Published in final edited form as:

Curr Opin Genet Dev. 2017 June ; 44: 110–116. doi:10.1016/j.gde.2017.01.014.

Structural Variants in *SNCA* Gene and the Implication to Synucleinopathies

Ornit Chiba-Falek^{1,2,*}

¹Department of Neurology, Duke University Medical Center, Durham, NC 27710, USA

²Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC 27710, USA

Abstract

Synucleinopathies are a group of neurodegenerative diseases that share a common pathological lesion of intracellular protein inclusions largely composed of aggregates of alpha-synuclein protein. Accumulating evidence, including genome-wide association studies, has implicated the alpha-synuclein (*SNCA*) gene in the etiology of synucleinopathies and it has been suggested that *SNCA* expression levels are critical for the development of these diseases. This review focuses on genetic variants from the class of structural variants (SVs), including multiplication of large genomic segments and short (<50bp) genomic variants such as simple sequence repeats (SSRs), within the *SNCA* locus. We provide evidence that *SNCA*-SVs play a key role in the pathogenesis of synucleinopathies via their effects on gene expression and on regulatory mechanisms including transcription and splicing.

Keywords

Synucleinopathies; *SNCA*; Structural Variants; CNV; SSR; regulation of gene expression

Introduction

Structural Variants (SVs) are genomic variants other than single nucleotide variant (SNVs), are often multiallelic and include deletions, insertions, microsatellites or simple sequence repeats/short tandem repeats (SSRs/STRs), insertion/deletions (indels), copy number variation (CNV), block substitutions, and inversions¹. Recently, there has been increased support for the idea that SVs affect many human phenotypes and complex traits^{2–6}. We are interested in the functional consequences of SVs and their causative role in neurodegenerative diseases of aging.

To whom correspondence should be addressed: Ornit Chiba-Falek, Dept of Neurology, DUMC Box 2900, Duke University, Durham, North Carolina 27710, USA, Tel: 919 681-8001, Fax: 919 684-6514, o.chibafalek@duke.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Our research strategy expands on the concept that changes, even subtle, in expression levels of normal (wild-type) proteins in the brain can lead to neurodegenerative diseases of the aging brain. The genetic contribution of *SNCA* (*SNCA*, MIM #163890) to synucleinopathies is a prominent example. Synucleinopathies are a group of neurodegenerative disorders that share a common pathological lesion composed of protein inclusions in the cytoplasm of select populations of neurons and glia (*i.e.*, oligodendrocytes), known as Lewy bodies (LBs) and Lewy neurites, and glial cytoplasmic inclusions (GCIs), respectively^{7–11}. Aggregates of the insoluble α -syn protein are the major component of LBs¹², Lewy neurites, and GCIs. This group of disorders include Parkinson's Disease (PD), dementia with Lewy bodies (DLB), Lewy body variants of Alzheimer disease (LBV/AD), Neurodegeneration with Brain Iron Accumulation (NBIA) type I, pure autonomic failure (PAF), and multiple system atrophy (MSA). Over the last decade, genome wide association studies (GWAS) and candidate gene-based approaches¹³ have implicated *SNCA* as a highly significant genetic risk factor for synucleinopathies including sporadic PD^{14–27}, DLB²⁸, MSA^{29,30} and LBV/AD^{31,32}. In addition, accumulating evidence from *in vitro* systems and *in vivo* models, suggesting that the α -syn expression levels are critical for the development of synucleinopathies (reviewed in³³). In this review, we will describe the contribution of both long and short SVs in the *SNCA* locus to the regulation of *SNCA* gene expression profiles, mRNA levels and splicing, in relation to their possible role in the etiology of synucleinopathies.

Long Structural Variants (SV)- Copy number variations of the *SNCA* locus

Long SV are defined as DNA variants ≥ 50 bp. A recent study based on the analysis of an integrated SV map of 2504 human genomes showed that SVs are enriched in haplotypes identified by genome wide association studies (GWAS) and exhibit enrichment for expression quantitative trait loci (eQTLs)⁵.

