

OPEN

Common genetic variants associated with Parkinson's disease display widespread signature of epigenetic plasticity

Amit Sharma^{1,2}, Naoki Osato³, Hongde Liu⁴, Shailendra Asthana⁵, Tikam Chand Dakal⁶, Giovanna Ambrosini⁷, Philipp Bucher⁷, Ina Schmitt¹ & Ullrich Wüllner^{1,8*}

Parkinson disease (PD) is characterized by a pivotal progressive loss of substantia nigra dopaminergic neurons and aggregation of α -synuclein protein encoded by the *SNCA* gene. Genome-wide association studies identified almost 100 sequence variants linked to PD in *SNCA*. However, the consequences of this genetic variability are rather unclear. Herein, our analysis on selective single nucleotide polymorphisms (SNPs) which are highly associated with the PD susceptibility revealed that several SNP sites attribute to the nucleosomes and overlay with bivalent regions poised to adopt either active or repressed chromatin states. We also identified large number of transcription factor (TF) binding sites associated with these variants. In addition, we located two docking sites in the intron-1 methylation prone region of *SNCA* which are required for the putative interactions with DNMT1. Taken together, our analysis reflects an additional layer of epigenomic contribution for the regulation of the *SNCA* gene in PD.

Parkinson's disease (PD) is a heterogeneous age-associated incurable neurodegenerative syndrome, occurring in both sporadic and familial forms. Although phenotypically diverse, the defining pathological hallmark of PD is the loss of midbrain dopaminergic neurons in the substantia nigra pars compacta. Surviving dopaminergic neurons display characteristic extensive aggregate confirmations (Lewy bodies, LB), which among other proteins contain alpha-synuclein (α -SYN). Although it is unclear why exactly neurons degenerate in PD, it is striking that the *SNCA* gene (encoding α -SYN) has repeatedly been identified in several genome-wide association studies (GWAS)¹⁻³. The importance of α -SYN in the pathogenesis of PD is highlighted by the fact that a multiplication of *SNCA* lead to severe Parkinsonism, which suggests that *SNCA* expression level, determines the severity of the PD pathology⁴.

In addition to genetic changes, understanding of epigenetic alterations is equally important to explain the dynamics of histone modifications and transcription factor binding sites that strongly regulate the gene expression. Among genetic regulators, microRNAs and transcription factors (TF) plays an essential role for regulating gene expression. Furthermore, the alterations induced by environmental factors also contribute to the disease pathology. In previous experiments our team demonstrated that decreasing methylation of intron-1 of *SNCA* increased the expression of α -SYN⁵.

Herein, we focused especially on a putative impact of the genetic variants (SNPs) distributed throughout the *SNCA* gene on the epigenetic landscape (chromatin mark diversity, nucleosome occupancy, transcription factors motifs, chromatin organization at splice sites). Furthermore, we provide structural insights into the methylation prone intron-1 region of *SNCA* as binding site of DNMT1 (DNA methyltransferase 1).

¹Department of Neurology, University Clinic Bonn, Bonn, Germany. ²Department of Ophthalmology, University Clinic Bonn, Bonn, Germany. ³Department of Bioinformatics Engineering, Osaka University, Osaka, Japan. ⁴State Key Laboratory of Bioelectronics, Southeast University, Nanjing, China. ⁵Drug Discovery Research Centre (DDRC), Translational Health Science and Technology Institute (THSTI), Haryana, 121001, India. ⁶Genome & Computational Biology Lab, Department of Biotechnology, Mohanlal Sukhadia University, Udaipur, 313001, Rajasthan, India. ⁷EPFL and Swiss Institute of Bioinformatics, Lausanne, Switzerland. ⁸German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. *email: ullrich.wuellner@ukbonn.de

