Discovery of functional non-coding conserved regions in the α-synuclein gene locus [version 2; peer review: 3 approved]

Lori Sterling¹, Michael Walter², Dennis Ting¹, Birgitt Schüle¹

¹Parkinson's Institute and Clinical Center, Sunnyvale, CA 94085, USA
²Institute of Human Genetics, Eberhard-Karls-University Tübingen, Tübingen, 72076, Germany

Abstract

Several single nucleotide polymorphisms (SNPs) and the Rep-1 microsatellite marker of the α-synuclein (SNCA) gene have consistently been shown to be associated with Parkinson's disease, but the functional relevance is unclear. Based on these findings we hypothesized that conserved cis-regulatory elements in the SNCA genomic region regulate expression of SNCA, and that SNPs in these regions could be functionally modulating the expression of SNCA, thus contributing to neuronal demise and predisposing to Parkinson's disease.

In a pair-wise comparison of a 206kb genomic region encompassing the SNCA gene, we revealed 34 evolutionary conserved DNA sequences between human and mouse. All elements were cloned into reporter vectors and assessed for expression modulation in dual luciferase reporter assays. We found that 12 out of 34 elements exhibited either an enhancement or reduction of the expression of the reporter gene. Three elements upstream of the SNCA gene displayed an approximately 1.5 fold (p<0.009) increase in expression. Of the intronic regions, three showed a 1.5 fold increase and two others indicated a 2 and 2.5 fold increase in expression (p<0.002). Three elements downstream of the SNCA gene showed 1.5 fold and 2.5 fold increase (p<0.0009). One element downstream of SNCA had a reduced expression of the reporter gene of 0.35 fold (p<0.0009) of normal activity.

Our results demonstrate that the SNCA gene contains cis-regulatory regions that might regulate the transcription and expression of SNCA. Further studies in disease-relevant tissue types will be important to understand the functional impact of regulatory regions and specific Parkinson's disease-associated SNPs and its function in the disease process.
Keywords
α-synuclein, transcriptional regulation, evolutionary conserved genomic region, SNCA, Parkinson's disease

Corresponding author: Birgitt Schüle (bschuele@thepi.org)
Competing interests: No competing interests were disclosed.
Grant information: This work was supported by a pilot grant of NIEHS-CCPDER 1U54ES012077 to B. Schüle (PI: J.W. Langston) and Parkinson’s Unity Walk.
Copyright: © 2014 Sterling L et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).
How to cite this article: Sterling L, Walter M, Ting D and Schüle B. Discovery of functional non-coding conserved regions in the α-synuclein gene locus [version 2; peer review: 3 approved] F1000Research 2014, 3:259 https://doi.org/10.12688/f1000research.3281.2
Introduction

An emerging hypothesis is gaining increasing interest and is based on the concept that subtle overexpression of α-synuclein (α-syn) over many decades can either predispose or even cause the neurodegenerative changes that characterize Parkinson’s disease (PD). Neurons subjected to higher, non-physiological levels of α-syn might be more likely to be damaged by oligomerization or aggregation of this protein, eventually leading to the formation of α-synuclein-based neuropathological features of the disease.  

It is now well established that both point mutations and large genomic multiplications of the α-synuclein (SNCA) gene can cause an autosomal-dominant form of PD. Furthermore, several association studies investigating genetic variants in the SNCA gene have found an increased risk for PD. The finding that both qualitative and quantitative alterations in the SNCA gene are associated with the development of a parkinsonian phenotype indicates that amino acid substitutions as well as overexpression of wild-type α-syn are capable of triggering a clinicopathological process that is very similar to sporadic PD. Nevertheless, the precise mechanisms leading to α-syn-related pathology in sporadic PD in the absence of any α-syn mutations remain elusive.

The best characterized polymorphism in the SNCA gene is the Rep-1 mixed dinucleotide repeat which has been shown to act as a modulator of SNCA transcription. The DNA binding protein and transcriptional regulator PARP-1 showed specific binding to SNCA-Rep1. These data were confirmed by a transgenic mouse model and demonstrated regulatory translational activity.

Functionally, SNCA expression levels in postmortem brains suggest that the Rep-1 allele and SNPs in the 3′ region of the SNCA gene have a significant effect on SNCA mRNA levels in the substantia nigra and the temporal cortex.

The promoter region of the SNCA gene has been recently examined in more detail in cancer cell lines and also in rat cortical neurons. Regulatory regions in intron 1 and the 5′ region of exon 1 have been shown to exhibit transcriptional activation as well as the NACP-Rep-1 region upstream of the SNCA gene. Several transcription factors have been identified such as PARP-1, GATA-3, ZIPRO1, and ZNF219 to have an effect on regulating the SNCA promoter region.