A sub-class of the long SV are the copy number variations (CNV), segmental multiplication of genomic sequences larger than 1kbp. Two CNVs in the *SNCA* gene, triplication and duplication, have been identified in only a few families with an early onset, autosomal dominant form of PD^{34–38}. Genomic triplication of the *SNCA* – containing region results in four fully functional copies of *SNCA*, and 2-fold over-expression of *SNCA* mRNA and protein. The triplication leads to high penetrance of an early-onset PD phenotype with cognitive impairment and autonomic dysfunction^{34,39,40}. Duplications of the wild-type *SNCA* gene result in a 1.5-fold elevation of *SNCA* expression and, compared with the triplication, a slightly later onset of heritable PD that is characterized by a lower penetrance and a 'milder' phenotype with slower progression^{35–38}, demonstrating the dose-dependent effect of *SNCA* on disease etiology.

The molecular mechanism that mediates the pathogenic effect of *SNCA*-CNV has been also investigated *in vitro* using isogenic induced Pluripotent Stem Cells (iPSC)-derived models. Differentiating of iPSCs from a patient with *SNCA* triplication into dopaminergic neurons confirmed the vulnerability of the dopaminergic neurons to overexpression of *SNCA*^{41,42}. The iPSC-derived dopaminergic neurons carrying the *SNCA* triplication showed an approximately two-fold increase in α -syn levels compared to the control cells and the

accumulation of α -syn protein in this iPSC-derived model coincided with increased oxidative stress markers, and conferred increased vulnerability to oxidative stress-induced cell death⁴³. Another study that investigated the *SNCA* triplication iPSCs, reported that iPSCs-derived Neural Precursors Cells (NPCs) displayed overall normal cellular and mitochondrial morphology, but showed substantial changes in growth, viability, cellular energy metabolism and stress resistance when challenged by starvation or toxicant challenge. These phenotypic changes were reversible upon knockdown of *SNCA* overexpression⁴⁴. A recent study demonstrated that dopaminergic neurons derived from *SNCA* triplication iPSCs recapitulate several PD-related phenotypes and exhibit cellular and molecular characteristic of neurodegeneration⁴⁵. Collectively, these studies introduced the iPSCs-derived dopaminergic neurons as a model to investigate the effects of *SNCA*-CNV mutations on PD related cellular phenotypes, and demonstrated that the phenotypic effect of the *SNCA*-CNV mutation is mediated by upregulation of *SNCA* expression levels. Patients carrying *SNCA*-CNV mutation- triplications and duplications also manifest clinical and pathological features similar to DLB. While the above studies focused on iPSC-derived dopaminergic neurons to model PD, it is also important to gain insight into the effect of *SNCA*-CNVs in cortical neurons to model DLB. In unpublished work, we established a model system of iPSCs-derived neurons from a normal subject and a patient with *SNCA*-triplication to evaluate the effect of *SNCA*-triplication and up regulation of *SNCA*-mRNA. Those iPSC lines were differentiated into dopaminergic and cholinergic neurons to model PD and DLB, respectively. *SNCA*-mRNA (Fig. 1A) and protein (Fig. 1B) expression in the *SNCA*-triplication cells exhibited two-fold increase compared to the cells from a normal control for each neuronal type across differentiation, as shown in Fig. 1 an example for iPSCs and the derived precursor and matured cholinergic neurons. These results recapitulate the *in vivo* observations in human tissues. Notable, *SNCA*-mRNA expression in both the *SNCA* triplication and the control iPSCs lines increased along the maturation process of the cholinergic neurons (Fig. 1A). These observations warrant a deeper characterization of the differentiated neuronal lines derived from each of the isogenic-iPSCs in the context of the implication to susceptibility to PD compared to DLB.

Short Structural Variants (SSV) within noncoding regions of *SNCA* locus

Short SVs (SSVs) are short genomic variants (<50bp) other than SNVs, and generally present multi alleles. SSVs are thought to affect phenotype by altering the regulation of gene transcription^{46–51}, splicing⁵², and translation, and it is by these mechanisms that SSVs may play a role in the etiology of human diseases, including complex disorders. A new study identified >2000 expression STRs (eSTRs) in the human genome and found that eSTRs contribute to ~10–15% of the *cis* heritable variation in gene expression attributed to common variants⁵³. This study also showed that these eSTRs are enriched in genomic regions associated with clinically relevant phenotypes⁵³.

