response to a range of stresses and in the NGDs has been linked to a range of pathological processes (Figure 1 and Table S1). For instance, Baleriola *et al.* (2014) reported moderate eIF2 α activation by A β and a greater frequency of ATF4 (the eIF2 α effector) transcripts were identified in axons in the brain of AD patients [27]. ATF4 is also known to activate the Parkin gene in PD and also to be related to stress pathways [5]. ATF4 therefore induces transcriptional expression of genes mediated by the UPR, including genes involved in amino acid metabolism, resistance to oxidative stress, and the proapoptotic transcription factor CHOP, which are related to diseases processing [27, 28]. However, rare coding variation in ATF4 has been also found with pathway impairment in patients with sporadic cervical dystonia [29]. Additionally, Chaudhry *et al.* (2015), using gene-based analyses for late onset AD, revealed associations with the WT1, ZC3H12C, DLGAP2, and GPR1 genes, suggesting a possible role in AD pathogenesis [30].

Genetic variants therefore produce a broad spectrum of biological effects. This can be seen in analysis of natural variants affecting LOAD, where more than 20 genes involved in a range of processes including metabolism, inflammation, synaptic activity and intracellular trafficking have been identified [46]. CNVs are also associated with variation in NGD phenotypes, with, for example, a rare duplication of the amyloid- β protein precursor linked to early onset AD [47]. Thus, the question how genetic variants affect mechanisms in diseases such as AD and PD remains challenging. Here, model species like the nematode *C. elegans* might offer a solution and provide more insight into the mechanism underlying individual variation in NGD disease phenotypes.

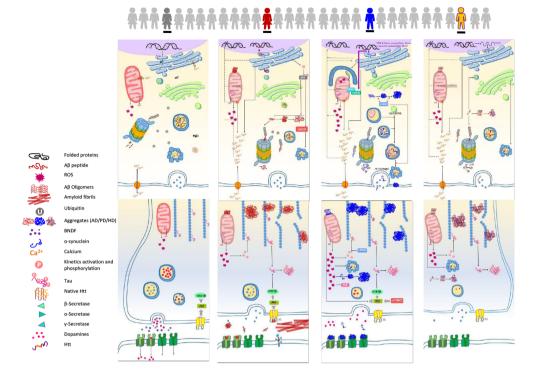


Figure 2: The molecular processes implicated in neurodegeneration in the neurons of a normal healthy individual (grey) and in AD (red), PD (blue) and HD (vellow) patients. The schematic neuron is divided into the soma and axonal terminal bouton. For simplicity, postsynaptic/dendritic events are not included. Misfolded proteins aggregate first into oligomers, and then into higher-molecular-weight insoluble protofibrils and further aggregates [14, 31, 32]. In AD patients with mutations in APP, PSEN1, PSEN2 or APOE, the JNK pathway is activated, increasing levels of phospho-JNK in neurons. This mediates the phosphorylation of APP and FoxO-dependent autophagy [33, 34]. Moreover, the soluble A β oligomers activate the mTOR pathways again promoting autophagy [33]. Further phosphorylation of tau and impaired A β activate the IIS/Akt pathways and affects cognitive function and synaptic plasticity [35]. In PD, mutations in PINK and Parkin (related to EOPD) or α-synuclein mutations (LOPD) lead to the inhibition of α -synuclein degradation as well as accumulation of autophagic vacuoles, which result in neuronal death [36, 37]. Misfolded α -synuclein also interacts with membranes and mitochondria, causing calcium dysregulation and a reduction of mitochondrial activity. This results in mtDNA damage as well as impairments to the ubiquitin-proteasome system (UPS) and mitophagy. The significant pathological aetiology of HD is the enlargement of the polyglutamine (polyQ) domain within the HTT protein's N-terminus [38]. MLH1 (MutL homolog 1) and an SNP within a nuclear factor-kB binding site (Nf-KB) in the HTT promoter play a role in the altered onset of HD. In comparison with AD and PD, proteasome efficiency is strongly reduced in HD patients. Meanwhile, the polyQ domain of mutant HTT contributes increased toxicity by attracting and binding to other cytoplasmic and nuclear structures that contain polyglutamine (in the review [39]). Additionally, a major loss of brain-derived neurotrophic factor (BDNF) protein has been shown in HD, may due to the deficits in BDNF delivery and/or loss of BDNF gene transcription by mutant Htt [40, 41].

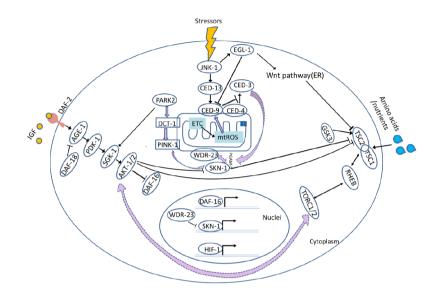


Figure 3: C. elegans cellular pathways, and their crosstalk, relating to aging and the stress response. Shown (from left to right) are the Insulin/Insulin like growth factor (IIS) signalling pathway, the mitochondrial signalling pathways, and the mechanistic target of rapamycin (TOR) pathway. Arrows indicate positive regulatory events and bars indicate inhibitory interactions. Purple block arrows represent interactions between the different pathways, whereas dashed purple block arrows indicate possible indirect interactions. The oval molecules and their corresponding mammalian homologs involved in IIS [42-45] are as below: AGE-1/PI3K, phospatidylinositol-3-kinase; PDK-1, 3-phosphoinositide-dependent kinase 1 ortholog; SGK-1, a serine/threonine protein kinase that is orthologous to the mammalian serum- and glucocorticoid-inducible kinases (SGKs); Akt/PKB, the serine/threonine kinase; DAF-16/FOXO, forkhead box O (FOXO) transcription factor. Mitochondrial dysfunctions [90-96] are associated with apoptotic/programmed cell death (PCD), aberrant autophagic regulation, endoplasmic reticulum dysfunction, and intracellular calcium, including: c-Jun N-terminal kinase (JNK) subgroup of mitogen-activated protein (MAP) kinases; CED-4, CED-9, and EGL-1 belongs to a conserved genetic pathway to regulate apoptosis during C. elegans development [34]; PINK-1, a predicted serine/threonine kinase which is similar with human PINK1, PTEN-induced kinase-1; SKN-1/Nrf, skin in excess transcription factor 1/NF-E2-related factor; mtROS, mitochondrial reactive oxygen species; ATP, adenosine-5'-triphosphate; HIF-1, hypoxia-inducible transcription factor 1. Major molecules in TOR pathways include: TSC1/2, tuberous sclerosis complexes 1 and 2; RHEB, Ras homolog enriched in brain; TOR, target of rapamycin kinase; GSK3, glycogen synthase kinase ortholog. See text in Box 1 for further details.

Natural variation associated with complex traits in C. elegans NGD models

C. elegans is a globally distributed nematode and its natural genetic diversity is similar to human genetic variation [48]. Studying NGDs in a genetically tractable model species, such as *C. elegans*, allows detailed insight into the molecular pathogenesis. C. elegans models of AD, PD, HD and other several NGDs have been established [49–53]. These models involve the transgenic expression of human genes under the control of a C. elegans promotor, with the resulting protein often linked to GFP or YFP. Applying fluorescent proteins (e.g. YFP) allows tracking of target proteins over time and the visualization of aggregation in vivo via the observation of the fluorescent foci (e.g. [54]). Such studies of AD using transgenic *C. elegans* have been carried out for many years. More recently, McColl et al. (2012) have generated an improved model with human DA- A β 1-42 under the control of the muscle-specific promoter *unc-54* promotor. In this line, the full-length A β 1-42 (the predominant A β species in human brain) is expressed in worm body wall muscle cells, and its oligomerizations and aggregations develop and result in severe. This fully penetrant, age progressive-paralysis also shows more rapid than that caused from A β 3-42 expression [20].

These transgenic models have all been generated in the N2 strain of *C. elegans* and have been widely used in screens for gene function analysis. However, such analyses in a single background have constrained the analysis and detection of natural allelic variants associated with complex traits. The most striking demonstration of this is a comparison of the RNAi phenotypes of \sim 1,400 genes between N2 and CB4856 that showed that \sim 20% of genes differ in the severity of phenotypes between just these two genetic backgrounds [55]. Crucially, the natural variation demonstrated in *C. elegans* is also sufficient to cause significant changes in signalling pathways both at the gene expression (transcript and protein abundance) and phenotypic levels [56]. Accordingly, these could contribute to understanding how allelic variation affects gene expression, at the level of translation, in the multiple pathways and/or networks of complex traits involved in development and disease progression. For comparative investigations on mammalian genetics (including metabolism, aging, cancer and neurodegeneration diseases), we therefore suggest that taking the different genetic backgrounds of the worm into consideration could be valuable. This approach would be likely to clarify proteostasis mechanisms, and more fully reveal the relationships between genotype and phenotype. For example, in Drosophila, Chow et al. (2013) showed 114 lines from the sequenced Drosophila Genetic Reference Panel of wild derived inbred strains exhibit the high heterogeneity in survival under ER stress conditions, and showed that 17 of 25 tested candidate genes were active in the putative response to ER stress [57].

Genetic and phenotypic differences between *C. elegans* populations have been identified with many traits shown to be variable between isolates (*e.g.* [58–63], for review see [64]). Next to the canonical wild type Bristol N2, the most extensively characterized *C. elegans* isolate is the Hawaiian strain CB4856. This isolate is extensively diverged from N2, with a recent de novo assembly of the CB4856 genome suggesting that many variations have been maintained by balancing selection over long evolutionary timescales [65]. Wild *C. elegans* populations have also been shown to display significant local adaptation to their environment at multiple levels from the genotype and transcriptome [60]. Given this genetic and phenotypic variation, it is likely that variable natural genetic backgrounds of *C. elegans* will harbor abundant genetic variation that will modify the severity of NGD phenotypes. Wider analysis of more recently isolated wild isolates is also important as recent research has revealed specific adaptations to laboratory conditions that exist in the canonical wild type Bristol N2 [66–68].

Studies into genetic background effects on NGDs in *C. elegans* have focused on a polyglutamine model. Here, Gidalevitz *et al.* (2013) looked at the effects of natural genetic variation on susceptibility to aggregation and to toxicity [69]. This work introgressed a polyQ40 transgene from an N2 background into three wild genetic backgrounds, including the California-derived isolate DR1350, the Madeira isolate JU258, and the genetically distant Hawaiian isolate CB4856. A series of markers of pathology progression, such as onset of toxicity aggregation, and the differential cell-specific susceptibility to aggregation, showed wide variation among the new introgression lines and between polyQ40 carrying RILs derived from DR1350 and N2. This indicates that natural variation in genetic background can control resistance to misfolded protein aggregations, and act to bind its associated cellular dysfunctions, which encourages further dissection on natural genetic variation susceptibility to age-related protein homeostasis in disease mechanisms.

Li *et al.* (2004) mapped a large fraction of the *C. elegans* protein-protein interaction network. This was extended when the initial version of a human interactome map came out, adding more than 300 new connections to over 100 disease-associated proteins, including proteins related to NGD [70, 71]. Lejeune *et al.* (2012) accomplished a large-scale RNA interference screen in *C. elegans* strains that express N-terminal huntingtin in touch receptor neurons [72]. Then, a subset of high-confidence modifier genes in pathways of

interest in HD were identified by network-based analysis, including metabolic, neurodevelopmental and pro-survival pathways. These results, and those of similar large-scale analyses, support the investigation of human disease pathways using *C. elegans* as a model (Table S1).

C. elegans offers the opportunity to investigate the genetic background modifiers affecting complex disease pathways. For example, comparisons of RAS/MAPK signaling – a pathway critical in many complex human diseases, including AD [73] – across two different genetic backgrounds, N2 and CB4856, as well as their derived recombinant inbred lines (Box 2) identified the polymorphic monoamine oxidase amx-2 (MAOA) as a negative regulator of RAS/MARK [74]. It was found that MAOA's effects on RAS/MAPK signaling are produced by its effects on the level of the serotoninmetabolite 5 hydroxyindoleacetic acid (5-HIAA), the first endogenous small molecule identified to act as a systemic inhibitor of RAS/MAPK signaling [74].

Genetic networks in stress and aging leverage the detection of NGD modifiers

NGDs are associated with other complex diseases (*e.g.* cancers). In part, this is a consequence of the involvement of the cellular systems that deal with various types of stress. For instance, variation in the genes that regulate the brain's molecular response to oxidative stress are associated with differential neural vulnerability to the damaging effects of amyloid- β [75]. Here, oxidative phosphorylation takes place within mitochondria, to meet the elevated energy demands of neurons (Figures 1, 2, and 3). However, this can also accelerate the accumulated oxidative damage, which could trigger impaired mitochondrial energy production, upregulate oxidative phosphorylation causing further DNA damage and significant levels of neuronal apoptosis [76].

A large number of genome wide association studies (GWAS) in humans have identified pathways modulating the rate of aging and simultaneously influencing multiple age-related diseases such as NGDs [77, 78]. These GWAS studies have been invaluable in identifying and characterizing genetic variants associated with variation in disease phenotypes. For example, Seshadri *et al.* (2010) identified 38 SNPs within 10 loci by integrating data from multiple studies and also identified CLU, PICALM, and CR1 as novel genes for late-onset AD (LOAD) [79]. A similar meta-analysis of GWAS studies of PD identified six more risk loci associated with the disease [80], whilst a combination of GWAS and whole-exome sequencing (WES) identified variants in the CEL gene/locus associated with PD [81]. Similarly, N'Songo *et al.* (2017) identified ABCA7 missense variants that play a role in conferring AD risk in African Americans by performing WES with previously identified AD GWAS loci [82]. According to their highlighted allelic heterogeneity at this locus, they also suggested the presence of additional AD-risk variants in MS4A6A, PTK2B, and ZCWPW1.

NGDs are also strongly associated with aging, indicating that there is a link between protein misfolding and aging. This is very clear when the ILS (FOXO) pathway, a representative aging-related pathway, is considered. In *C. elegans*, this pathway ultimately controls the localization of DAF-16 – a FOXO transcription factor transcription factor that regulates a large number of genes involved in abiotic and biotic stress resistance, metabolism, and longevity – with dephosphorylation of DAF-16 allow it to enter the nucleus [19]. The phosphorylation of DAF-16 is controlled by the insulin/IGF-1 transmembrane receptor ortholog DAF-2, with reductions in daf-2 activity resulting in daf-16 dependent increases in lifespan. Loss of function daf-2 mutations also increase lifespan in PD mutants of C. elegans, delaying the accumulation of small aggregates of alpha-synuclein in the body wall muscle, and rescuing deficiencies in resistance to different stresses [31]. In humans, the age of onset of HD has a strong association with the length of polyQ expansion in the huntingtin protein, but also varies between individuals with the same repeat length. In laboratory mouse strains, different genetic backgrounds can induce differential somatic expansion of CAG repeat.

Perspectives: What is the genetic architecture of variation in NGD responses in C. elegans?

C. elegans quantitative genetics (Box 2) has yielded considerable insights into understanding complex human disease pathways, but there has been limited work in extending this to the analysis of NGDs. To this end, powerful and serviceable datasets relating to causal genetic variants exploration based on genomic analysis have been collected and can be obtained from WormQTLHD (Human Disease) www.WormQTL-HD.org [83]. WormQTLHD allows for systematically investigating phenotypic expression of *C. elegans* at levels equivalent to those of human diseases, by catalyzing integration of reported disease candidate gene associations, gene orthologue data, molecular profiles, phenotypic variations and QTL results [83, 84]. Thus, this will support relevant available meta datasets for human-worm studies and database exploration.

Taken together, *C. elegans*, a powerful model organism, can be used to study how variation affects the onset and progression of protein misfolding disease

and how the susceptibility to proteotoxicity performs in genetically diverse but phenotypically general individuals. This may explain the higher propensity to aggregation of the mutant disease-related proteins, in order to uncovering onset and phenotypes of disease patterns [69]. Regardless of the toxicity and aggregation of misfolded proteins, other factors leading to the variable complex traits could possibly include multiple additive or allele interactions, with the consequence of underlying intervention strategies at onset and/or in progression of disease. In addition, Paaby et al. (2014) revealed cryptic genetic variations (CGVs) in the gene networks of C. elegans embryogenesis [85]. CGVs are silent alleles in general and can be only activated to influence phenotype, when other functional genes is perturbed. The seemingly omnipresent cryptic-effect loci segregate at intermediate frequencies in the wild. However, they reveal low developmental pleiotropy. In the specific perturbations, *e.g.* changes of the molecular, cellular, or developmental processes that govern its phenotypic expression in complex metabolic human diseases, CGVs are required to be revealed [43, 59], for reviews see [86, 87]. Above all, the natural variation play an important role in neurodegeneration, involving inherently plastic genetic and molecular pathways, and might allow for description of complex etiology and implementation in eluding the harmful influences.

Conclusions

Given the molecular conservation in neuronal signalling between *C. elegans* and vertebrates, including humans, this nematode is a valuable model species for studying NGD pathways and the alleles that affect them. Despite progress in identification of several AD and PD related genes, the effect of natural alleles underlying protein misfolding in these diseases remains mostly unclear. Here, we reviewed how natural variation could influence the severity of disease phenotypes. Experiments that rely solely on induced mutants in Bristol N2 limit the ability to explore how naturally varying alleles alter signalling pathways. Thus, further research should go beyond classic mutant screens on the genetic pathway analysis of complex traits, *i.e.* phenotypic differences among individuals, to also consider the allelic interactions in different genetic backgrounds. Consequently, the predictive nematode models of human genetic diseases could provide a more complete genetic and molecular understanding of how genetic variation shapes gene expression and cell biology for personalized genomic medicine.

List of abbreviations

AD	Alzheimer's disease
ALS	amyotrophic lateral sclerosis
APP	amyloid precursor protein
Αβ	beta amyloid peptide
BDNF	Brain-derived neurotrophic factor
CGVs	cryptic genetic variations
CNV	copy number variants
EOAD	early onset AD
eQTLs	expression quantitative trait loci
ER	endoplasmic reticulum
FOXO	forkhead box O protein
FTD	frontotemporal dementia
GWAS	genome wide association studies
HD	Huntington's disease
HTT	huntingtin
IIS	Insulin/Insulin like growth factor
ILs	introgression lines
JNK	c-Jun N-terminal kinase
LOAD	late-onset alzheimer's disease
LRRK2	leucine-rich repeat kinase 2
MAOA	monogeamine oxidase amx-2
MAPK	mitogen-activated protein kinases
MLH1	MutL homolog 1
mROS	mitochondrial reactive oxygen species
mTOR	mechanistic target of rapamycin
mTORC2	mammalian TORC2
NF-kB	Nuclear factor kB
NGDs	Neurodegenerative diseases
PCD	programmed cell death
PD	Parkinson's disease
PESN1/2	presenilin-1/2 gene
polyQ	polyglutamine
PrD	prion diseases
QTLs	quantitative trait locus
QTNs	quantitative trait nucleotides
RILs	Recombinant inbred lines
ROS	reactive oxygen species
S6K	subunit S6 kinase
SNP	single nucleotide polymorphism
UPR	unfolded protein response
UPS	ubiquitin/proteasome system
WES	whole-exome sequencing
5-HIAA	serotoninmetabolite 5 hydroxyindoleacetic acid
	, , , , , , , , , , , , , , , , , , ,

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: Not applicable.

Competing interests: The authors declare that they have no competing interests.

Funding: YW is supported by a PhD Scholarship from Canterbury Christ Church University. JK is supported by the Human Frontiers Science Program. SH is supported by the Leverhulme Trust.

Authors' contributions: YW, JK and SH wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Mark Sterken, Basten Snoek, Jana Stastna, Katharina Jovic, Yu Nie and the members of our laboratories for helpful advice and discussions. We also thank Bei Tian, Jun Qiu, Yongran Ji, Chunyue Wei and Dianfan Zhou for assistance.

Endnotes: Not applicable.

Supplementary material

 Table S1: Pathways regulating longevity, stress, and disease responses, which is deposited at:

 https://humgenomics.biomedcentral.com/articles/10.1186/s40246-017-0108-4/tables/1

Table S2: Quantitative genetic studies of genomic and phenotypic variation, which is deposited at:

https://humgenomics.biomedcentral.com/articles/10.1186/s40246-017-0108-4/tables/2

References

- Wang, Y. A., J. E. Kammenga, and S. C. Harvey, Genetic variation in neurodegenerative diseases and its accessibility in the model organism *Caenorhabditis elegans*. *Hum Genomics*, 2017. 11(1): p. 12.
- 2. Soto, C., Protein Misfolding in Neurodegenerative Diseases: The Key Pending Questions. *Journal of Neurology & Translational Neuroscience*, 2013. 1: p. 1010.
- 3. Ukraintseva, S. V. *et al.*, Trade-offs between cancer and other diseases: do they exist and influence longevity? *Rejuvenation Res*, 2010. 13(4): p. 387–396.
- 4. Del Tredici, K. and H. Braak, Lewy pathology and neurodegeneration in premotor Parkinson's disease. *Mov Disord*, 2012. 27(5): p. 597–607.
- Mercado, G., P. Valdés, and C. Hetz, An ERcentric view of Parkinson's disease. *Trends Mol Med*, 2013. 19(3): p. 165–175.
- 6. Schneider, L. and J. Zhang, Lysosomal function in macromolecular homeostasis and bioenergetics in Parkinson's disease. *Mol Neurodegener*, 2010. 5: p. 14.
- Ciechanover, A. and Y. T. Kwon, Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. *Experimental & Molecular Medicine*, 2015. 47: e147.
- 8. Saito, T. and J. Sadoshima, Molecular mechanisms of mitochondrial autophagy/mitophagy in the heart. *Circ Res*, 2015. 116(8): p. 1477–1490.
- 9. Chakraborty, S. et al., Oxidative stress mechanisms underlying Parkinson's disease-associated neurodegeneration in C. elegans. Int J Mol Sci, 2013. 14(11): p. 23103–23128.
- 10. Schumacher, B. et al., Translational repression of *C. elegans* p53 by GLD-1 regulates DNA damage-induced apoptosis. *Cell*, 2005. 120(3): p. 357–368.
- 11. Schrag, A. and J.M. Schott, Epidemiological, clinical, and genetic characteristics of early-onset parkinsonism. *Lancet Neurol*, 2006. 5(4): p. 355–363.
- 12. **Lu**, **J.-X.** *et al.*, Molecular structure of β-amyloid fibrils in Alzheimer's disease brain tissue. *Cell*, 2013. 154(6): p. 1257–1268.
- 13. **Stöhr, J.** *et al.*, Distinct synthetic Aβ prion strains producing different amyloid deposits in bigenic mice. *Proc Natl Acad Sci U S A*, 2014. 111(28): p. 10329–10334.
- 14. **Watts, J. C.** *et al.,* Serial propagation of distinct strains of Aβ prions from Alzheimer's disease patients. *Proc Natl Acad Sci U S A*, 2014. 111(28): p. 10323–10328.
- 15. **Bressler, J.** *et al.*, Genetic variants associated with risk of Alzheimer's disease contribute to cognitive change in midlife: The Atherosclerosis Risk in Communities Study. *Am J Med Genet*, 2017. 174(3): p. 269–282.
- 16. **Kammenga, J. E.**, The background puzzle: how identical mutations in the same gene lead to different disease symptoms. *FEBS J*, 2017. 284(20): p. 3362–3373.
- 17. Albert, F. W. and L. Kruglyak, The role of regulatory variation in complex traits and disease. *Nat Rev Genet*, 2015. 16(4): p. 197–212.
- Dawson, T. M., H. S. Ko, and V. L. Dawson, Genetic animal models of Parkinson's disease. *Neuron*, 2010. 66(5): p. 646–661.
- 19. **Hesp, K., G. Smant, and J.E. Kammenga**, *Caenorhabditis elegans* DAF-16/FOXO transcription factor and its mammalian homologs associate with age-related disease. *Exp Gerontol*, 2015. 72: p. 1–7.
- McColl, G. *et al.*, Utility of an improved model of amyloid-beta (Aβ₁₋₄₂) toxicity in *Caenorhabditis elegans* for drug screening for Alzheimer's disease. *Mol Neurodegener*, 2012. 7: p. 57.
- 21. Stranger, B. E. *et al.*, Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, 2007. 315(5813): p. 848–853.

- 22. **Zhang, F.** *et al.*, Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*, 2009. 10: p. 451–481.
- 23. Bertram, L. and R. E. Tanzi, The genetic epidemiology of neurodegenerative disease. *J Clin Invest*, 2005. 115(6): p. 1449–1457.
- 24. Serrano-Pozo, A. et al., Neuropathological alterations in Alzheimer disease. Cold Spring Harbor Perspectives in Medicine, 2011. 1(1): a006189.
- 25. **Choi, A. M. K., S. W. Ryter, and B. Levine**, Autophagy in human health and disease. *N Engl J Med*, 2013. 368(19): p. 1845–1846.
- Monfrini, E. and A. Di Fonzo, Leucine-Rich Repeat Kinase 2 (LRRK2). In: ed. by H. J. Rideout. Advances in Neurobiology. Springer, 2017. Chap. Leucine-Rich Repeat Kinase (LRRK2) Genetics and Parkinson's Disease, p. 3–30.
- 27. **Baleriola**, J. *et al.*, Axonally synthesized ATF4 transmits a neurodegenerative signal across brain regions. *Cell*, 2014. 158(5): p. 1159–1172.
- Fusakio, M. E. *et al.*, Transcription factor ATF4 directs basal and stress-induced gene expression in the unfolded protein response and cholesterol metabolism in the liver. *Mol Biol Cell*, 2016. 27(9): p. 1536–1551.
- Rittiner, J. E. *et al.*, Functional Genomic Analyses of Mendelian and Sporadic Disease Identify Impaired eIF2α Signaling as a Generalizable Mechanism for Dystonia. *Neuron*, 2016. 92(6): p. 1238–1251.
- Chaudhry, M. et al., Genetic variation in imprinted genes is associated with risk of late-onset Alzheimer's disease. J Alzheimers Dis, 2015. 44(3): p. 989–994.
- Cooper, J. F. *et al.*, Delaying aging is neuroprotective in Parkinson's disease: a genetic analysis in , javax.xml.bind.JAXBElement@7edef250, models. *NPJ Parkinson's Disease*, 2015. 1: p. 15022.
- 32. **Imarisio**, S. *et al.*, Huntington's disease: from pathology and genetics to potential therapies. *Biochem J*, 2008. 412(2): p. 191–209.
- Ma, X. M. and J. Blenis, Molecular mechanisms of mTOR-mediated translational control. Nat Rev Mol Cell Biol, 2009. 10(5): p. 307–318.
- 34. **Oh, S. W.** *et al.*, JNK regulates lifespan in *Caenorhabditis elegans* by modulating nuclear translocation of forkhead transcription factor/DAF-16. *Proc Natl Acad Sci U S A*, 2005. 102(12): p. 4494–4499.
- 35. **Jack, C. R.** *et al.*, Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol*, 2013. 12(2): p. 207–216.
- 36. **Collier, T. J.** *et al.,* Is Alpha-Synuclein Loss-of-Function a Contributor to Parkinsonian Pathology? Evidence from Non-human Primates. *Front Neurosci,* 2016. 10: p. 12.
- Triplett, J. C. *et al.*, Quantitative expression proteomics and phosphoproteomics profile of brain from PINK1 knockout mice: insights into mechanisms of familial Parkinson's disease. *J Neurochem*, 2015. 133(5): p. 750–765.
- Zhang, X. et al., Potential Transfer of Polyglutamine and CAG-Repeat RNA in Extracellular Vesicles in Huntington's Disease: Background and Evaluation in Cell Culture. *Cell Mol Neurobiol*, 2016. 36(3): p. 459–470.
- 39. **Bečanović, K.** *et al.*, A SNP in the HTT promoter alters NF-κB binding and is a bidirectional genetic modifier of Huntington disease. *Nature Neuroscience*, 2015. 18(6): p. 807–816.
- 40. **Ferrer, I.** *et al.*, Brain-derived neurotrophic factor in Huntington disease. *Brain Res*, 2000. 866: p. 257–261.
- Jeon, J. *et al.*, Gene therapy by proteasome activator, PA28γ, improves motor coordination and proteasome function in Huntington's disease YAC128 mice. *Neuroscience*, 2016. 324: p. 20–28.