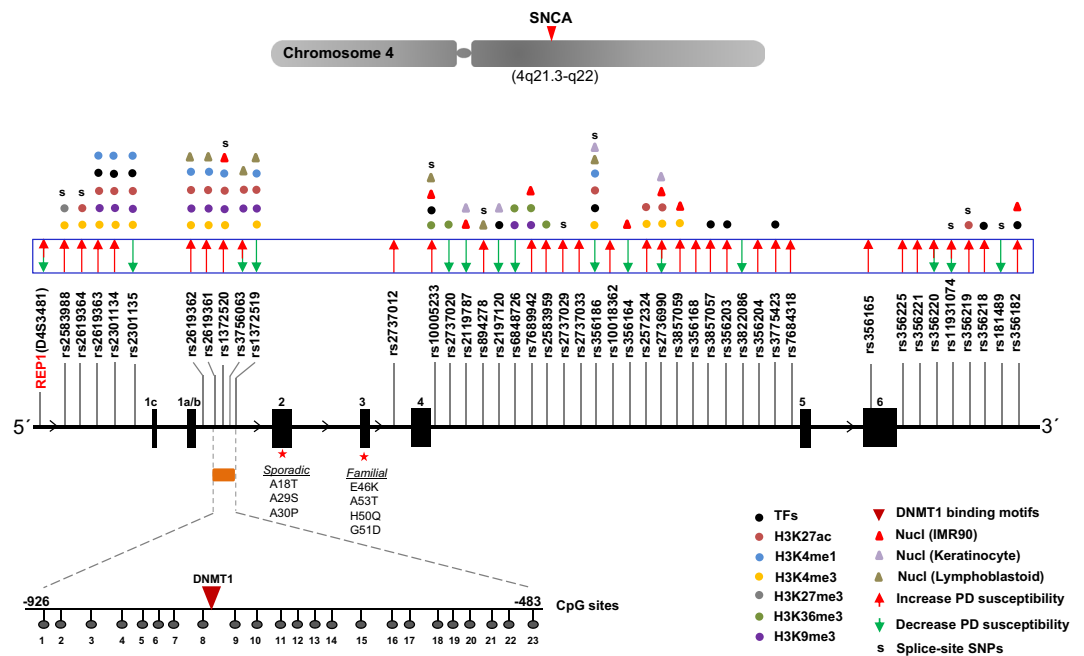


Figure 1. Sequence variants in *SNCA* and associated epigenetic landscape. (A) Figure showing exon/intron structure of *SNCA* gene at chromosome 4 of human genome. The relative positions of SNPs are plotted and high/low susceptibility of Parkinson's disease is shown with arrows. Transcription factors and multiple histone modification marks overlays SNP sites are shown. SNPs enriched at nucleosomes are marked with triangles. DNMT1 interaction site corresponding to 23 CpG sites at intron-1 region of *SNCA* gene is demonstrated. Sporadic/familial known mutations at exon 2–3 are described for the orientation of *SNCA* gene.

Results

Nucleosomes occupancy affinity for transcription factors in the vicinity of risk variants. All the SNPs selected for this study were previously linked to increased or decreased PD susceptibility (29 SNPs: high susceptibility, 10 SNPs: low susceptibility, 4 SNPs: variability in clinical data) (Fig. 1, adapted from Campelo *et al.*⁶). The distribution of PD associated SNPs appeared to be non-random as most of them were found to be embedded in the intron 4 region. To address how the sequence variation is reflected in the occupancy profiles of nucleosomes, we categorized them on the basis of nucleosome occupancy i.e. SNPs located within nucleosomes and SNPs lacking nucleosomes in their vicinity. In order to control the processing for the sequenced datasets, we first checked the nucleosome occupancy profiles near transcription start sites (TSSs). The profiles showed a typical distribution pattern of nucleosome-depleted region (NDR) flanked by two well positioned nucleosomes, followed by equally spaced nucleosome array downstream of TSSs. The nucleosome occupancy around the selective SNPs was measured in three different cell types (keratinocytes, IMR90, lymphoblastoid; Fig. 2, Supplementary Fig. 1B). As nucleosome positioning varies in different cell types, in our analysis we show only the SNPs with significant read distribution ($P \leq 0.05$) in every cell type. We also performed the analysis on undifferentiated human iPS cells (hiPS) and iPS cells differentiated to neural progenitor cells (NPC). The data clearly demonstrate that most of SNPs were located on the nucleosomes, while few among them (rs2197120, rs2737033, rs7684318) were also occupied by the nucleosome after differentiation into the NPC (Supplementary Fig. 3).

Nucleosomes impair the binding of transcription factors (TFs), thus SNPs in the TF binding sites may exert effects upon binding affinity of TFs^{7,8}. Our results indicate that the effect of SNPs on TFs binding varies according to the cell types. SNP having a variable nucleosome occupancy e.g. rs1372519 was occupied in keratinocytes and IMR90, but not in lymphoblastoid (Fig. 2C,D), which also means that the SNP effect on TF binding will be in lymphoblastoid, but not in keratinocytes and IMR90. We found no correlation for genomic distance (striking periodicities across nucleosomal regions) between these SNPs. In our cumulative analysis, we identified 11 SNPs in TF binding sites, among them 5 in nucleosome positions and 3 lacked any enrichment of chromatin marks (Fig. 1, Supplementary Table 2). Among former some TFs such as FOXA1, NFATC2; Prrx2, SP1 have previously been linked with PD^{9–12}.

A polymorphic microsatellite REP1 (D4S3481), located in the promoter region (~10 kb upstream) of *SNCA* also contains SNPs but they have not yet been linked to PD risk or phenotype. REP1 may however, play a role in the regulation of *SNCA* expression as its longest (263 bp) and intermediate length (261 bp) alleles are associated with an increased risk for PD¹³. Previously, the binding to transcription factor poly (ADP-ribose) transferase/polymerase-1 (PARP-1) with REP1 has been described¹⁴. Herein, using computationally predicted binding sites, we described additional transcription factor motifs enriched in this tandem repeats region (Fig. 1, Supplementary Fig. 2). As the Rep1 sequence contains only (TG)_n and (GA)_n motifs, but no CG sequences, a role for CpG-methylation can be excluded.