Furthermore, recent studies proposed a potential mechanism whereby SSRs affect transcription (illustrated in Fig 2). These studies showed that certain repetitive DNA sequences, when present in the flanking regions of specific transcription factor (TF) binding sites, can have a magnitude effect on the intensity of TF-DNA binding, through a mechanism we termed “non-consensus binding”^{54,55}.

Using our newly developed SSV Evaluation System⁵⁶, we screened for SSVs within a ~215 Kb segment of the *SNCA* genomic region, +/-50kb (hg19, chr4: 90,595,250–90,809,447), with the goal of identifying candidate causal variants for synucleinopathies. We identified 555 SSVs, of which 75 are from the SSR category. Below we describe a few SSVs that have been studied experimentally in relation to their association with disease phenotypes and mode of function.

(1) *SNCA*-CT rich haplotype

Combined *in silico* and wet lab approaches led to the identification of an intronic polymorphic CT-rich region in *SNCA* that is a candidate regulatory SSV. The SSV evaluation system provided high scores (representative Total Impact Score=42) in a CT-rich, low-complexity region of intron 4 of *SNCA*. We cloned and sequenced this region in case (LBV/AD) and control (AD) subjects and identified four distinct haplotypes within this region, with specific haplotype-conferred risk to develop LB pathology in AD patients³¹. We further demonstrated that the risk haplotype was significantly associated with elevated levels of *SNCA*-mRNA in human brain tissues relevant to the LB pathology and suggested that the CT-rich site acts as an enhancer element of *SNCA* transcription³¹.

(2) *SNCA*-Rep1

Rep10⁴⁷ is a polymorphic dinucleotide complex repeat site located ~10 kb upstream of the *SNCA* transcription start site^{57,58}. The length of Rep1 appears to be associated with increased risk of PD⁵⁹. The overwhelming majority of the reported association studies, including a large meta-analysis, have shown that the extended alleles of *SNCA*-Rep1 confer increased risk to develop late-onset, 'idiopathic' PD, while the shorter allele is protective^{24,60–63}. We investigated the effect of the Rep1 polymorphism on *SNCA* expression, and discovered that Rep1 regulates *SNCA* transcription in human brain tissues⁶⁴. These results have been confirmed using luciferase reporter assay^{47,65} and in a humanized mouse model⁶⁶. The PD-risk Rep1 allele led to increased *SNCA*-mRNA levels, providing further support to the pathogenic effect of *SNCA* overexpression^{47,64,66}. In support of this, we identified a factor, poly(ADP-ribose) transferase/polymerase-1 (PARP-1), that binds specifically to Rep1 and modulates *SNCA* transcription⁶⁷, suggesting that the association of Rep1 alleles with sporadic PD may be mediated, in part, by the effect of PARP-1 on *SNCA* expression. Retrospectively, the Rep1 site was assessed using the SSV evaluation system and had a high score (Total Impact Score=31) relative to the complete list of SSRs within *SNCA* genomic regions including +/-50kb flanking regions in SSV evaluation system.

However, these results conflict with the recent findings that showed no effect of Rep1 on *SNCA*-mRNA level using genome edited iPSC-derived neurons⁶⁸. It is possible that difficulties in quantifying the total *SNCA* transcripts levels affected the validity of the reported conclusions. In fact, we recognized that the allele-specific assay used to measure *SNCA*-mRNA was designed to target only the long 3'UTR isoform of *SNCA* transcript, and therefore the method Soldner et al. used to quantify *SNCA*-mRNA levels did not capture all *SNCA* transcript species. It is crucial to note that the long 3'UTR isoform of *SNCA* is not as abundant as the short 3'UTR isoform and represents only a small fraction of *SNCA*

transcripts. Therefore, we cannot rule out the possibility of a false negative for the *SNCA*-Rep1 finding reported by Soldner et al⁶⁸.