There is mounting evidence that SNCA expression levels could be crucial for maintenance, survival of neurons and its misregulation could play a key role in the development of PD. Thus, the importance of thoroughly investigating the SNCA gene to fully understand its cis- and trans-acting elements and factors and for the functional interpretation of the PD-diseased associated risk alleles is becoming increasingly clear.

The goal of this study was to investigate transcriptional regulation of the SNCA region using a complementary approach, under the hypothesis that conserved non-coding regions of the SNCA gene are comprised of transcriptional enhancers or silencers and thus modulate gene expression. This would mean that single nucleotide polymorphisms (SNPs) in these regions could influence the transcriptional pattern of the SNCA gene.

Materials and methods

Comparative genomics

Using comparative genomics, we searched for highly conserved non-coding sequences between human and mouse and identified 34 evolutionary conserved non-coding genomic regions (ncECRs) within the SNCA gene that are conserved between human and mouse.

We utilized two complementary browsers (Vista browser (http://pipeline.lbl.gov/cgi-bin/gateway2) and ECR browser (http://ecrbrowser.dcode.org/) to generate a conservation profile by aligning the human SNCA gene with its mouse counterpart in a pair-wise fashion. We applied established selection parameters for our search with >100bp in length and >75% identity. In addition to the 11.4kb SNCA gene region, we included a 44.5kb upstream and a 50kb downstream intergenic region to also capture surrounding regulatory elements.

We identified 34 ncECRs in the SNCA genomic region of 206kb on chromosome 4q21 (Chr. 4: 90,961056-91,167082, UCSC Genome Browser on Human Mar. 2006 Assembly) by pair-wise comparison between human and mouse (Figure 1). Ten of these DNA sequences were located downstream of the SNCA gene, 17 were intrinsic between exon 4 and 5, which is 92kb in length, and five were upstream of the SNCA gene (Figure 1). None of the selected sequences overlapped with known expressed sequence tags (ESTs) or had an open reading frame of more than 20 amino acids in length, suggesting that these ncECRs are non-coding.

Cloning and luciferase assays

To test, if the ncECRs exhibit enhancer or silencer activity, we cloned all identified regions in specific reporter vectors and measured their luciferase activity after transfection into neuroblastoma cells. For our studies, we used the pGL3 luciferase reporter vectors (Promega, Cat. No. E1751, E1741, E1771, E1761) and the human neuroblastoma cell line SK-N-SH. ncECRs identified through the comparative analysis (Supplementary Table 1) were cloned upstream of a SV-40 promoter in the pGL3 promoter construct, transfected in SK-N-SH cells and assayed with the Dual-Luciferase® Reporter Assay System (Promega, Cat. No. E1910).

Some of these regions were combined in one vector because of their close proximity to each other. Primers with specific restriction sites (KpnI, BglII or XhoI from New England Biolabs Inc.) were designed to amplify the conserved elements, and PCR products with specific restriction sites were directly cloned into the pGL3 promoter vector to ensure correct orientation of the genomic elements (Supplementary Table 1). All constructs were sequenced to ensure that no point mutations were introduced through the amplification and/or cloning process.

For transfection experiments, we used a 96-well format (Nunc, Cat. No. 167008). Cells were plated one day before transfection at a
density of 3000–5000 cell/well to reach 90–95% confluency at the time of transfection. Luciferase assays were performed 24 hrs after transfection. SK-N-SH cells were maintained in HyClone DMEM media (High Glucose, Fisher Scientific, Cat No. SH30081.02) with 10% HyClone fetal bovine serum (Fisher Scientific, Cat No. SH30910.03) in 1x glutamine (Life Technologies, Cat. No. 25030-081) and 1x penicillin/streptomycin (Life Technologies, Cat. No. 15140-122). For SK-N-SH cells, we used 1:2 ratio of nucleic acid to transfection reagent (Lipofectamine® 2000 Transfection Reagent, Life Technologies, Cat No. 11668-019). For the luciferase assay, we used the Dual-Luciferase® Reporter (DLR™) Assay System (Promega, Cat. No. E1910) according to the manufacturer’s instructions in 96-well white plates, flat bottom (E&K Scientific, Cat. No. EK-25075). In this assay, activities of firefly and Renilla luciferases were measured sequentially in one sample. All assays were performed in quadruplicate and each experiment was repeated three times. Altogether, 12 data points were ascertained for each conserved region/construct.

**Statistical analysis**

Differences among means were analyzed using two-samples student’s t-test. For differences in transcriptional activation of the luc+ gene, ncECRs were tested in quadruplicates in three independent experiments. Differences were considered statistically significant at p<0.05.