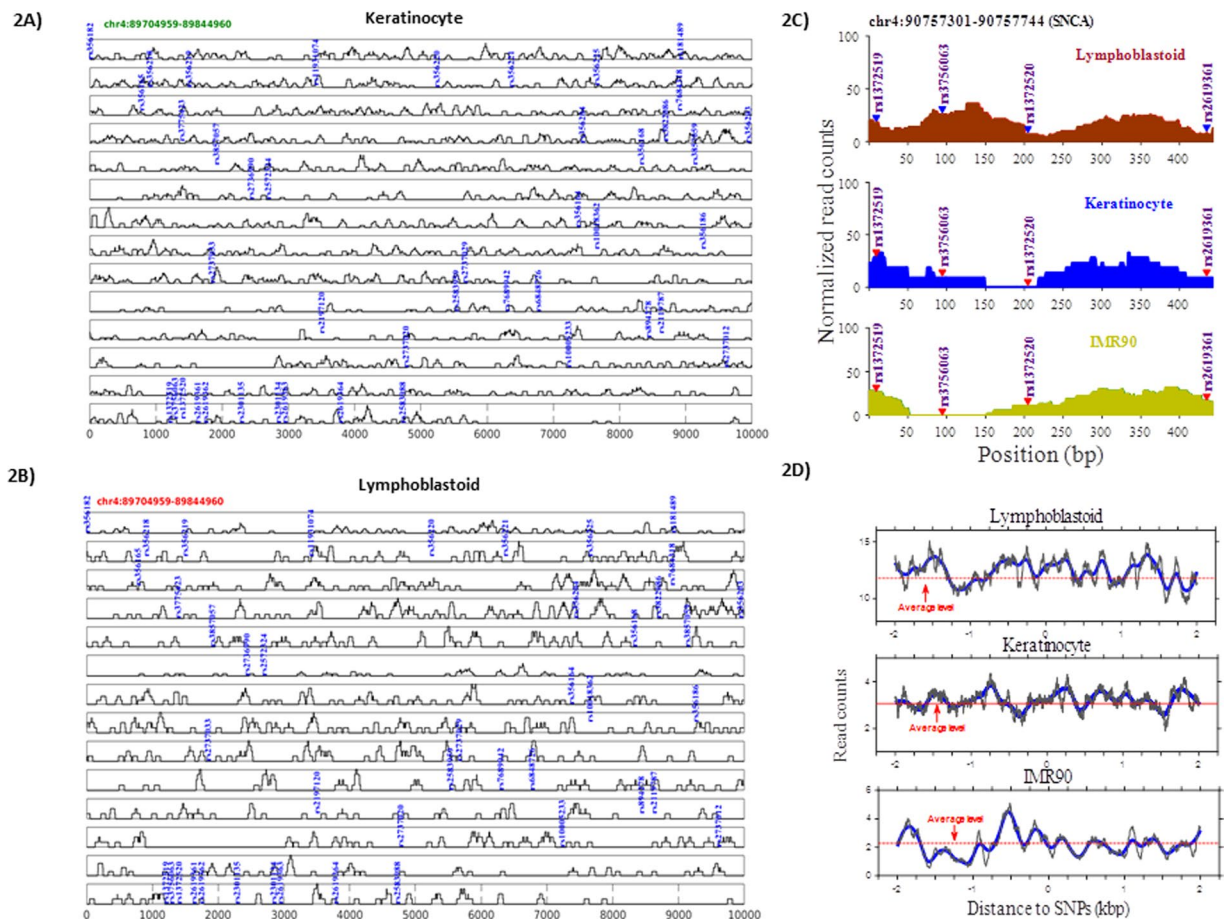


Figure 2. Nucleosome occupancy in vicinity of *SNCA* sequence variants. Nucleosome occupancy around the SNPs is summarized for lymphoblastoid (A) and keratinocyte (B). Nucleosome occupancy differs around the *SNCA* sequence variants (SNP sites) in three cell types with normalized read counts is drawn (C). Nucleosome occupancy profiles near the *SNCA* sequence variants with the profiles aligned by SNP sites and averaged by dividing number of the SNP sites is evaluated (D).