(3) Intron 2 poly-T

The *SNCA*126 splicing variant may have a protective role because the in-frame deletion of exon 3 leads to the interruption of the N-terminal protein-membrane interaction domain which may lead to less aggregation^{69,70}. *SNCA*126 levels were decreased in the prefrontal cortex of DLB patients⁷¹. In contrast, *SNCA*126 expression was increased in the frontal cortex of PD brains and no significant differences in MSA⁷².

A polyT variant in intron 2 of *SNCA* gene comprises three alleles (5T, 7T, an12T) and the length of the polyT stretch is directly associated with *SNCA*126 expression levels in the normal brain, influencing the splicing efficiency of *SNCA* exon 3. Whereas the shortest 5T allele was associated with lower expression of *SNCA*126-mRNA, the longest 12T led to the highest *SNCA*126-mRNA levels⁷³. The same study also reported that the 12T-allele-carrying genotypes accumulated with increasing age in the normal population, while the frequency of 5T-allele-carrying genotypes decreased in successive age groups until reaching zero in the oldest group. Collectively these observations imply that the longest poly-T allele has a protective effect in aging, presumably via its association with higher *SNCA*126-mRNA levels.

Summary

We have reviewed evidence of the causative role of *SNCA* SVs in the etiology of synucleinopathies in general, and PD in particular, and described the molecular mechanisms that underlie their pathogenic impact (summarizes in Fig 3). We proposed that a 1.5- to 2-fold increase in expression caused by CNVs can lead to early onset familial disease, while even subtle changes in the *SNCA* expression are able to trigger the onset of non-Mendelian synucleinopathies. Subtle alterations in the *SNCA* expression are driven by different mechanisms and factors, and here we focused on noncoding *cis*-SSV and their corresponding regulatory functions. This implies that common and distinct SSVs and regulatory mechanisms through which they act might be involved in the etiology of the different synucleinopathies. Noteworthy, as summarized in Fig 3, selective vulnerability of each cell types involved in the different synucleinopathies to overexpression and/or particular *SNCA*-isoform will contribute, among other factors, to the development of a particular disease in the synucleinopathies spectrum disorders.

While it is widely accepted that up-regulation of the wild-type *SNCA* plays a causative role³³, it is still not completely understood if *SNCA* produces toxic isoforms and, if so, their precise identity. Current technologies and novel model systems, including iPSCs –derived neurons and CRISPR/Cas9 genome editing, will provide the opportunities to look at the functional and phenotypic consequences of each candidate SVs to fill in those gaps. In return, this important knowledge will result in the development of precision medicine including, biomarkers for pre-clinical diagnosis and effective treatment approaches for synucleinopathies in general and for a specific disease in the spectrum.

Acknowledgments

This work was funded in part by the National Institutes of Health/National Institute of Neurological Disorders and Stroke (NIH/NINDS) [R01 NS085011 to O.C.].