**Bioinformatic search for transcription factor binding sites (TFBS) with MatInspector (Genomatix)**

To estimate the number of potential TFBSs and the number of interacting transcription factors (TFs) that could represent potential candidate proteins for our positive ncECRs, we used MatInspector in an *in silico* approach. We chose two elements for this bioinformatic analysis with MatInspector. The MatInspector software utilizes a large library of matrices for TFBSs to locate matching DNA sequences. The program assigns quality rating to matches and allows quality-based filtering and selection of matches. MatInspector can group similar or functionally related TFBSs into matrix families.

In addition to the original human-mouse comparison, we added the sequences for dog and cow for comparisons. Only the TFBSs were considered that were present in all four species, in the same orientation, and similar distance to each other. We ran two analyses with 10 and 15 nucleotides distance, respectively. We accepted only models in which at least four TFs can bind in a concerted way. Each TFBS can potentially bind several TFs.

We also computationally tested all possible TFs for interactions with the SNCA promoter region, which were retrieved from the proprietary ElDorado database (Genomatix, Munich, Germany). In this database, promoters are defined and ranked by transcription
start sites, corresponding known mRNA or EST sequences and by orthologous conservation.

**Results**

**Functional non-coding conserved elements within the SNCA genomic locus**

Overall, 12 of 34 conserved non-coding elements exhibited either an increase or reduction of the expression of the luciferase reporter gene (Figure 2 and Dataset 1). Three elements upstream of the SNCA gene (U3, U4-1, and U4-3) displayed a significant approximately 1.5 fold (p<0.009) increase in expression (Figure 2A). Of the intronic regions, three showed a 1.5 fold increase (I2, I6, I8) and two others showed a 2 and 2.5 fold increase in expression (p<0.002), I5 and I12, respectively (Figure 2B). Two elements downstream of the SNCA gene showed approximately 2 fold (D1 and D2) and 2.5 fold (D3) increase (p<0.0009) (Figure 2C). One element D6 downstream of SNCA had a reduced expression of the reporter gene of 0.35 fold (p<0.0009) of normal activity (Figure 2C, green) that was also confirmed after cloning the D6 element in a pGL3 control vector (Figure 2C, insert). The pGL3 control vector contains the SV-40 promoter and a SV-40 enhancer element. The D6 element reduced the expression of the pGL3 control construct by ~50%, confirming that this element represents a repressor. Between 4 and 12 replicates were performed per ncECR.

These data provide experimental evidence that a significant proportion of the ncECRs show a regulatory function in the luciferase reporter assay.

**In silico analysis reveals potential binding of midbrain transcription factors to regulatory conserved regions**

We performed MatInspector (Genomatix) analysis on two elements (I12: chr4:90940532-90940786 and D6: chr4:90855871-90856339, Human Genome assembly NCBI36/hg18) with the highest fold change in the luciferase assay. In addition to the original human-mouse comparison to identify the ncECRs, we added the sequences from dog and cow. Only TFBSs that were present in all four species, in the same orientation, and similar distance to each other were considered. We ran two analyses with 10 and 15 nucleotides distance, respectively. We accepted only models in which at least four TFs can bind in a concerted way. Each TFBS can potentially bind several TFs. Interestingly, using this more restricted model, five factors showed an interaction with the SNCA promoter as well as with the ncECRs (Figure 3A). These factors were the Paired-like homeodomain transcription factor 3 (PITX3), the Homolog of Drosophila orthodenticle 2 (OTX2), the Nuclear receptor subfamily 3, group c, member 1 (NR3C1) or glucocorticoid receptor (GCCR), the Androgen receptor (AR), and the general transcription initiation factor TATA box-binding protein (TBP).

It is intriguing to note that by searching for TFs that bind to both the promoter and the functional ncECR, several DNA-binding proteins were found that are linked to dopaminergic regulation and susceptibility for nigrostriatal impairment. Two of these TFs (PITX3 and OTX2) implicated in determination of a dopaminergic phenotype in the substantia nigra emerged from this preliminary search. PITX3 has shown to be regulated in a negative feedback circuit through the microRNA mi-133b to fine-tune maintenance of dopaminergic neurons. In an association study, a SNP in the

---

**Figure 2.** Non-coding conserved elements within the SNCA genomic locus show changes in luciferase assays. Panels A–C show the luciferase assay results of ncECRs upstream (A), intragenic (B), and downstream (C) of the SNCA gene. The Y-axis shows the ratio of luciferase to renilla expression as percentage. Bas=pGL3 basic, Con=pGL3 control, prom=pGL3 promoter construct. All red or green box plot elements represent ncECRs that modulate expression significantly. The box plots show the median (horizontal line within box), the 25 and 75% tiles (horizontal borders of box), and the whiskers show the minimal and maximal values. Panel C, insert: Luciferase assay results of D6 element cloned into the pGL3 control vector construct.
Figure 3. In silico analysis reveals midbrain transcription factors binding to two ncECRs. A. MatInspector network view of SNCA promoter interaction with TFs that also potentially bind to two ncECRs (I12 and D6) within the SNCA gene. M=gene product is part of metabolic pathway, IN=input gene, TF=transcription factor, ST=gene product is part of signal transduction pathway, green line=matches target promoter B. UCSC Genome browser custom track of PD associated SNPs (based on PD Gene metaanalysis), Rep1 allele and functional ECR regions on chromosome 4 (Human Genome Assembly Feb. 2009, GRCh37/hg19).