Distinct chromatin structure at SNPs in different genomic regions. Enhancers and repressors, which are sometimes located far from TSSs, have been found to affect the expression of genes significantly^{15–17}. In our analysis, we found SNP rs1372519 within a peak of H3K27ac (enhancer and promoter marks) and within the DNA binding motif sequence of NeuroD1 (NDF1), which had been involved in a specific neurogenic program, including differentiation and migration earlier¹⁸. Likewise, SNP rs2301134 is located within a peak of H3K9me3 histone modification mark (repressor and heterochromatin marks) and the DNA binding motif sequence of PRDM1 (BLIMP1), which has been shown previously to modulate the divergence of neural or germline fates through repression of SOX2 during human development¹⁹. SNP rs2619364 was located near a peak of H3K27me3 histone modification mark (repressor and heterochromatin marks) and a DNA binding motif sequence of RE1-Silencing Transcription factor (REST), also known as Neuron-Restrictive Silencer Factor (NRSF). NRSF is a zinc-finger transcription factor initially described as a nuclear negative regulator of differentiation²⁰, now known to play a role in neuronal cells. Furthermore, we found that selective SNP sites lay within bivalent regions poised to adopt either active or repressed chromatin states. Overall, SNPs located in promoter region and intron 4 site were found to be enriched with histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3). To undermine the bivalency of histone marks, we further analyzed H3K4me1, H3K4me3, H3K36me3, H3K9me3, H3K27me3, H3K27ac and Input DNA in human monocyte, stomach, fetal muscle leg, H9 human embryonic stem cell (ESC), neural progenitor populations of neuroepithelial (NE), early radial glial (ERG) and mid radial glial (MRG). The bivalency of H3K4me3 and H3K27me3 around the transcriptional start site (TSS) of *SNCA* was observed in all tissue and cell types, while the co-existence of H3K27ac and H3K27me3 was limited to few sample types. Interestingly, the co-existence of H3K4me1 and H3K27me3 was also noticed in all the tissue and the cell types (Supplementary Fig. 4).

Intron-1 of *SNCA*, a docking site for DNMT1. *SNCA* is expressed as distinct transcripts variants with at least three alternative transcription start sites/exons1. Previously, some activity of core promoter in the upstream of exon 1b (variant 2) was predicted²¹. But the CAGE analysis, enrichment of CpG sites and several putative transcription factor binding sites point towards an additional promoter region in the intron 1. Importantly, the