References

1. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009; 10:241–51. [PubMed: 19293820]
2. Mirkin SM. Expandable DNA repeats and human disease. *Nature.* 2007; 447:932–40. [PubMed: 17581576]
3. Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet.* 2005; 6:729–42. [PubMed: 16205713]
- 4*. Willems T, et al. The landscape of human STR variation. *Genome Res.* 2014; 24:1894–904. A genome-wide analysis of STR variation in the human population using sequencing data from Phase 1 of the 1000 Genomes Project. Elucidates the scale of human genetic variation beyond the classical SNPs. [PubMed: 25135957]
5. Sudmant PH, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2015; 526:75–81. [PubMed: 26432246]
- 6*. Huddleston J, Eichler EE. An Incomplete Understanding of Human Genetic Variation. *Genetics.* 2016; 202:1251–4. Brief overview that highlights the importance of structural variants (SVs) in interpreting the “missing heritability” of human disease. Suggests that SVs have a large impact on human phenotypes, and discusses that SVs require a more systematic assessment of genomes at deeper coverage and alternate sequencing and mapping technologies. [PubMed: 27053122]
7. Galvin JE, Lee VM, Trojanowski JQ. Synucleinopathies: clinical and pathological implications. *Arch Neurol.* 2001; 58:186–90. [PubMed: 11176955]
8. Marti MJ, Tolosa E, Campdelacreu J. Clinical overview of the synucleinopathies. *Mov Disord.* 2003; 18(Suppl 6):S21–7. [PubMed: 14502652]
9. Spillantini MG. Parkinson’s disease, dementia with Lewy bodies and multiple system atrophy are alpha-synucleinopathies. *Parkinsonism Relat Disord.* 1999; 5:157–62. [PubMed: 18591134]
10. Jellinger KA. Neuropathological spectrum of synucleinopathies. *Mov Disord.* 2003; 18(Suppl 6):S2–12. [PubMed: 14502650]
11. Goedert M, Jakes R, Anthony Crowther R, Grazia Spillantini M. Parkinson’s Disease, Dementia with Lewy Bodies, and Multiple System Atrophy as alpha-Synucleinopathies. *Methods Mol Med.* 2001; 62:33–59. [PubMed: 21318766]
12. Spillantini MG, et al. Alpha-synuclein in Lewy bodies. *Nature.* 1997; 388:839–40. [PubMed: 9278044]
13. Han W, et al. Alpha-synuclein (SNCA) polymorphisms and susceptibility to Parkinson’s disease: a meta-analysis. *Am J Med Genet B Neuropsychiatr Genet.* 2015; 168B:123–34. [PubMed: 25656566]
14. Pankratz N, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet.* 2009; 124:593–605. [PubMed: 18985386]
15. Myhre R, et al. Multiple alpha-synuclein gene polymorphisms are associated with Parkinson’s disease in a Norwegian population. *Acta Neurol Scand.* 2008; 118:320–7. [PubMed: 18485051]
16. Ross OA, et al. Familial genes in sporadic disease: common variants of alpha-synuclein gene associate with Parkinson’s disease. *Mech Ageing Dev.* 2007; 128:378–82. [PubMed: 17531291]
17. Pals P, et al. alpha-Synuclein promoter confers susceptibility to Parkinson’s disease. *Ann Neurol.* 2004; 56:591–5. [PubMed: 15455394]
18. Mueller JC, et al. Multiple regions of alpha-synuclein are associated with Parkinson’s disease. *Ann Neurol.* 2005; 57:535–41. [PubMed: 15786467]
19. Mizuta I, et al. Multiple candidate gene analysis identifies alpha-synuclein as a susceptibility gene for sporadic Parkinson’s disease. *Hum Mol Genet.* 2006; 15:1151–8. [PubMed: 16500997]
20. Winkler S, et al. {alpha}-Synuclein and Parkinson disease susceptibility. *Neurology.* 2007