PITX3 promoter was reported to be associated with PD and might dysregulate expression of PITX3 suggesting that transcription factors play a critical role not only in the development and differentiation of dopaminergic neurons, but also for cell maintenance and survival of dopaminergic neurons.

GCCR and AR belong to a class of nuclear receptors called activated class I steroid receptors. GCCR is a cytosolic ligand-activated transcription factor that regulates the expression of glucocorticoid-responsive genes. GCCR shows strong anti-inflammatory and immunosuppressive effects. Interestingly, impaired GCCR expression in a mouse model shows a dramatic increase in the vulnerability of the nigrostriatal dopaminergic neurons to a toxic insult of MPTP.

Taken together, this preliminary in silico screen resulted in very intriguing new candidates that might directly regulate SNCA expression and could play a role in the pathological processes that underlie PD.

Dataset 1. Combined normalized raw datasets of Luciferase assays on SNCA conserved elements

http://dx.doi.org/10.5256/f1000research.3281.d37452

Data are ratios of luminometer readings for firefly luciferase and renilla luciferase. Ratios were normalized to Prom. Each non-coding element is labeled and data are presented under each element. Elements are organized according to Figure 2A–C.

Discussion

A major focus in PD research has been on post-translational modification of α-syn. The alterations seen in PD that were linked to disease pathogenesis were nitrated α-syn and α-syn phosphorylated at serine 129 identified in Lewy bodies and Lewy neurites, however, the gene transcription as a control point and its regulation in particular cell types or upon cellular signals has only been touched fairly recently in PD-relevant genes.

Our results show that potential regulatory regions are not restricted to the promoter region of the SNCA gene as discussed in the introduction, but are likely to be located also in other intronic and intergenic regions (Figure 3B). Comparing our results to similar screens, where conserved regions range from 8–45 elements, we found a similar number of functional elements in our screen that show a high evolutionary conservation.

Not only the promoter region of a gene drives the transcription/expression of a gene. Also other cis-acting genomic regions within a certain gene, up to several hundred kb away, can serve as enhancers, silencers, or modifiers to ensure the accurate temporal and spatial expression of a gene by recruiting transcription activating or silencing factors that bind to them. There is ample precedence for this approach to analyze genomic regions of genes implicated in human disease. Mutations in those conserved elements were found to cause human genetic syndromes, for example SALL1/Townes-Brocks syndrome or SHH/preaxialpolydactyly. Other groups have investigated the non-coding regulatory elements within disease
genes such as RET (Ret proto-oncogene) and MECP2 (Methyl-CpG binding protein 2) and found multiple regulatory enhancer and silencer elements.

Transcriptional regulation of dopaminergic neurons

Computationally determining transcription factor binding sites is a challenging process and multiple prediction algorithms have been developed over the last decade (Cartharius 2005, Wu 2009, Mathelier 2013). Therefore our preliminary data should solely open the discussion and drive novel hypotheses for potential transcription factors that regulate transcription of the SNCA locus. Specific TFs seem to be directly involved in neurodegeneration and models of PD. TFs have been shown to be critical regulators for the development, maintenance and survival of dopaminergic neuronal populations.

E.g. forkhead transcription factor (Foxa2) is responsible for early development of endoderm and midline structures. Foxa2 is specifically expressed in postmitotic dopaminergic neurons. Genetically engineered mice that are null for Foxa2 are not viable, whereas heterozygotes for Foxa2 develop major motor abnormalities starting at 18 months with an asymmetric posture, rigidity, and bradykinesia.

Conclusion

This screen of evolutionary conserved genomic elements in the SNCA locus showed a number of functionally elements that in an in vitro assay modulated the expression of a reporter gene. Furthermore, we identified very intriguing new candidate transcription factors that could directly regulate SNCA expression and could, if binding is altered by genetic variants, play a role in the pathological processes that underlie PD. This is the first step to systematically analyze the SNCA locus to understand its transcriptional regulation in more detail. Further studies are needed in neuronal tissues (e.g. dopaminergic neurons derived from patient-specific induced pluripotent stem cells) to confirm these findings and expand the analysis to identify SNCA-regulating transcription factors. By defining the transcription factors that regulate expression and potentially overexpression of α-synuclein that can lead to neurodegeneration, we will be able to identify targets for novel therapeutic approaches for α-synucleinopathies including Parkinson’s disease.