- 42. **Gil**, **E**. **B**. *et al.*, Regulation of the insulin-like developmental pathway of *Caenorhabditis elegans* by a homolog of the PTEN tumor suppressor gene. *Proc Natl Acad Sci U S A*, 1999. 96(6): p. 2925–2930.
- 43. **Mihaylova**, **V. T.** *et al.*, The PTEN tumor suppressor homolog in *Caenorhabditis elegans* regulates longevity and dauer formation in an insulin receptor-like signaling pathway. *Proc Natl Acad Sci U S A*, 1999. 96(13): p. 7427–7432.
- 44. **Ogg, S. and G. Ruvkun**, The *C. elegans* PTEN homolog, DAF-18, acts in the insulin receptor-like metabolic signaling pathway. *Mol Cell*, 1998. 2(6): p. 887–893.
- 45. **Rouault, J. P.** *et al.*, Regulation of dauer larva development in *Caenorhabditis elegans* by daf-18, a homologue of the tumour suppressor PTEN. *Curr Biol*, 1999. 9(6): p. 329–332.
- 46. Lambert, J. C. *et al.*, Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*, 2013. 45(12): p. 1452–1458.
- 47. **Hooli, B.V.** *et al.*, Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol Psychiatry*, 2014. 19(6): p. 676–681.
- 48. **Cutter, A. D.**, Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer Caenorhabditis elegans. *Genetics*, 2006. 172(1): p. 171–184.
- Faber, P. W. et al., Glutamine/proline-rich PQE-1 proteins protect *Caenorhabditis elegans* neurons from huntingtin polyglutamine neurotoxicity. *Proc Natl Acad Sci U S A*, 2002. 99(26): p. 17131–17136.
- 50. **Harrington, A. J.** *et al., C. elegans* as a model organism to investigate molecular pathways involved with Parkinson's disease. *Dev Dyn*, 2010. 239(5): p. 1282–1295.
- 51. **Hassan**, **W. M.** *et al.*, Identifying Aβ-specific pathogenic mechanisms using a nematode model of Alzheimer's disease. *Neurobiol Aging*, 2015. 36(2): p. 857–866.
- 52. Link, C. D., C. *elegans* models of age-associated neurodegenerative diseases: lessons from transgenic worm models of Alzheimer's disease. *Exp Gerontol*, 2006. 41(10): p. 1007–1013.
- Settivari, R. et al., The Nrf2/SKN-1-dependent glutathione S-transferase π homologue GST-1 inhibits dopamine neuron degeneration in a *Caenorhabditis elegans* model of manganism. *Neurotoxicology*, 2013. 38: p. 51–60.
- 54. **Ham, T. J. van** *et al., C. elegans* model identifies genetic modifiers of alpha-synuclein inclusion formation during aging. *PLos Genet,* 2008. 4(3): e1000027.
- 55. **Vu**, **V.** *et al.*, Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. *Cell*, 2015. 162(2): p. 391–402.
- 56. Singh, K. D. *et al.*, Natural Genetic Variation Influences Protein Abundances in *C. elegans* Developmental Signalling Pathways. *PLoS One*, 2016. 11(3): e0149418.
- Chow, C.Y., M.F. Wolfner, and A.G. Clark, Using natural variation in Drosophila to discover previously unknown endoplasmic reticulum stress genes. *Proc Natl Acad Sci U S A*, 2013. 110(22): p. 9013–9018.
- 58. Hodgkin, J. and T. Doniach, Natural variation and copulatory plug formation in Caenorhabditis elegans. *Genetics*, 1997. 146(1): p. 149–164.
- 59. Kammenga, J. E. *et al.*, Beyond induced mutants: using worms to study natural variation in genetic pathways. *Trends Genet*, 2008. 24(4): p. 178–185.
- 60. Volkers, R. J. M. *et al.*, Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations. *BMC Biol*, 2013. 11: p. 93.
- 61. Andersen, E. C. *et al.*, A Powerful New Quantitative Genetics Platform, Combining *Caenorhabditis elegans* High-Throughput Fitness Assays with a Large Collection of Recombinant Strains. *G3*, 2015. 5(5): p. 911–920.
- 62. **Stastna, J. J.** *et al.*, Genotype-dependent lifespan effects in peptone deprived Caenorhabditis elegans. *Sci Rep*, 2015. 5: p. 16259.

7

- Kamkina, P. et al., Natural Genetic Variation Differentially Affects the Proteome and Transcriptome in Caenorhabditis elegans. *Molecular & Cellular Proteomics*, 2016. 15(5): p. 1670–1680.
- 64. **Phillips, P.C.**, Epistasis-the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*, 2008. 9(11): p. 855–867.
- 65. **Thompson, O. A.** *et al.*, Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. *Genetics*, 2015. 200(3): p. 975–989.
- 66. **Balla, K. M.** *et al.*, A wild *C. elegans* strain has enhanced epithelial immunity to a natural microsporidian parasite. *PLoS Pathog*, 2015. 11(2): e1004583.
- 67. McGrath, P. T. *et al.*, Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. *Nature*, 2011. 477(7364): p. 321–325.
- Petersen, C., P. Dirksen, and H. Schulenburg, Why we need more ecology for genetic models such as C. elegans. *Trends Genet*, 2015. 31(3): p. 120–127.
- 69. **Gidalevitz**, **T**. *et al.*, Natural genetic variation determines susceptibility to aggregation or toxicity in a *C*. *elegans* model for polyglutamine disease. *BMC Biol*, 2013. 11: p. 100.
- 70. Li, S. *et al.*, A map of the interactome network of the metazoan C. elegans. *Science*, 2004. 303(5657): p. 540–543.
- 71. **Rual**, J.-F. *et al.*, Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 2005. 437(7062): p. 1173–1178.
- 72. Lejeune, F.-X. *et al.*, Large-scale functional RNAi screen in *C. elegans* identifies genes that regulate the dysfunction of mutant polyglutamine neurons. *BMC Genomics*, 2012. 13: p. 91.
- 73. Kim, E.K. and E.-J. Choi, Pathological roles of MAPK signaling pathways in human diseases. *Biochim Biophys Acta*, 2010. 1802(4): p. 396–405.
- 74. Schmid, T. *et al.*, Systemic Regulation of RAS/MAPK Signaling by the Serotonin Metabolite 5-HIAA. *PLos Genet*, 2015. 11(5): e1005236.
- 75. **Hohman, T. J.** *et al.*, Genetic variation modifies risk for neurodegeneration based on biomarker status. *Front Aging Neurosci*, 2014. 6: p. 183.
- 76. **Driver, J. A.**, Inverse association between cancer and neurodegenerative disease: review of the epidemiologic and biological evidence. *Biogerontology*, 2014. 15(6): p. 547–557.
- 77. **Ramanan, V. K. and A. J. Saykin**, Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders. *American Journal of Neurodegenerative Disease*, 2013. 2(3): p. 145–175.
- 78. Sanchez-Mut, J. V. *et al.*, Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. *Transl Psychiatry*, 2016. 6: e718.
- 79. **Seshadri, S.** *et al.*, Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*, 2010. 303(18): p. 1832–1840.
- Nalls, M. A. et al., Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet, 2014. 46(9): p. 989–993.
- Siitonen, A. et al., Genetics of early-onset Parkinson's disease in Finland: exome sequencing and genome-wide association study. *Neurobiol Aging*, 2017. 53: 195.e7–195.e10.
- N'Songo, A. et al., Comprehensive Screening for Disease Risk Variants in Early-Onset Alzheimer's Disease Genes in African Americans Identifies Novel PSEN Variants. J Alzheimers Dis, 2017. 56(4): p. 1215–1222.
- 83. Velde, K. J. van der *et al.*, WormQTLHD–a web database for linking human disease to natural variation data in C. elegans. *Nucleic Acids Res*, 2014. 42: p. D794–D801.
- 84. Swertz, M. A. *et al.*, XGAP: a uniform and extensible data model and software platform for genotype and phenotype experiments. *Genome Biol*, 2010. 11(3): R27.
- 85. **Paaby, A. B.** *et al.,* Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *eLife*, 2015. 4.

- 86. **Gaertner, B.E. and P.C. Phillips**, *Caenorhabditis elegans* as a platform for molecular quantitative genetics and the systems biology of natural variation. *Genetics Research*, 2010. 92: p. 331–348.
- 87. **Gaertner, B. E.** *et al.*, More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in Caenorhabditis elegans. *Genetics*, 2012. 192(4): p. 1533–1542.

Chapter 3

Genetic background modifies phenotypic and transcriptional responses in a C. elegans model of α -synuclein toxicity

> Yiru A. Wang L. Basten Snoek Mark G. Sterken Joost A.G. Riksen Jana J. Stastna Jan E. Kammenga Simon C. Harvey

This chapter has been published as:

Y. A. Wang *et al.*, Genetic background modifies phenotypic and transcriptional responses in a *C. elegans* model of α -synuclein toxicity. *BMC Genomics*, 2019. 20(1): p. 232

Abstract

Background

Accumulation of protein aggregates are a major hallmark of progressive neurodegenerative disorders such as Parkinson's disease and Alzheimer's disease. Transgenic *Caenorhabditis elegans* nematodes expressing the human synaptic protein α -synuclein in body wall muscle show inclusions of aggregated protein, which affects similar genetic pathways as in humans. It is not however known how the effects of α -synuclein expression in *C. elegans* differs among genetic backgrounds. Here, we compared gene expression patterns and investigated the phenotypic consequences of transgenic α -synuclein expression in five different *C. elegans* genetic backgrounds.

Results

Transcriptome analysis indicates that α -synuclein expression effects pathways associated with nutrient storage, lipid transportation and ion exchange and that effects vary depending on the genetic background. These gene expression changes predict that a range of phenotypes will be affected by α -synuclein expression. We confirm this, showing that α -synuclein expression delayed development, reduced lifespan, increased rate of matricidal hatching, and slows pharyngeal pumping. Critically, these phenotypic effects depend on the genetic background and coincide with the core changes in gene expression.

Conclusions

Together, our results show genotype-specific effects and core alterations in both gene expression and in phenotype in response to α -synuclein expression. We conclude that the effects of α -synuclein expression are substantially modified by the genetic background, illustrating that genetic background needs to be considered in *C. elegans* models of neurodegenerative disease.

Keywords: Natural variation, Gene expression profile, Protein aggregation, α-synuclein, Genetic background, *Caenorhabditis elegans*.

Background

Dementia is a growing global problem affecting large numbers of people. The most common causes of dementia are associated with neurodegeneration, such as that resulting from Alzheimer's disease (AD) and Parkinson's disease (PD) [2–4]. These neurodegenerative dementias share neuroanatomical and biochemical similarities and result from protein misfolding. A major difficulty in determining the mechanisms that produce neurodegenerative dementia is that the underlying cellular-level pathology varies greatly among patients (reviewed in [5, 6]. These differences in pathology allow for genome-wide association studies (GWAS) of PD patients, identifying genetic risk-variants and demonstrate the complex genetic architecture of neurodegenerative dementia [7–9].

Model organisms, such as the nematode *Caenorhabditis elegans*, are of great value for studying neurodegenerative diseases [10–14]. This is due to their experimental tractability and to the broad and general conservation of genetic pathways across species. Importantly, *C. elegans* allows for sophisticated genome-wide genetic screens, which are much less complex than similar studies on mammals. In *C. elegans*, most of these screens have sought to identify loci that represent potential candidate diagnostic and therapeutic targets, or to understand the underlying pathological processes. For example, analysis of transgenic *C. elegans* expressing the human A β peptide – the main component of the amyloid plaques found in AD – linked toxicity to insulin/insulin-like growth factor (IGF), dietary restriction and the heat shock response [10, 11, 13, 15–17]. Similarly for PD, analysis of worms expressing a-synuclein identified associations with ageing and insulin-like signalling [18] and with autophagy and lysosomal function [19].

Most *C. elegans* studies are however limited to the canonical Bristol N2 genetic background, and therefore do not provide insight into how genetic variation between individuals might affect disease-associated traits. This is a major issue as it is clear, both specifically in *C. elegans* (*e.g.* [20–23] and more generally in other systems [24–26], that the phenotype of any given mutation, transgene or allele can vary depending on the genetic background [27]. For *C. elegans*, natural genetic variation is known to result in extensive phenotypic variation (*e.g.* [28, 29] and see [30, 31] for reviews for older studies) and differentially affects both the proteome [32, 33] and transcriptome [23, 32, 34–38]. Therefore, only studying mutational effects in a single genetic background biases our understanding of disease phenotypes, and represents a missed opportunity to elucidate disease mechanisms.

Exemplifying this pattern of a reliance on a single genetic background, only one study has, to date, looked at protein misfolding disease in multiple wild isolate genetic backgrounds of *C. elegans* [39]. Crucially, this showed that natural genetic variation can uncouple different phenotypic effects of polyglutamine (polyQ40) expression [39]. This strongly suggests that similar important variation between genetic backgrounds will be found for other protein misfolding diseases. Given that for protein misfolding diseases the nature of the modifying alleles segregating within human populations remain largely elusive [40], the experimental tractability of *C. elegans* in combination with genetic variation makes the species an excellent system in which to address this issue.

In *C. elegans*, expression of an α -synuclein and yellow fluorescent protein (YFP) fusion in the body wall muscle results in an age-dependent accumulation of inclusions [18]. These inclusions of α -synuclein form aggregates in aging worms that are similar to the pathological inclusions seen in humans with PD [18]. This therefore represents an appropriate model of α -synuclein toxicity [41]. To investigate the effect of genetic background on the consequences of α -synuclein expression, we have created introgression lines (ILs) containing this α -synuclein and YFP transgene in the background of four wild isolates of *C. elegans*. Our analyses of these new ILs, and of α -synuclein in an N2 genetic background, identify both general and genotype-specific changes in gene expression. These changes predict a range of phenotypic effects that we then experimentally confirm. Importantly, given the reliance on N2 in C. elegans research, we show both that some effects are N2-specific and that for other phenotypes the analysis of other genetic backgrounds uncovers substantial variation not seen in N2. Our approach therefore evaluates how genetic background conditions transgene effect(s), and specifically illustrates how the consequences of ectopic overexpression of α -synuclein depends on genetic background.

Results

Introgression line construction and validation

We introgressed the *pkIs2386* transgene [*unc-54*p:: α -synuclein::YFP + *unc-119*(+)] from NL5901 [18], which has an N2 genetic background, into the genetically divergent wild isolates JU1511, JU1926, JU1931, and JU1941 (see [37] for information on genetic distance between these lines). The *pkIs2386* transgene results in α -synuclein expression in the body wall musculature and the vulval muscles. After back-crossing and selfing, four new ILs were obtained, SCH1511,

SCH1926, SCH1931, and SCH1941, with for example, SCH1511 containing the transgene in a JU1511 background. In combination with N2 and NL5901 as controls, we were therefore able to investigate the phenotypic and genomic effects of α -synuclein in five genetic backgrounds (N2, JU1511, JU1926, JU1931, and JU1941). An additional IL, SCH4856, was subsequently created in which the transgene has been introgressed into a CB4856 background. SCH4856 was used to test for the effects of N2 alleles (see below).

We sought to identify the site of the introgression and to determine how much of the N2 genome surrounding the transgene had also been introgressed into the wild isolates. PCR-based genotyping located the transgene in chromosome IV, indicating that the homozygous introgressions in the new ILs spanned between 4.2 and 13.2Mb of chromosome IV (Table S1). By using the set of genetic markers from [37], we identified significantly differential expressed genes on chromosome IV of the α -synuclein lines, and also identified an additional introgression on chromosome V in SCH1931 (Figure S1). PCR-based genotyping together with the transcriptomic analysis indicated a consistent genomic location for the α -synuclein transgene in each of the genetic backgrounds, but did not allow detection of the precise introgression boundaries.

Transcriptome effects of the α -synuclein introgression depend on the genetic background.

The effects of α -synuclein expression on gene expression was measured in five different genetic backgrounds, with three biological replicates per genotype (see Additional file 1 and ArrayExpress accession E-MTAB-6960 for details). Analysis of genome-wide transcriptional changes in the α -synuclein expressing worms indicated that there were differences in developmental rate between the ILs. Variation between ILs in developmental rate was also observed during IL construction. We therefore estimated the age of the samples by their gene expression profile (as in [42] and [43]) and used principal component analysis (PCA) on genome-wide expression levels to investigate differences between isolates and the effects of α -synuclein. This analysis revealed separation between the wild isolates and their corresponding transgenic ILs (Figure 1A). Although genotypes were more scattered on these first two PCA axes, this indicated that developmental delay and reduced lifespan were associated with the introgressed α -synuclein transgene (Figure 1A). A large part of the global gene expression differences are therefore explained by slower development in the α -synuclein lines. Hence, many of the genes identified as responding to α -synuclein in the different backgrounds were likely to be associated with their differential age. Using a threshold of $-\log_10(p) > 3.4$ (FDR <0.05) we found 3521 genes affected by Age, 1509 affected by Genotype, 646 by α -synuclein and 428 by an interaction between Genotype and α -synuclein (Figure 1B, Table S2). Genes affected by α -synuclein or the interaction between α -synuclein and age were most often also affected by genotype. The overlap among these four factors yielded 78 genes highly specific for α -synuclein relative to both Age and Genotype (Table S2). This showed that both genotype-specific and universal (genotype independent) changes in gene expression were induced by α -synuclein expression, *i.e.* we identified a core set of genes of which the expression was altered in all genetic backgrounds and others that were genotype-specific. Strikingly, many genes were not expressed and/or regulated in the same way in wild isolate genetic backgrounds compared with N2. These data therefore indicate that the transcriptome effects of the α -synuclein introgression depend on the genetic background.

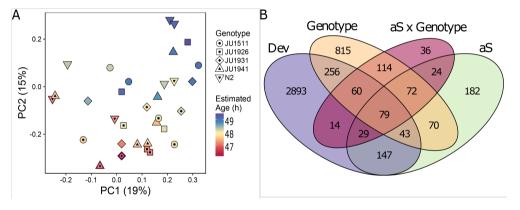


Figure 1: Gene expression varies with genetic background in *C. elegans* ILs constitutively expressing α -synuclein. A) Principal component analysis on gene expression differences, with genotypes represented by shape and α -synuclein (aS) expressing ILs indicated with a black point. Estimated age difference is shown by the colour gradient. Here, developmental age was determined based on the expression levels of class I age responsive genes from [42], which show a linear increase in expression during development between 46 and 54 hours. This indicated a maximum developmental difference of a 2.5 hours between genotypes. B) Venn diagram of differentially expressed genes that are specific to aS, Genotype, Age and aS x Genotype including both induced and repressed genes.

Gene expression enrichment analysis

Gene ontology analysis (GO) of differentially expressed genes revealed the molecular, cellular, and biological processes affected by development, α -synuclein and genetic background (Table 1, with full results in Table S2). Genes that change expression in response to genetic background were enriched for genes involved in the innate immune response and oxidation-reduction process (Table S2) as was previously found by [37]. Widespread changes in genes related to muscle function were observed (Table 1), an expected response given the changes in cellular environment induced by expression of α -synuclein in the body wall muscle. This analysis also identified changes in genes involved in pharyngeal pumping (Table 1).

As expected there were also changes in various pathways associated with cellular stress responses (Table 1), and protein homeostasis (Table S2). Taking the effect of both α -synuclein and age into consideration, genes associated with metabolic processes, transporters of ions and lipids, and kinase activities for ATP were enriched (Table S2).

Phenotypic effects of α -synuclein expression vary among genetic backgrounds

Expression of α -synuclein slows development in some genetic backgrounds

Gene expression analysis indicated that α -synuclein lines were developmentally delayed. To test this directly, we scored development time, *i.e.* the time to the first appearance of eggs. Analysis of these data indicated that development was affected by α -synuclein (aS, p <2e-8), genetic background (Genotype, p <2e-8) and the interaction between α -synuclein and genetic background (aS x Genotype, p <0.0002). In most α -synuclein expressing lines development was delayed compared to the corresponding wild isolate (Figure 2A), with SCH1941 and SCH1926 showing significantly longer larval development time periods than the corresponding wild type controls (p <1e-6, and p <0.008 for SCH1941 and SCH1926, respectively) (Figure 2A).

As found in the phenotypic assay, the transcriptomic samples of the α -synuclein lines were estimated to be younger than the corresponding wild isolate and so showed delayed development (Figure 2B, ANOVA model Estimated_Age \sim aS*Genotype; aS *p* <6e-5), which was independent of the genetic background (Genotype, *p* = 0.72) and of the interaction between α -synuclein and genetic background (aS x Genotype, *p* = 0.84). Comparing the WT and α -synuclein lines for each individual line showed that SCH1931 and JU1931 again displayed the smallest developmental difference as found in the phenotypic assay. Overall, we therefore conclude that the α -synuclein introgression had a genotype-specific impact, differentially decreasing the developmental rate in each genetic background.

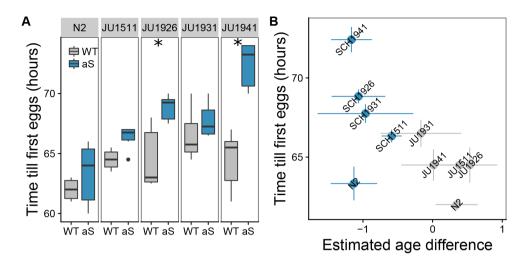
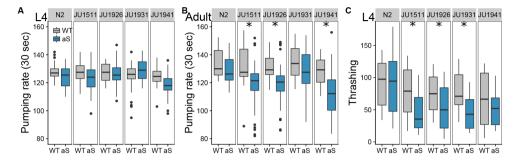


Figure 2: Development time varies with genetic background in *C. elegans* ILs constitutively expressing α -synuclein. A) Boxplot of development measured by time until first egg-laying showing developmental delay of α -synuclein expressing lines and variation between genotypes (ANOVA: aS effect *p* <2e-8; Genotype *p* <2e-8; aS x Genotype Interaction *p* <0.0002; Tuckey HSD: N2 *p* = 0.96; JU1511 *p* = 0.88; JU1926 *p* <0.008; JU1931 *p* = 0.98; JU1941 *p* <1e-6, with significant differences between genotypes indicated by the stars). B) Scatterplot of the relation between estimated age at transcriptomics sampling and the development time. Estimated age difference was determined based on the expression levels of class I age responsive genes from [42], which show a linear increase in expression during development between 46 and 54 hours. The points represent the mean and the lines indicate the standard error.

Expression of α -synuclein decreases pharyngeal pumping rate in some genetic backgrounds

Given the observation of differential expression of genes with a function in pharyngeal pumping (Table 1), we measured pharyngeal pumping rate in all lines 48 and 72 hours after recovery from L1 arrest (Figure 3). No significant differences were seen at 48 hours, but 72 hours after recovery from L1 arrest pharyngeal pumping was affected by α -synuclein (aS, p <3e-16), by the genetic background (Genotype, p <3e-10) and by the interaction of α -synuclein with genetic background (aS x Genotype, p <0.004; Figure 3B). At this point, the pharyngeal pumping rate at 72 hours in the α -synuclein expressing lines had slowed down compared to their corresponding wild isolate (Figure 3B; N2 p = 0.93; JU1511 p <5e-4; JU1926 p <2e-5; JU1931 p = 0.54; JU1941 p <1e-8). These data therefore indicated that the presence of an α -synuclein introgression had a noticeable genotype-specific impact on pharyngeal pumping rate. As pumping rate primarily reflects pharyngeal muscle activity, a tissue in which the α -synuclein transgene is not expressed, these changes reflect an indirect



systemic effect of α -synuclein expression.

Figure 3: Muscle function varies with genetic background in *C. elegans* ILs constitutively expressing α -synuclein. Boxplots of the number of pharyngeal pumps over a 30s period in worms at **A**) 48 and **B**) 72 hours after recovery from L1 arrest. Black dots represent outliers (two times interquartile range). Analysis shows no significant effects at 48 hours. At 72 hours however, pharyngeal pumping is affected by α -synuclein expression, by the genetic background and by the interaction of α -synuclein expression with the genetic background (ANOVA: aS, *p* <3e-16, Genotype, *p* <3e-10, aS x Genotype, *p* <0.004; Tucky HSD: N2 *p* = 0.93; JU1511 *p* <5e-4; JU1926 *p* <2e-5; JU1931 *p* = 0.54; JU1941 *p* <1e-8). C) Boxplots of the activity rates of worms in liquid, with analysis indicating that locomotion pumping is affected by α -synuclein expression with the genetic background (ANOVA: aS *p* <1.953e-12, Genotype *p* <4.4e-13, aS x Genotype *p* = 0.0040; Tucky HSD: N2 *p* = 1.00; JU1511 *p* = 2.4e-6; JU1926 *p* = 0.0016; JU1931 *p* = 1.3e-3; JU1941 *p* = 0.51). Significant differences between genotypes are indicated by stars.

Expression of α -synuclein decreases movement in some genetic backgrounds

Given the site of α -synuclein expression and the detection of changes in the expression of genes involved in muscle structure and function (Table 1) we tested for differences between lines in movement. Here, we assayed movement of worms in liquid (Figure 3C). These data indicate that N2 and NL5901 show similar movement rates and that the α -synuclein expressing lines expression show reduced activity in comparison to their corresponding wild isolate. These data therefore indicate that the presence of an α -synuclein introgression produces a genotype-specific impact, likely a direct consequence of α -synuclein expression in the body wall muscles.

	•		0		5	
GO_ID	Description	All	α-synuclei	nType	<i>p</i> -Value	Adjusted
		genes	Genes			р
GO:0005576	extracellular region	329	54	CC	3.99E-06	8.29E-04
GO:0030246	carbohydrate binding	288	48	MF	8.12E-06	1.13E-03
GO:0005861	troponin complex	8	5	CC	1.21E-05	1.26E-03
GO:0055120	striated muscle dense body	93	21	CC	1.77E-05	1.48E-03
GO:0045087	innate immune response	300	47	BP	5.16E-05	3.06E-03
GO:0010332	response to gamma radiation	10	5	BP	7.77E-05	3.59E-03
GO:0006869	lipid transport	26	8	BP	2.74E-04	1.12E-02
GO:0098869	cellular oxidant	52	12	BP	5.12E-04	1.29E-02
	detoxification					
GO:0030017	sarcomere	22	7	CC	4.10E-04	1.29E-02
GO:0006936	muscle contraction	17	6	BP	3.93E-04	1.29E-02
GO:0016810	hydrolase activity, acting	13	5	MF	5.02E-04	1.29E-02
	on carbon-nitrogen (but not peptide) bonds					
GO:0071949	FAD binding	18	6	MF	5.94E-04	1.37E-02
GO:0008652	cellular amino acid	20	6	BP	1.23E-03	2.57E-02
	biosynthetic process					
GO:0043050	pharyngeal pumping	16	5	BP	1.85E-03	2.96E-02
GO:0016787	hydrolase activity	818	96	MF	2.24E-03	3.45E-02
GO:0016311	dephosphorylation	63	12	BP	3.36E-03	3.49E-02
GO:0030170	pyridoxal phosphate	48	10	MF	2.88E-03	3.49E-02
	binding					
GO:0031430	M band	28	7	CC	2.46E-03	3.49E-02
GO:0003993	acid phosphatase activity	28	7	MF	2.46E-03	3.49E-02
GO:0099132	ATP hydrolysis coupled	23	6	BP	3.08E-03	3.49E-02
	cation transmembrane					
	transport					

Table 1: The top-20 enriched GO-terms with 5 or more genes with an α -synuclein effect.

CC: cellular component; MF: molecular function; BP: biological process.

Expression of α -synuclein decreases lifespan in some genetic backgrounds

Given the age-related changes in pharyngeal pumping rate and enrichment of genes associated with aging and stress response pathways in response to α -synuclein expression, we hypothesised that the introgressed α -synuclein transgene could affect longevity. Therefore, we measured lifespan in all lines. Comparison of N2 and NL5901 (α -synuclein in an N2 background) showed that lifespan was not affected by α -synuclein expression in the N2 genetic background (Figure 4; N2 log-rank p = 0.14, mean age p = 0.99). However, in the other genetic backgrounds, lines containing the α -synuclein introgression displayed significantly accelerated death (Figure 4A) and a shortened lifespan (Figure 4B), a global effect of α -synuclein expression on animal physiology and function. These data also indicated that the α -synuclein introgression in the wild isolate genetic backgrounds resulted in increased rates of maternal hatching (bagging) (Table S3, Figure 5). Given that the α -synuclein transgene is expressed in the vulval muscles this result mirrors the direct effect of α -synuclein expression on muscle function seen for thrashing (Figure 3C). These observations indicate that α -synuclein expression lowers lifespan in some genotypes, but not others.

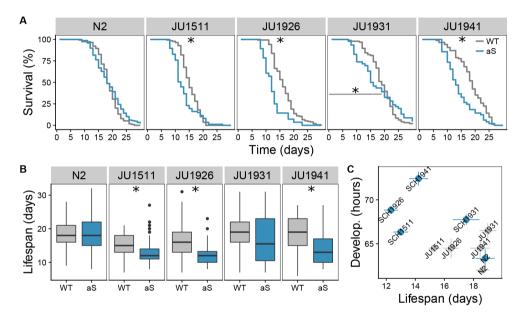


Figure 4: *C. elegans* ILs constitutively expressing α -synuclein exhibit significantly shorter lifespan. A) Survival curves for wild-type (WT) and transgenic nematodes expressing α -synuclein (aS) at 20°C (log-rank test ('survival' package, R): N2 p = 0.14; JU1511 p <0.003; JU1926 p < 4*1e-7; JU1931 p = 0.79 (first 20 days p < 0.0004); JU1941 p < 7e-5, with significant differences between genotypes indicated by stars). B)Box-plot of the lifespan for WT and aS lines. Black dots denote outliers (two times interquartile range), with analysis indicating that lifespan is affected by α -synuclein expression, by the genetic background and by the interaction of α -synuclein expression with the genetic background (ANOVA: as p < 3e-5; Genotype p < 1e-16; aS x Genotype p < 0.0002; Tucky HSD: N2 p = 0.99; JU1511 p < 0.008; JU1926 p < 0.0004; JU1931 p = 0.71; JU1941 p < 1e-5, with significant differences between genotypes indicated by stars). C) Relationship between lifespan and development time. The points represent the mean and the lines indicate the standard error.