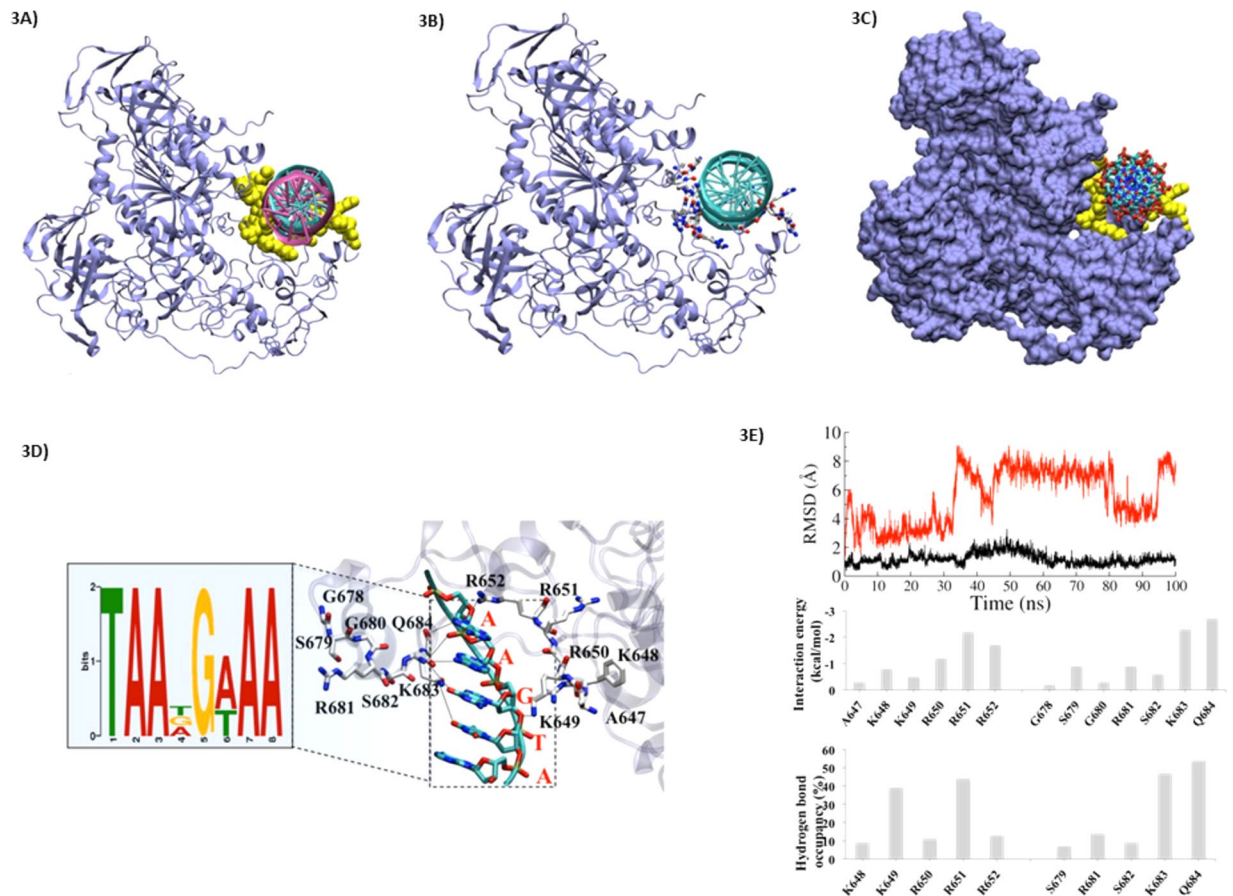


Figure 3. Characterization of binding site of interaction for DNMT1 with SNCA intron-1. (A) DNMT1 interaction with identified motifs (8 bp). The nucleotides are shown in cyan, and key residues are highlighted in yellow and shown in in VdW sphere. The crystalized DNA (in magenta) is shown to check the binding site. (B) The DNMT1 and DNA interacting zone are shown by residues shown in licorice and coded by atom wise: Carbon: white, Oxygen: red, Nitrogen: blue and Sulphur: yellow. (C) The DNMT1 in surface view and DNA are shown in licorice view. (D) The interaction site between nucleotides and key amino acids. The identified motif is highlighted in zoom-out view. (E) Root mean square deviation (RMSD) of APO (DNMT1 without nucleotide in red) and complex (DNMT1 with nucleotide in black) DNMT1. The dotted line is showing the average value. Residue-wise interaction energy values in kcal/mol and residue-wise hydrogen bonding occupancy of key residues in percentage wise (%) is shown.

methylation levels in the intron-1 of SNCA have been associated with expression of α -SYN in PD^{5,21,22}. Therefore, we searched for potential docking sites required for putative interactions with DNMT1 in intron-1. Based on our previously published methodology, the molecular docking was conducted between the motifs identified from nucleotides and DNMT1²³.

In order to understand how the extensive positive potential created by DNMT1 influences DNA binding, we also performed rigid-body dockings with DOT (a computational docking tool for macromolecular interactions) by using our established method²⁴. Docking of linear B-DNA (flexible) 8 bp from the identified motif to DNMT1 (stationary) identified the similar distinct DNA-binding sites in the 109 top-ranked conformations (Fig. 3A–C). The B-DNA conformations, which are docked (39 of the 100 solutions) showed significant interactions with the protein residues of DNMT1 protein. The side chains of highly conserved amino acids of DNMT1 (three Arginines: R650, R651 and R652) were found to be inserted into the DNA major groove, while the backbone of some other residues A647, K648, and K649 were also found to have high binding affinity. The localization of binding site constitutes the region from G678 to A685, among them the key residues which contribute heavily in the stability and high affinity of DNA motif were R681, K683 and Q684.

Because of the observed high binding affinity, we took the docked position as prime site and used it for further molecular dynamic simulations. We observed that the complementary interactions between DNMT1 and DNA fit well along with the basic patches of DNA. From the crystal structure 3SWR, it appears that the region near to DNA binding site might play a role to hold the DNA which we observed in different orientation and in different crystals. This region also shows very low electron density and highest B-factor, indicating its structural flexible.

From the crystal structure 3SWR, it appears that the missing link near DNA binding site might play a role to hold the DNA as we observed in different orientation in different crystals. This region also shows very low electron density and highest B-factor, indicating its structural flexible.

Molecular Interaction of conserved DNA motif and DNMT1. We performed the interaction studies by using random genes VHL-150, RB1-445, WT1-390, WT1-394, and WT1-396 (including the “AGCGA” motif), according to our previously shown methodology²⁵. Our data showed that the residues K649, R650, R651 and R652 make important contributions in increasing the stability of DNA-protein complexes (Fig. 3D). It was found that the key interaction of identified motif may take place with two amino acids Q684 and K683 i.e. A@Q684 and A/G/T@K683. In case of Q684 this interaction persists throughout the simulation, while K683 showed interaction with three different nucleotides. This outcome also supports the relevance of dynamics analysis rather than only static analysis through docking. From the interaction energy analysis these two interactions claimed -3.4 and -3.2 kcal/mol energy, respectively. Other key interactions were G@R650 and A@R652 (Fig. 3E, upper). The residue analysis also helps to explain the stability of the complex (Fig. 3D). The interaction energy graph revealed that 76% of binding affinity contribution is derived from the key residues. The residues A647, K648, K649, R650, R651, R652, G678, S679, G680, R681, S682, K683 and Q684 contributed > -0.5 kcal/mol. However, the maximum contribution (above ~ 1.5 kcal/mol) was exerted by amino acids R562, R681, K683 and Q684 (Fig. 3E, middle).