Data availability


Author contributions

BS conceived the study and designed the experiments, and drafted the manuscript. LS carried out the experiments. DT analyzed data. MW carried out in silico analysis for transcription factors. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by a pilot grant of NIEHS-CCPDER 1U54ES012077 to B. Schüle (PI: J.W. Langston), and by the Parkinson’s Unity Walk.

Acknowledgements

Part of the content of this manuscript has been presented as a poster at the Annual Meeting of The American Society of Human Genetics 2007:


Supplementary Table

Supplementary Table 1. Primer sequences and design for cloned ncECRs.

<table>
<thead>
<tr>
<th>ECRs in SNCA locus on chromosome 4</th>
<th>Human Genome assembly NCBI36/hg18 (March 2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIndIII: CCCAAGCTTT</td>
<td>XhoI: CCGCTCGAG</td>
</tr>
<tr>
<td>KpnI: CCGGTACC</td>
<td>BglII: GGAAGATCT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ECR</th>
<th>Length</th>
<th>Identity</th>
<th>Location</th>
<th>Primsers</th>
<th>PCR product length</th>
<th>Ann. Temp</th>
<th>Restriction sites within PCR product</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>146bp</td>
<td>78.10%</td>
<td>chr4:90833665-90833810</td>
<td>CGGGGTACC CACGAATCTGTCCTCTGAAAAAT</td>
<td>601bp</td>
<td>no RE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GGAAGATCTaagtcacaaggtgagccctt</td>
<td>60C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>239bp</td>
<td>74.50%</td>
<td>chr4:90844830+90845413</td>
<td>CCGGGTACC tgcagaaatccacacacacat</td>
<td>584bp</td>
<td>no RE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GGAAGATCTTCAGCAAGATGGGATGGAATA</td>
<td>60C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3-1/2</td>
<td>143bp</td>
<td>72%</td>
<td>chr4:90848813-90848955</td>
<td>CGGGGTACC AAGGGCTAACATTGAATTT</td>
<td>60C</td>
<td>no RE</td>
<td></td>
</tr>
</tbody>
</table>
### ECRs in SNCA locus on chromosome 4

Human Genome assembly NCBI36/hg18 (March 2006)

<table>
<thead>
<tr>
<th>ECR</th>
<th>Length</th>
<th>Identity</th>
<th>Location</th>
<th>Primers</th>
<th>PCR product length</th>
<th>Ann. Temp</th>
<th>Restriction sites within PCR product</th>
</tr>
</thead>
</table>
| HindIII: CCCAAGCTT

| XhoI: CGGCCTCGAG

| KpnI: CGGGGTACC

| BglII: GGAAGATCT |
|---|---|---|---|---|---|---|
| D4 | 99bp | 75.80% | chr4:90849405-90849503 | GGAAGATCTCGCGCTCTGAAATAAGCAA | 989bp | 60C | |
| D5-1 | 241bp | 75.90% | chr4:90853634-90853874 | CGGGGTACCAGGCTGAAATGGCCAAATGCAAT | no RE | |
| D5-2 | 114bp | 75.40% | chr4:90854429-90854654 | CGGGGTACCAGGGCTGAAATGGCCAAATGCAAT | BglII at 1143 | |
| D6-1/2 | 234bp | 73.60% | chr4:90855871-90856104 | GGAAGATCTCCATTCCTTTACCTCCTAAATG | 582bp | 60C | |
| D7 | 167bp | 72.50% | chr4:90859690-90859856 | GGAAGATCTCGGCGGCTCTGAAATAAGCAA | no RE | |
| D8-1/2 | 100bp | 75% | chr4:90860722-90860821 | GGAAGATCTCCGCGCTCTTCTGCTTTTCATTCCTCC | no RE | |
| I1 | 192bps | 81.80% | chr4:90871989-90872180 | CGGGGTACCGAGGATTGCGATAATGGCCAAATGCAAT | 840bp | 60C | BglII at 571 |
| I2 | 154bp | 74% | chr4:90878220-90878373 | GGAAGATCTCGGCGGCTCTGAAATAAGCAA | 393bp | 60C | |
| I3-1/2 | 276bps, | 77.50% | chr4:90887100-90887375 | GGAAGATCTCGGCGGCTCTGAAATAAGCAA | 986bp | 60C | no RE |
| I4-1/2 | 194bps | 75.80% | chr4:90891860-90892053 | GGAAGATCTCGGCGGCTCTGAAATAAGCAA | 899bp | 60C | no RE |
| I5-1/2 | 109bps | 76.10% | chr4:90893684-90893792 | GGAAGATCTCGGCGGCTCTGAAATAAGCAA | 705bp | 60C | no RE |
| I6 | 130bps | 74.60% | chr4:90894785-90894914 | GGAAGATCTCGGCGGCTCTGAAATAAGCAA | 355bp | 60C | no RE |
| I7 | 187bps | 75.90% | chr4:90897558-90897744 | GGAAGATCTCGGCGGCTCTGAAATAAGCAA | 432bp | 60C | no RE |
| I8 | 112bps | 76.80% | chr4:90901290-90901401 | CGGGGTACCGGCGGCTCTGAAATAAGCAA | 474bp | 60C | BglII at 34 |
### ECRs in SNCA locus on chromosome 4
Human Genome assembly NCBI36/hg18 (March 2006)