Chapter 3

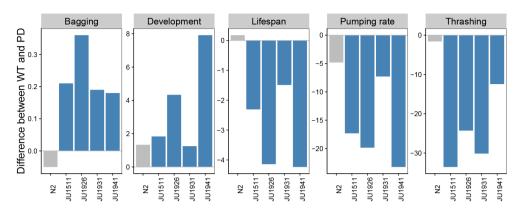


Figure 5: Phenotypic effects of α **-synuclein expression across different genetic backgrounds.** Shown are differences between wild type and α -synuclein expressing lines in bagging (% difference in matricidal hatching), in development (difference in hours from L1 arrest to first egg lay), in lifespan (difference in mean lifespan in days), in pumping rate (difference in number of pumps recorded over a 30s period) and in trashing (activity counts in liquid recorded over a 30m period).

Genetic background affects the outcome of α -synuclein expression

Comparison of phenotypic effects across the genetic backgrounds tested suggests that the phenotypes associated with the α -synuclein transgene introgression vary between genotypes in a consistent manner (Figure 5). Multiple phenotypes were affected by α -synuclein in the JU1511, JU1926, and JU1941 background, whereas they were less affected in the JU1931 and N2 backgrounds. This suggests a shared genetic component across many of the phenotypes we have observed, but does also indicate that there is genotype-specific variation, *e.g.* no pair of lines shows the same pattern of phenotypic effects (Figure 5). These data also indicate that there are both local effects – *e.g.* bagging and trashing, where effects can be directly attributed to expression of α -synuclein in the muscles – and global effects on the whole animal.

As alleles from different *C. elegans* backgrounds can produce a range of synthetic deleterious effects resulting in full or partially genomic incompatibilities between their genomes [44, 45], we sought to test if the N2 region alone replicated the phenotypes we observed here. We therefore introgressed the *pkIs2386* transgene into a CB4856 genetic background, generating the SCH4856 line and undertook comparisons of this line with CB4856 and CBN93, a line with an introgression of the N2 genome spanning the 3.3-12.8Mbp region on chromosome IV in an CB4856 background (Table S1). This control was

undertaken as deleterious interactions between alleles from N2 and those from CB4856 are well characterized [44, 45]. Comparisons of pumping rate and of development time between these lines indicated that the expression of α -synuclein in SCH4856 produced effects not seen in CBN93 (Figure 6) and that these effects mirror those seen in the other genetic backgrounds. This provides strong support for the view that the phenotypic effects we observe were a consequence of α -synuclein expression and not the introgression of the N2 region.

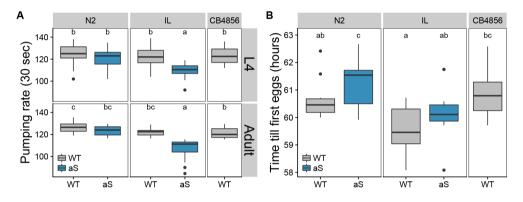


Figure 6: Phenotypic effects observed in a CB4856 genetic background depend on α -synuclein expression rather than N2 alleles. Here, comparisons are made between N2 (N2 WT and AS) and CB4856 (CB4856 and IL AS) genetic backgrounds with and without the α -synuclein introgression and with CBN93 (IL WT), a line with an introgression of the N2 genome spanning the 3.3-12.8Mbp region on chromosome IV in an CB4856 background. A) A box-plot of the average pumping rate in 30s of worms 48 and 72 hours after recovery after L1 arrest, and B)Time from recovery of L1 arrest to first egg lay. The black dots represent outliers (outside two times the interquartile range). Letter codes denote lines that differ significantly (p < 0.05).

Discussion

We have introgressed a transgene that results in the expression of α -synuclein in the body wall and vulval muscles into different genetic backgrounds of *C. elegans*. Analysis of these newly created α -synuclein ILs indicates that genetic background effects both the response of the animals to α -synuclein expression at the level of gene expression and in terms of the phenotypic consequences. Further, this can be seen at both a local and global level, with effects seen in traits associated with muscles in which α -synuclein is expressed (bagging and thrashing) and traits that represent a global, whole organism, effects (pumping, development rate and lifespan). Our gene expression analysis identifies a range of changes associated with α -synuclein expression. Many of these – particularly those involved in muscle function and in various stress responses – matched expectations given the site of expression and the known effects of α -synuclein aggregation on cellular function. We also identify changes in lipid metabolism (Tables 1 and S2), an important result given that this has been previously linked to α -synuclein pathology in a yeast model [46] and that there is evidence of a direct association between α -synuclein and lipid droplets [47]. Changes in lipid levels of *C. elegans* expressing α -synuclein in the muscles have previously been observed [48], our results therefore provide a set of candidate genes to investigate how α -synuclein expression may mediate lipid metabolism. We also show that such effects can be expected to vary across *C. elegans* genetic backgrounds.

On a phenotypic level we found that, in comparison to the corresponding wild isolates, the α -synuclein expressing lines: developed more slowly; suffered an increased rate of decline in pharyngeal pumping; exhibited a shorter lifespan; and died from matricidal hatching of eggs at an increased rate. Similar phenotypic effects are observed in other *C. elegans* models of protein misfolding disease. For example, transgenic worms expressing polyQ proteins and A β both show a significantly shortened lifespan [39, 49]. A reduced lifespan could therefore be a general toxicity phenotype, providing an indirect measure of organismal dysfunction caused by misfolded protein aggregations in the body wall.

Many of the phenotypic changes we see are either only found in wild isolate genetic backgrounds or are more pronounced in these backgrounds (Figure 5). This demonstrates that there is variation between isolates that appears to differently affect local and global consequences of α -synuclein expression. Previously, only a single *C. elegans* study has directly investigated the effect of genetic background in the context of protein misfolding disease. That study found complex variation in polyglutamine (polyQ40) aggregation and toxicity in three wild isolate backgrounds and in a panel of 21 recombinant inbred lines (RILs) [39]. The RIL analysis also showed that the various effects of polyQ40 expression could be uncoupled [39]. In this study, we also show some evidence that natural genetic variation in *C. elegans* can uncouple different effects of α -synuclein expression (Figure 5), suggesting that this may be a general pattern.

A caveat with this interpretation is however that a range of incompatibilities between alleles from different *C. elegans* genomes have been identified [44, 45] and that the region surrounding the site of the α -synuclein transgene contains a number of mapped quantitative trait loci affecting various life history traits.

For example, ILs with introgressed regions of chromosome IV from CB4856 in an N2 genetic background have identified QTLs affecting lifespan and pumping rate that partially overlap our introgression [50]. Similarly, a complex interaction between alleles from the N2 and CB4856 genomes has been found on chromosome IV [45]. However, little is known about the frequency of such synthetic deleterious effects between other genotypes of *C. elegans*. Proteomic analysis of age-dependent changes in protein solubility by [51] also indicates that the genes encoding proteins that become insoluble with age are enriched for modifiers of lifespan. Additionally, a heat-stress specific QTL for recovery [52] as well as an eQTL-hotspot have been identified on the left arm of chromosome IV (IV: 1.0–2.5 Mb) [38], while a QTL for maternal hatching rate is at the position of the introgression (chromosome IV \sim 6M) [53]. That such effects might alter disease-related processes complicates the interpretation of both our results here and the previous work of Gidalevitz et al. (2013) [39]. To experimentally address this caveat, we introgressed the α -synuclein transgene into a CB4856 genetic background and undertook comparisons with a line that contains a comparable region of the N2 genome, but no transgene, in a CB4856 background. These comparisons indicated that the N2 introgression alone does not recapitulate the phenotypic effects seen in the α -synuclein expressing line (Figure 6). As these comparisons again show that the expression of α -synuclein in a CB4856 genetic background results in different and more severe effects than those seen in an N2 background, this strongly supports the view that genetic background needs to be considered more generally when *C. elegans* is used as a disease model. A move to the use of more defined modifications in a range of genetic backgrounds that limited the introduction of other alleles – perhaps via the use of CRISPR to introduce the relevant transgenes – would therefore be ideal. It would also be of interest to determine the extent to which genotype-specific were observed in other models, for example where α -synuclein is expressed in neurons [46]. As there is extensive variation between isolates of *C. elegans* in lifespan and in various stress responses and in lifespan, and that variation has now been identified for two protein misfolding diseases such a development would facilitate the systematic analysis of the role of natural variation in such diseases.

Methods

C. elegans maintenance and growth conditions

Worms were maintained at 20°C on nematode growth medium (NGM) plates seeded with *Escherichia coli* OP50 [54] and all assays were undertaken at 20°C. Assays were initiated using eggs isolated from gravid adults treated with sodium hypochlorite and NaOH [55]. For the microarray analysis of gene expression and for the pharyngeal pumping assays, these synchronized eggs were allowed to hatch on NGM plates seeded with *E. coli* overnight at 20°C, i.e. these worms were not arrested at the L1 stage. After 48 hours, worms in all strains had reached the L4 stage. For the lifespan, development, and thrashing activity assays, the synchronized eggs were hatched on NGM plates without E.coli, in order to obtain synchronized L1 larvae. When they reach L4 stage, these assay were then measured or setup will all lines.

C. elegans lines

The *C. elegans* wild types JU1511, JU1926, JU1931, JU1941, CB4856 (Hawaii), canonical strain N2 (Bristol), and the transgenic strain NL5901, which contains the *pkIs2386* transgene [*unc-54p*:: α -synuclein::YFP + *unc-119*(+)], were obtained from the *Caenorhabditis* Genetics Center. To introgress the α -synuclein::YFP transgene into the wild backgrounds, NL5901 hermaphrodites were mated with wild type males, then five to ten F1 fluorescent males were mated with wild type hermaphrodites. Fluorescent males were then isolated in the F2 progeny of this cross and used to backcross to the relevant wild type. This backcrossing to the wild type background was repeated until generation F7 was reached. At this point, worms were allowed to self-fertilise to obtain homozygotes lines and then cryopreserved. This crossing design was used to produce the strains SCH1511, SCH1926, SCH1931, SCH1941, and SCH4856, with for example, SCH1511 containing the transgene in a JU1511 background.

Five genetic backgrounds – N2, JU1511, JU1926, JU1931, and JU1941 – were used to investigate the phenotypic and genomic effects of α -synuclein expression. SCH4856 was used in separate assays to test the effect of N2 alleles surrounding the transgene integration site. Here, comparisons were made to the CBN093 strain which contains a comparable region of the N2 genome introgressed into CB4856, but no transgene. CBN093 was constructed by back-crossing WN071 [56] with CB4856. The strain was back-crossed followed by segregation for 11 generations until a single homozygous region on chromosome IV was obtained. Thereafter, the strain was inbred for 11 generations and the genotype confirmed by sequencing (SRP154243; https://www.ncbi.nlm.nih.gov/sra) using the approach [57, 58].

IL genotyping

PCR-based genotyping: DNA was isolated from the lysates of 5 individual adults from each line, and then was used for genotyping PCRs. Genotyping primers utilized insertions/deletions between the CB4856 and N2 genomes, with 41 primer pairs covering the genome used (see Table S1 for details of primers and marker locations) [57]. PCR was carried out with the GoTaq DNA polymerase kit (Promega) according to the manufacturer's recommendations. The sizes of amplified products were assessed by electrophoresis in 1.5% agarose gels stained with Ethidium Bromide. This allowed the size of the introgressions to be determined, with these regions shown in Table S1.

Hybridization markers: Marker genes found by DNA hybridizations in [37] were used to find the genomic position of the α -synuclein introgression in each genetic background. We tested which genomic region was missing markers for the wildtype background, this region then must be from the NL5901 background.

Sample preparations and RNA microarray analysis

mRNA microarrays: In total three independent replicates of the four JU strains and their corresponding transgenic ILs, as well as N2 and NL5901, were analysed. Worms were assayed at 48 hours from egg isolation (see details above) and all strains were at the L4 stage. Worms were therefore the same chronological age, and at the same larval stage, but this design would allow the detection of effects of genetic background, α -synuclein expression and the interaction of these effects on development. At this point worms were generated by hatching alkaline hypochlorite-purified eggs and then harvested by centrifugation, washed with M9 buffer, frozen in liquid N2, and stored at -80 °C until use. The Maxwell®16 Tissue LEV Total RNA Purification Kit was used for mRNA isolation, following the manufacturer's protocol with a modified lysis step. In addition to the lysis buffer, proteinase K was added and the sample was incubated at 65°C while shaking at 1000 rpm for 10 minutes. Thereafter the standard protocol was followed. PolyA RNA was used to generated Cy3 and Cy5-labeled cRNA samples, which were then hybridized to 4X44K slides V2 (Agilent) C. elegans whole genome GeneChips, processed, and scanned (full microarray data in ArrayExpress with accession E-MTAB-6960). RNA microarray statistical analysis and data processing were performed using the Limma package for the R software environment (htt ps://www.bioconductor.org/packages/release/bioc/html/limma.html). To find the genes affected by α -synuclein, genotype and age, these terms were used as explanatory factors in a linear model (gene expression \sim age + aS * genotype). Significance thresholds were determined by permutations of all spots on the array. In the permutations, the RNA hybridization intensities were randomly distributed over the genotypes and batches. Therefore, the p-value that gave a min ratio of false positives/true positives of 0.05 (= -log10(*p*) >3.4) was set for convenience, *i.e.* an FDR at 0.0186 for α -synuclein and genotype.

Enrichment: Enrichments were done on genes groups divined by their significance from the linear model a $-\log 10(p) > 2$ was used, then a hyper geometric test in R was used to test each GO term for enrichment.

PCA: To partition the variation in gene expression a PCA was done on the transcription profiles (the log2 ratios with the mean) of all samples. The first two axis were used for visualisation.

Age estimation: Age estimation of the worms sampled for transcriptomics was done by comparing the class I age responsive genes from [42], as done by [42, 43, 59]. These genes show a linear increase during development between 46 and 54 hours and can therefore be used to estimate the relative differences in development between samples/populations by their transcriptomes. The α -synuclein expressing SCH lines developed more slowly than the JU lines and none of the lines had started moulting at 48 hours, so there were no adults on the plates.

Phenotypic assays

Lifespan: Nematodes were cultured on OP50 bacteria without fluorouracil deoxyribose (FUdR) from synchronised L1 juveniles until young adults. When setting up the experiment, in total 100 worms per strain were randomly selected from the population and transferred onto 10 plates (*i.e.* each plate contains 10 worms). Then, they were transferred away from their progeny every day during the vigorous reproductive period and every other day during

the reduced reproductive period. From the young adult stage, worms were examined for signs of life daily. Individuals that were not moving or twitching after gentle stimulation, followed by vigorous stimulation, or that did not exhibit pharyngeal pumping for 30 seconds, were considered dead. Individual worms that had died from internal hatching of progeny (bagging), or that had crawled off the plates, were censored from lifespan result but were used for maternal hatching analysis. Two biological replicates were done for lifespan assay. Data analysis was performed in R using the 'survival' package. The rates of bagging from these assays were also compared to determine if this was affected by α -synuclein expression.

Development time: Synchronised L1 juveniles were obtained by allowing eggs to hatch in M9 buffer after three washes in M9 after hypochlorite treatment. Arrested L1s were then transferred to NGM plates, which fed with *E. coli* OP50 and were incubated at 20°C. Tracking observations and inspections were done at regular time intervals. Development time was defined as the period between worm inoculation and the moment at which the first appearance of eggs and the period until the reproductions reach peak level. Three biological replicates were done for this assay.

Pharyngeal pumping: To avoid the effect of short term starvation after inoculation on pharyngeal pumping rate, synchronized eggs, isolated as described above, were allowed to hatch on NGM plates with E.coli (t=0 hours). Individual worms were then isolated and observed at the L4 stage (t=48 hours, day 2) and as young adults (t=72 hours, day 3). The number of contractions in the terminal bulb of pharynx was counted for 30s (n = 19 per genotype) for the L4s, and for 60s for 3 day-old worms (n = 30 per genotype). Three biological replicates were done for this assay.

Thrashing: Synchronized starved L1 juveniles were inoculated on NGM plates with E.coli at 20°C. When N2 worms were observed to have reached the L4 stage, worms were transferred into the wells of a 96-well-plate, where each well contained 50µl of M9 buffer [55]. Three biological replicates were undertaken, with 10 worms per well for 3-5 wells per genotype in each of the replicates. Activity was then measured in an 30 minute period using a WMicroTracker (PhylumTech). Here, worm movements in the wells are detected as they interfere with an array of micro beams of infrared light and the number of interference events is counted – hence a higher score represents more activity.

Statistical analysis: Phenotypic differences between the lines were tested using an ANOVA (model: phenotype~ aS * GB) to test for overall effects and Tukey HSD to test for lines specific differences. Survival curves were tested by log-rank test from the "survival" package in R.

Declarations

Ethics approval: Not required.

Consent for publication: Not applicable.

Availability of data and material: All data generated or analysed during this study are available, either in this published article and its supplementary information files or ArrayExpress (accession E-MTAB-6960).

Competing interests: The authors declare that they have no competing interests.

Funding: This work is supported by a Leverhulme Trust Research Grant (RPG-2016-040) awarded to SH, and by BLS NWO (project no. 823.01.001). Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). These funding bodies had no role in the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript.

Authors' contributions: YW constructed the introgression lines and undertook the phenotypic analyses with assistance from JS. YW and JR performed the microarray. YW, LBS and MGS analysed the data. SH, JK and YW conceived the project. SH, YW, LBS and JK wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements: We thank Yu Nie for comments and assistance with assays. We also thank the CGC for worm strains.

References

- 1. **Wang, Y. A.** *et al.*, Genetic background modifies phenotypic and transcriptional responses in a *C. elegans* model of α-synuclein toxicity. *BMC Genomics*, 2019. 20(1): p. 232.
- 2. **Kivipelto**, **M**. *et al.*, Obesity and vascular risk factors at midlife and the risk of dementia and Alzheimer disease. *Arch Neurol*, 2005. 62(10): p. 1556–1560.
- 3. Aarsland, D. and M. W. Kurz, The epidemiology of dementia associated with Parkinson disease. *J Neurol Sci*, 2010. 289: p. 18–22.
- 4. **Aybek**, **S.** *et al.*, Hippocampal atrophy predicts conversion to dementia after STN-DBS in Parkinson's disease. *Parkinsonism & Related Disorders*, 2009. 15(7): p. 521–524.
- 5. **Seelaar, H.** *et al.*, Distinct genetic forms of frontotemporal dementia. *Neurology*, 2008. 71(16): p. 1220–1226.
- 6. Gratwicke, J., M. Jahanshahi, and T. Foltynie, Parkinson's disease dementia: a neural networks perspective. *Brain*, 2015. 138: p. 1454–1476.
- 7. **Chen**, **Y**. *et al.*, No association between 5 new GWAS-linked loci in Parkinson's disease and multiple system atrophy in a Chinese population. *Neurobiol Aging*, 2018. 67: 202.e7–202.e8.
- Nalls, M. A. et al., Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet, 2014. 46(9): p. 989–993.
- 9. Satake, W. et al., Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet*, 2009. 41(12): p. 1303–1307.
- Teschendorf, D. and C. D. Link, What have worm models told us about the mechanisms of neuronal dysfunction in human neurodegenerative diseases? *Mol Neurodegener*, 2009. 4: p. 38.
- Calahorro, F. and M. Ruiz-Rubio, *Caenorhabditis elegans* as an experimental tool for the study of complex neurological diseases: Parkinson's disease, Alzheimer's disease and autism spectrum disorder. *Invert Neurosci*, 2011. 11(2): p. 73–83.
- 12. Rodriguez, M. et al., Worms under stress: *C. elegans* stress response and its relevance to complex human disease and aging. *Trends Genet*, 2013. 29(6): p. 367–374.
- 13. Sin, O., H. Michels, and E. A. A. Nollen, Genetic screens in *Caenorhabditis elegans* models for neurodegenerative diseases. *Biochim Biophys Acta*, 2014. 1842(10): p. 1951–1959.
- 14. Wang, Y. A., J. E. Kammenga, and S. C. Harvey, Genetic variation in neurodegenerative diseases and its accessibility in the model organism *Caenorhabditis elegans*. *Hum Genomics*, 2017. 11(1): p. 12.
- Fonte, V. et al., Suppression of in vivo beta-amyloid peptide toxicity by overexpression of the HSP-16.2 small chaperone protein. J Biol Chem, 2008. 283(2): p. 784–791.
- Steinkraus, K. A. *et al.*, Dietary restriction suppresses proteotoxicity and enhances longevity by an hsf-1-dependent mechanism in Caenorhabditis elegans. *Aging Cell*, 2008. 7(3): p. 394–404.
- McColl, G. *et al.*, Utility of an improved model of amyloid-beta (Aβ₁₋₄₂) toxicity in *Caenorhabditis elegans* for drug screening for Alzheimer's disease. *Mol Neurodegener*, 2012. 7: p. 57.
- 18. **Ham, T. J. van** *et al., C. elegans* model identifies genetic modifiers of alpha-synuclein inclusion formation during aging. *PLos Genet,* 2008. 4(3): e1000027.
- 19. Hamamichi, S. *et al.*, Hypothesis-based RNAi screening identifies neuroprotective genes in a Parkinson's disease model. *Proc Natl Acad Sci U S A*, 2008. 105(2): p. 728–733.
- 20. **Duveau, F. and M.-A. Félix**, Role of pleiotropy in the evolution of a cryptic developmental variation in Caenorhabditis elegans. *PLoS Biol*, 2012. 10(1): e1001230.
- 21. Elvin, M. et al., A fitness assay for comparing RNAi effects across multiple *C. elegans* genotypes. *BMC Genomics*, 2011. 12: p. 510.

- 22. Schmid, T. et al., Systemic Regulation of RAS/MAPK Signaling by the Serotonin Metabolite 5-HIAA. *PLos Genet*, 2015. 11(5): e1005236.
- 23. Sterken, M. G. *et al.*, Ras/MAPK Modifier Loci Revealed by eQTL in *C. elegans*. *G*3, 2017. 7(9): p. 3185–3193.
- 24. **Coe, T. S.** *et al.*, Genetic variation in strains of zebrafish (Danio rerio) and the implications for ecotoxicology studies. *Ecotoxicology*, 2009. 18(1): p. 144–150.
- 25. **Justice**, M. J. and P. Dhillon, Using the mouse to model human disease: increasing validity and reproducibility. *Disease Models & Mechanisms*, 2016. 9(2): p. 101–103.
- Wade, C. M. and M. J. Daly, Genetic variation in laboratory mice. *Nat Genet*, 2005. 37(11): p. 1175–1180.
- 27. **Kammenga**, **J. E.**, The background puzzle: how identical mutations in the same gene lead to different disease symptoms. *FEBS J*, 2017. 284(20): p. 3362–3373.
- 28. **Reynolds, R. M. and P. C. Phillips**, Natural variation for lifespan and stress response in the nematode Caenorhabditis remanei. *PLoS One*, 2013. 8(4): e58212.
- 29. Martin, N., J. Singh, and A. Aballay, Natural Genetic Variation in the *Caenorhabditis elegans* Response to *Pseudomonas aeruginosa*. *G3*, 2017. 7(4): p. 1137–1147.
- Kammenga, J. E. et al., Beyond induced mutants: using worms to study natural variation in genetic pathways. *Trends Genet*, 2008. 24(4): p. 178–185.
- Gaertner, B. E. and P. C. Phillips, *Caenorhabditis elegans* as a platform for molecular quantitative genetics and the systems biology of natural variation. *Genetics Research*, 2010. 92: p. 331–348.
- Kamkina, P. et al., Natural Genetic Variation Differentially Affects the Proteome and Transcriptome in Caenorhabditis elegans. *Molecular & Cellular Proteomics*, 2016. 15(5): p. 1670–1680.
- 33. Singh, K. D. *et al.*, Natural Genetic Variation Influences Protein Abundances in *C. elegans* Developmental Signalling Pathways. *PLoS One*, 2016. 11(3): e0149418.
- 34. Li, Y. *et al.*, Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans. *Genetics*, 2010. 186(1): p. 405–410.
- 35. Rockman, M. V., S. S. Skrovanek, and L. Kruglyak, Selection at linked sites shapes heritable phenotypic variation in C. elegans. *Science*, 2010. 330(6002): p. 372–376.
- 36. **Viñuela**, **A**. *et al.*, Genome-wide gene expression regulation as a function of genotype and age in C. elegans. *Genome Res*, 2010. 20(7): p. 929–937.
- 37. Volkers, R. J. M. *et al.*, Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations. *BMC Biol*, 2013. 11: p. 93.
- 38. **Snoek, B. L.** *et al.*, Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans. BMC Genomics*, 2017. 18(1): p. 500.
- 39. **Gidalevitz**, **T**. *et al.*, Natural genetic variation determines susceptibility to aggregation or toxicity in a *C*. *elegans* model for polyglutamine disease. *BMC Biol*, 2013. 11: p. 100.
- Kearney, J. A., Genetic modifiers of neurological disease. *Curr Opin Genet Dev*, 2011. 21(3): p. 349–353.
- 41. **Jansen, I. E.** *et al.*, Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome Biol*, 2017. 18(1): p. 22.
- 42. Snoek, L. B. *et al.*, A rapid and massive gene expression shift marking adolescent transition in C. elegans. *Sci Rep*, 2014. 4: p. 3912.
- 43. **Bent**, **M. L. van der** *et al.*, Loss-of-function of β-catenin bar-1 slows development and activates the Wnt pathway in Caenorhabditis elegans. *Sci Rep*, 2014. 4: p. 4926.
- 44. Seidel, H. S., M. V. Rockman, and L. Kruglyak, Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science*, 2008. 319(5863): p. 589–594.

- 45. **Snoek, L. B.** *et al.*, Widespread genomic incompatibilities in Caenorhabditis elegans. *G3* (*Bethesda*, *Md*.) 2014. 4(10): p. 1813–1823.
- 46. **Cooper, A. A.** *et al.*, Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science*, 2006. 313(5785): p. 324–328.
- 47. **Cole**, **N. B.** *et al.*, Lipid droplet binding and oligomerization properties of the Parkinson's disease protein alpha-synuclein. *J Biol Chem*, 2002. 277(8): p. 6344–6352.
- 48. **Jadiya, P.** *et al.*, Anti-Parkinsonian effects of Bacopa monnieri: insights from transgenic and pharmacological *Caenorhabditis elegans* models of Parkinson's disease. *Biochem Biophys Res Commun*, 2011. 413(4): p. 605–610.
- 49. **Fong, S.** *et al.*, Energy crisis precedes global metabolic failure in a novel *Caenorhabditis elegans* Alzheimer Disease model. *Sci Rep*, 2016. 6: p. 33781.
- 50. **Doroszuk, A.** *et al.*, A genome-wide library of CB4856/N2 introgression lines of Caenorhabditis elegans. *Nucleic Acids Res*, 2009. 37(16): e110.
- 51. **Reis-Rodrigues**, **P.** *et al.*, Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan. *Aging Cell*, 2012. 11(1): p. 120–127.
- 52. **Rodriguez**, **M.** *et al.*, Genetic variation for stress-response hormesis in *C. elegans* lifespan. *Exp Gerontol*, 2012. 47(8): p. 581–587.
- 53. **Stastna**, **J. J**. *et al.*, Genotype-dependent lifespan effects in peptone deprived Caenorhabditis elegans. *Sci Rep*, 2015. 5: p. 16259.
- 54. Brenner, S., The genetics of Caenorhabditis elegans. Genetics, 1974. 77(1): p. 71–94.
- 55. **Stiernagle, T.**, WormBook. In: ed. by **D. Fay**. The *C. elegans* Research Community, 2006. Chap. Maintenance of C. elegans.
- 56. Li, Y. *et al.*, Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. *PLos Genet*, 2006. 2(12): e222.
- 57. **Sterken, M. G.**, Building towards a multi-dimensional genetic architecture in Caenorhabditis elegans. PhD thesis. Wageningen University, 2016.
- Gao, A. W. et al., Natural genetic variation in *C. elegans* identified genomic loci controlling metabolite levels. *Genome Res*, 2018. 28(9): p. 1296–1308.
- 59. Jovic, K. *et al.*, Temporal dynamics of gene expression in heat-stressed Caenorhabditis elegans. *PLoS One*, 2017. 12(12): e0189445.