21. 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; 41:1303–7. [PubMed: 19915576]
22. Simon-Sanchez J, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet.* 2009; 41:1308–12. [PubMed: 19915575]
23. Edwards TL, et al. Genome-Wide Association Study Confirms SNPs in SNCA and the MAPT Region as Common Risk Factors for Parkinson Disease. *Ann Hum Genet.*
24. Maraganore DM, et al. Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *Jama.* 2006; 296:661–70. [PubMed: 16896109]
25. Spencer CC, et al. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Hum Mol Genet.* 20:345–53.
26. Simon-Sanchez J, et al. Genome-wide association study confirms extant PD risk loci among the Dutch. *Eur J Hum Genet.*
27. Mata IF, et al. SNCA variant associated with Parkinson disease and plasma alpha-synuclein level. *Arch Neurol.* 67:1350–6.
28. Bras J, et al. Genetic analysis implicates APOE, SNCA and suggests lysosomal dysfunction in the etiology of dementia with Lewy bodies. *Hum Mol Genet.* 2014; 23:6139–46. [PubMed: 24973356]
29. Scholz SW, et al. SNCA variants are associated with increased risk for multiple system atrophy. *Ann Neurol.* 2009; 65:610–4. [PubMed: 19475667]
30. Al-Chalabi A, et al. Genetic variants of the alpha-synuclein gene SNCA are associated with multiple system atrophy. *PLoS One.* 2009; 4:e7114. [PubMed: 19771175]
- 31**. Lutz MW, et al. A cytosine-thymine (CT)-rich haplotype in intron 4 of SNCA confers risk for Lewy body pathology in Alzheimer's disease and affects SNCA expression. *Alzheimers Dement.* 2015; 11:1133–43. Discovery of a CT-rich haplotype in SNCA intron 4 that contains a cluster of SSVs, and demonstrates that a specific haplotype confers risk to develop LBV/AD and acts as an enhancer element to increase SNCA-mRNA expression. [PubMed: 26079410]
32. Linnertz C, et al. The genetic contributions of SNCA and LRRK2 genes to Lewy Body pathology in Alzheimer's disease. *Hum Mol Genet.* 2014; 23:4814–21. [PubMed: 24777780]
- 33*. Tagliafierro L, Chiba-Falek O. Up-regulation of SNCA gene expression: implications to synucleinopathies. *Neurogenetics.* 2016 An excellent review of mechanisms mediate SNCA overexpression and the causative impact on synucleinopathies.
- 34**. Singleton AB, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science.* 2003; 302:841. The first report of the discovery of *SNCA* triplication in PD patients. [PubMed: 14593171]
35. Fuchs J, et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. *Neurology.* 2007; 68:916–22. [PubMed: 17251522]
36. Ross OA, et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. *Ann Neurol.* 2008; 63:743–50. [PubMed: 18571778]
37. Chartier-Harlin MC, et al. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet.* 2004; 364:1167–9. [PubMed: 15451224]
38. Ibanez P, et al. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet.* 2004; 364:1169–71. [PubMed: 15451225]
39. Farrer M, et al. Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. *Ann Neurol.* 2004; 55:174–9. [PubMed: 14755720]
40. Miller DW, et al. Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. *Neurology.* 2004; 62:1835–8. [PubMed: 15159488]
41. Devine MJ, et al. Parkinson's disease induced pluripotent stem cells with triplication of the alpha-synuclein locus. *Nat Commun.* 2011; 2:440. [PubMed: 21863007]
42. Byers B, et al. SNCA triplication Parkinson's patient's iPSC-derived DA neurons accumulate alpha-synuclein and are susceptible to oxidative stress. *PLoS One.* 2011; 6:e26159. [PubMed: 22110584]
43. Byers B, Lee HL, Reijo Pera R. Modeling Parkinson's disease using induced pluripotent stem cells. *Curr Neurol Neurosci Rep.* 2012; 12:237–42. [PubMed: 22538490]