<table>
<thead>
<tr>
<th>ECR</th>
<th>Length</th>
<th>Identity</th>
<th>Location</th>
<th>Primers</th>
<th>PCR product length</th>
<th>Ann. Temp</th>
<th>Restriction sites within PCR product</th>
</tr>
</thead>
<tbody>
<tr>
<td>I9</td>
<td>199 bps</td>
<td>75.40%</td>
<td>chr4:90906237-90906435</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>561bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>I10</td>
<td>269 bps</td>
<td>74.30%</td>
<td>chr4:90926832-90927100</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>899bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>I11</td>
<td>108 bps</td>
<td>74.10%</td>
<td>chr4:90929480-90929587</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>669bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>I12</td>
<td>255 bps</td>
<td>77.30%</td>
<td>chr4:90940532-90940786</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>578bp</td>
<td>60°C</td>
<td>BglII at 328</td>
</tr>
<tr>
<td>I13</td>
<td>127 bps</td>
<td>76.50%</td>
<td>chr4:90945579-90945706</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>578bp</td>
<td>60°C</td>
<td>BglII at 328</td>
</tr>
<tr>
<td>I14</td>
<td>100 bps</td>
<td>75.00%</td>
<td>chr4:90958054-90958153</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>701bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>I15</td>
<td>329 bps</td>
<td>76.30%</td>
<td>chr4:90961985-90962223</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>490bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>I16</td>
<td>106 bps</td>
<td>75.50%</td>
<td>chr4:90976615-90976720</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>651bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>U1</td>
<td>261 bps</td>
<td>81.60%</td>
<td>chr4:90977921-90978181</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>883bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>U2-1</td>
<td>105 bps</td>
<td>76.20%</td>
<td>chr4:90980743-90980847</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>860bp</td>
<td>59.5°C</td>
<td>no RE</td>
</tr>
<tr>
<td>U2-2</td>
<td>100 bps</td>
<td>75.00%</td>
<td>chr4:90981402-90981501</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>837bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>U3</td>
<td>329 bps</td>
<td>73.90%</td>
<td>chr4:91004670-91004998</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>751bp</td>
<td>60°C</td>
<td>no RE</td>
</tr>
<tr>
<td>U4-1</td>
<td>131 bps</td>
<td>74.00%</td>
<td>chr4:91009155-91009285</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>698bp</td>
<td>58°C</td>
<td>no RE</td>
</tr>
<tr>
<td>U4-3</td>
<td></td>
<td></td>
<td>chr4:91010061-91010758</td>
<td>GGAAGA(T)C(T)GCAATGAAAAACCACAATGG</td>
<td>698bp</td>
<td>58°C</td>
<td>no RE</td>
</tr>
</tbody>
</table>
Open Peer Review

Current Peer Review Status:  ✔  ✔  ✔

Version 2

Reviewer Report 23 December 2014

https://doi.org/10.5256/f1000research.6309.r6588

© 2014 Stefanis L et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Leonidas Stefanis
Division of Basic Neurosciences, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

Georgia Dermentzaki
Biomedical Research Foundation of the Academy of Athens, Athens, Greece

The manuscript is thorough, and uses an array of approaches to tackle the issue of the regulation of SNCA expression, an issue that regrettably has received little attention, despite its obvious importance. This analysis will undoubtedly be very helpful to the scientific community.

We have some minor comments:

1. For the in silico analysis with MatInspector they used the two elements with the highest activity in the luciferase assay (the I12 and D6). However in Figure 2 the D3 element had even higher activity and was not used in the analysis. Explain why?

2. For the luciferase experiments they measured the luciferase activity after 24 hrs after transfection. Why so early? Did they check later time points? Usually 36-40 hrs are necessary for pGL3 vectors.

3. Regarding the possible involvement of GR in SNCA expression, it may be worthwhile referencing the paper by Park et al. (2013).

Competing Interests: No competing interests were disclosed.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
Ornit Chiba-Falek
Department of Neurology, Duke University, Durham, NC, USA

The authors replied adequately to my suggestions. I have no further comments.