Supplementary material

The supplementary files and figures are deposited at:

https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-5597-1

Additional file 1: Microarray data ("Pvalues_aS_paper.txt").

Table S1: Markers used in the genotyping of the α -synuclein ILs.

Table S2: Gene ontology analysis (GO) of genes identified as affected by Age, Genotype (GB), α -synuclein (aS) and the interaction between genotype and aS, and the details of genes shown in Figure 1B including the p values and average log_2 ratios with the mean per genotype for the 78 genes highly specific for aS relative to both Age and GB.

Table S3: Rates of matricidal hatching in α -synuclein ILs and relevant wild isolates.

Table S4: Lifespan assays in α -synuclein ILs and relevant wild isolates.

Figure S1: Gene expression markers across the genome. Determination of the α -synuclein introgression in the four wild isolate backgrounds. Marker genes with different expression between N2 and the wild isolates (All) were used to detect the α -synuclein introgression and N2 border regions. The marker genes missing in the α -synuclein lines indicate the N2 border regions and position of the α -synuclein introgression (aS Missing). Different genetic backgrounds are indicated by the different colours. The position(s) where all lines have missing markers show the likely α -synuclein locus, the extra missing markers on chromosome V show a possible extra introgression in SCH1931.

 \mathbf{m}

m

Chapter 4

The accumulation of α -synuclein aggregates and associated toxicities in different natural genetic backgrounds of C. elegans

Yiru A. Wang Lisa van Sluijs Yu Nie Mark G. Sterken Simon C. Harvey Jan E. Kammenga

Abstract

The α -synuclein protein encoded by the human SNCA gene is associated with Parkinson's disease (PD) pathogenesis. The oligomeric complexes of α -synuclein that accumulate with advancing age mediate disruption of cellular homeostasis, resulting in neuronal death. Polymorphic regulatory elements harboured in variable genetic backgrounds can differently affect this complex pathologic mechanism. We study six genetically diverse backgrounds of C. *elegans* containing a human α -synuclein transgene. Previoulsy, gene expression analysis revealed that the expression of α -synuclein differentially affects genome-wide transcription due to genotypic variation. Here, we focus on α -synuclein gene expression and protein accumulation levels, in worms with six different genetic backgrounds. Quantitative polymerase chain reaction (qPCR) indicated similar expression levels of the α -synuclein transgene among these backgrounds. Significantly different α -synuclein accumulation levels were observed and measured by fluorescent microscopy and by assays using the COPAS Biosorter. Because of the implied lipid binding of the α -synuclein protein structure, we also measured fatty acids using Nile Red staining of lipid droplets in the α -synuclein transgenic worms as well as their corresponding wild-type strains. Variable lipid content and body-length changes were observed caused by genotypic variation but also α -synuclein introgression. This suggests that the molecular responses to its toxicity could be affected by natural variation. Taken together, we suggest that variable genetic backgrounds offer insight into individual genetic differences regarding misfolded proteins aggregates and their cytotoxic effects.

Introduction

The progressive accumulation of α -synuclein is a key step in the pathogenic mechanism of Parkinson's disease (PD). In patients and model organisms, the pathogenic mechanism of abnormal α -synuclein aggregation is determined by genetic, biochemical, and biophysical variables [1]. The protein α -synuclein is a 140 amino acids long, encoded by the human gene *SNCA* located on chromosome 4q21.3-q22, and spans a region of 111 kb [2, 3]. This small soluble protein has three distinct domains: an N-terminal part that adopts an α -helical structure upon preferentially binding to negatively-charged lipids [4]; the central hydrophobic non-amyloid- β component (NAC) that forms the β -sheet structure [5], and an acidic negatively charged C-terminus that contributes to forming fibril structures [6]. The aggregated α -synuclein can form soluble

oligomers, which serve as intermediates in insoluble fibril formation [7, 8]. The process of oligomerization and fibril growth induces perturbations of molecular homeostasis. Subsequent binding of α -synuclein to the lipid membrane has led to the assumption that the toxicity of α -synuclein oligomers is mediated via membrane disruption [9]. This leads to lipid-metabolism-associated lysosomal dysfunction, mitochondrial deficit and oxidative stress, and the changed concentrations of fatty acids, phospholipids, and metal ions (see reviews [10, 11]).

Research on α -synuclein-induced toxicity in various animal models has revealed that accumulation of α -synuclein affects endoplasmic reticulum integrity, ROS accumulation, ATP synthesis, and cytoplasmic Ca₂⁺ homeostasis [12–15]. These changes can trigger pathways linking synaptic α -synuclein pathology to abnormal deposition and mitochondrial dysfunction. Oxidative stress and consequently mitochondrial damage are also strongly linking the changes in lipid homeostasis [16]. Natural genetic variation occurring in DNA affect differential gene expression and translates into variation in protein abundance. Thereby, the associated molecules can impact complex traits and/or disease-related traits. Multiple pathways and networks may be affected by natural genetic variants. For example, the genetic factor sls was identified as a transcriptional hub in a co-regulated module associated with mitochondrial respiration, and its regulation of natural variation in mitochondrial function in the fruit fly Drosophila melanogaster [17]. In the nematode Caenorhabditis briggsae structural variation in mitochondrial DNA, in the form of a heteroplasmic *nad5* gene deletion (*nad5* Δ), was known to negatively correlate with organismal fitness. Several mitochondrial phenotypes were also found to be associated with $nad5\Delta$ leading to extensive phenotypic variation [18]. Furthermore, common or rare variants can affect the lipid and lipoprotein metabolic pathways [19].

Different *C. elegans* models of α -synuclein have been developed. One of the models, the strain NL5901, expresses human α -synuclein in the body wall muscle, which allows for visualizing aggregations and inclusions in vivo. The identified phenotypes in this transgenic strain include age-dependent aggregation, reduced pharyngeal pumping, and deficits in movement [20, 21]. A genome-wide RNA interference screen performed in NL5901 contributed to identification of genetic suppressors and enhancers of the aggregation [21].

So far, these investigations on α -synuclein aggregates have always been conducted in the Bristol N2 genetic background, which limits understanding the genotype-phenotype relationship in PD disease pathogenesis [22]. We recently introgressed the *pkIs2386* transgene (transgene inserted in NL5901)

into four more different genetic backgrounds (aS-ILs). Compared to the corresponding wild type strains, the aS-ILs and NL5901 exhibited a series of phenotypic traits induced by α -synuclein toxicity, related to the genetic backgrounds. Genome-wide patterns of gene expression were determined across different genetic backgrounds in response to the expression of the introgressed α -synuclein. Transcriptome analyses of these worms revealed genotype-specific effects on α -synuclein associated life-history phenotypic traits (Chapter 3, *i.e.* [23]).

To further investigate whether α -synuclein expression was affected by the genetic background, we measured α -synuclein transgene expression in those α -synuclein worms. For NL5901, it was shown that size and abundance of α -synuclein inclusions increased with age [21]. Therefore, we also tracked the α -synuclein aggregation by fluorescent imaging in these transgenic worms over age. Furthermore, we quantified these accumulations to present the impact of the genetic background on inclusion formation. In addition, according to the transcriptome analysis results (Chapter 3, *i.e.* [23]), the pathways associated with nutrient storage and lipid transportation are affected by an α -synuclein introgression. Therefore, we measured lipid droplet content as the outcome of α -synuclein aggregate induced toxicity. Our results confirm the influence of natural variation on susceptibility to proteotoxicity. We conclude that studies on natural variation beyond the N2 genetic background could facilitate *C. elegans* as model of α -synuclein for understanding α -synuclein toxicity.

Materials and Methods

C. elegans Strain and Maintenance

Next to N2 and NL5901, four wild type strains JU1511, JU1926, JU1931, JU1941, and their corresponding ILs carrying α -synuclein introgression (aS-ILs), SCH1511, SCH1926, SCH1931, SCH1941 were used (Chapter 3; [23]). These worms contain transgene *pkIs2386* [*unc-54*p:: α -synuclein::YFP + *unc-119*(+)], and more details have been showed in Chapter 3. Standard techniques were used in culturing for all worms as previously described in Chapter 3 [24].

RT-qPCR

Over 2000 young-adult worms, which were growing at 20°C for 48 hours old after age-synchronizing the population, were washed off NGM plates by M9

buffer (KH₂PO₄, 22mM; K₂HPO₄, 34mM; NaCl, 86mM; MgSO₄, 1mM). RNA was isolated from worm pellets by using Maxwell®16 LEV simply RNA tissue kit (Promega). One µg of total RNA was reverse transcribed with oligo dT primers using Superscript II reverse transcriptase (Invitrogen), according to the manufacturer's protocol. Realtime quantitative polymerase chain reaction (qPCR) was performed on a Biorad iCycler using iQ SYBR green with the following primers: α-synuclein forward 5'-ATGGATGTATTCATGAAAGG-3', α-synuclein reverse 5'-TTCAGGTTCGTAGTCTTGA-3'; eGFP forward 5'-TTT CTGTCAGTGGAGAGGGT-3', eGFP reverse 5'- CCTGTACATAACCTTCGG GC-3'; one of the reference genes, Y37E3.8, forward 5'-ATCCTGGAGGTCGCG GTAAC-3', reverse 5'-GCGCCAAGATAGGAGCGGAT-3'; another reference gene, *rpl-6*, forward 5'- AGTGCTCCGCTTCTCTGCTT-3', reverse 5'- AGGTGA CTCTGGAACCTCGTT-3'. The experiments were conducted in three biological replicates and each replicate included technical duplicates of each sample.

Fluorescent imaging

The formation of aggregates per individual was tracked by imaging fluorescent foci in live animals of: SCH1511, SCH1926, SCH1931, SCH1941, and NL5901, using fluorescent microscopes (Olympia 800). Nematodes were age-synchronized and cultured as described above to 2, 3, 4, 5, 6, and 7 days. Around 6~8 individuals were mounted on slides and exposed to 5x dilution of 1M NaN3 (by M9 buffer). Then, images were taken using the GFP channel with 600ms expose time. Notably, worms over 4 days old required less time in NaN3 in imaging. Meanwhile, correct developmental stages were verified based on morphological features.

Nile red assay

In *C. elegans*, intracellular fat droplets can be stained with the fluorescent dye Nile Red (a lipophilic stain that fluoresces in a lipid environment). Nile Red stock (500 μ g/mL in acetone) was diluted 1:500 in the OP50 solution used as food for each experiment. The freshly diluted Nile Red *E. coli* diet 200 μ l was seeded on 6 cm diameter NGM plates. Subsequently, L1 larvae obtained from bleaching were transferred to the plates and allowed to grow for 48 hours (stage L4 larvae). Plates containing Nile Red as well as feeding worms were protected from the light by wrapping in aluminium foil. The stained worms were observed using the fluorescent microscopes (Olympia 800) equipped with a G filter (excitation at 510-560 nm wavelength; emission at 590 nm wavelength)

for Nile Red. Images were acquired using identical settings and 400 ms exposure times to allow direct comparisons. A minimum of 9 animals was measured for each strain/line.

Worm sorter

Besides fluorescent imaging, an automated worm sorter, COPAS[™] BIOSORT (Union Biometrica Inc., Somerville, MA, USA) was used for automated fluorescence quantification. Briefly, the 48 hours or 72 hours old worms were washed off plates by M9 buffer and then were immobilized on ice. To quantify the intensities of the yellow fluorescence signal, the reading parameters used were time-of-flight (TOF) for the x-axis and GREEN/Red peak height for the y-axis. The mean number of measurements per sample was over 450. Both 48 hours and 72 hours old worms were analysed. Where the 72 hours old populations were a mixture of mainly adults and some embryos. Average body size of each animal in a population was analysed by ratio of extinction by TOF.

Statistical analysis

All statistical analyses were performed in R (version 3.5.1 64x). For the worm sorter assay, raw data were read in, processed, and plotted using the COPASutils R package [25]. Noise in the data, *e.g.* air bubbles, were filtered out according to normalization of worm body parameters (Supplementary File 1 and 2 and Supplementary Figure 1 and 2). The qPCR measurements were transformed and processed for data normalization, according to the method described in the study of Sterken *et al.*, 2014. Subsequently, pairwise testing was done by using a two-sample independent t-test not assuming equal variances. Testing over multiple samples was done by ANOVA.

Results

Stable α -synuclein transgene expression in different genetic backgrounds

Previous transcriptomic analysis of the aS-ILs and NL5901, together with their corresponding wild-type strains demonstrated the effects of the introgressed α -synuclein transgene on genome-wide transcription levels in different backgrounds (Chapter 3, [23]. As *C. elegans* does not encode a α -synuclein

homolog, whole genome microarray profiling did not detect expression of α -synuclein transgene in those five different genetic backgrounds. Therefore, we measured the expression level of the transgene *pkIs2386* [*unc-54*p::α-synuclein::YFP+*unc-119*(+)] in 48 hour-old worms (after hatching; similar to Chapter 3; [23]) by quantitative PCR (qPCR). First, the reference genes used in the qPCR showed stable expression in each genotype (Supplementary File 3). After ascertaining the stability of reference genes between the genotypes, we characterized the gene expression of the transgene by using primers specific to both α -synuclein and YFP. The simultaneous amplification and quantification of both sequences within transgene *pkIs2386* showed similar expression levels, which also were retained across different genetic backgrounds (Figure 1). This result indicated that the transgene *pkIs2386* was indeed introgressed from NL5901 into the five different genetic backgrounds for these aS-ILs by inter-strain crossings. There were no copy number differences of α -synuclein transgene among these worms. Therefore, we suggest that different genetic backgrounds tend to have no significant influence on the expression level of α -synuclein transgene.

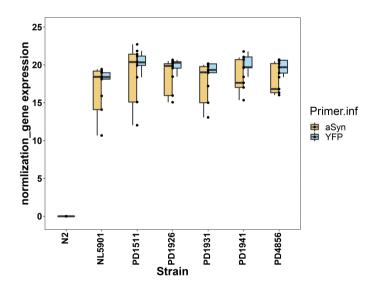


Figure 1: Gene expression levels of α -synuclein transgene in five different genetic **backgrounds by qPCR**. The relative expression of α -synuclein and GFP measured by RT-qPCR for these 48hrs old worms is shown, in total three independent repeats in duplo per experiment. No significant difference was found between the expression of α -synuclein and GFP for each strain (t-test, *p*>0.05), meanwhile none was among the strains (ANOVA, *p*>0.05).

77

The different accumulation levels of α -synuclein in different genetic backgrounds

We further investigated the formation and accumulation of α -synuclein protein inclusion with age in those worms by fluorescent imaging and COPAS Biosorter measurements. The α -synuclein protein was fused to yellow fluorescent protein (YFP) and its increasing accumulation was observed as visible inclusions in body wall muscle cell (Figure 2). The increasing size and abundance of fluorescent α -synuclein protein aggregation with age were present in NL5901 (Figure 2), which, as a characteristic of α -synuclein aggregation, was repeatedly documented before [21, 26]. Interestingly, we observed a relatively stronger fluorescent signal in worms with wild-isolate JU backgrounds compared to NL5901 in the early stage of worms (Figure 2). In older worms it became more apparent that protein inclusions accumulated in NL5901, but also in the JU strains (Figure 2).

To quantify the YFP-tagged α -synuclein protein accumulating in different genetic backgrounds, live worms of all four aS-ILs, as well as NL5901, were quantified with the COPAS Biosorter (Union Biometrica) where they were measured at 48 hour- and 72 hour-old (Figure 3). Notably, 48 hour-old worms showed significantly different fluorescent green signals among different genetic background (ANOVA, *p* <2.2e-16, Supplementary File 5), even though these worms showed similar expression levels of the α -synuclein transgene (Figure 1). Moreover, older worms obviously presented higher fluorescent signal than young worms (ANOVA, p <2.2e-16, Supplementary File 5). SCH1941 displayed the strongest increase, while SCH1931 showed the lowest increase (Figure 3). Notably, in both 48 hour-old worms and 72-hour-old worms, NL5901 exhibited the lowest levels over all strains. This confirmed the fluorescent observations shown in Figure 3. Together, the accumulations of α -synuclein among the aS-ILs and NL5901 worms measured by Biosorter and microscopy suggests that genetic background could strongly impact processing of α -synuclein accumulation.

Notably, extensive protein aggregation occurs close to the vulva in α -synuclein JU-derived lines (Figure 2B), the relative susceptibility of vulva muscle cells to the aggregates is markedly different compared to the N2 background (Figure 2B). The head area exhibited more visible aggregation than body wall muscles in NL5901, whereas the inverses was observed in the aS-IL animals, with the vulva muscles exhibiting the high accumulations already.

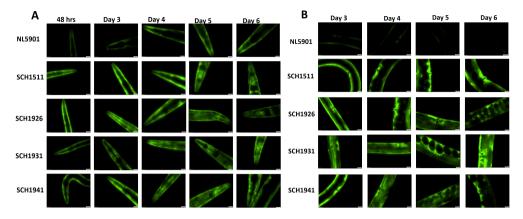


Figure 2: Fluorescent tracking focusing on head area **(A)** and vulva area **(B)** over time among the four transgenic ILs and NL5901. From left to right are different ages of worms. Exposure time of all imaging was 600ms.

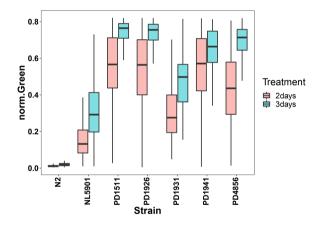


Figure 3: Quantifications of α -synuclein-YFP fusion protein by the COPAS Biosorter, with N2 as negative control. The six transgenic strains contain α -synuclein but differed among genetic backgrounds. The fluorescent signal levels were significantly different among 48 hours-old worms (ANOVA, *p*<2.2e-16), which similarity was among the 72 hours-old transgenic worms (ANOVA, *p*<2.2e-16). Post-hoc Tukey tests can be found in Supplementary File 4.

α -synuclein aggregation affects lipid droplet formation

Previous findings show that α -synuclein caused synaptic pathologies in several NGD, *e.g.* PD, that are supposed to be involved in lipid membrane interactions [27]. Alterations in α -synuclein, including structural binding and lipid specificity, can cause disruptions of the complex network in the synaptic machinery [28]. Based on that, the effect of aggregates on lipid formation in *C. elegans* was expected to be representative of α -synuclein proteotoxicity.

Thus, we visualized and measured lipid droplets in the aS-ILs and the five wild-type strains as well as N2 and NL5901 by Nile Red staining. Fluorescent intensity of Nile Red staining was tracked by both fluorescent microscopy and COPAS Biosorter. Nile Red stained vesicular structures (*e.g.* lysosome-like compartments) in the intestine; the main fat storing organ of *C. elegans* [29–31]. Thus, the lysosome-related Nile Red staining could offer the insight into lysosomal degradation of α -synuclein in *C. elegans*.

All 48 hour-old late L4 larvae displayed little differences in body sizes. 72 hour-old worms obviously had a larger body size than 48 hour-old worms (Supplementary File 6). 48 hour-old NL5901 did not show difference in body size compared to N2, whereas a significant difference was observed for the Nile Red signal (Figure 4A and 4B) (false discovery rate corrected Tukey test, q-value = 1 and 0, respectively). Furthermore, 72-hour-old NL5901 had much larger body size than N2 (q-value = 0), and its lipid droplet amounts were also higher than that of 72 hour-old N2 (q-value = 0) (Figure 4A and 4B, Supplementary File 5). aS-ILs exhibited variable differences of lipid droplet growth compared to NL5901 (ANOVA, p<2e-6) (Figure 4A and 4B). Moreover, Nile Red stained lipid accumulation varied among wild-type JU and N2 worms but also were significantly different from those changes in aS-ILs and NL5901 (ANOVA, p-value = 0.021). Therefore, our results suggest that α -synuclein affects lipid accumulation depends on the genetic background. Hence, regulators associated with body growth and/or lipid storage in variable backgrounds could be differently affected by α -synuclein resulting into differential alterations in lipid metabolism.

Discussion

Polymorphic regulators affect aggregate accumulation

Our investigation on α synuclein aggregation and accumulation used a *C. elegans* model of α synuclein (NL5901, transgene *pkIs2386* [*unc-54*p:: α -synuclein::YFP + *unc-119*(+)]) with different genetic backgrounds beyond N2 (see also Chapter 3; [23]). Under expression of a *unc-54* promoter, α -synuclein fused with YFP is expressed in the body wall [21] and accumulates into aggregate inclusions with age (Figure 2). Some studies point out that α synuclein monomers might play a role in the complex molecular mechanism leading to cellular dysfunction or disruption of other molecular or signaling pathways (see review [1]). Moreover, oligomeric and phosphorylated α synuclein has mainly been viewed toxic and essential for PD [32]. Oligomeric

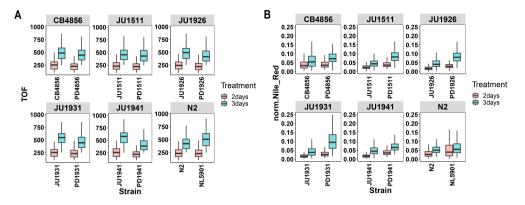


Figure 4: Variation in the body length and stained lipid droplets between wild type worms and their corresponding aS-ILs worms. Both 48 hours- and 72 hours-old worms were scanned and sorted by COPAS Biosorter. In total, five different genetic backgrounds were included. **A)** Measurements of individual body length (time of flow, TOF). **B)** Measurements of individual lipid formation stained by Nile Red (normalized fluorescent Red signal). For each treatment and strain, approximately 500 worms were studied. (Supplementary File 5).

 α -synuclein aggregates, which are scattered as unevenly distributed fluorescent spots in *C. elegans* (Figure 2; [21]), could represent the primary cytotoxic α -synuclein species present in post-mortem human PD brain tissue [33]. The accumulation of α -synuclein-YFP fusion shown in Figure 2 (fluorescent spots) might include several α synuclein aggregate structures in the process of clump formation and accumulation.

More α synuclein-YFP remains diffusely localized in the aS-ILs compared to NL5901. Further accumulation of α synuclein-YFP into foci has been seen in all elderly α synuclein worms due to intrinsic properties of the α -synuclein protein, e.g. propagation to misfolded α -synuclein [34]. The inclusions that α synuclein forms with age are clearly visualized in the head area in all adult α synuclein genotypes. In the aS-ILs worms, the number of inclusions with age clearly increases in head muscles surrounding the pharynx. These inclusions initially appear as encircling the isthmus and then spread to the metacorpus and terminal bulb (Figure 1A). A Similar spread is found in the strain NL5901, but much less and at a later stage. This could explain a lower pumping deficit observed in NL5901 compared to the aS-ILs (Chapter 3). Meanwhile, numerous aggregates appear at the vulva of young adult aS-ILs and significantly increas in the vulva muscles, which are vivid cross-connection on toroid (Figure 2B). In contrast, NL5901 does not show vulval aggregates (Figure 2B). Ultimately, the accumulations could explain the high internal hatching rate in the aS-ILs (Chapter 3).

The variable accumulation of α synuclein in the aS-ILs and NL5901 is confirmed by quantification of the fluorescent signals from the fusion protein. Especially, NL5901 indeed has the lowest α synuclein protein levels compared to the other strains. Interestingly, the transcription level of α synuclein transgenes in NL5901, with the N2 genetic background, is not significantly different compared to the other five genetic backgrounds tested. However, the exact folding and oligomerization pathway leading to these α synuclein aggregates is still not known.

Proteins with variable abundance are likely to be associated with the biological function of polymorphic variants elsewhere on the genome [35]. Our findings match the previous transcription analysis of four aS-ILs and NL5901 (Chapter 3, [23]), as expression changes are detected enriched for genes related to regulating protein coding and synthesis. Hence, these transgenic α synuclein encoding *C. elegans* could provide a more reliable platform to elucidate the molecular interactions involved in α synuclein aggregation by using different natural genetic backgrounds.

α -synuclein proteotoxic effects on lipid regulation varied

The extreme N-terminus of α synuclein is found to play a role in lipid interaction and membrane-induced helix formation (from review [4]). In addition to the N-terminal region, two more regions in α -synuclein (the non-amyloid- β component and C-terminus) participate in the membrane binding process as well [36]. When α -synuclein is in the unfolded monomeric state, it binds to small unilamellar vesicles composed of negatively charged lipids [37]. As vesicles containing negatively charged lipids are impermeable to protein adsorption in general, further binding of more toxic oligomeric α -synuclein induces membrane permeabilization leading to membrane defects, which could result in more vulnerability to oligomeric α -synuclein binding [38].

Interestingly, SCH1931 shows much lower α synuclein accumulation than other aS-ILs (Figure 2), while it has the highest accumulation of lipid droplets of all four (Figure 4). This suggest that lipid storage regulation could be affected by α synuclein, while the variable effects rely on the natural variation in different genetic backgrounds. However, no significant difference in body length is found between 48 hour-old worms, with or withoud α -synuclein presence (Figure 4A). Development has been affected by α -synuclein, like SCH1941 showed over 7 hrs developmental delay compared to JU1941 (Chapter 3, [23]).

Importantly, lipid storage and lipid membrane growth are branches of lipid metabolism, which are very complex and poorly understood. Our results cannot directly point out how α -synuclein aggregates interact with lipid metabolism, as the results are too coarse for understanding the lipid deposition in response to α -synuclein protein misfolding. However, our investigation on lipid droplets associated with both natural variation and introgressed α -synuclein expression uncovered that lipid metabolism can be affected by α -synuclein expression and aggregation. Importantly, it also interacts with genetic background both via the transcriptome and in measurable phenotypes.

Conclusion

In this study, genetic divergence in response to the α -synuclein transgene contributes to finding genotype-phenotype relationship related to α -synuclein-induced traits or toxicity. These results reveal the effect of genetic backgrounds on the accumulation of α -synuclein and the potential regulation on lipid by α -synuclein. The earlier onset and more accumulation of aggregates in aS-ILs suggest variable effects of α -synuclein compared to that in NL5901, including lipid droplet formation. A reduced lifespan (shown in Chapter 3) could therefore be a general toxicity phenotype due to accumulation of α -synuclein, as well as a decreased pumping rate, and reduced hatching found in Chapter 3 [23]. These phenotypic changes arevariable among different genetic backgrounds and they were more pronounced in wild isolate JU backgrounds than in the N2 background. This demonstrates that natural genetic variation appears to affect phenotypic traits in the aggregation processing.

Supplementary Material

The supplementary files and figures are deposited at: https://ldrv.ms/u/s!AqhbSv9gNbn2ac0H oR1de29x0ms?e=LwgTMX

Supplementary figure 1. Normalization of worm body parameters with all YFP measurement per strain per sample. The x-axis is the time-of-flight (TOF), while the y-axis is the ratio of Extinction by TOF (rExtTOF).

Supplementary figure 2. Normalization of worm body parameters with all measurement of Nile Red fluorescence intensity per strain samples. The x-axis is the time-of-flight (TOF), while the y-axis is the ratio of Extinction by TOF (rExtTOF).

Supplementary file 1. The dataset of YFP measurement obtained by filtering out technical noise.

Supplementary file 2. The dataset of Nile Red staining lipid content measurement obtained by filtering out technical noise.

Supplementary file 3. The expression levels of two reference genes over all strains. Data are expressed as mean ± standard deviation. Data were analyzed using one-way analysis of variance (ANOVA), followed by Tukey's test.

Supplementary file 4. The quantification of YFP signal over all worms (based on Supplementary file 1). Each strain was measured at 48 hours- and 72 hours-old. Data are expressed as mean ± standard deviation. Data were analyzed using one-way analysis of variance (ANOVA), followed by Tukey's test.

Supplementary file 5. The quantification of lipid content stained by Nile Red over all worms (based on Supplementary file 2). Each strain was measured for at 24 hours- and 72 hours-old. Data are expressed as mean ± standard deviation. Data were analyzed using one-way analysis of variance (ANOVA), followed by Tukey's test.

Supplementary file 6. The measurement of time-of-flight (TOF) over all measured worms (based on data from Supplementary file 1 & 2). Each strain was measured at 24 hours- and 72 hours-old. Data are expressed as mean ± standard deviation. Data were analyzed using one-way analysis of variance (ANOVA), followed by Tukey's test.