44. Flierl A, et al. Higher vulnerability and stress sensitivity of neuronal precursor cells carrying an alpha-synuclein gene triplication. *PLoS One*. 2014; 9:e112413. [PubMed: 25390032]
45. Lin L, et al. Molecular Features Underlying Neurodegeneration Identified through In Vitro Modeling of Genetically Diverse Parkinson's Disease Patients. *Cell Rep*. 2016; 15:2411–26. [PubMed: 27264186]
46. Akai J, Kimura A, Hata RI. Transcriptional regulation of the human type I collagen alpha2 (COL1A2) gene by the combination of two dinucleotide repeats. *Gene*. 1999; 239:65–73. [PubMed: 10571035]
- 47*. Chiba-Falek O, Nussbaum RL. Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system. *Hum Mol Genet*. 2001; 10:3101–9. The first report that illustrates the functional role of Rep1 SSV in the transcriptional regulation of *SNCA*-mRNA levels. [PubMed: 11751692]
48. Okladnova O, et al. A promoter-associated polymorphic repeat modulates PAX-6 expression in human brain. *Biochem Biophys Res Commun*. 1998; 248:402–5. [PubMed: 9675149]
49. Peters DG, Kassam A, St Jean PL, Yonas H, Ferrell RE. Functional polymorphism in the matrix metalloproteinase-9 promoter as a potential risk factor for intracranial aneurysm. *Stroke*. 1999; 30:2612–6. [PubMed: 10582986]
50. Searle S, Blackwell JM. Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility. *J Med Genet*. 1999; 36:295–9. [PubMed: 10227396]
51. Shimajiri S, et al. Shortened microsatellite d(CA)21 sequence down-regulates promoter activity of matrix metalloproteinase 9 gene. *FEBS Lett*. 1999; 455:70–4. [PubMed: 10428474]
52. Hefferon TW, Groman JD, Yurk CE, Cutting GR. A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. *Proc Natl Acad Sci U S A*. 2004; 101:3504–9. [PubMed: 14993601]
- 53*. Gymrek M, et al. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet*. 2016; 48:22–29. Highlights the contribution of STRs to the genetic architecture of quantitative human traits and identified eSTR on a whole genome scale. [PubMed: 26642241]
54. Afek A, Cohen H, Barber-Zucker S, Gordan R, Lukatsky DB. Nonconsensus Protein Binding to Repetitive DNA Sequence Elements Significantly Affects Eukaryotic Genomes. *PLoS Comput Biol*. 2015; 11:e1004429. [PubMed: 26285121]
- 55*. Afek A, Schipper JL, Horton J, Gordan R, Lukatsky DB. Protein-DNA binding in the absence of specific base-pair recognition. *Proc Natl Acad Sci U S A*. 2014; 111:17140–5. Identification of a new protein-DNA binding mechanism induced by DNA sequence symmetry (repeat sequences), and demonstration that this mechanism governs protein-DNA binding preferences and is relevant to binding affinities of transcription factors to DNA. [PubMed: 25313048]
- 56*. Saul R, Lutz MW, Burns DK, Roses AD, Chiba-Falek O. The SSV Evaluation System: A Tool to Prioritize Short Structural Variants for Studies of Possible Regulatory and Causal Variants. *Hum Mutat*. 2016; 37:877–83. Presents the SSV evaluation system as a new powerful tool to guide genetic investigations aiming to uncover SSVs that underlie human complex diseases including neurodegenerative diseases in aging. [PubMed: 27279261]
57. Xia Y, et al. Genetic studies in Alzheimer's disease with an NACP/alpha-synuclein polymorphism. *Ann Neurol*. 1996; 40:207–15. [PubMed: 8773602]
58. Touchman JW, et al. Human and mouse alpha-synuclein genes: comparative genomic sequence analysis and identification of a novel gene regulatory element. *Genome Res*. 2001; 11:78–86. [PubMed: 11156617]
59. Goldman SM, Umbach DM, Kamel F, Tanner CM. Head injury, alpha-synuclein Rep1 and Parkinson's disease: a meta-analytic view of gene-environment interaction. *Eur J Neurol*. 2015; 22:e75. [PubMed: 26060929]
60. Kay DM, et al. Genetic association between alpha-synuclein and idiopathic Parkinson's disease. *Am J Med Genet B Neuropsychiatr Genet*. 2008; 147B:1222–30. [PubMed: 18404644]
61. Farrer M, et al. alpha-Synuclein gene haplotypes are associated with Parkinson's disease. *Hum Mol Genet*. 2001; 10:1847–51. [PubMed: 11532993]