**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

**Jinglan Liu**
Clinical and Molecular Cytogenetics, St. Christopher's Hospital for Children, Philadelphia, PA, USA

The article by Sterling et al. has described the identification and functional analysis of evolutionally conserved non-coding elements that might be involved in the transcriptional regulation of the gene SNCA, mutations in which were associated with Parkinson's disease. This is a very interesting, proof-of-concept article, with an attempt to provide pathogenic insight from the point of view of regulatory genomics for a complex human disease. I endorse the indexing of this manuscript.

It is now well recognized that ~98% of human genome do not code for proteins. Comparative genomics studies revealed that the majority of evolutionally conserved regions consist of non-coding elements that that might be involved in regulating gene expression. Genome-wide association studies (GWAS) have showed that the majority (~93%) of SNPs contributing to human diseases or susceptibility lie outside protein-coding regions, and there are many non-coding SNPs have been demonstrated to be associated with common diseases and traits.

By identifying functionally significant non-coding elements for SNCA, Sterling et al.’s work could lend a new perspective to study the genetic architecture of Parkinson's disease, and promote further investigations on the pathogenic impact of non-coding elements and their regulatory networks on the clinical courses of Parkinson's disease.
**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

**Reviewer Report 06 November 2014**

https://doi.org/10.5256/f1000research.3521.r6592

© 2014 Chiba-Falek O et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ornit Chiba-Falek  
Department of Neurology, Duke University, Durham, NC, USA  

Lidia Tagliafierro  
Duke University, Durham, NC, USA

The paper by Dr. Schüle's team describes the identification of evolutionary conserved non-coding regions (ncECRs) in the α-synuclein (SNCA) gene and their assessment as candidate regulatory elements. The work coupled *in silico* and cell-based studies. By using a comparative genomic screen between human and mouse, the authors identify 32 ncECRs, out of which 11 regions exert an effect on expression level using a luciferase reporter assay approach. Their findings add on previous reports in the field that have shown, using both luciferase reporter system and human brain tissues, that the SNCA gene contains cis-regulatory sites across the 3’ and the 5’ LD blocks that regulate its expression levels.

The study was well designed and thoroughly executed, the results are of interest to the scientific community of PD-genetics, and provide seeds for follow up studies. The paper is nicely written, logically flows and summarizes the literature in the field. However, the authors should make major revisions according to the following comments:

1. There is some inconsistency regarding the number of the ncECRs identified in the initial screen between the different sections of the article (32, 34, 37). Please make the corrections where needed.

2. Additional necessary control for the Luciferase experiments is a pGL-(SV40) promoter vector harboring an insert of a scrambled sequence that its size range mimics the average insert size of the tested ECRs. This is required to control for the 'spacer' effect of ECR lengths.

3. What method was used for the statistical analysis? It is also not clear in the text whether all significant changes were calculated in comparison to the SV-40 promoter-only vector. That should be described in details in the method section.

4. To demonstrate the important implication of this study the authors are recommended to follow up on an event as an example. That is to say, to evaluate the effect of a genetic variation, a PD-associated SNP, on the regulatory function of the corresponding ECR using...
the luciferase system established in this work. Figure 3 demonstrates overlap between PD associated SNPs and ncECR, connecting these dots will be of high significance.

5. Supp Table: there is a typo in the coordinates of D2. In the footnote include the human genome assembly of the coordinates.

6. Figure 2A X-axis: modify title to ‘upstream....’

7. Omit Figure 3A. Instead include a new panel to figure 3B that indicates the position of the putative binding sites of these TFs within SNCA locus.

8. The identification of Transcription Factor Binding Sites (TFBS) is an important step required in order to evaluate the transcriptional regulation network of the SNCA gene. To this end, the computational prediction of TFBS is a classic approach that gives preliminary data but should be interpreted with caution. Integration of the classic approach with new models described in Mathelier & Wasserman (2013) is highly recommended. The relation between TF motifs and in vivo binding sites is far from simple. The analysis lacks of information about the context of the identified sequences. TF are highly context-specific, and the same TF typically binds to different genomic binding sites in different conditions. Obtaining information about the context could be helpful in better understanding the possible involvement of the predicted sites as TFBS. While this is beyond the scope of this study, this topic should be thoroughly discussed in the discussion section.

**Competing Interests:** No competing interests were disclosed.

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.**

Author Response 26 Nov 2014

**Birgitt Schuele,**

We very much appreciate the careful review and excellent comments, suggestions and future directions of the reviewers. We hope to have addressed all of the comments to the reviewers’ satisfaction.

There is some inconsistency regarding the number of the ncECRs identified in the initial screen between the different sections of the article (32, 34, 37). Please make the corrections where needed.