These results are supported by hydrogen bond (HB) occupancy analysis as well, in which we observed a similar trend of contribution from the residues in terms of their establishment of HBs (Fig. 3E, lower). In addition, the binding affinity of a conserved motif also showed similar kind of interaction pattern. Furthermore, the analysis from molecular dynamics also indicate that the common key residues of B-DNA (mostly G and A) interacted significantly. The results from crystal structure clearly show that the newly identified motif of nucleotides interacts with DNMT1, however, the experimental data is required to validate them. Overall, we identified two docking sites required for putative interactions with DNMT1 in intron-1, which were near to CpG8-CpG9 (Fig. 1A). It is well documented that the enzyme DNMT1 binds to DNA and adds methyl groups at the C residue of CpG islands. In this context, we can propose that the high levels of methylation at CpG8-CpG9 among 23 CpG sites observed by Jowaed *et al.*⁵ might be the result of DNMT1 interactions.

To determine DNA methylation at complete *SNCA* gene, we checked the correlation of *SNCA* expression and DNA methylation around *SNCA* in iPSC-derived dopaminergic neurons using public available data (GSE51921)²⁶. The analysis showed that during the differentiation there is a slight increase in DNA methylation at upstream of *SNCA*, however, the gene expression was also increased [Fold change (PD/control) = 1.13; P-value (t-test) (log10) = 2.49; q-value = 0.053] (Supplementary Fig. 5).

Discussion

The last few years have seen expanding knowledge of the genetic-epigenetic insights of PD pathology^{27–29}. There is sufficient evidence that *SNCA*, which has been identified as first PD gene plays a central role in this disease. Therefore, in this study, we focused on the PD associated SNPs embedded within *SNCA* gene and investigated their epigenetic landscape. Herein, we could show that the risk-associated SNPs are enriched at the nucleosomes and are associated with multiple histone modifications. Among repressive histone marks (H3K27me3, H3K9me3), H3K9me3 was uniformly present in most SNP sites located at promoter region, while H3K27me3 correlates only to rs2583988. Similarly, histone marks for active chromatin (H3K27ac, H3K4me1, H3K4me3, H3K36me3) were also found to be overlapping with certain variants. In intron 4 specifically, we could define 2 regions (Chr4:89821930–89800236; 89772075–89753230), where most of PD related SNPs were located. How the interplay of active/repressive histone marks contributes towards *SNCA* regulation through these SNP sites is unclear. However, previously studies provided a hint that the genetic variants might engage in crosstalk with sequence-specific transcription factors and epigenetic patterns (including DNA methylation^{30,31}) thereby impacting the transcription of genes locally or remotely^{32,33}. Importantly, each histone modification is unique to regulate the predictive chromatin signature at promoters, enhancer and TSS, still they can also share overlapping chromatin³⁴ signatures³⁵. This is what we have observed in several SNP sites, where bivalent signature (co-existing “activating” and “silencing” nucleosomal modifications) poised to adopt either active or repressed chromatin states. It is noteworthy to mention that even though the bivalency of H3K4me3 and H3K27me3 around the transcriptional start site (TSS) of *SNCA* was observed in all tissue and cell types, the low *SNCA* expression was also observed in some samples (e.g. stomach and fetal muscle). Thus, it can be argued that the observed bivalency at the given locus simply reflects the cellular heterogeneity (mosaic signal from the mixed population of cells). Another possibility might be the repression of *SNCA* antisense overlapping with *SNCA* (at their 5' ends) by H3K27me3 histone mark. Therefore, in the future, the emerging single-cell epigenomic methods hold the potential to add deeper insights into *SNCA* gene regulation.

Apart from histone medications, disease variants are also known to alter transcriptional levels. Specifically, the variants in coding regions may confer disease risk through altered protein sequences, however, variants in non-coding regions or regulatory variants (enhancer, sequence/TF binding sites)^{36–38} can also contribute to disease susceptibility by changing the gene transcription³⁹. Brenner *et al.* identified and confirmed the bindings of two zinc finger proteins ZSCAN21 (intron-1) and GATA2 (intron-2) in *SNCA*, but did not observe any SNP or mutations within these binding sites⁴⁰. Recently, two RNA-binding proteins (ELAVL1, TIAR) that target *SNCA* 3'UTRs and controlling its expression were also discussed⁴¹. Herein, we specifically investigated TF motifs overlapping the SNPs and found that 11 SNPs shared the affinity towards specific TFs. Interestingly, 2 of them were splice site SNPs and rs356186 among them emerged as the one with preferential sites for TFs as well as for various chromatin modifications. We could also show many of these TFs (FOXA1, NFATC2; Prrx2, SP1) have previously been implicated in PD. We extended our analysis for repetitive microsatellite D4S3481 (called REP-1) located upstream of *SNCA* gene and found several associated TFs motifs in and around the repetitive sequences. Recently,