References

- 1. Lashuel, H. A. *et al.*, The many faces of α-synuclein: from structure and toxicity to therapeutic target. *Nat Rev Neurosci*, 2013. 14(1): p. 38–48.
- Stefanis, L., α-Synuclein in Parkinson's disease. Cold Spring Harbor Perspectives in Medicine, 2012. 2(2): a009399.
- 3. **Venda**, **L. L.** *et al.*, *α*-Synuclein and dopamine at the crossroads of Parkinson's disease. *Trends Neurosci*, 2010. 33(12): p. 559–568.
- 4. **Bartels, T.** *et al.*, N-alpha-acetylation of α-synuclein increases its helical folding propensity, GM1 binding specificity and resistance to aggregation. *PloS O ne*, 2014. 9(7): e103727.
- 5. **Bisaglia**, **M**. *et al.*, Structure and topology of the non-amyloid-beta component fragment of human alpha-synuclein bound to micelles: implications for the aggregation process. *Protein Sci*, 2006. 15(6): p. 1408–1416.
- 6. **Ruzza**, **P.** *et al.*, Ceftriaxone blocks the polymerization of α-synuclein and exerts neuroprotective effects in vitro. *ACS Chem Neurosci*, 2014. 5(1): p. 30–38.
- Lam, H. T. *et al.*, Stabilization of α-Synuclein Fibril Clusters Prevents Fragmentation and Reduces Seeding Activity and Toxicity. *Biochemistry (Mosc)*, 2016. 55(4): p. 675–685.
- 8. Lee, H.-J. et al., Clearance of alpha-synuclein oligomeric intermediates via the lysosomal degradation pathway. *J Neurosci*, 2004. 24(8): p. 1888–1896.
- 9. **Fusco**, **G.** *et al.*, Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers. *Science*, 2017. 358(6369): p. 1440–1443.
- Angelova, P. R. *et al.*, Ca²⁺ is a key factor in α-synuclein-induced neurotoxicity. *J Cell Sci*, 2016. 129(9): p. 1792–1801.
- 11. **Pieri, L., K. Madiona, and R. Melki**, Structural and functional properties of prefibrillar α-synuclein oligomers. *Sci Rep*, 2016. 6: p. 24526.
- 12. **Bellucci**, M. *et al.*, Predicting protein associations with long noncoding RNAs. *Nat Methods*, 2011. 8(6): p. 444–445.
- Devi, L. et al., Mitochondrial import and accumulation of alpha-synuclein impair complex I in human dopaminergic neuronal cultures and Parkinson disease brain. J Biol Chem, 2008. 283(14): p. 9089–9100.
- 14. **Ludtmann**, **M. H. R.** *et al.*, *α*-synuclein oligomers interact with ATP synthase and open the permeability transition pore in Parkinson's disease. *Nat Commun*, 2018. 9(1): p. 2293.
- 15. **Perfeito, R.** *et al.,* Linking alpha-synuclein phosphorylation to reactive oxygen species formation and mitochondrial dysfunction in SH-SY5Y cells. *Mol Cell Neurosci,* 2014. 62: p. 51–59.
- 16. **Paillusson, S.** *et al.*, α-Synuclein binds to the ER-mitochondria tethering protein VAPB to disrupt Ca, javax.xml.bind.JAXBElement@215089da, homeostasis and mitochondrial ATP production. *Acta Neuropathol (Berl)*, 2017. 134(1): p. 129–149.
- 17. **Jumbo-Lucioni**, **P.** *et al.*, Nuclear genomic control of naturally occurring variation in mitochondrial function in Drosophila melanogaster. *BMC Genomics*, 2012. 13: p. 659.
- 18. Hicks, K.A. *et al.*, In vivo quantification reveals extensive natural variation in mitochondrial form and function in Caenorhabditis briggsae. *PLoS One*, 2012. 7(8): e43837.
- 19. Dron, J.S. and R.A. Hegele, Genetics of Lipid and Lipoprotein Disorders and Traits. *Current Genetic Medicine Reports*, 2016. 4(3): p. 130–141.
- 20. **Bodhicharla**, **R**. *et al.*, Effects of α-synuclein overexpression in transgenic *Caenorhabditis elegans* strains. *CNS & Neurological Disorders Drug Targets*, 2012. 11(8): p. 965–975.
- 21. **Ham, T. J. van** *et al., C. elegans* model identifies genetic modifiers of alpha-synuclein inclusion formation during aging. *PLos Genet,* 2008. 4(3): e1000027.

- 22. Wang, Y. A., J. E. Kammenga, and S. C. Harvey, Genetic variation in neurodegenerative diseases and its accessibility in the model organism *Caenorhabditis elegans*. *Hum Genomics*, 2017. 11(1): p. 12.
- Wang, Y. A. *et al.*, Genetic background modifies phenotypic and transcriptional responses in a *C. elegans* model of α-synuclein toxicity. *BMC Genomics*, 2019. 20(1): p. 232.
- 24. Brenner, S., The genetics of Caenorhabditis elegans. Genetics, 1974. 77(1): p. 71–94.
- 25. Shimko, T. C. and E. C. Andersen, COPASutils: an R package for reading, processing, and visualizing data from COPAS large-particle flow cytometers. *PLoS One*, 2014. 9(10): e111090.
- Shamsuzzama, L. Kumar, and A. Nazir, Modulation of Alpha-synuclein Expression and Associated Effects by MicroRNA Let-7 in Transgenic, javax.xml.bind.JAXBElement@64f70ea4, . Front Mol Neurosci, 2017. 10: p. 328.
- 27. Schulz-Schaeffer, W.J., The synaptic pathology of alpha-synuclein aggregation in dementia with Lewy bodies, Parkinson's disease and Parkinson's disease dementia. *Acta Neuropathol*, 2010. 120(2): p. 131–143.
- 28. **Viennet, T.** *et al.*, Structural insights from lipid-bilayer nanodiscs link α-Synuclein membrane-binding modes to amyloid fibril formation. *Communications Biology*, 2018. 1: p. 44.
- 29. **Bonilla**, E. and A. Prelle, Application of nile blue and nile red, two fluorescent probes, for detection of lipid droplets in human skeletal muscle. *J Histochem Cytochem*, 1987. 35(5): p. 619–621.
- 30. **O'Rourke**, **E.J.** *et al.*, *C. elegans* major fats are stored in vesicles distinct from lysosome-related organelles. *Cell Metab*, 2009. 10(5): p. 430–435.
- 31. Fowler, S.D. *et al.*, Use of nile red for the rapid in situ quantitation of lipids on thin-layer chromatograms. *Journal of lipid research*, 1987. 28(10): p. 1225–1232.
- 32. **Majbour**, **N. K.** *et al.*, Oligomeric and phosphorylated alpha-synuclein as potential CSF biomarkers for Parkinson's disease. *Mol Neurodegener*, 2016. 11: p. 7.
- Xin, W. et al., Toxic Oligomeric Alpha-Synuclein Variants Present in Human Parkinson's Disease Brains Are Differentially Generated in Mammalian Cell Models. *Biomolecules*, 2015. 5(3): p. 1634–1651.
- Recasens, A. *et al.*, Lewy body extracts from Parkinson disease brains trigger α-synuclein pathology and neurodegeneration in mice and monkeys. *Ann Neurol*, 2014. 75(3): p. 351–362.
- 35. **Roumeliotis, T.I.** *et al.,* Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. *Cell Reports,* 2017. 20(9): p. 2201–2214.
- 36. **Fusco**, **G**. *et al.*, Direct observation of the three regions in α-synuclein that determine its membrane-bound behaviour. *Nat Commun*, 2014. 5: p. 3827.
- Fecchio, C., L. Palazzi, and P. P. de Laureto, α-Synuclein and Polyunsaturated Fatty Acids: Molecular Basis of the Interaction and Implication in Neurodegeneration. *Molecules*, 2018. 23(7).
- Rooijen, B. D. van, M. M. A. E. Claessens, and V. Subramaniam, Membrane Permeabilization by Oligomeric α-Synuclein: In Search of the Mechanism. *PLoS One*, 2010. 5(12): e14292.

Chapter 5

eQTL mapping in a newly constructed CB4856/N2 recombinant inbred line populations harbouring human α-synuclein gene

Yiru A. Wang Vivian Tegelbeckers Yuzhi Chen Demi Vogels Joost A.G. Riksen Jan E. Kammenga Simon C. Harvey Mark G. Sterken

Abstract

 α -synuclein is genetically and neuropathologically linked to Parkinson's disease (PD). Genetic factors can influence disease phenotypic outcomes in variable ways due to the contribution of genetic variation. However, how genetic variation is related to α -synuclein-associated disorders still remains unclear. The nematode C. elegans as a model organism is used in studies of α -synuclein pathology. Our previous transcriptome analysis indicates that effects of α -synuclein expression vary depending on the genetic background. To further investigate the underlying genotypic-phenotypic relationship of α -synuclein expression, we constructed a specific panel of recombinant inbred lines (RILs) derived from SCH4856 and NL5901 harbouring α -synuclein introgression (*i.e.* aS-RILs). We performed an expression quantitative trait locus (eQTL) analysis using the newly constructed genetic map of the population. To detect the eQTLs associated with α -synuclein expression, an estimated α -synuclein effect was determined by the gene expression changes between measurements of N2 and NL5901, which was used as a filter in the detection of aS-RILs. In the eQTL analysis, we found that 2274 genes act as *cis*-eQTL and 2941 genes as *trans*-eQTL, while 528 of these 5215 genes (as eQTL) were also detected in both heritability and transgression analysis. We detected six out of nine eQTL *trans*-bands specific for the interaction between the genetic background and the introgressed α -synuclein, including genes functioning in a variety of biological molecular pathways, such as innate immune system, lifespan, muscle function & locomotion, cellular components (e.g. lysosome), variable neuron activities, and protein syntheses. Taken together, we detected eQTL responding to α -synuclein. The transcriptome analysis contributes to dissect the complex genetic architectures associated with the expression and accumulation of α -synuclein.

Introduction

Individual genetic variation can play a significant role in disease-development in humans [1]. However, untangling the genetic factors from environmental contributions is extremely hard in human studies. For example, genetic variation affects the risk of developing Parkinson's disease (PD). The accumulation of α -synuclein, encoded by the gene *SNCA*, is thought to play a critical role in the pathogenesis of Parkinson's disease (PD) [2, 3]. Furthermore, common genetic variation in *SNCA* gene is associated with an elevated risk of developing PD. However, multiple factors are thought to contribute, such as the communication of metabolic constituents and numerous molecular signalling processes. Hence, we sought to uncover genetic variation associated with α -synuclein accumulation in the tractable genetic model *Caenorhabditis elegans*.

The nematode *C. elegans* is widely used as model organism to elucidate molecular pathways, such as the insulin-pathway, apoptosis-related genes, and so on. Importantly, stress-related signalling pathways are conserved from nematode to human, such as for heat shock, oxidative stress, hypoxia, and osmotic stress (from Review [4]). Furthermore, *C. elegans* models of α -synuclein were established. Transgenic expression of allogeneic α -synuclein in specific tissues has already uncovered and characterize genetic factors and metabolic compounds that attenuate neurodegeneration [5]. Moreover, *C. elegans* has been used in other neurodegenerative disease study, *e.g.* Alzheimer's disease (AD), Huntington's disease (HD), as well as multifactorial cancer diseases [6]. Here, we use genetic variation in *C. elegans* to map loci involved in α -synuclein accumulation. Thereto, we use two divergent strains, namely the canonical strain Bristol N2 and the polymorphic strain Hawaii CB4856 [7, 8].

In a previous study, we showed that we could associate genetic variation in *C. elegans* to gene-expression, α -synuclein accumulation and fitness-related traits. Also in human studies, gene-expression differences have been implicated in α -synuclein accumulation. For example, the transcriptional regulation of *SNCA* gene is thought to be important in the pathogenesis of PD [9]. Besides the disorder causal variants in *SNCA* gene, systemic responses to α -synuclein gene expression and aggregation are predicated on the communication of numerous molecular signalling processes and metabolic constituents, such as the mitochondrial dysfunction [10], oxidative stress [11], and age-dependent features [12]. *C. elegans* is very suitable to transcriptome studies, as clonal populations of completely homozygous animals can be established [8, 13]. Furthermore, animals can be developmentally synchronized and its small genome-size makes transcriptomics and genetic analysis relatively cheap.

We used two wild-types strains of *C. elegans* that contain an α -synuclein introgression. The use of natural variants is beneficial for investigating phenotypic trait variation. Trait variation among natural populations can be governed by multiple loci [14, 15]. This makes quantitative traits inherently difficult to study, but some approaches, such as recombinant inbred lines (RILs) populations, can make headway. After N2, the Hawaii strain CB4856 is often used for constructing RIL populations. RILs derived from CB4856 and N2 [16, 17] and introgression lines [18] have been widely used in the molecular dissection of their phenotypic differences stemming from their genotypic

differences. Many QTL studies, *e.g.* pathogenic resistance/susceptibility, lifespan, fitness, and stress responses have been performed with different panels of RILs derived from CB4856 and N2 [19–23].

Genetic differences can also result in changes in gene expression levels, which might reflect phenotypic outcomes [24]. *C. elegans* RIL panels also allow for the genetic mapping of gene expression and detection of expression quantitative trait loci (eQTL) [16, 21, 22, 25, 26]. The detection of eQTL contributes to determine gene regulatory networks underlying molecular phenotypes, and is especially suitable for studying complex diseases [27]. Functional loci can be characterized as either *cis*- or *trans*-acting, which represents their regulation of genes located at the same locus or genes located elsewhere [28]. Panels of RILs containing transgenic integration/mutations are a valuable community resource to investigate the links between the transgene and genetic backgrounds, such as the CB4856 x MT2124 RILs (*let-60* in N2 background, [26] and the CX12311xLSJ2 RIL panel (wild type strains, *glb-5 & npr-1* in N2 background, respectively, [29]).

In our previous study, we found developmental delay in the α -synuclein introgression lines, which affects the identification of genetic variation affecting α -synuclein. Moreover, we could not pin-point the origin of polymorphic modifiers of α -synuclein. Therefore, we generated the aS-RILs carrying an α -synuclein introgression by reciprocal crosses between NL5901 and SCH4856 (Chapter 3; [30]). Here, we present the genetic map of the panel. We measured gene-expression and estimated the heritability and transgressive segregation of gene expression to reveal the allelic combination produced higher or lower expression level in each RIL than in either parent. Furthermore, we characterize the effect of local (*cis*) and distant (*trans*) genetic variation on gene expression changes. To identify α -synuclein-specific eQTL, we estimated a α -synuclein effect by comparing NL5901 with N2 to filter out similar effect in the identified *trans*-eQTL hotspots in this *C. elegans* aS-RILs population. We found that two third of *trans*-eQTL hotspots left after filtering, which are mainly associated with the α -synuclein expression and/or aggregation.

Material and Methods

α -synuclein RIL construction

Strain culture

All nematodes were grown and maintained on nematode growth medium (NGM) seeded with *Escherichia coli* OP50 bacteria [8]. Strains used were wild-type Bristol N2, Hawaii CB4856, and strain NL5901 carrying the transgene *pkIs2386* ([*unc-54::* α -synuclein::YFP *unc-119*(+)]). Previously, NL5901 was backcrossed to both N2 and CB4856. The N2-background transgenic strain is named NL5901 throughout the manuscript. The CB4856-background transgenic strain is named as SCH4856 and contains an introgression of NL5901 carrying the α -synuclein transgene insertion (Chapter 3, *i.e.* [30]).

Crossing

The RILs containing an α -synuclein transgene within a background of CB4856 and N2, called as aS-RILs, were generated in two steps (Supplementary Figure 1). In the first step, NL5901 was crossed with SCH4856 in two directions: NL5901 hermaphrodites crossed with SCH4856 males and SCH4856 hermaphrodites crossed with NL5901 males. This resulted in offspring with mitochondria from both backgrounds and a heterozygous F1 population. Subsequently, the F1 hermaphrodites were back-crossed with SCH4856 males to obtain a mix of alleles at the incompatibility *peel-1/zeel-1* locus on chromosome I [31]. From these crosses, single hermaphrodites were selected and propagated, resulting in six individual populations as intermediates. These six populations, the offspring of 21 populations were interbred, ultimately. After the next eight generations of inbreeding by self-fertilization, 93 RILs were obtained. However, these RILs showed a low rate of crossover and displayed a skewed N2/CB4856 contribution to the genetic background, probably because of poor male performance during mating.

We selected eight (of 93) genetically diverse RILs and NL5901 and SCH4856 were used for a new round of interbreeding. First, the reciprocal crosses between these 10 strains yielded an F1 population including both heterozygous males and hermaphrodites. Each of the 45 F1 populations were interbred with all other populations, resulting in 200 offspring-populations, which were randomly crossed in a round-robin design, for three subsequent generations. After this step, a singled out hermaphrodite of each population was then self-fertilized for

subsequent eight generations by random selection. At the end, 200 homozygous aS-RILs were obtained with approximately an equal genetic contribution of each parental strain. These aS-RILs are available upon request from the Harvey laboratory (CCCU, UK) and the Kammenga laboratory (WUR, NL). After construction of the aS-RILs, we froze the 200 lines at minus 80°C for long term-storage. A set of 94 aS-RILs was randomly selected for genotyping (out of the population of 200 lines).

PCR-based Genotyping

From each strain, DNA was collected in parallel with cryopreservation of the strains. The DNA samples were then used for PCR-based genotyping with a set of primers which detected insertions/deletions (Indels) between the CB4856 and N2 genomes [30, 32]. PCR was done using the GoTaq®DNA polymerase kit according to the manufacturer's recommendations (Promega).

PCR-based genotyping was performed with in total 41 primer pairs covering the arms and central area of each chromosome (Supplementary File 1). PCR products were visualized on a 1.5% agarose gel for genotyping. Meanwhile, genotypes of ancestral strains NL5901 and SCH4856 were confirmed as well. The intermediate set of 93 RILs produced were genotyped for the entire genome by 41 primer pairs. The subsequent selected 94 RILs of the final 200-RILs panel was genotyped using 20 primer pairs. The parental strains SCH4856 and NL5901 together with reference strain N2 and CB4856 were involved into the PCR-based genotyping as control.

Cis-eQTL genotyping

We used gene-expression data to further genotype 70 aS-RILs (out of 94). First, we obtained data from previous eQTL studies using the same microarray platform [21, 26]. Combined, these studies represent eQTL measured in four different environments. From these studies, we selected one *cis*-eQTL per block of 500,000 bases. This *cis*-eQTL had to meet two criteria: (i) detected in at least three of the four environments, (ii) explain the highest amount of variance of the eQTL within the specific region. Subsequently, the expression of the corresponding spots was tested in the four ancestral strains: N2, CB4856, NL5901, and SCH4856. We did not include markers that incorrectly genotyped more than 2 of the 16 parental replicates (Supplementary Figure 2). This procedure added an additional 124 markers to the population.

Genotypes of aS-RILs were individually inspected per marker and then were assigned automatically when unambiguous. For the marker which includes ambiguous genotype calling for the individual RIL, the line genotype could still be inferred by both vertical (among the aS-RILs, parental and reference strains) and horizonal (with the adjacent markers within the same chromosome) predictions. In total, 124 mishybridization markers were generated in this assay, which are reliable with the following required properties: N2 and NL5901 were assigned different from CB4856 and SCH4856. Furthermore, the genotypes at the ends of the chromosomes were inferred from the distal most-assigned marker, which were the predictive *cis*-eQTL. Subsequently, the correlations and assigned genotypes of these 70 aS-RILs can be found in Supplementary Figure 2 and Supplementary File 2. Together, we generated a complete genetic map of the 70 aS-RILs by integrating *cis*-eQTL markers with the insertions/deletions markers.

eQTL experiment

Strains

Microarray assay (MA) was performed in a smaller panel of the aS-RILs for the genome-wide expression profiling. Based on PCR-based genotyping, 70 out of the total aS-RILs were selected (Supplementary Figure 3 and Supplementary Figure 4) for the assay together with the two parental strains, NL5901 and SCH4856, as well as two reference strains, N2 and CB4856. In the assay, all strains were assorted in a random order. Furthermore, the two parental strains and two reference strains were included and had more biological replicate. For the later eQTL mapping quality control, these four strains had two more biological replicates.

Experiment setup

A trial MA experiment was done on N2 and NL5901 with their different development stages (Supplementary Figure 5), which indicated that mature adult worms present relatively stable expression levels. Therefore, 120 hours old nematodes of both strains (from synchronized embryo onward) were used to avoid the development effect on gene expression levels. All the incubation steps were performed in a temperature incubator set at 20°C.

All strains were synchronized by bleaching solution (with sodium hypochlorite and NaOH; Stiernagle, Wormbook), which is recorded as t=0. Each strain had

two 9cm NGM plates seeded with E. coli, and the eggs hatched and grown for two days. When the worms reached L4, (t=50-54 hours), around 400 worms were transferred, by washing off with M9 buffer, to new NGM plates containing FUdR (5 mg/ml). After an additional 48 hours (t=98 hours) growing, the worms were transferred to a new FUdR plate to avoid starvation. When t=120 hours, each strain was harvested by being washed off worms on both plates into one Eppendorf tube and flash frozen by liquid nitrogen. The pellets were then stored at -80°C and used for following RNA isolation.

RNA isolation

The pellets were firstly prepared by home-made lysis buffer, and RNA isolation were then done by using the Maxwell®16 AS2000 instrument with the Maxwell®16 LEV simply RNA tissue kit (Promega), according to manufacturer's protocol. Total RNA samples were quantified and qualified by using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Using 100 or 200 ng RNA as input, we generated Cy3 and Cy5-labeled cRNA samples. The specific activity of the labelled cRNA was quantitated by NanoDrop after which samples were paired for hybridization (Supplementary File 3). Samples were hybridized with to 4X44K V2 *C. elegans* slides (Agilent®) following the 'Two-Color Microarray-Based Gene Expression Analysis' protocol from Agilent (version 6.0). Microarrays were scanned by an Agilent High Resolution C Scanner on the recommended settings. Obtained data were extracted with Agilent Feature Extraction Software using (version 12.1.1.1) based on the developers guidelines.

Microarray normalisation

Within-array normalization was done with the Loess method, and between-array normalization was done with the Quantile method [33, 34]. These steps were performed using the Limma package in "R" (3.3.5 x64). Hereafter, the log_2 transformed intensities were used for further analysis.

Data analysis

Genetic map analysis

Our estimated genetic map for these 70 aS-RILs employed the rigorous selection of 124 mishybridization markers to ensure uniform coverage of

the *C. elegans* genome. All the *cis*-eQTL were sorted by physical position, we further estimated genetic distances in aS-RILs. The data set includes: a) the counts of crossover breakpoints observed between each pair of markers in the genetic map; b) the unassigned genotype at the certain marker was predicted by the genotype of its closest surrounding marker. The statistical analyses of recombination frequencies were performed in R (version 3.5.3 x64). Notably, most of chromosome IV was assigned as predominantly N2 due to both parental strains carrying the α -synuclein introgression [30].

Principal component analysis (PCA)

The gene expression data was subjected to PCA aiming to explore the structure of the phenotypic variation in the aS-RILs as well as parental and reference strains. The PCA was conducted using log_2 ration with the mean transformed gene expression data,

$$R_{i,j} = \log_2(\frac{y_{i,j}}{mean\ yi})$$

where R is the log_2 relative expression of spot *i* (*i* = 1, 2, ..., 45,220) in strain *j* (aS-RIL), and *y* is the intensity (not the log_2 -transformed intensity) of spot *i* in strain *j*.

Subsequently, the *prcomp* function in "R" was used to calculate the principal components.

eQTL mapping

eQTL mapping was done in "R" (version 3.5.3 x64), following the procedure as described in [21, 26]. In short, we used a linear model to map gene expression QTL using a single marker model,

$$y_{i,j} \sim x_j + e_j$$

where *y* is the log_2 -normalized intensity as measured by microarray of spot *i* (*i* = 1, 2, ..., 45220) of aS-RIL *j*. This is explained by the genotype (either CB4856 or N2) on marker location *x* (*x* = 1, 2, ..., 124) of aS-RIL *j*.

Permutations and calculations were performed to determine an empirical false discovery rate [21, 22]. We randomly permuted the log_2 -normalized intensities of each line per gene over the genotypes, which we used for mapping with the single marker model using the original genetic map. This was then repeated for ten randomized datasets as described in the study of Snoek *et al.*, 2017. The threshold was determined by a false discovery rate,

$$\frac{FDS}{RDS} \le \frac{m_0}{m} q \times \log(m)$$

where FDS (false discoveries) is the permutation outcome and RDS (real discoveries) comes from the eQTL mapping at a specific significance level. The value of m0 and m were taken from the number of microarray spots (45220), representing true null hypotheses and hypotheses tested respectively. The FDR adjusted p-value (*i.e.* q-value) was set at 0.05 threshold.

eQTL analysis

The variation amount explained per microarray spot with an eQTL was calculated by ANOVA. For all spots, also those with multiple peaks, this analysis of the gene expression explained over the peak-marker was performed assuming a single-marker.

Trans-bands (an enrichment of *trans*-eQTL) were identified by counting a Poisson distribution of the mapped *trans*-eQTL [21, 26, 35]. The number of *trans*-eQTL per 0.5 Mb block were counted, and then we calculated how many *trans*-eQTL were expected to be found. This we used to identify if an overrepresentation according to the Poisson distribution was present (p <0.001). Thereby, *trans*-eQTL hotspots at specific markers were determined.

Heritability analysis

The log_2 -transformed intensities were used to calculate board-sense heritability (H^2) . H^2 estimates were calculated as the fraction of phenotypic variance that can be explained by strain by estimating the technical variance from the repeats of the two parental and two genetic background strains. By estimating the genotypic variance and the remaining variance (*e.g.* measurement error) in the parental lines (as in [36]), heritability was calculated:

$$H_{RIL}^2 = \frac{V_{RIL} - V_e}{V_{RIL}}$$

where V_{RIL} is the variance within the RIL population and V_e is the pooled variance of both parental and reference strains.

To establish whether the heritability was significant and not outlier driven, we applied a permutation approach (as in [23]). The trait values were randomized over the line designations and the heritability calculation were repeated. This was done 1000 times for each transcript to generate a by-chance distribution. The 50^{th} highest value was used as the FDR = 0.05 threshold.

Transgression analysis

Transgression was calculated by counting the number of lines with expression levels beyond three times the standard deviations of the mean from the parental strains (as in [37]). The lower boundary was established by the parental line with the lowest mean, and the upper boundary was established by the parental line with the highest mean. The standard deviation (σ) was calculated as the pooled standard deviation of the two parental lines (n = 4 for both). We set the boundary for transgression at 3 × σ . Significance of the transgression was calculated by permutation. The expression values were randomized over the line designations and the same test as above was conducted. This was repeated 1000 times for each transcript, so the obtained values could be used as the by-chance distribution. The 50th highest value was used as the false discovery rate (FDR) = 0.05 threshold.

Results

Construction of the aS-RILs

In a previous study, we showed the effect of the genetic background on α -synuclein introgression. This introgression affected the transcriptome and life-history traits [30]. To be able to localize modifier loci, we took the α -synuclein introgression line (IL) SCH4856 and crossed it with NL5901. The crossing scheme was designed to avoid influence of peel-1/zeel-1, as well as to increase recombinations (see Materials and Methods for details). We generated a population of 438 RILs that carried the α -synuclein transgene in an N2 and CB4856 mosaic background. These were genotyped with a select

number of insertions/deletions markers, after which 70 aS-RILs were selected for microarray analysis. These 70 lines were further genotyped using *cis*-eQTL based gene-expression markers. In total, 124 gene-expression markers were validated and used. This led to a genetic map of the 70 aS-RILs, covering all six chromosomes of the *C. elegans* genome (Figure 1).

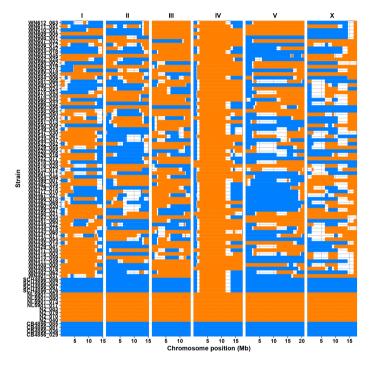


Figure 1: Genetic map of the NL5901xSCH4856 aS-RIL population. The x-axis indicates the genomic position with separate chromosomes indicated on the top of the graph. Each row on the y-axis represents a strain. The genotype shown is a combination of PCR-based genotyping and *cis*-eQTL based gene expression markers measrued by microarray. Orange represents the N2 genotype and blue represents the CB4856 genotype, while regions with a uncertain genotype are depicted in white.

The genotyped RIL population showed ample recombination events and followed recombination patterns expected from an N2 x CB4856 cross. In total, 620 breakpoints were identified (transition from one genotype to the other), with an average of 103 breakpoints per chromosome. The average number of recombination events was 8.9 per RIL and 1.5 per chromosome. The total length of the genetic map was 810 cM, with an average genetic distance between two adjacent markers of 6.8 cM from 5.02 cM (chromosome V) to 10.19 cM (chromosome X) (Supplementary File 4). Lower recombination rates were observed at the chromosome centers compared to the arms of genetic

map (Figure 2A), thus the genetic lengths of the centers were shorter than that of the arms (Figure 2A). Moreover, the parental allele frequencies were roughly equal over most of the genome (Figure 2B; Supplementary File 5). However, on chromosome IV no recombination events occurred from 2.1 Mbp up to 11.4 Mbp (Figure 1; Figure 2). As this was observed for all aS-RILs as well as SCH4856, it indicates the location of the α -synuclein transgene introgression.