*Thanks so much for the comment. We made changes to reflect the correct number of 34 ncECRs. We combined counts for ncECRs that were located very closely in the luciferase assay to one ncECR therefore different numbers appeared in the text. That has been addressed.*

Additional necessary control for the Luciferase experiments is a pGL-(SV40) promoter vector harboring an insert of a scrambled sequence that its size range mimics the average insert size of the tested ECRs. This is required to control for the ‘spacer’ effect of ECR lengths.
We have included in our analysis three controls: 1. The pGL3-Basic Vector which lacks eukaryotic promoter and enhancer sequences should not show any transcription activity. 2. The pGL3-Enhancer Vector contains an SV40 enhancer located downstream of luc+ and the poly(A) signal and is showing transcription at very high levels (enhancer element is 246bp in length). 3) pGL3-Promoter Vector contains an SV40 promoter upstream of the luciferase gene (promoter is 202bp in length).

Even though we have not directly included a control with scrambled sequence, we think that the ncECR elements that do not change transcription of luc+ provide enough evidence that the experimental system is valid. Of a total of 34 in silico determined elements, only 12 show an effect of transcriptional regulation. 22 elements did not change expression compared to pGL3-Promoter Vector.

What method was used for the statistical analysis? It is also not clear in the text whether all significant changes were calculated in comparison to the SV-40 promoter-only vector. That should be described in details in the method section.

A description of the analysis of luciferase assays was lacking and has now been added as a paragraph at the end of Method section Cloning and luciferase assays and reads as follows:

“Statistical analysis:
Differences among means were analyzed using two-samples student’s t-test. For differences in transcriptional activation of the luc+ gene, ncECRs were tested in quadruplicates in three independent experiments. Differences were considered statistically significant at p<0.05.”

To demonstrate the important implication of this study the authors are recommended to follow up on an event as an example. That is to say, to evaluate the effect of a genetic variation, a PD-associated SNP, on the regulatory function of the corresponding ECR using the luciferase system established in this work. Figure 3 demonstrates overlap between PD associated SNPs and ncECR, connecting these dots will be of high significance. This is an excellent suggestion and will definitely be conquered in future work with this system as this is the basis for the understanding of transcriptional regulation of the SNCA locus for potential translational applications. The presented study was intended to understand the basic changes in transcriptional regulation within the SNCA locus.

Supp Table: there is a typo in the coordinates of D2. In the footnote include the human genome assembly of the coordinates.

We corrected the coordinates for D2 which was a duplicate of D1 with the correct genomic location chr4:90844830+90845413 and added in the header the corresponding Human Genome assembly NCBI36/hg18 (March 2006).

Figure 2A X-axis: modify title to ‘upstream…’

Correction has been made. It reads now in Figure 2A “Upstream SNCA conserved elements”. We also changed for consistency Figure 2B to “Intronic SNCA conserved elements” and capitalized Figure 2C “Downstream SNCA conserved elements”.

Omit Figure 3A. Instead include a new panel to figure 3B that indicates the position of the putative binding sites of these TFs within SNCA locus.

We have modified Figure 3 according to the MatInspector network view with respective changes in the legend. We also included which genomic sequences have been analyzed in the text. Since
this is a preliminary in silico analysis, we feel that the overview is sufficient and has to be validated in functional studies. As pointed out below by the reviewer, these analyses have to be taken with care and a grain of salt.

The identification of Transcription Factor Binding Sites (TFBS) is an important step required in order to evaluate the transcriptional regulation network of the SNCA gene. To this end, the computational prediction of TFBS is a classic approach that gives preliminary data but should be interpreted with caution. Integration of the classic approach with new models described in Mathelier & Wasserman (2013) is highly recommended. The relation between TF motifs and in vivo binding sites is far from simple. The analysis lacks of information about the context of the identified sequences. TF are highly context-specific, and the same TF typically binds to different genomic binding sites in different conditions. Obtaining information about the context could be helpful in better understanding the possible involvement of the predicted sites as TFBS. While this is beyond the scope of this study, this topic should be thoroughly discussed in the discussion section.

Thank you very much for this suggestion. Indeed, further studies are necessary to provide experimental evidence for the binding of predicted transcription factors. The analysis provided in this article was only a first step to model potential transcription factor binding sites and should stimulate further studies.

The reference Mathelier and Wasserman has been now included in the Discussion of the manuscript and reads as follows:

“Computationally determining transcription factor binding sites is a challenging process and multiple prediction algorithms have been developed over the last decade (Cartharius 2005, Wu 2009, Mathelier 2013). Therefore our preliminary data should solely open the discussion and drive novel hypotheses for potential transcription factors that regulate transcription of the SNCA locus.”

**Competing Interests:** None.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com