Afek *et al.* claimed that two TF families (GATA and ELK) have consensus binding sites in a very close proximity to REP-11 region, which appears to be potentially relevant for SNCA transcriptional regulation⁴². Till date, apart from the expansion of repetitive sequence, none of the SNP inside this locus have been proven to be predictive for PD. Previously, It was shown that the alternative splicing also contribute towards the dysregulation of SNCA⁴³, therefore, our findings of chromatin regulation at splice site SNPs provide useful insights towards the molecular characteristics of SNCA gene.

To evaluate further evidence for genetic-epigenetic regulation of SNCA gene, we investigated promoter specific CpG rich region overlapping these SNP sites. We identified a DNMT1 binding site specifically at CpG8-CpG9 region of intron-1 region of SNCA, where a high level of methylation was previously observed by Jowaed *et al.*³. The control of DNMTs in SNCA locus is also confirmed by one recent study in which authors by using CRISPR-deactivated Cas9 (dCas9) fused with the catalytic domain of DNA-methyltransferase 3A (DNMT3A) claimed to reduce the expression levels of SNCA in human-induced pluripotent stem cell (hiPSC)-derived dopaminergic neurons from a PD patient with SNCA triplication⁴⁴. Therefore, we propose that the sequence specific recognition of DNMT1 we identified in this study may impose the intrinsic control of SNCA specific gene regulation. Apart from DNMT1, other unknown factor(s) may also have an additional role in SNCA regulation e.g. some risk variant in a non-coding distal enhancer element were predicted to regulate the expression of α -synuclein⁴⁵. Therefore, we checked the correlation of SNCA expression and DNA methylation in iPSC-derived dopaminergic neurons and observed an increase in both DNA methylation and expression at this particular locus. Hence, the contribution of multiple factors towards the expression of SNCA cannot be excluded. On a broader view, the multiplication of SNCA gene, hypomethylation of SNCA intron1, the A53T mutation and mutation of miRNA binding sites in the 3'UTR of SNCA, at least in part, may participate in the regulatory mechanisms of SNCA expression.

Taken together, we report an in-depth characterization of the epigenetic landscape of PD linked risk polymorphisms distributed across SNCA gene. The blueprint (Fig. 1: increased/decreased PD susceptibility) we have presented here will help to determine the clinical outcome of newly screened patients. Our data suggest that the chromatin marks overlaying the disease associated SNPs in SNCA, might alter the genomic architecture and contribute towards the disease susceptibility. Hence, on molecular level, we pose an open question, whether these disease risk variants can disproportionately influence the gene expression among different tissues/anatomical regions of the brain to cause phenotypic heterogeneity in PD.

Material and Methods

Classification and retrieval of SNPs. We investigated the panel of 34 SNPs previously found in GWAS studies and from the literature in association to SNCA gene and PD susceptibility. The genomic locations were retrieved from dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP/>) and NCBI for computational analysis. For predicting the splice acceptor or donor sites and exonic enhancer and silencer elements, we used Human Splicing Finder tool (<http://www.umd.be/HSF3/HSF.shtml>). The input sequence used was 50 bp up- and downstream of the SNP position related to the disease associated ten rsIDs. From the output, only predictions with score more than 70 percent were considered in results. Transcription factor enrichment analysis was performed by using SNP2TFBS (<http://cgg.vital-it.ch/snp2tfbs/>) tool which selects and visualizes user defined variants that affect single or multiple transcription factors. Apart from SNCA gene, Rep1 (Accession no. U46895; D4S3481), the polymorphic microsatellite repeat located approximately 10 kb upstream of the SNCA gene, was also checked for 9 SNP sites and TF binding sites.