In order to use the map for QTL mapping, we conducted a correlation analysis to detect linkage over the map (Supplementary Figure 6). Overall, this confirmed local linkage (*i.e.* the markers are arranged conform their physical position) and showed no excessive linkage between chromosomes. Hence, the populations formed a solid basis for eQTL analysis.

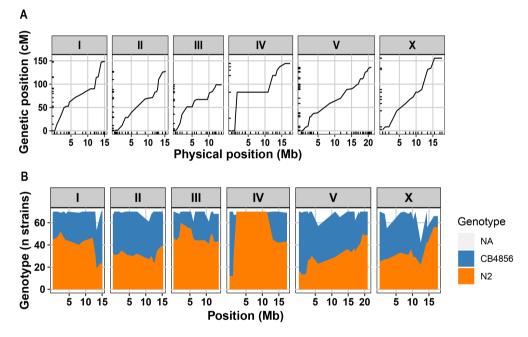


Figure 2: The characteristics of recombination events shown in genetic map of 70 aS-RILs. A) The genetic length in centimorgans (cM) and the physical length in Mbp of the 124 markers along the six chromosomes. **B)** Genotype distribution of the 124 markers along the six chromosomes. Blue and orange colors indicate the N2 and the CB4856 genotype occurrence within the aS-RILs population, respectively. White indicates the marker regions that have unassigned genotypes (NA).

Heritability and transgression of gene expression indicate ample genetic variation

The genome-wide gene-expression was measured in 70 aS-RILs and CB4856, N2, NL5901, and SCH4856 using microarrays. Initial investigation by principal component analysis (PCA) showed that the expression profiles of N2, NL5901, CB4856, and SCH4856 were clearly separated based on their genetic background. Furthermore, the individual RILs were found in-between the SCH4856 and NL5901 strains (Supplementary Figure 7).

To determine the proportion of gene-expression variance caused by genetic variation in the population, we estimated the broad-sense heritability (H^2) per microarray spot. We used the pooled variance of the four replicates of the four ancestral strains to estimate the technical variance (see Materials and methods for details). The majority of significant H^2 values were in the range of 0.60–1.00 (Figure 3). In total, 4227 heritable genes were found (permutation, FDR = 0.05; Supplementary File 6). This showed that the expression of many genes was heritable, hence there was a broad genetic basis underlying the variance in gene expression.

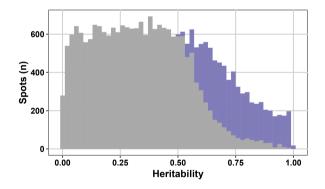


Figure 3: A histogram of the estimated broad-sense heritability (H²**) values for gene expression in the aS-RILs.** The purple bars represent significantly heritable genes (permutation, FDR=0.05).

To further understand the genetic complexity underlying gene expression variance within the RIL population, we analysed transgression by comparing gene-expression of the aS-RILs with both parental strains (NL5901 and SCH4856), identifying 1468 genes showing transgression (permutation, FDR = 0.05). This indicated that multiple loci from both parental strains were contributing to gene-expression variance. Interestingly, 855 genes were both heritable and transgressive (Supplementary Figure 8). They were associated with the heritable expression quantitative trait exhibited in the aS-RILs and

also showed transgressive expression patterns compared to their parental strains. These results together revealed that genetic variations response the α -synuclein expression and/or interact with α -synuclein aggregation.

eQTL analysis To identify loci associated with gene-expression variation, we used eQTL mapping. Using a single-marker model, we identified 7587 eQTL, representing 4805 genes (Figure 4 and Supplementary File 7; threshold log10(p) >3.6; permutation, FDR = 0.05). Among the eQTL, 2274 genes (represented by 3386 spots) were categorized as local-acting, *i.e. cis*-eQTL. The threshold for *cis*-eQTL was set at a distance of 1.0 Mb around the affected gene or an overlap between the QTL confidence interval and the affected gene. We also found many distant-acting eQTL, affecting the expression of 2941 genes (represented by 4201 spots). Hence, these *trans*-eQTL, affected 61.2% of all eQTL found. The contribution of each chromosome to the number of *cis*-eQTL approximately equal and *cis*-eQTL mainly occurred on the chromosome arms (Figure 4). The *trans*-eQTL were also found on all six chromosomes but with an unequal distribution (Figure 4A); the majority of *trans*-eQTL localized on chromosome V (Figure 4B).

The target genes with either *cis*- or *trans*-eQTL were significantly enriched in specific functional categories (Supplementary File 8). The genes with *cis*-eQTL were enriched for genes in the gene classes *clec*, *fbxa*, *fbxb*, *cyp*, *bath*, *math*, *pals*, and *btb* as well as for genes in the gene association of oxidoreductase activity. The genes with *trans*-eQTL with higher expression linked to germline, pharyngeal muscle cell, sperm proteins, outer labial lateral (OLL) and PVD neurons, and hypodermis.

We continued by analyzing the occurrence of *trans*-bands (or eQTL-hotspots) across the genome. Trans-bands are frequently found in eQTL mapping and indicate a locus affecting the expression of multiple distant genes. To identify the *trans*-bands, we calculated the chance of overrepresentation of *trans*-eQTL within a region using a Poisson distribution (see Materials and Methods). The genome was investigated per 0.5 Mb large bin. If *trans*-eQTL were overrepresented in adjacent regions, these were considered a single *trans*-band. In total, nine *trans*-bands were identified (Poisson distribution, p <0.0001, Supplementary File 9). Except for chromosome III, each chromosome was found to contain *trans*-eQTL hotspots, with three occurring on chromosome V. These enrichments found within the *trans*-eQTL were likely to stem from the *trans*-bands also consisted of genes were enriched for genes associated with other functions, such as neuron, protein synthesis, cellular structure, nucleus activities, reproduction and/or development (Supplementary File 10).

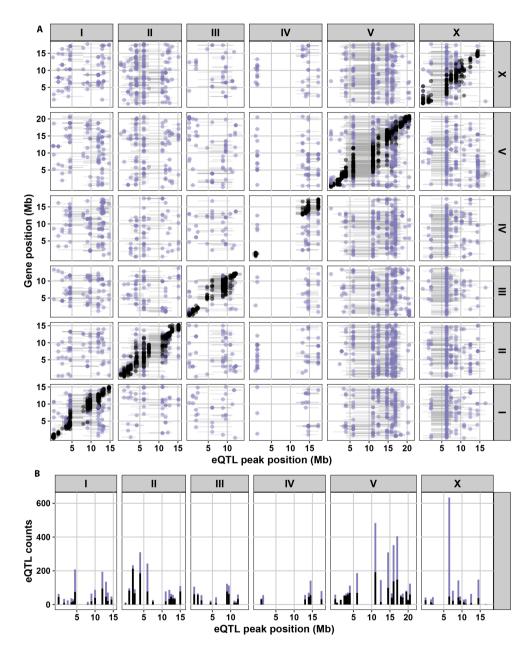


Figure 4: Identified eQTL in microarray data of the 70 NL5901xSCH4856 aS-RILs. With a threshold of $-\log 10(p) > 3.6$ (permutation, FDR ≤ 0.05). The eQTL peak position is shown on the x-axis and gene position is shown on the y-axis. The *cis*-eQTLs (within 1 Mb of the gene) are shown in black and the *trans*-eQTLs in purple. **A)** Plot of both *cis*- and *trans*-eQTLs. The horizontal bars indicate the confidence interval of the eQTL (1.5 LOD-drop). The chromosomes are indicated on the top and right of the plot. **B)** The histogram of the eQTL density per 0.5 Mb block.

Heritability, eQTL, and transgressive segregation

By combining the results of the heritability, transgression, and eQTL we could further our understanding of the genetic variation underlying gene expression variance in the aS-RILs. In total, 528 genes were identified within each analysis which represented 24.7%, 24.5%, 5.4% of a total number of each group genes. (three groups of genes identified function in transgression, heritability, and eQTLs, respectively) (Figure 5). Many eQTL were not detected as heritable (4485), which is probably due to the relatively low statistical power to detect broad-sense heritability in this study. Furthermore, we found 1822 heritable genes for which eQTL were detected. These probably represent *cis*-acting eQTL resulting from the genotypic variation within the aS-RILs. Then, 145 genes were both transgressive and had associated eQTL indicating that they could represent multiple *trans*-acting variants. Focusing on the subset of heritable genes exhibiting transgressive variation and were heritable, we found 327 genes. These could be hallmarks of genotypic variation influencing complex phenotypic traits.

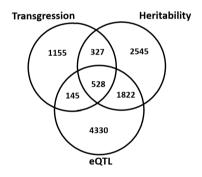


Figure 5: Venn diagram of significantly regulated genes that belong to transgression, heritability and eQTL. Combination of all groups and their overlaps.

We found that the genes associated with a *cis*-eQTL exhibited a stronger linear relationship with heritability (Figure 6A) than did the genes with only *trans*-eQTL (Figure 6B). Compared to the *trans*-eQTL, *cis*-eQTL explained more of the total gene expression variance as shown by corresponding R2 values. The variance explained by *cis*-eQTLs correlated with the heritability, while the estimated heritability was less correlated with *trans*-eQTLs. The proportions of genes that had significant heritability and also act as significant *cis*-eQTL support a close correlation between transcription variation and sequence polymorphism.

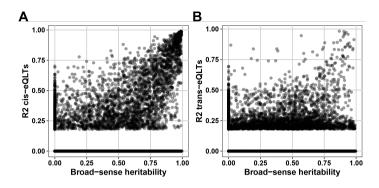


Figure 6: Relationship between the variances explained by eQTLs and their estimated heritability (H²). The x-axis represents the H² value of the genes, and y-axis showed the genes determined as eQTLs (R2 is the variance explained by eQTL). A) was focusing on the *cis*-eQTLs while B) showed the genes with *trans*-eQTLs.

α -synuclein specific trans-bands

To further explore the *trans*-eQTL hotspots in relation to the effect of introgressed α -synuclein, we determined which genes were affected by the α -synuclein introgression by comparing gene expression difference between N2 and NL5901 (Supplementary File 11). We identified 9928 genes (37509 spots) that were affected by the transgene (-log10(p) >3.51; FDR = 0.05). We filtered the mapped eQTL for occurrence within this set of genes, which resulted in a set of 7588 eQTL representing 2407 genes. Furthermore, we also investigated whether genes had eQTL in previous studies (see Materials and Methods for details; Supplementary File 6).

Based on the set of genes affected by the α -synuclein transgene, six *trans*-bands were identified as shown in Supplementary File 10. We performed enrichment analysis on these *trans*-bands (Supplementary File 12). The *trans*-bands that were significantly enriched, included genes functioning in the innate immune system, lifespan, muscle function & locomotion, cellular components (*e.g.* lysosome), variable neuron activities, apoptotic, gonad functions, protein syntheses, and reproduction (*e.g.* vulva, germline and embryo development).

Discussion

aS-RILs and genetic markers used in mapping the eQTLs associated with α -synuclein

In this study, we present a new *C. elegans* RIL population carrying an introgressed α -synuclein transgene derived from NL5901 and SCH4856 (aS-RILs). This population of aS-RILs carries alleles of two divergent genotypes Bristol N2 and Hawaii CB4856 [7, 38]. The construction of the RIL population was optimized in three ways: i) males of each parental strain were used for reciprocal crosses, which contributes to balance their mitochondrial and X chromosome from both parental strains [39]; ii) genetic allele frequencies were under the control for a relatively equal allele fraction at most of loci to avoid strong skews; iii) two-step-procedure to ensure frequent recombination. Although not all 239 and 199 lines were selected for genotype assignment and gene expression analysis, this new aS-RILs population, with a proper panel size, enable studies of quantitative traits variation for α -synuclein associated traits. Thereby, this population complements existing *C. elegans* models of α -synuclein.

A hybrid approach was taken to construct the genetic map for 70 aS-RILs. The genetic markers in this research were insertions/deletions-based [32] and *cis*-eQTL-based. This is a reliable and effective approach [40] to the determination of genetic map and eQTL mapping in *C. elegans* RILs. Gene expression data from microarray were adequately used and concerned within genotype assignment rather than only as molecular phenotypic traits for eQTL mapping. Moreover, mishybridization markers contribute to fill in the gaps and dissert the fragments with more details; vice versa PCR-based genotyping contributes to estimate the markers, avoiding the measurement noise in calling induced by mishybridization of probes [41]. The final marker set provided an accurate genetic map for the aS-RILs, which covers the whole genome and enabled the identification of recombination events within the aS-RILs population.

As expected, every RIL has an N2 genotype-specific segment on chromosome IV and this specific area in the genome also had this for NL5901 and SCH4856. The microarray-based genetic map showed that the α -synuclein gene was inserted at chromosome IV from 2.1Mbp up to 11.4 Mbp. In previous research, the α -synuclein transgene was located in the range of 4.2 Mbp and 13.2 Mbp in chromosome IV [30]. The similar location but with small differences, is mainly due to the less accurate estimation that could previously be provided.

In summary, the resulting population and genetic map provide enough power and resolution to investigate which loci contribute to gene-expression variance related to α -synuclein.

eQTL mapping in the aS-RILs

Our eQTL mapping experiment shows the genetic architecture of gene expression in response to an α -synuclein introgression, by identifying α -synuclein-associated gene-expression variation in the constructed aS-RILs panel. A two-step procedure was performed to detect α -synuclein-specific eQTLs. First, we did a standard eQTL analysis for the gene expression data with analysis of heritability and transgression, which detected 7533 microarray spots, *i.e.* 4773 genes, with either a *cis-* or *trans-*eQTL (Supplementary File 13). Meanwhile, nine *trans*-bands were identified (Supplementary File 9). To detect how much of the expression variation is due to α -synuclein introgression, we estimated and predicted the effect of α -synuclein expression by comparison analyses between N2 and NL5901. It indeed leads to a lower amount of genes as α -synuclein-specific eQTL (Supplementary File 13).

Numerous genes were detected to have *cis*-eQTLs, which was much less than those with *trans*-eQTLs (Supplementary File 8). The *cis*-eQTL were detected within a +/- 1.0 Mb window relative the transcription starting site of each gene (see Materials and Methods), thereby explaining expression variation of genes that are physically located at the locus. Our mapping results identified that the *cis*-eQTL are enriched for genes in the gene classes clec, fbxa, fbxb, cyp, bath, math, pals, and btb. These classes were also identified in previous eQTL studies of *C. elegans* [16, 21, 26, 35, 38]. These detected *cis*-eQTL were found to be highly enriched for polymorphic genes, which confirms the variation among the RILs from genetic divergence of parental strains as well as the mapping quality.

More genes were found to have a *trans*-eQTL, which controls more distant gene expression (Supplementary File 8). However, it is known that external factors can affect *trans*-eQTL by interacting with multiple genes, like development, age, stress response, and cryptic variation [42]. According to previous studies in *C. elegans* with thermal/developmental treatment, cryptic variation of the transcriptional architecture is mainly determined by *trans*-eQTL [16, 21, 22]. Hence, it is likely that gene-expression differences related to α -synuclein can mainly be detected in trans.

Estimating the heritability of complex and polygenic traits was done as an important component of defining the genetic basis of expression variation. Moreover, transgression occurred due to the presence of eQTL with opposing parental allelic effects [43](Ansel *et al.*, 2008), which more likely indicates the presence of interacting alleles, *i.e. trans*-eQTL hotspot. Therefore, multiple eQTLs with opposite allelic effects and exhibited higher heritability had been identified to control the variable gene expression levels in the aS-RILs.

Detecting α -synuclein-specific eQTLs

The α -synuclein proteins expressed are aggregating with age in *C. elegans*, which mimics α -synuclein inclusion bodies seen in PD [44]. The present experimental setup designed for microarray analyses was optimized compared to the previous gene expression study of α -synuclein ILs ([30], as showed in Chapter 3). Adult worms (t=120 hrs incubation after synchronized eggs) were used in this study instead of L4 larva (t=48 hrs) with concerning both gene expression and protein accumulation aspects. Firstly, aging and development effect on gene expression were avoided. As shown in gene expression of those 48 hrs old larvae worms, α -synuclein affected the genes related to development to cause significant delay. Therefore, the use of mature adult worms avoided the impact of development on gene expression changes. The trial microarray had been done in N2 and NL5901 for different ages, which indicated relative stable gene expression level in 120 hrs old adults (data were not shown). Moreover, in addition to expression, α -synuclein aggregates in adults were more than that of larva worms which were showed in Chapter 4.

Heritability and transgression analysis contribute to feature out those genes function as eQTL. The approach is beneficial for *trans*-eQTL architecture determination. However, in our case, it is difficult to determine the transcript regulation of those genes in eQTL that attribute to introgressed α -synuclein or only to genetic variants in the RIL panel. Even though the aS-RILs always have introgressed α -synuclein, these variable expression changes (*Vtotal*) observed from the RILs cannot simply be deemed to α -synuclein. Thus, post-processing on all eQTLs of the whole dataset was done based on the "filtering- α -synuclein-effects". Owing to the technique replicates and biological replicates of the parental and reference strains, a reference base on gene expression changes induced by α -synuclein were gained by comparison of N2 and NL5901. Matching the estimated Ve leaded to the determination of eQTLs highly relative to the effect of α -synuclein. Interestingly, all we measure is in an α -synuclein context, so that plays a role in the eQTL patterns we observed. The introgressed segment always surrounding α -synuclein transgene still gave poor understanding on most other configurations, making identification of unique effect difficult. Hence, further study will try to address and prove if a *trans*-band is mainly linked to aS. It could be addressed by crossing an IL covering that region of SCH4856 into NL5901 and find the *trans*-band back. Our eQTL analysis in this panel of aS-RILs indeed allows us to learn more details of genetic architecture associated with α -synuclein introgression.

Polymorphic regulators associated with α -synuclein in different genotypes

Interestingly, a majority of these genes as *cis*-eQTL belonged to protein coding type. Their encoded domains and proteins in human homologs were associated with nuclear plasma and tissue enriched (The Human Protein Atlas, www.proteinatlas.org; Uniprot, www.uniprot.org). Moreover, one *cis*-eQTL enriched the genes in group of oxidoreductase activity regulator (Supplementary File 8). We believe that α -synuclein induced the oxidative stress even though its expression and aggregation present in the body wall rather than dopaminergic neurons. Under control of the *unc-54* promoter, expression of α -synuclein was driven to the body wall muscle cells and accumulating inclusion with age. Striking, genes as *trans*-eQTL function in PVD and OLL neuron, possibly highlighting the α -synuclein toxicity and revealing the α -synuclein transmission. However its role in transmitter release and underlying mechanism were complex and unclear, which could be a further research question to be addressed in this aS-RILs.

Identified genes in this eQTL analysis will enable us to connect their regulatory variation to the complex phenotypic traits. Besides the appearance of genomic region which was α -synuclein introgression in chromosome IV, a *trans*-eQTL hospots were detected very close to the area, chrIV:14.5-15.0Mb, even associated with an α -synuclein-effect (Supplementary File 9 and 10). In this *trans*-band, genes related to regulation of cadmium hypersensitivity were identified, which is associated with a variety of human diseases [45]. Interestingly, after filtering based on α -synuclein regulation in an N2 genetic background, this enrichment was not present. In addition, genes in the chrX:6.0-6.5Mb *trans*-band were associated with pharyngeal muscle cell function (Supplementary File 8), which could associate to the pumping deficit observed in aS-ILs [30]. Although this group of genes was filtered out by α -synuclein-effect analogy, their associated biological and molecular phenotypic trait, *i.e.* pharynx pumping, presented in different genetic variation. Therefore, it could be caused by the polymorphic

transcription response to α -synuclein introgression. These results indicate the power of our eQTL analysis to detect true eQTL responding to α -synuclein.

Reported regulated genes enriched in our results were also found in previous microarray analysis of other genetic background carrying α -synuclein introgression [30]. 37 genes (of 78) found in all six genotypes (JU1511, JU1926, JU1931, JU1941, CB4856, N2); 242/644 genes are associated with α -synuclein expression in young and old worms; 203/427 genes are associated with the interaction of genotypes and α -synuclein in four aS-ILs which are also mapped in this NL5901xSCH4856 RILs. However, the genetic architecture of genetic backgrounds in response to α -synuclein is complex and remains unclear.

Conclusion

In this study, we investigated gene expression in newly constructed NL5901xSCH4856 aS-RILs to identify *cis-* and *trans*-effects of common variants on differential gene expression levels induced by α -synuclein aggregation and expression. This improved our understanding in the context of sample size, allele frequency and the estimated effect size, as well as design methods for adjusting α -synuclein effect. This eQTL analysis provides insight into the underlying genetic architecture relate to gene expression changes of the N2 and CB4856 genetic variations in response to introgressed α -synuclein.

Supplementary figures and files

The supplementary files and figures are deposited at: https://ldrv.ms/u/s!AqhbSv9gNbn2ac0H oR1de29x0ms?e=LwgTMX

Supplementary Figure 1. Construction of NL5901xSCH4856 recombinant inbred lines (aS-RILs) by two-step-crossings. α -synuclein transgene [P(*unc-54*):: α -synuclein::YFP+*unc-119*] has been introgressed into Hawaii CB4856 genetic backgrounds, which is SCH4856 (strain details can be found in Chapter 3).

Supplementary Figure 2. Genotype assignment per individual RIL per marker with the *log*² expression ratio. The manual corrections of genotypic map were performed based on it, constructing a list of reliable genetic markers as well. If a specific genotype was very unlikely for an individual RIL at a marker and it was determined by a "potentially flawed marker" and was corrected. Genotypes were inferred based on genotypes at surrounding markers.

Supplementary Figure 3. The selected aS-RILs used for PCR-based genotyping. 93 lines out of 293 RILs generated after the first step crossing, while 94 out of 199 RILs which were gained after the second step crossing. Blue means the genotype is from Hawaii CB4856 and orange represents Bristol N2.

Supplementary Figure 4. The 70 aS-RILs selected to perform microarray assay for mapping eQTL. Blue means the genotype is from Hawaii CB4856 and orange represents Bristol N2.

Supplementary Figure 5. Test of gene expression in N2 and NL5901 at different timepoints from larvae to mature worms.

Supplementary Figure 6. The correlation between the validated markers used to genotype assignment as well as eQTL mapping.

Supplementary Figure 7. Principal component analysis of gene expression in all 70 aS-RILs and two reference strains (N2 and CB4856), and two parental strains (NL5901 and SCH4856). The amount of variation explained is indicated on the axis annotation. The orange represent N2 genotypic samples (replicates), while dark orange is wild-type N2 and slight orange is NL5901. The blue represent CB4856 genotypic samples (replicates), including CB4856 and SCH4856 (dark blue and slight blue, respectively).

Supplementary Figure 8. Venn diagram of genes identified in heritability analysis and transgression analysis.

Supplementary File 1. The primer sequences for the insertion/deletion markers. The markers are organized per chromosome and location to cover the genome-wide.

Supplementary File 2. The genetic map of the NL5901xSCH4856 aS-ILs, as determined by the insertion/deletion markers. The genotypes are as follows: CB4856 (-1), N2 (1), None (0).

Supplementary File 3. Target file of the microarray design with all the samples of aS-RILs, references strains (N2, CB4856, their replicates in the experiments) and parental strains (NL5901, SCH4856, their replicates in the experiments).

Supplementary File 4. The physical distance of this aS-RILs genetic map based on the validated marker set.

Supplementary File 5. Distribution of the at each marker over the chromosomal domains covering the whole genome.

Supplementary File 6. Outcome of the heritability analysis in the aS-RILs, including the broad-sense heritability (H^2) values, genotypic variation (Vg), environmental variation, *i.e.* expression of introgressed α -synuclein (Ve). FDR set at 0.05, location and size of the gene, the expression up/down changes, gene identity (WBID, sequence_name, public_name)

Supplementary File 7. Outcome of the single marker mapping. The mapped trait is annotated by spot on the microarray (trait), location of the gene (Chr_gene and BP_gene), gene identity (WBID, sequence_name, public_name). The eQLT is described by location (Chr, BP, marker, Peak), the boundaries of the location (marker_left, BP_left, marker_right, BP_right). The eQLT is also described by effect (qtl_effect), significance (qtl_significance) and its relation to the type (qtl_type). Furthermore the relationship of eQTL with α -synuclein (qtl_aS_unique, PL_alpha_significance_fdr, PL_alpha_effect), as well as the explained variation (r2_sm) and the *trans*-band which belongs to are given.

Supplementary File 8. Enrichment analysis on the genes function as eQTL in aS-RILs. The database used for enrichment (Annotation) and the category (Group), and the number of genes on the array that are in the group (Genes_in_group) is also indicated. Furthermore, the overlap with the locus pair (Overlap) and the significance of that overlap (sig), Bonferroni test, FDR adjusted p-value are shown.

Supplementary File 9. Summary and enrichment of detected *trans*-banding in original eQTL analysis.

Supplementary File 10. Summary and enrichment of detected *trans*-banding in eQTL analysis with filtering " α -synuclein-effect".

Supplementary File 11. Comparison of gene expression dataset of NL5901 and N2 to determine the " α -synuclein-effect". Liner model were used for N2 verse NL5901.

Supplementary File 12. Enrichment analysis on the *trans*-eQTL hotspots, *i.e. trans*-bands in eQTL analysis with filtering " α -synuclein-effect".

Supplementary File 13. Summary of eQTL detected in original eQTL analysis compared with the one after filtering the " α -synuclein-effect". Both *cis*- and *trans*-eQTL are showed as well as what kind of " α -synuclein-effect" impact on them.

References

- 1. **Timpson**, **N. J.** *et al.*, Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Genet*, 2018. 19(2): p. 110–124.
- Redenšek, S., M. Trošt, and V. Dolžan, Genetic Determinants of Parkinson's Disease: Can They Help to Stratify the Patients Based on the Underlying Molecular Defect? *Front Aging Neurosci*, 2017. 9: p. 20.
- 3. **Stefanis**, L., α-Synuclein in Parkinson's disease. *Cold Spring Harbor Perspectives in Medicine*, 2012. 2(2): a009399.
- 4. Rodriguez, M. *et al.*, Worms under stress: *C. elegans* stress response and its relevance to complex human disease and aging. *Trends Genet*, 2013. 29(6): p. 367–374.
- 5. Gaeta, A. L., K. A. Caldwell, and G. A. Caldwell, Found in Translation: The Utility of *C. elegans* Alpha-Synuclein Models of Parkinson's Disease. *Brain Sciences*, 2019. 9(4).
- 6. Kyriakakis, E., M. Markaki, and N. Tavernarakis, *Caenorhabditis elegans* as a model for cancer research. *Molecular & Cellular Oncology*, 2015. 2(2): e975027.
- 7. Andersen, E. C. *et al.*, Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet*, 2012. 44(3): p. 285–290.
- 8. Brenner, S., The genetics of Caenorhabditis elegans. Genetics, 1974. 77(1): p. 71–94.
- 9. **Brenner, S., C. Wersinger, and T. Gasser**, Transcriptional regulation of the α-synuclein gene in human brain tissue. *Neurosci Lett*, 2015. 599: p. 140–145.
- Rocha, E. M., B. De Miranda, and L. H. Sanders, Alpha-synuclein: Pathology, mitochondrial dysfunction and neuroinflammation in Parkinson's disease. *Neurobiol Dis*, 2018. 109: p. 249–257.
- 11. **Dias, V., E. Junn, and M. M. Mouradian**, The role of oxidative stress in Parkinson's disease. *Journal of Parkinson's Disease*, 2013. 3(4): p. 461–491.
- 12. **McCormack**, **A. L., S. K. Mak**, and **D. A. Di Monte**, Increased α-synuclein phosphorylation and nitration in the aging primate substantia nigra. *Cell Death & Disease*, 2012. 3: e315.
- 13. **Bent, M. L. van der** *et al.*, Loss-of-function of β-catenin bar-1 slows development and activates the Wnt pathway in Caenorhabditis elegans. *Sci Rep*, 2014. 4: p. 4926.
- 14. Mackay, T.F.C., Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet*, 2014. 15(1): p. 22–33.
- 15. **Rockman, M. V.**, The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, 2012. 66(1): p. 1–17.
- 16. Li, Y. *et al.*, Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. *PLos Genet*, 2006. 2(12): e222.
- 17. Rockman, M. V. and L. Kruglyak, Recombinational landscape and population genomics of Caenorhabditis elegans. *PLos Genet*, 2009. 5(3): e1000419.
- 18. **Doroszuk, A.** *et al.*, A genome-wide library of CB4856/N2 introgression lines of Caenorhabditis elegans. *Nucleic Acids Res*, 2009. 37(16): e110.
- 19. **Ghosh, R.** *et al.*, Natural variation in a chloride channel subunit confers avermectin resistance in C. elegans. *Science*, 2012. 335(6068): p. 574–578.
- 20. **Gutteling**, E.W. *et al.*, Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in Caenorhabditis elegans. *Heredity*, 2007. 98(1): p. 28–37.
- 21. **Snoek, B. L.** *et al.*, Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans. BMC Genomics*, 2017. 18(1): p. 500.
- 22. Viñuela, A. *et al.*, Genome-wide gene expression regulation as a function of genotype and age in C. elegans. *Genome Res*, 2010. 20(7): p. 929–937.
- 23. Viñuela, A. *et al.*, Aging Uncouples Heritability and Expression-QTL in Caenorhabditis elegans. *G3*, 2012. 2(5): p. 597–605.