62. Mizuta I, et al. Meta-analysis of alpha synuclein/NACP polymorphism in Parkinson's disease in Japan. *J Neurol Neurosurg Psychiatry*. 2002; 73:350. [PubMed: 12185186]
63. Mellick GD, Maraganore DM, Silburn PA. Australian data and meta-analysis lend support for alpha-synuclein (NACP-Rep1) as a risk factor for Parkinson's disease. *Neurosci Lett*. 2005; 375:112–6. [PubMed: 15670652]
64. Linnertz C, et al. Genetic regulation of alpha-synuclein mRNA expression in various human brain tissues. *PLoS One*. 2009; 4:e7480. [PubMed: 19834617]
65. Chiba-Falek O, Touchman JW, Nussbaum RL. Functional analysis of intra-allelic variation at NACP-Rep1 in the alpha-synuclein gene. *Hum Genet*. 2003; 113:426–31. [PubMed: 12923682]
66. Cronin KD, et al. Expansion of the Parkinson disease-associated SNCA-Rep1 allele upregulates human alpha-synuclein in transgenic mouse brain. *Hum Mol Genet*. 2009; 18:3274–85. [PubMed: 19498036]
67. Chiba-Falek O, Kowalak JA, Smulson ME, Nussbaum RL. Regulation of alpha-synuclein expression by poly (ADP ribose) polymerase-1 (PARP-1) binding to the NACP-Rep1 polymorphic site upstream of the SNCA gene. *Am J Hum Genet*. 2005; 76:478–92. [PubMed: 15672325]
68. Soldner F, et al. Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature*. 2016; 533:95–9. [PubMed: 27096366]
69. Beyer K. Alpha-synuclein structure, posttranslational modification and alternative splicing as aggregation enhancers. *Acta Neuropathol*. 2006; 112:237–51. [PubMed: 16845533]
70. Lee HJ, Choi C, Lee SJ. Membrane-bound alpha-synuclein has a high aggregation propensity and the ability to seed the aggregation of the cytosolic form. *J Biol Chem*. 2002; 277:671–8. [PubMed: 11679584]
71. Beyer K, et al. Low alpha-synuclein 126 mRNA levels in dementia with Lewy bodies and Alzheimer disease. *Neuroreport*. 2006; 17:1327–30. [PubMed: 16951579]
72. Beyer K, et al. Differential expression of alpha-synuclein, parkin, and synphilin-1 isoforms in Lewy body disease. *Neurogenetics*. 2008; 9:163–72. [PubMed: 18335262]
73. Beyer K, et al. A variable poly-T sequence modulates alpha-synuclein isoform expression and is associated with aging. *J Neurosci Res*. 2007; 85:1538–46. [PubMed: 17387688]

Highlights

- Structural Variants (SVs) have a large impact on many human phenotypes including complex diseases such as synucleinopathies, and also contribute significantly to variation in gene expression in human as exemplified here for *SNCA* gene.
- Copy number variations (CNVs) of the *SNCA* gene cause familial Parkinson's disease (PD), where disease severity and the levels of wild-type *SNCA* mRNA and protein are correlated with *SNCA* gene dose.
- A specific haplotype at a CT-rich region in *SNCA*-intron 4 that consists of a cluster of SSVs confers risk to develop LBV/AD and acts as an enhancer element to increase *SNCA*-mRNA expression.
- Rep1, a SSV ~10kb upstream of *SNCA* gene that comprises of a cluster of SSRs, has been associated with PD-risk and has a role in the regulation of *SNCA* transcription
- A polyT variant in *SNCA*-intron 2 influences the splicing efficiency of exon 3, deletion of this exon is differently expressed in synucleinopathies and has been suggested to lead to less protein aggregation.

Figure 1. SNCA-mRNA and α-syn protein in iPSC-derived neurons

(A) Levels of *SNCA*-mRNA were measured by real-time RT-PCR using TaqMan expression assay and *SNCA*-mRNA fold levels were calculated relative to the geometric mean of *GAPDH*-mRNA and *PPIA*-mRNA reference controls using the 2^{-CT} method. (B) Immunofluorescence (IF) labeling for α-syn protein (green) was performed using rabbit recombinant monoclonal anti-alpha synuclein primary antibody (ab138501 ABCAM) and AlexaFluor 488 goat anti-Rabbit secondary antibody (A11034 ThermoFisher). Signal intensities were analyzed using ImageJ. Example are shown for the iPSC-derived basal forebrain cholinergic neurons (BFCNs) differentiation.

Figure 2. Schematic illustration of a putative mechanism whereby SSRs affect differential transcription

The overall hypothesis is that differential non-consensus binding of transcription factors (TFs) to the variable alleles at repeat sequence sites is responsible, at least in part, for differential expression of the *cis*-regulated genes, which can contribute to disease pathogenesis. For alleles of different lengths, some SSRs will have differential non-consensus binding free-energy, and thus differential TF binding strength.

Figure 3. Schematic representation of the role of SNCA SVs in synucleinopathies

The effect of *SNCA*-CNVs is mediated via increases in overall *SNCA*-mRNA and protein levels. The effects of *SNCA*-SSVs can be mediated via several RNA-based mechanisms. The cell-type selective vulnerability to overexpression and/or particular *SNCA*-isoform will contribute, among other factors, to the development of a particular disease in the spectrum of synucleinopathy disorders.