- 24. **Guo**, Y. *et al.*, Networks Underpinning Symbiosis Revealed Through Cross-Species eQTL Mapping. *Genetics*, 2017. 206(4): p. 2175–2184.
- 25. Li, Y. *et al.*, Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans. *Genetics*, 2010. 186(1): p. 405–410.
- 26. **Sterken, M. G.** *et al.*, Ras/MAPK Modifier Loci Revealed by eQTL in *C. elegans*. G3, 2017. 7(9): p. 3185–3193.
- 27. **Cookson, W.** *et al.*, Mapping complex disease traits with global gene expression. *Nat Rev Genet*, 2009. 10(3): p. 184–194.
- 28. Nica, A. C. and E. T. Dermitzakis, Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*, 2013. 368(1620): p. 20120362.
- 29. Large, E. E. *et al.*, Selection on a Subunit of the NURF Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of Caenorhabditis elegans. *PLos Genet*, 2016. 12(7): e1006219.
- 30. **Wang**, **Y. A.** *et al.*, Genetic background modifies phenotypic and transcriptional responses in a *C. elegans* model of α-synuclein toxicity. *BMC Genomics*, 2019. 20(1): p. 232.
- 31. **Seidel**, **H.S.** *et al.*, A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in C. elegans. *PLoS Biol*, 2011. 9(7): e1001115.
- 32. Sterken, M. G., Building towards a multi-dimensional genetic architecture in Caenorhabditis elegans. PhD thesis. Wageningen University, 2016.
- 33. Smyth, G.K. and T. Speed, Normalization of cDNA microarray data. *Methods (San Diego, Calif.)* 2003. 31(4): p. 265–273.
- 34. Zahurak, M. et al., Pre-processing Agilent microarray data. BMC Bioinf, 2007. 8: p. 142.
- 35. Rockman, M.V., S.S. Skrovanek, and L. Kruglyak, Selection at linked sites shapes heritable phenotypic variation in C. elegans. *Science*, 2010. 330(6002): p. 372–376.
- 36. **Keurentjes**, J. J. B. *et al.*, Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci U S A*, 2007. 104(5): p. 1708–1713.
- 37. Brem, R.B. and L. Kruglyak, The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A*, 2005. 102(5): p. 1572–1577.
- 38. **Thompson, O. A.** *et al.*, Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. *Genetics*, 2015. 200(3): p. 975–989.
- 39. Andersen, E. C. *et al.*, A Powerful New Quantitative Genetics Platform, Combining *Caenorhabditis elegans* High-Throughput Fitness Assays with a Large Collection of Recombinant Strains. *G3*, 2015. 5(5): p. 911–920.
- 40. **Zych, K.** *et al.*, reGenotyper: Detecting mislabeled samples in genetic data. *PLoS One*, 2017. 12(2): e0171324.
- 41. **Tu, Y., G. Stolovitzky, and U. Klein**, Quantitative noise analysis for gene expression microarray experiments. *Proc Natl Acad Sci U S A*, 2002. 99(22): p. 14031–14036.
- 42. **Paaby, A. B.** *et al.,* Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *eLife*, 2015. 4.
- 43. **Ansel, J.** *et al.*, Cell-to-Cell Stochastic Variation in Gene Expression Is a Complex Genetic Trait. *PLoS Genetics*, 2008. 4(4). Ed. by **J. Flint**: e1000049.
- 44. **Ham, T. J. van** *et al., C. elegans* model identifies genetic modifiers of alpha-synuclein inclusion formation during aging. *PLos Genet,* 2008. 4(3): e1000027.
- 45. **Cui**, **Y**. *et al.*, Toxicogenomic analysis of *Caenorhabditis elegans* reveals novel genes and pathways involved in the resistance to cadmium toxicity. *Genome Biol*, 2007. 8(6): R122.

Chapter 6

General Discussion: The importance of different genetic backgrounds in α -synuclein pathology studies in C. elegans

Yiru A. Wang

Introduction

Neurodegenerative disorders are to a large extent age-dependent diseases caused by genetic and molecular changes where many disease related phenotypes differ among individuals. As neurodegenerative disorders are progressive diseases, identification of onset and the stage to control progression among individuals are very important for disease diagnosis and treatment.

The underlying molecular processes and pathology have been studied intensively for neurodegenerative disease. In particular Parkinson's disease (PD), Alzheimer's diseases (AD), dementia, Amyotrofische Laterale Sclerose (ALS), and Huntington's disease (HD) [1]. The aggregation, deposition and dysfunction of α -synuclein proteins are common events in PD and are also strikingly shared across other neurodegenerative disorders. A diminished control over voluntary movement and progressive large-scale loss of dopamine-producing neurons appear as symptoms in PD patients. This mainly stems from the deposition of α -synuclein protein within dopaminergic neurons of Lewy bodies - cytoplasmic inclusions consisting mainly of α -synuclein inclusions [2]. Dementia can also be induced by the α -synuclein inclusions caused neuron loss, which has also been defined as a key feature of PD [3]. It has become increasingly clear that the pathological mechanisms leading to disease traits can be affected by many factors such as aging, the ambient environment, and the presence of background genetic variants (modifiers) that predispose the individual to the pathogenesis of PD. Causative mutations and genetic variants have been identified in PD patients [4], However, variation in the observed symptoms indicates that the genetic background can impact on the molecular and cellular processes underlying neurodegeneration [5-7]. As of yet, although studies of the underlying causes of PD have strongly increased, knowledge about the effect of genetic background modifiers is limited.

The use of model species has yielded a wealth of information regarding pathways associated with PD using mutant screening. Indeed, these studies have revealed neuroprotective mechanisms that effectively counter pathogenic cellular processes [8].Investigations using *C. elegans* nematode models of PD have proved highly effective in the search for clues to the underlying cause of the PD pathology [9, 10]. This has facilitated the discovery of novel treatments for the motor symptoms of PD [11, 12]. A number of *C. elegans* models of α -synuclein have been constructed (see review [13]). Due to the fact that α -synuclein is not present in *C. elegans*, transgene insertions have been generated to study their role in the underlying processes. So far almost

all studies on α -synuclein transgenics in *C. elegans* have been conducted in a single genetic background, Bristol N2. Hence, these studies do not allow investigating genetic background effects of modifier genes.

With an increasing number of wild strains being identified that are genetically divergent to Bristol N2, genome editing in combination with natural genetic variation in these other wild type strains offers a new potential for understanding AD associated mechanisms. Overall, this chapter highlights the importance of the variable genetic backgrounds in *C. elegans* model of α -synuclein studies from different angles, which is a small stepping stone to understanding human disease using nematode models.

Genetic backgrounds harbour the genetic factors associated with complex disease traits

Genetics plays a key role in many common diseases that burden the human population globally. In humans, increasingly more loci or genes have been identified by GWAS and can be defined as potential "risk" genes. However, are these causal genes – or risk factors – for complex disease traits causal by themselves, or only in a particular genetic background? Because it has become clear that causal disease genes are tightly connected to numerous functional genes in the genome-wide background [14].

The allelic variation of background genes are located at multiple positions, or genetic loci, of the genome. They can affect any aspect of a disease pathway, including transcriptome, proteome, or metabolism, resulting in disease symptoms (as biological phenotypic traits shown in Figure 1). Interestingly, the effect of genetic variants could be little or differential on the phenotypic outcomes. It could be due to opposing allelic effect potentially in the backgrounds [14]. The effect could tend to impact the genes associated only with complex disorders and/or the genes involved into functional relevance in the protein network. The characteristics of genetic variation responsible for this heritable phenotypic variability can be defined as the genetic architecture of the variable symptoms or complex diseases traits [18].

Our genome-wide gene expression profile of four wild-type strains capturing local genetic variation carrying human α -synuclein presented transcription-level changes by comparing corresponding wild-type strains, *e.g.* the genes associated with functional muscle construction and developmental delay. The locomotion deficit and development delay matched the transcription changes, meanwhile specific transcription factors and molecular activities in



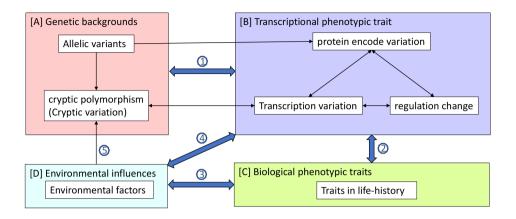


Figure 1: A framework of parameters that are causal determinants of disease traits, indicating the complex genome-transcriptome-phenome associations as well as the potential environmental interactions. The "genetic background" (Box A), which harbours the disease gene of interest, is defined as the inclusion of all other genes that may interact with the gene of interest (also described in [15]. Allelic variations encode for variable protein abundances (arrow ① from Box A to B) which are involved in transcriptional regulatory networks (Box B). These complex networks are also defined as intermediate phenotypes [16] or transcriptional phenotypes, e.g. transcript, protein or metabolite levels. Transcriptional variation potentially influences specific life-history traits (arrow 2) from Box B to C, e.g. translations after gene expression). In terms of levels of gene expression, it is possible to conceive of biologically relevant gene-trait associations [17], which is from Box C to B (arrow 2), e.g. post-translational modification) and/or B to A (arrow ①, e.g. post-transcriptional regulation). Gene-trait associations can be affected by environmental factors (Box D). Environmental factors may sensitize the genome-transcriptome-phenome dynamics by influencing the transcription (arrow (3)) and/or its subsequent phenotypic variations (arrow (4)) including cryptic variation in different backgrounds (arrow ⑤). The genetic architecture which describes the characteristics of genetic variation that is responsible for heritable disease phenotypic variation, thus depends on the number of genetic variants affecting a trait, the magnitude of their effects and their interactions with each other and the environmental effects.

the regulation were revealed (shown in Chapter 3). Transcription analysis is used to study the complex genetic regulatory network by detectable transcript abundance leading to the genetic architecture underlying gene-expression variance [19]. Such as current approaches used to study functional genomics, expression quantitative trait loci (eQTL) studies has been proven to be useful for elucidating the molecular mechanisms of human diseases [20–22].

Researches uncovered the genetic variation of gene expression in humans associated with PD using GWAS. Variable loci associated with risk of PD onset and progression are not always present in different populations (reviewed in Chapter 2). Hence, the genetic architecture of a given PD trait can be more complex, which implies the potential roles of background variants (*e.g.* cryptic variants) relevant to individual-level risk prediction. If sequence variation occurs within introns, or 5'/3' extremities, the genetic structural features could correspond to different regulatory elements involved in the pathologic responses. For example, regulatory sequence variations which are often at the extremities of key genes, could also affect the function of disease causative genes. Moreover, a background modifier that alters expression of a human gene could be at different locus in the genome, which, in turn, causes the gene-related disease or influence the developing of disease.

The power of genetic variation in *C. elegans* model for complex disease pathway analyses

The transparent nematode *C. elegans* has emerged as a powerful model organism for more than 50 years. Characteristics of this model organism have contributed to its successful career in the laboratory, including its genetic tractability, well-studied developmental and life cycle details and well-characterized genome. These characteristics have led to an increased use of *C. elegans* in mechanistic studies of pathology associated with complex diseases as well as in high-throughput screening approaches. The range covers human-pathogen infections, cancers, muscle disorders, diabetes and obesity, mitochondrial-associated disorders, aging and neurodegenerative diseases.

C. elegans displays a low worldwide genetic diversity but a relatively high local genetic diversity of populations. This local diversity is mainly due to the immigration of new alleles rather than by mutation [23]. This *C. elegans* natural variation can contribute to genotype-specific alterations associated with introgressed human disease genes. In this thesis, one of the α -synuclein transgenes was introduced into wild-type strains collected from Orsay and Santeuil [24]. These divergent genetic backgrounds were differentially affected by the pathological α -synuclein aggregation and expression in both molecular and life-history phenotypic traits (Chapter 3 and 4). Moreover, the variable genetic backgrounds also contribute to genetic mapping in *C. elegans* which has proven useful to identify genes involved in interaction between allelic variation and phenotypic traits, such as those involved in pathogen resistance, metabolism, mutualism, and α -synuclein expression (Chapter 5).

The Hawaiian wild-type CB4856 is the second well studied strain next to Bristol N2 aiding to understanding natural genetic effects. Populations of introgression

lines (ILs) and recombinant inbred lines (RILs) have been constructed based on the polymorphic differences between Bristol N2 and CB4856. Also, RILs were derived from CB4856- and N2-transgenic strains which were generated by deliberately introducing genes of interest, like let-60, and α -synuclein. On the basis of these recombinant inbred lines, many QTL/eQTL associated with specific phenotypic traits were found. In Chapter 5, eQTLs associated with α -synuclein were mapped in a newly constructed aS-RILs This contrasts human transcriptome analysis, where estimates of FDR and sensitivity of QTL mapping are more difficult to set due to a relatively low power for statical analysis [25].

Excess of rare variants and a high haplotype homozygosity have been observed in more recently discovered wild type strains. Global collaborative sampling efforts have provided more and more additional Caenorhabditis species and *C. elegans* wild-type strains in culture and many more are still being isolated [26–28]. Meanwhile, the diversity of these globally distributed wild type strains is collected and analyzed by the Andersen Lab (see the free-accessible online source CeNDR [26]). CeNDR is a comprehensive database and offers a set of tools for examining natural diversity.. These sources enrich the genetic analysis based so far mostly on the divergence of CB4856 with N2. In addition, other wild-type strains are compared with the laboratory strain N2 and the Hawaii CB4856 strain contributing to functional genomic studies, such as characterization of genomic and phenotypic diversity [24] and this thesis; and a newly constructed multi-parental RIL population originating from four wild-types by Snoek *et al.* 2018, as well as a multiparental experimental evolution (CeMEE) panel [29].

Multifactorial aetiology associated with α -synuclein

 α -synuclein belongs to the synuclein family with two other members: β -synuclein and γ -synuclein but α -synuclein is the only one found in Lewy bodies. A number of key features potentially describe the onset and progression of α -synuclein-associated disorders, including the increased incidence with age and the clear influence of both genetic background and environmental exposures, like multifactorial carcinogenesis [30]. Aging is one of the important parameters due to its large risk factor for neurodegenerative progression and accompanying age-dependent changes in cellular function. Intrinsically, aging is a multifactorial process and is heavily influenced by genetics and environmental factors [31]. The identification of genes associated with longevity as well as their functional roles in key metabolic pathways has been intensively studied by using *C. elegans* mutagenesis. Moreover, these signalling pathways participate into a complex network underlying phenotypic traits. Natural variation may cause subtle phenotypic changes due to genetic polymorphisms and increase our understanding of genes/molecular pathway functions associated with α -synuclein pathological mechanism like the variable factors and their dynamics in Figure 1.

Kinases, phosphatases, as well as general metabolic homeostasis are responsible for regulating α -synuclein aggregation. They can induce abnormal translation of α -synuclein messenger RNA (mRNA) leading to α -synuclein misfolding and aggregation. Taking LRRK2 as an example, it is one of casual PD genes Genetic evidence collected by Dorval and Hébert (2012) [32], points out that LRRK2 is a large, widely expressed, multi-domain and multifunctional protein that acts as a potentially important player in transcriptional and translational control. It could interact with core components of the miRNA pathway as a "molecular hub" for gene expression and protein synthesis together with other factors. Interestingly, two missense variants (R1628P and G2385R) of LRRK2 are unlikely to cause PD but increase susceptibility to PD in Han Chinese and East-Asians. Overall, genetic modifiers can potentially impact on transcription and protein translation of these genes which is also associated with α -synuclein disease progression.

Additionally, many other factors can also impact on biological processes interacting with abnormal protein aggregation. For example, the C-terminal of the α -synuclein monomer protein is homologous with small heat-shock proteins, which are required for maintaining their native stabilities [33]. According to the gene expression analysis in Chapter 5, some genes associated with thermal sensitivity have been detected which might suggest a linkage with the aggregation of α -synuclein.

Found in translation: the natural variation of model organisms contributes to uncover α -synuclein pathology

α -synuclein pathology variation by genetic background

The utility of the model organism *C. elegans* for studying α -synuclein pathology consists of a variety of models [13]. Among them, multiple *C. elegans* models revolve around the α -synuclein protein which takes an important and central role in PD. In this study we concentrated on one of these models, namely NL5901 which is a transgenic strain containing a human α -synuclein copy (h ttps://www.ncbi.nlm.nih.gov/nuccore/KR710288.1). The transgene *pkIs2386*

[P_{unc54} :: α -syn::YFP, *unc-119*(+)] encodes an α -synuclein-YFP fusion protein expressed under an unc-54 promotor in body-wall muscle cells (van Ham *et al.*, 2008). Meanwhile, other models have been established to express α -synuclein specifically in dopaminergic (DA) neurons using different promoters [34].

Although the transgenic strain NL5901 (*pkIs2386*) does not express in DA neurons, it still could be associated with neuropathology. Compared to the other *C. elegans* models expressing α -synuclein in DA neurons, NL5901 exhibits mature fluorescent misfolded protein aggregates. According to the fluorescent imaging of aS-ILs shown in Chapter 4, animals carrying this transgene show visible inclusions and widely-distributed non-aggregated fluorescence. Especially, these aggregations and fluorescence are more clearly observed in the worms that carry the transgene in wild-type genetic backgrounds. This could affect neuron cells in the head and/or around the body wall through cell-cell contact.

Strikingly, the gene expression profiling analysis in Chapter 5 reveal that neuronal-associated pathways were affected by expression of this α -synuclein transgene. This could coincide with the cell-to-cell transmission characteristics of α -synuclein protein which was pointed out by studies in mammal models. It suggests that α -synuclein protein can be released from cells via apoptosis, apparently in the absence of serious membrane damage. This transmission may compromise lysosomal function or vice versa is antagonized against by lysosomal function [35]. There is more evidences in this thesis: the enrichment of genes that were detected in α -synuclein-associated trans-banding (Chapter 5) such as the genes function in lysozyme activity (chrIV:14.5-15.0Mb), apoptotic dynamics and receptor mediated endocytosis defective (chrX:6.0-6.5Mb). The similarity was showed in L4 stage α -synuclein worms gene expression profiles: groups of genes were involved in α -synuclein-associated membrane transportation such as lipid transport, phospholipid scrambling, extracellular space and recycling endosome membrane (Chapter 3). The genes that function in lysosome/apoptosis were found in mature α -synuclein worms instead of L4 stage worms, these transcription changes and aggregates distribution suggest the dynamics of α -synuclein-associated pathological processing and its novel property of transmission in this *C. elegans* model of α -synuclein. Moreover, our results also indicate that variations of α -synuclein-associated pathological processing in different genetic backgrounds. By using qPCR, the expression of α -synuclein transgene were quantitated as steady-state (*i.e.* no significant differences) in different genetic backgrounds. However, the aggregation of α -synuclein-YFP fusion varied among the α -synuclein worms with different genetic backgrounds (Chapter 4). This similarity is also showed in transcriptome analysis in Chapter 3 and 5. The genotypic-specific

effects in response to α -synuclein expression was identified in L4 stage aS-ILs (Chapter 3). With avoiding the gene expression dynamic effect of development, the detected trans-eQTL hotspots affected by α -synuclein in mature worms include such groups of genes that are involved in cell membrane organization biogenesis variant, endosome/endocytosis organization variant (chrX:6.0-6.5Mb), life span variant (chrIV:14.5-15.0Mb) and so on. Furthermore, the lipid contents linked to α -synuclein proteotoxicity also presented the differential varieties of accumulation in wild-type genetic backgrounds rather than N2 background (Chapter 4). Therefore, the genetic/molecular factors involved into the α -synuclein pathology may vary among the genetic backgrounds, which could also imply the cell-cell interaction may account for both age dependence and regional specificity of neurodegeneration.

Advantages and limitations of transcriptome analysis in this thesis

Transcriptome analysis reveals differences in regulation mechanisms by uncovering the functional genome [36]. Identification of genetic interactions with α -synuclein in human genome studies focused on the sequencing data with transcriptional profiles. For example, identifiable SNP variants that alter coding regions and transcription factors were determined to assemble into units of defined functional roles in protein-protein interaction along with the genes of target, *i.e.* a trans-acting [37]. Overexpression of two microRNAs, mir-7 and mir-153, significantly reduces endogenous α -synuclein levels as regulating α -synuclein levels post-transcriptionally [38].

However, transcriptomic studies of human genome involve a number of challenges: differentially expressed genes that undergo moderate or high fold-changes are difficult to be annotated due to the genomic complexity [39]; dynamic expression is sensitive to environment exposure which is difficult to identify genes strongly relative to environmental adaptation [40] and a majority mainly depend on individual transcriptomes rather than those of population[41].

In *C. elegans*, genome-wide expression profiles can provide a comprehensive overview of the genetic and molecular interaction for a particular trait [42]. Especially, comparative expression within different genetic backgrounds could offer a clue about the genotypic variation underlying traits. Therefore, our eQTL mapping analysis (Chapter 5) offers an opportunity to investigate α -synuclein-associated molecular phenotypic traits; *i.e.* gene expression changes. By measuring gene-expression levels in response to introgressed α -synuclein expression, the mRNA intermediates driving protein synthesis

could offer insights to identify the underlying polymorphic genes. More eQTL studies are performed which collect more database resources and more efficient technologies. These findings will be expected to offer the insight of genetic architecture of complex phenotypic traits with genetic variation which will also contribute to the uncover genotypic-phenotypic relationship underlying α -synuclein pathology in *C. elegans* model of human disease studies.

Several factors that potentially benefit to eQTL detection and mapping resolution were taken into consideration of this study (Chapter 5): 1) proper sample size and marker density; 2) microarray data and its correspondence with variable external data sources, *e.g.* Gene Ontology (GO) system; 3) appropriate correction focus, *i.e.* FDR control and permutation (see also [43, 44]). Furthermore, these studies built on previous eQTL studies in *C. elegans*, which allow comparative analysis of eQTL profiles.

One of the limitations of the eQTL study is the co-occurrence of an introgressed N2 segment together with the α -synuclein transgene. On the one hand, the eQTL mapping without any filtering is likely to also include non- α -synuclein linked genetic variation between N2 and CB4856 to be identified and hence is too liberal. On the other hand, mapping trans-eQTL hotspots specific associated with α -synuclein effect is probably too cautious and misses some interactions (Chapter 5), since the potential allelic trans-actions of CB4856 variants within that region with α -synuclein may have been filtered out. Meanwhile, a series of the health span traits of *C. elegans* (e.g. lifespan, development, locomotion, liposome, pharynx pumping) [45] and traits specific to cytotoxic α -synuclein accumulation/aggregation [46] have been identified and differ within the tested genetic backgrounds (Chapter 3 and 4). Yet, the accurate segment of this introgression in those wild-type genetic backgrounds is still unclear. Its relationships with the α -synuclein transgene as well as genetic backgrounds are difficult to pinpoint. Therefore, a better understanding on the genetic details of these wild-type strains is expected.

Novel interaction between α -synuclein and development/aging

Genes and/or eQTL identified in Chapter 3 and 5 suggest that development and aging probably have a strong relation with α -synuclein expression. This fits the proposition mentioned by a review of J Hou *et al.* that a complex modifier landscape underlies genetic background effects [47] suggesting that multiple sets of genomic regions underlie the phenotypic outcomes.

Neurodegeneration is age-dependent with the influences of cellular senescence, cellular stress and other cellular processes in aging (reviewed in Chapter 2; [48–50]. Many of these cellular pathways are conserved in *C. elegans* which makes it widely used for studies of ageing and age-related diseases. Most of *C. elegans* models of NGD show a striking age-dependent aggregation [46, 51]. The investigation on natural variation of *C. elegans* has been found to affect the genotypic-phenotypic relationship via age-dependent pathways [52]. In Chapter 3, both transcriptome analysis and phenotypic assays revealed that the effect of α -synuclein on development and lifespan varied among different genetic backgrounds. Although aging and its association on α -synuclein pathology is not one of the topics of this study, our identifications indicate the influence of α -synuclein on biological pathways related to development/aging and vice versa the genetic variants of these pathways affect the α -synuclein phenotypic outcomes. This developmental variety could also enrich the model for studying environmental exposure and even its associated epigenetic mechanisms on α -synuclein-induced developmental delay and age-dependence. A better understanding on the genetic and molecular bases of aging still is essential not only for functional genomics of aging but also for age-dependent diseases diagnosis/treatment.

Conclusion & Remarks

 α -synuclein associated disorders can be conceptualized in different ways like "synucleinopathies" [53] or "neurodegenerative mitochondriopathies" [54]. These may suggest the key flaws in the pathological system but also highlight the complexity and multifactorial aetiology of α -synuclein-induced phenotypic outcomes. α -synuclein pathological mechanisms underlying neurodegeneration still needs more intensive studies with variable models as well as human genome analysis. Successful breakthroughs are expected from further collaborations and more subject intercross species -amalgamation.

For disease studies with *C. elegans*, more recent targeted transposon insertion and CRISPR/Cas9 genome editing approaches can further enhance the ability to perform reverse genetics. The relative simplicity of this tool in a model organism is convenient for studying more complex biological mechanisms in different genetic backgrounds. More wild-type strains are going to be isolated and sequenced yielding a global calibration that further promotes the forward genetics in other backgrounds. The combination of both forwards and reverse genetics with natural variation in nematode *C. elegans* will contribute to uncover the genetic architecture of complex disease pathological associates.

References

- 1. Gitler, A. D., P. Dhillon, and J. Shorter, Neurodegenerative disease: models, mechanisms, and a new hope. *Disease Models & Mechanisms*, 2017. 10(5): p. 499–502.
- Taylor, J. P., J. Hardy, and K. H. Fischbeck, Toxic proteins in neurodegenerative disease. Science (New York, N.Y.) 2002. 296(5575): p. 1991–1995.
- 3. **Pienaar, I. S., J. Götz, and M. B. Feany**, Parkinson's disease: insights from non-traditional model organisms. *Prog Neurobiol*, 2010. 92(4): p. 558–571.
- 4. Klein, C., K. Lohmann, and A. Ziegler, The Promise and Limitations of Genome-wide Association Studies. *JAMA*, 2012. 308(18): p. 1867.
- 5. **Gasser, T.**, Mendelian forms of Parkinson's disease. *Biochim Biophys Acta*, 2009. 1792(7): p. 587–596.
- 6. Hauser, S.L. and J.R. Oksenberg, The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. *Neuron*, 2006. 52(1): p. 61–76.
- Labbadia, J. and R. I. Morimoto, Huntington's disease: underlying molecular mechanisms and emerging concepts. *Trends Biochem Sci*, 2013. 38(8): p. 378–385.
- 8. Garden, G. A. and A. R. La Spada, Intercellular (mis)communication in neurodegenerative disease. *Neuron*, 2012. 73(5): p. 886–901.
- 9. Hsu, C. H. *et al.*, MKK6 binds and regulates expression of Parkinson's disease-related protein LRRK2. *J Neurochem*, 2010. 112(6): p. 1593–1604.
- Zheng, M. et al., A liquid phase based *C. elegans* behavioral analysis system identifies motor activity loss in a nematode Parkinson's disease model. *J Neurosci Methods*, 2012. 204(2): p. 234–237.
- 11. **Dehay**, **B**. *et al.*, Targeting α-synuclein for treatment of Parkinson's disease: mechanistic and therapeutic considerations. *The Lancet. Neurology*, 2015. 14(8): p. 855–866.
- 12. **Jadiya, P.** *et al.,* Anti-Parkinsonian effects of Bacopa monnieri: insights from transgenic and pharmacological *Caenorhabditis elegans* models of Parkinson's disease. *Biochem Biophys Res Commun,* 2011. 413(4): p. 605–610.
- 13. Gaeta, A. L., K. A. Caldwell, and G. A. Caldwell, Found in Translation: The Utility of *C. elegans* Alpha-Synuclein Models of Parkinson's Disease. *Brain Sciences*, 2019. 9(4).
- 14. **Kammenga**, J.E., The background puzzle: how identical mutations in the same gene lead to different disease symptoms. *FEBS J*, 2017. 284(20): p. 3362–3373.
- 15. **Chandler, C. H., S. Chari, and I. Dworkin**, Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in genetics : TIG*, 2013. 29(6): p. 358–366.
- 16. **Civelek, M. and A. J. Lusis**, Systems genetics approaches to understand complex traits. *Nat Rev Genet*, 2014. 15(1): p. 34–48.
- Nachtomy, O., A. Shavit, and Z. Yakhini, Gene expression and the concept of the phenotype. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2007. 38(1): p. 238–254.
- 18. **Timpson, N. J.** *et al.*, Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Genet*, 2018. 19(2): p. 110–124.
- 19. **Kirst**, **M.** *et al.*, Genetic architecture of transcript-level variation in differentiating xylem of a eucalyptus hybrid. *Genetics*, 2005. 169(4): p. 2295–2303.
- 20. Franke, L. and R.C. Jansen, eQTL analysis in humans. *Methods Mol Biol*, 2009. 573: p. 311–328.
- 21. **Guo, H.** *et al.*, Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum Mol Genet*, 2015. 24(12): p. 3305–3313.

- 22. Vandiedonck, C., Genetic association of molecular traits: A help to identify causative variants in complex diseases. *Clin Genet*, 2018. 93(3): p. 520–532.
- 23. Barrière, A. and M.-A. Félix, High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Current biology* : *CB*, 2005. 15(13): p. 1176–1184.
- 24. Volkers, R. J. M. *et al.*, Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations. *BMC Biol*, 2013. 11: p. 93.
- 25. **Huang, Q. Q.** *et al.*, Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Res*, 2018. 46(22): e133.
- 26. **Cook, D. E.** *et al.,* The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics,* 2016. 204(1): p. 371–383.
- 27. Frézal, L. and M.-A. Félix, C. elegans outside the Petri dish. eLife, 2015. 4.
- 28. Kiontke, K.C. *et al.*, A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. *BMC Evol Biol*, 2011. 11: p. 339.
- 29. **Noble, L. M.** *et al.*, Polygenicity and Epistasis Underlie Fitness-Proximal Traits in the *Caenorhabditis elegans* Multiparental Experimental Evolution (CeMEE) Panel. *Genetics*, 2017. 207(4): p. 1663–1685.
- 30. **Patrick, K.L.** *et al.*, Exploring the "Multiple-Hit Hypothesis" of Neurodegenerative Disease: Bacterial Infection Comes Up to Bat. *Front Cell Infect Microbiol*, 2019. 9: p. 138.
- Rodríguez-Rodero, S. et al., Aging genetics and aging. Aging and Disease, 2011. 2(3): p. 186–195.
- 32. **Dorval, V. and S.S. Hébert**, LRRK2 in Transcription and Translation Regulation: Relevance for Parkinson's Disease. *Front Neurol*, 2012. 3: p. 12.
- Cox, D. *et al.*, Small Heat-shock Proteins Prevent α-Synuclein Aggregation via Transient Interactions and Their Efficacy Is Affected by the Rate of Aggregation. *J Biol Chem*, 2016. 291(43): p. 22618–22629.
- 34. Lakso, M. *et al.*, Dopaminergic neuronal loss and motor deficits in *Caenorhabditis elegans* overexpressing human alpha-synuclein. *J Neurochem*, 2003. 86(1): p. 165–172.
- Lee, S.-J. *et al.*, Cell-to-cell transmission of α-synuclein aggregates. *Methods Mol Biol*, 2012. 849: p. 347–359.
- 36. Lowe, R. et al., Transcriptomics technologies. PLoS Comput Biol, 2017. 13(5): e1005457.
- 37. Kreimer, A. and I. Pe'er, Variants in exons and in transcription factors affect gene expression in trans. *Genome Biol*, 2013. 14(7): R71.
- Doxakis, E., Post-transcriptional regulation of alpha-synuclein expression by mir-7 and mir-153. *The Journal of biological chemistry*, 2010. 285(17): p. 12726–12734.
- 39. **Wu, P.-Y., J. H. Phan, and M. D. Wang**, Assessing the impact of human genome annotation choice on RNA-seq expression estimates. *BMC Bioinf*, 2013. 14 Suppl 11: S8.
- 40. **Evans, T. G.**, Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. *J Exp Biol*, 2015. 218: p. 1925–1935.
- 41. Kara, E. *et al.*, Genetic and phenotypic characterization of complex hereditary spastic paraplegia. *Brain*, 2016. 139: p. 1904–1918.
- 42. Williams, R. B. H. *et al.*, The influence of genetic variation on gene expression. *Genome Res*, 2007. 17(12): p. 1707–1716.
- 43. **Snoek, B. L.** *et al.*, Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans. BMC Genomics*, 2017. 18(1): p. 500.
- 44. **Sterken, M. G.** *et al.*, Ras/MAPK Modifier Loci Revealed by eQTL in *C. elegans*. G3, 2017. 7(9): p. 3185–3193.
- 45. **Keith, S. A.** *et al.*, The *C. elegans* healthspan and stress-resistance assay toolkit. *Methods*, 2014. 68(3): p. 476–486.

- 46. **Ham, T. J. van** *et al., C. elegans* model identifies genetic modifiers of alpha-synuclein inclusion formation during aging. *PLos Genet,* 2008. 4(3): e1000027.
- 47. **Hou, J.** *et al.*, Complex modifier landscape underlying genetic background effects. *Proc Natl Acad Sci U S A*, 2019. 116(11): p. 5045–5054.
- 48. Calabrese, V. *et al.*, Nitric oxide in the central nervous system: neuroprotection versus neurotoxicity. *Nat Rev Neurosci*, 2007. 8(10): p. 766–775.
- 49. Chinta, S. J. *et al.*, Cellular Senescence Is Induced by the Environmental Neurotoxin Paraquat and Contributes to Neuropathology Linked to Parkinson's Disease. *Cell reports*, 2018. 22(4): p. 930–940.
- 50. **Hindle, J. V.**, Ageing, neurodegeneration and Parkinson's disease. *Age Ageing*, 2010. 39(2): p. 156–161.
- 51. **Gidalevitz, T.** *et al.,* Natural genetic variation determines susceptibility to aggregation or toxicity in a *C. elegans* model for polyglutamine disease. *BMC Biol,* 2013. 11: p. 100.
- 52. Gao, A. W. et al., Natural genetic variation in *C. elegans* identified genomic loci controlling metabolite levels. *Genome Res*, 2018. 28(9): p. 1296–1308.
- 53. **Peelaerts**, **W**. *et al.*, α-Synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature*, 2015. 522(7556): p. 340–344.
- Martin, L. J., M. Wong, and A. Hanaford, Neonatal Brain Injury and Genetic Causes of Adult-Onset Neurodegenerative Disease in Mice Interact With Effects on Acute and Late Outcomes. *Front Neurol*, 2019. 10: p. 635.

Summary Acknowledgments About the author Education statement

Summary

Neurodegenerative diseases (NGDs), such as Alzheimer's diseases (AD) and Parkinson's diseases (PD) are characterized by progressive degeneration in the human nervous system. The nematode *C. elegans* is an excellent model in which to study NGDs due to the high level of conservation of gene functions compared to humans. **Chapter 1** is a general introduction about the effect of variants harboured in different genetic backgrounds on disease onset and progression in studies with either humans or worms models. The role of the protein α -synuclein in neurodegeneration was briefly described mainly linked to the PD. Recent advances by studying natural genetic variation in *C. elegans* offer the opportunity to understand the genetics of complex diseases.

Chapter 2 provides an overview of the genetic variation in NGD and its accessibility in *C. elegans*. Genetic variation in the background can leverage the detection of NGD modifiers due to complex genetic interactions in response to protein misfolding and toxicity, such as induced pathways related to stress responses and aging. Currently, the effect of background alleles underlying protein misfolding in NGD disease formation remains unclear because until now *C. elegans* research has relied on a single worm genotype – the canonical N2 strain – limiting the ability to explore how naturally varying alleles alter pathological mechanisms in NGDs. Hence we propose that the genetic backgrounds of *C. elegans* harbour variants that shape the genetic architecture of disease-associated phenotypic outcomes.

Given the molecular conservation in disease signalling between the model organism *C. elegans* and humans, transgenic worms that express aggregating proteins have been created for studying the onset and progression of NGD and the functional alleles that affect them. In order to identify how genetic variation acts on α -synuclein associated-pathology we constructed transgenic *C. elegans* introgression lines (*i.e.* aS-ILs) that express aggregating human α -synuclein-YFP fusion associated with molecular pathogenic progression of PD in five backgrounds that are genetically divergent to N2 (**Chapter 3**). Measurements of life history phenotypic traits at various aspects and the microarray-based gene expression analysis indicate that transgene effects vary greatly depending on the genetic background. Functional enrichment analysis reveals that the background genes which significantly responded to the α -synuclein-effect link to developmental arrest, metabolic and cellular repair mechanisms.

The observed alterations in variable phenotypic traits, including global gene expression, were strongly associated with α -synuclein-induced pathology. As α -synuclein is an allogenic insertion in the *C. elegans* genome, the genome-wide transcriptional analysis did not present the influence of genetic backgrounds on the expression of α -synuclein. In **Chapter 4**, we measured both expression levels of α -synuclein transgene and its accumulation level of which encoded fluorescent protein fusions. Surprisingly, no significant difference of α -synuclein expression was found , whereas accumulated protein aggregation varied. Together with the quantified variable lipid contents in different genetic backgrounds, natural genetic variation appears to affect α -synuclein aggregation and proteotoxicity resulting in a variety of disease phenotypic outcomes.

Transcriptome analysis reveals differences in regulation mechanisms to uncover the functional genome. In **Chapter 5**, we mapped the eQTL related to α -synuclein aggregation and expression in the newly constructed panel of NL5901xSCH4856 recombinant inbred lines (RILs) harbouring the α -synuclein introgression. By detectable expression levels of the parental (N2&CB4856) genotypes of the RILs population, we identified cis- and trans-effects of variants. In conclusion, the N2 and CB4856 genetic variants were associated with α -synuclein-induced gene expression changes which provides insight into the underlying genetic architecture of transcriptional phenotype in response to introgressed α -synuclein.

Chapter 6 synthesised the findings presented in this thesis and discussed the importance of *C. elegans'* variable genetic backgrounds in human α -synuclein pathology studies. I discuss the series of efficient and powerful advantages of genetic variants in *C. elegans* used to uncover the genotypic-phenotypic relationship relative to complex α -synuclein-pathological mechanism. Genetic variants present differential alterations in global gene expression in response to α -synuclein and/or proteotoxicity, which could not be readily observed in the canonical N2 background. Despite still having inherent limitations, it is a necessary and important step to highlight the influence of natural variation in genetic background on the disease onset and progression mechanism.

Acknowledgments

In a world full of opportunities, the accumulated choices I did all work together over the past lifetime to take me to these various outcomes. Also there are times in life when encountering persons who stays with us for a long or short time but absolutely unforgettable in the rest of the life. During my PhD journey, which took four years and one month, I am super lucky and grateful to have many great people who bringme encouragements and supports. Therefore, I did enjoy each steps of the doctoral study experience, which also leads to my learning and growing in both personal and professional progress. Now, when achieving such a great milestone, I would like to appreciate this opportunity to express my deep sense of gratitude and sincere thanks to you all.

First of all, I would like to convey my sincere thanks to my supervisor and promoters. Jan, thank you a lot for having me join your group from a M.Sc. thesis student till being as a Ph.D. candidate. I enjoyed and benefited a lot from this creative, inspiriting and 'gezellig' environment. The tremendous freedom and professional approval you gave, kept me proactive, while the convincing guidance and supports you supplied behind it helped me out of the moment that I felt "drowning'. I also appreciated your introducing me to Simon, otherwise none of this would have been possible without the opportunity you gave me to do this cotutelle doctoral program. Simon, I am considered myself very lucky to join your group. I have been very impressed by your enthusiasm and dedication to work. You are full of energy and great ideas. I am grateful for your constructive supervision through the years, especially when I was baffled and confused in my work. I learned a lot from you that how valuable tenacity can be for overall success in scientific research, career and life. Then, Mark, I am glad that you are my co-promotor, and also a friend and a mentor. Ever since I did the M.Sc. major thesis under your supervision, I have been inspired by your enthusiasm to science. You were patient with my learning process, and I could go to you almost anytime to ask for favors when I was stuck or to share new results/ideas. I am grateful for your criticism throughout the years, especially in the last year. Your optimism and unbiased comments do help me gain confidence both in my work and myself.

This PhD thesis is the output of the concerted effort from many people members of both CCCU and WUR *C. elegans* groups. Joost, thanks for your always providing lab guidance, practical supports and materials within the last years, and I was often surprised to realize how fast you responded to my questions and how well you organized everyone's need for reassurance. Basten, without your help with the contribution of data analysis as well as its derived logical storyline, I would not have been able to complete the Microarray chapter/publication. Lisa, thanks for your 'driving' contribution on both qPCR and Biosorter, as well as the wonderful Spotify music library shared in the lab. Jana, thanks for helping me with fishing over hundreds worm in the lab, for providing the sweeties and earl grey tea in the office, as well as warm hugs in life. Nell(Nie), you are lovely, professional, patient and persistent in both science and life. Thanks for your constructive comments and practical advises, as well as for your accompanying through the narrow lane of Old Watling Street in the countless dark evenings. Katherina, we started our M.Sc. thesis at the same time and kept on doing Ph.D. project almost at the same time, which makes me so happy sharing the progresses with you. Yuqing, the little cutie, I am sincerely grateful for your presence and attention as well as practical assistances.

Also, I would like to thank the students who contributed this project: Demi, for your constructing the initial batch of RILs, Yuzhi, for your genotyping them, both of your works on RILs were crucial for the following construction and the eventual RIL population. Nondas, I would like to thank you for your interests on this project, although the work you did was not presented in this thesis. Viviane, I am so glad with your great contribution on selecting a set of markers which constructs a complete genetic map of this RIL population.

In additional to people who contribute to the thesis, Stefan, Lee, thanks for your reviewing on my CCCU thesis which includes two third of results presented in this WUR thesis, discussions in the viva and your comments makes me critically thinking about the final discussion chapter. Danielle, thanks for your training on western blotting. Naomi and Hannah, thanks for kindly offering NGM plates. Arthur, thanks for your feedback and discussion on the Ph.D. daily life. Rikus, thanks for your help with the trial western blotting by using various antibodies, and I was really impressed by your kindness, optimism and loud-laughing. Ester and Juliane, thanks for your kindly organization with using popular machines, CLARIOstar (Phytopathology)/COPAS BIOSORT (UU).

For sure, I would like to thank Anne and Iliana for your works on the final print version of this thesis book within a rush and limited schedule. Thanks Anne for your time, effort and efficiency in the layout setting. And Iliana, really thanks for the beautiful cover design as well as your patience and professional advises. Without these, this thesis paper book could not be so immensely wonderful and satisfying!

An important gratitude I would like to give to both graduate schools of CCCU and WUR for offering me the opportunity of this cotutelle doctoral program: thanks CCCU Graduate School for offering me the Ph.D. scholarship. Sheila, thanks for your kindly understanding and well processing of my CCCU viva and UK thesis. Paige and Marisa (CCCU), your supporting documents helped me for either British or Schengen visa processing. Claudius and Lennart (WUR), thanks for sharing news as well as the supports and suggestions you offered, which always makes me feel being involved in PE&RC.

Besides working hard, I also spent an unforgettable time with many warm people at the School of Life Science at CCCU. Emillia, Alex, Jaimie, Marie, Katie, Tyler, Hannah, Carol, Phil, Samit, Naeem, thanks for your warm greetings whenever I met you in the office, lab, coffee room, down town and, of course, pub evenings.

Also I would like to sincerely thank all people from Nematology. I always learned a lot from your professional fields not only about *C. elegans*. I am also glad that I have the privilege to blend into the group that brought me joy, happiness and humor. Jaap, you are exceptionally gentle and sedate to talk with, which I really respected, not just due to the chair position. Geert, the course you gave, host-parasite interaction, took me to C. elegan group as well as NEM family, thanks for that as well as for the thoughtful discussions during my presentations of the project. Aska, thanks for your critical discussions and advises on my results. Arjen, thanks for your inspire feedback and comments. as well as the great photos of landscape from your hiking at the coffee break. Hans, I was surprised by often seeing you around weekend, and thanks for your sharing knowledge of soil, nematodes, science as well as life. Casper, thanks for your always helpful supports in the lab and of course for the hilarious chatting/calling. Sven, thanks for the appreciated lab supports, primer organization and wonderful camping experience you provided to me. Hein, thanks for your energetic greetings. Lisebeth, thanks for your sweet smiles and cakes. Jose, thanks for your good ideas followed by presentations as well as warm greetings. Ruud, thanks for your professional advices on plasmid construction. Martijn, you are always gentle and optimise, thanks for sharing cool bioinformatic analysis. Jet, you are professional, lovely and cheerful, thanks a lot for having fun with you in yoga course and for the Dutch sentences you taught.

For sure I am so grateful to the proactive discussions with all NEMA Ph.D. members. Besides that, I so appreciated the interesting social activities together with you all in/away the campus: Koen, thanks for your suggestions on the hiking boots, which saves me feet indeed, for your great smiles, as well as for your sense of humour. Ava, thanks for your responses to my daily arm-waving-greetings with the same way back. Paula, you are bright and bubbly, thanks for sharing so much interesting stories and singings, 'lekker'.

Kim, it is really really enjoyable to talk with you, which has a magic power to inspire me a lot. Octavina, you are lovely, also thanks for the valuable comments and feedback on the progress of my project. Mattijs, you are full of enthusiasm to both science and life, thanks for your inspirational words. Jaap-jan, thanks a lot for your help with the CLARIOstar, hiking equipment, as well as the funs based on your daily critical comments. Joris, thanks a lot for your warm inspiration, especially the last phrase of my thesis writing, as well as the funny callings together with Casper. Nina, you are sweet and temperate, thanks for your sharing the Confocal technique. Sarah, thanks for the candies and chitchats inspiring and motivating me. Qi, thanks so much for your accompaniment, especially, within the last month before submitting the thesis manuscript. Katherina, Lisa and Yuqing, thanks you all again here, I am so lucky to work with you in the *C. elegans* group, sharing both lab experiences and heart-warming greetings from you.

Many thanks for our NEMA secretaries, thanks Lisette for your always valuable and important supports, especially the trouble shooting happened every year about my viva and enrolling procedures. Thanks Manouk for well-organizations, lovely Juicy cherries and sweeties shared in the office.

Also many thanks to all others who directly or indirectly helped me through the last years. Xiaoqian, I am so grad to invite you as my paranymph. How could only words express my feeling to you! I would like to hug you for all advices for my study and life. I am so happy that you have one more role in your life. Hannah and Xiaoyu, thanks for your inspirited words for me which encouraged me into a Ph.D. studying. Viv Griffiths, my landlord lady, when I firstly moved to UK to start my PhD, you kindly offered the room to me. You are so amiable and gracious, and helped through the really tough time. Luca and Silvia, thanks for living in an adorable and cozy flat sharing with you, thanks for the welcoming, warm hugs, as well as the super delicious Italian meals. Yu, Mengyao and Elyon, I am so lucky to meet you all in the UK, thanks for enriching my life with interesting stories.

To my dear beloved friends, many thanks for the sweet or weird or dramatic friendship and all supports you gave to me, you are together with me going through all laughs and tears within these four years. 周点凡,无论是当年一起 穿过的格纹还是条纹,一起操作的硕士还是博士,我们都在努力地追寻自己想 要的生活风格,坚守自己的不妥协。田北及其家属张睿麒,很奇妙我们的相识 熟识都带着有趣的戏剧性,多么严肃地讨论都可以因为你俩的妙语连珠变得有 趣起来,这也是每年心心念回到荷兰的理由之一。春悦,你精彩纷呈的生活体 验和韧性学霸属性总是让我佩服得很,最关键的是,句句在理又句句有梗。邱 俊,尊称一声邱总,专业性必是强,也总是知道瓦村什么啤酒,什么高科技点

亮生活。羽西, 喵喵, 萌妹属性还有学霸加持, 一起合住搭伙过日子的那段时 间,聊八卦聊人生聊态度,难忘的肆意。韩越,细想起来,每年两次的搬家, 你帮了好些忙,在瓦村求助,每每想到靠谱的也会有你,很感谢一直以来的帮 忙,也是因为你和花花(春月)的美妙缘分,认识了娴淑温柔的花花。静宜,作 为朋友的朋友那时,你就帮忙我搬家买家具各种安置,然后一百公里吃喝玩 乐外加洗涤心灵,如此厮混相熟,也谢谢你的分享,让我拓宽了对未来的选择 和可能。亢淼,铁瓷儿,十年同窗,硕士毕业以后,我们选择了不同的方向继 续,其实都因为适应而成长了,只有和你一起的时候,依然是酣畅淋漓地侃大 山,聊生活里的苦辣酸甜,谢谢你能在我迷茫困惑的时候一针见血。姜洋,还 好你没有等着亢同学排好日程,我们得以成功在伦敦会师,你拼事业,还把生 活过着如此精彩,每每都是给我动力让我期待博士之后的生活。霍喆和苏曼, 看到朋友圈里你们分享的快乐和幸福生活,真的很替你们开心,你们分别抢我 去北京和上海,可是让我偷笑了好久。还有好多老朋友, 王姝, 李悦, 燕妮, 翻洋,秀斌,一辈子的闺蜜们,窝心的温暖,乔乔,莫桐,纯邑,纯正发小, 即使偶尔才能浪一波,也互相懂得各自都在为了自己的梦在奋斗着。还有其他 不常联系但是我一直记着的老朋友们,每每想起你们也激励自己不能掉队下 来。

除了老朋友们,我也特别幸运地结识了很多有意思的新朋友们。熊炜,帮我搬 家真的是辛苦你啦。谭老师,白冰,方艺,张陈,刘壮,谢骊,谢谢你们带我 提高球技。宴因,靓姜,塘主阿金,包仔,余老哥,阿肖,伟达,谢田如三小 美,逸淳,何欣慰,刘睿喆,慧敏,徐东昊,小嘉,谢谢你们带我一起体会演 技的比拼,脑力的抗衡,速算的提升!

Special acknowledgement to my dear families - Thanks Minmin, my partner, my friend, my mentor, and my love. I can always share my emotional thoughts, my complaints, my worries, my happiness with you. Thanks for your understanding and accompaniment. I so appreciated that how brave we are for every decisions we did for the life, and I am looking forwards to the journeys and adventures that we will go together in the future. 爸爸妈妈,你 们一直都是我坚强的后盾,谢谢你们对我没有保留的爱和信任陪伴着我无数个 日日夜夜,硕士加博士的这段时间我们共同成长的过程,慢慢成了可以无话不 谈,辛辣点评,互相揶揄的朋友。云珍三姨,永武姨夫,佳佳姐和小松哥,每 年回国休假,大半的晚饭都去蹭你做的好吃的,还有云艳大姨和贾根儿姨夫, 时石榴姐夫和思卓姐,思群哥和司嫂子谢谢你们一直对我的关心,过去我人生 最重要的时刻都有你们,这个里程碑也是要有你们的出席,还有老刘一家亲的 大家们,从小到大你们对这一辈最小的我很疼爱,作为独苗苗的我出门在外, 没法儿照顾到爸爸妈妈,每次看你们聚餐聚会不是故意馋我就是气我,其实我 反而超级欣慰又安心。爷爷奶奶,你们对我的一直都是关心和支持的,心疼我 的时候没有时常言语出来,可我一直都是知道的。还有姥爷,对不起,我一直 很遗憾,也一直都很想你,生活苦闷的时候总会想起你这辈子的传奇故事,还

有临行前那最后一面,你对我说"在外面要好好的",我现在做到啦,愿无论 在哪儿,都安好,我一直都是你们的小骄傲。Thanks for your all supporting which has opened up my view to look at the world. It is your unconditional love, understanding supports, and spiritual companion always offered for me as the strongest backing, which accompanied me through those uncounted lonely nights.

I have collected here my appreciation for myself: being alone sometimes is painful, but also makes my mind touch its deep silence or happiness. After this journey, I realize that I have more inner strength and wisdom, but also I am so luck to meet so much warm and kind people. After suffering from the troubles and problems, I obtain the curiosity and passion to learn a lot and to look at more about the world.

And finally to all readers for their interests, time and comments. As a PhD thesis including over four years day-to-night study and efforts with the above people supports, it desired to be read and discussed, no matter that it brings serious criticism and/or meaningful thinking for other scientific researches or something else. Hope you do enjoy reading through it!

About the author

Yiru Wang was born in Liaoning, China, on May 29th 1990. In 2008, she began her B.Sc. studies in Plant Protection at the Shenyang Agricultural University (SAU), China. To complete her B.Sc., Yiru worked on a six-months thesis project entitled Genetic Diversity Analysis of *Thanatephorus cumcumeris* causing target spot disease in tobacco under the supervisor. After Bachelor study, Yiru continued her study in Plant science at Wageningen University with the specialization in Plant Pathology and Entomology. In 2014, she joined the group of Prof. Dr Jan Kammenga at Laboratory of Nematology for her M.Sc. major thesis project. Together with Dr Mark Sterken, she focused on quantitative trait mapping to identify the functional genes controlling antiviral response in *Caenorhabditis elegans* model organism. Afterwards, Yiru performed her M.Sc. minor thesis project in the group at the Plant Pathology. Under the supervision of Dr Hanna Rövenich, she used Yeast Two-Hybrid and Co-Immunoprecipitation methods to determine the interaction between the *Verticillium dahliae* effector Ave1 and the tomato immune receptor Ve1.

Later that year, Yiru started the cotutelle doctoral research project at the Laboratory of Nematology, Wageningen University & Research, The Netherlands, and the School of Human and Life Sciences at Canterbury Christ Church University, United Kingdom. This project was under the co-supervision of Prof. Dr Jan Kammenga and Prof. Dr Simon Harvey, in which Yiru focused on genetic dissection of complex traits associated with α -synuclein pathology in *C. elegans* and the importance of genetic background-dependent variation in study of the pathological mechanism by which protein misfolding and aggregation trigger neurodegeneration. The result of this project is presented in this thesis book, of which is also included in the UK thesis.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (4.5 ECTS)

- Genetic dissection of complex traits associated with neurodegenerative disease in *Caenorhabditis elegans* (2016)

Writing of project proposal (4.5 ECTS)

- Dissection of the genetic architecture in Caenorhabditis elegans

Post-graduate courses (4.9 ECTS)

- Introduction to R for Statistical Analysis; PE&RC (2016)
- Quantitative Research Methods; graduate School, CCCU, UK (2016)
- Principles of Ecological and Evolutionary Genomics; SENSE (2018)
- Electron microscopy course the basics from A to W; EPS (2019)
- Basic Statistics; PE&RC (2018)

Competence strengthening / skills courses (2.5 ECTS)

- Getting Started with Research Ethics & Governance; Graduate School, CCCU, UK (2015)
- Introduction to learning and teaching in High Education (2017)
- Venue course Writing propositions for your PhD (2018)
- Scientific publishing; WGS (2019)
- Brain friendly working and writing; WGS (2019)
- Supervising BSc & MSc thesis students; WGS (2019)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.8 ECTS)

- PE&RC First years weekend (2016)
- PE&RC Last years weekend (2018)
- PE&RC Day (2018, 2019)

Discussion groups / local seminars / other scientific meetings (5.5 ECTS)

- Dutch worm meeting (2015)
- Life science seminars; CCCU, UK (2015-2018)
- Fluorescent Microscopy Users Meetings; CCCU, UK (2016-2017)
- C. elegans London group meetings; UK (2017)
- R Users meeting; WUR (2018)
- Symposium towards ecological intensification; PE&RC (2019)



International symposia, workshops and conferences (8.8 ECTS)

- European Worm Meeting (EWM); poster presentation; Berlin, Germany (2016)
- The 21st International *C. elegans* conference; poster presentation; Los Angeles, US (2017)
- EMBO Workshop *C. elegans* development, cell biology and gene expression participant; poster presentation; Barcelona, Spain (2018)
- Workshop breeding for diversity; Wageningen, the Netherlands (2019)
- Life Congress; poster presentation; Utrecht, the Netherlands (2019)

Lecturing / supervision of practicals / tutorials (6.1 ECTS)

- Bioinformatics practicals; CCCU, UK (2015-2018)

Supervision of MSc students (5 ECTS)

- Determination of the genetic map of the newly constructed *C. elegans* α -synuclein recombinant inbred lines
- Quantitative analysis of α-synuclein gene expression and its protein accumulation in different genetic backgrounds of *C. elegans*

This research was conducted at both the Laboratory of Nematology at Wageningen University (Wageningen, The Netherlands) and the School of Human and Life Sciences at Canterbury Christ Church University (Canterbury, The United Kingdom). The research in this thesis was supported by the CCCU three-year PhD Scholarship (UK), a Leverhulme Trust Research Grant (RPG-2016-040, UK), the PE&RC of Wageningen University (The Netherlands) and the Human Frontiers Science Program (RGP0028-2014, The Netherlands) and Netherlands Organization for Scientific Research (NWO, The Netherlands) project No. 823.01.001. The above financial supports for this thesis project are gratefully acknowledged.

Cover design: Iliana Boshoven-Gkini | Agilecolor.com Cover *C. elegans* photograph by Yiru Wang & Lisa van Sluijs Layout design: Anne Morbach | Schlaugemacht.net Printed by GVO drukkers & vormgever B.V. | gvo.nl