Histone modifications and regulatory SNPs in the SNCA. To examine the overlap between SNPs and histone modification marks, FASTQ files of ChIP-seq data of H3K27ac, H3K4me1 and Input DNA in neuron were download from Sequence Read Archive (SRA) database referred from Gene Expression Omnibus (GEO) database (GSE71278), FASTQ files of ChIP-seq data of H3K4me3, H3K36me3, H3K9me3, H3K27me3 and Input DNA in fetal brain (GSE17312; GSM621457, GSM621410, GSM621393, GSM621427, GSM916054, GSM706851) were downloaded respectively. BAM files of ChIP-seq data of H3K4me1, H3K4me3, H3K36me3, H3K9me3, H3K27me3, H3K27ac and Input DNA in human monocyte, stomach, fetal muscle leg (ENCSR518BPP, ENCSR949WGV, ENCSR820MXK) were downloaded respectively from ENCODE project database (<https://www.encodeproject.org>). FASTQ files of ChIP-seq data of H3K27ac, H3K27me3, H3K4me1, H3K4me3 and Input DNA in H9 human embryonic stem cell (ESC) and neural progenitor populations of neuroepithelial (NE), early radial glial (ERG) and mid radial glial (MRG) were downloaded from SRA database referred from GEO database (GSE62193). Reads in FASTQ files of ChIP-seq data were aligned to the hg19 version of the human reference genome using BWA mem with default parameter⁴⁶. BAM files produced by BWA were converted to SAM files, sorted and indexed by Samtools⁴⁷. Peak calling of ChIP-seq data was performed by SICER with optional parameters 'hg19 1 200 150 0.74 600 0.01'⁴⁸. WIG files of ChIP-seq data were produced from the BAM files by bam2wig.pl with optional parameters '-extend-rpm-bin 200' in Biotoobox (<https://github.com/tjparnell/biotoobox>), and were displayed using custom track in UCSC Genome browser (<http://genome.ucsc.edu/>). The overlaps between SNPs and peaks of histone modification marks were investigated using intersectBed of Bed tools with default parameter⁴⁹. In context to check the overlap between SNPs and open chromatin regions, we obtained the Bed file of peaks of DNase-DGF data in fetal brain from NIH Epigenome roadmap website (<http://www.roadmapepigenomics.org>) and GEO database (GSM723021), and investigated the overlaps between SNPs and peaks of open chromatin regions using intersect Bed.

Repeat DNA sequences were searched for hg19 version of the human reference genome using RepeatMasker (<http://www.repeatmasker.org>) and RepBase RepeatMasker Edition (<http://www.girinst.org>). The overlap of repeat DNA sequences and SNPs was examined using intersectBed.

Nucleosome positioning around single nucleotide polymorphisms. To determine the nucleosome occupancy around SNPs, sequencing datasets of nucleosomes were retrieved from three cell types: lymphoblastoid (Gene expression omnibus (GEO) accession number: (GSM907783), keratinocytes (GSE65191), IMR90 (GSE44985) and neural progenitor cells (GSE117870). Reads in the datasets were determined with micrococcal nuclease digestion followed by next generation sequencing (MNase-Seq). Coordinates of single nucleotide polymorphisms (SNP) sites and transcription start sites (TSSs) were retrieved from UCSC using the Tables function (<http://genome.ucsc.edu>) in human genome assemble hg19 (GRCh7.p13). For nucleosome occupancy profile, the raw reads was mapped to human genome (hg19) using Bowtie⁵⁰. Only the uniquely mapped reads were used to further analysis. Nucleosome occupancy refers to reads count at each genomic locus which means the count of the reads covering the locus. Each read was extended to 73 bp in the 3' direction and shifted by 36 bp towards the 3' direction. The nucleosome occupancy profiles near special sites (e.g., SNP sites) were represented with the average nucleosome occupancy profile, which was calculated by summing the occupancy signal at each genomic site and then dividing the summed signal by the gene number. For the SNPs, the raw reads count data was normalized by dividing by the maximal reads count of whole genome and then multiplying by 100.

Significance of reads enrichment at each SNP sites was calculated with Poisson distribution

$P(k) = e^{-\lambda} \lambda^k / k!$, λ indicates the average reads count, k is reads count at SNP sites). A SNP site with a $P \leq 0.05$ was considered as being occupied by nucleosomes.

DNMT1-SNCA Intron-1 binding analysis. The intron-1 region of *SNCA*, previously evaluated for DNA methylation levels was used to find DNMT1 interacting sites⁵. The identification of motifs from nucleotide sequences, molecular docking and molecular dynamics simulation was performed according to the methods we have published previously^{23–25}.

Received: 3 May 2019; Accepted: 20 November 2019;

Published online: 05 December 2019

References

- 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* **46**, 989–993, <https://doi.org/10.1038/ng.3043> (2014).
- Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**, 1511–1516, <https://doi.org/10.1038/ng.3955> (2017).
- Blauwendraat, C. *et al.* Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and alpha-synuclein mechanisms. *Mov Disord*, <https://doi.org/10.1002/mds.27659> (2019).
- Chartier-Harlin, M. C. *et al.* Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169, [https://doi.org/10.1016/s0140-6736\(04\)17103-1](https://doi.org/10.1016/s0140-6736(04)17103-1) (2004).
- Jowaed, A., Schmitt, I., Kaut, O. & Wullner, U. Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains. *J Neurosci* **30**, 6355–6359, <https://doi.org/10.1523/jneurosci.6119-09.2010> (2010).
- Campelo, C. & Silva, R. H. Genetic Variants in SNCA and the Risk of Sporadic Parkinson's Disease and Clinical Outcomes: A Review. *Parkinsons Dis* **2017**, 4318416, <https://doi.org/10.1155/2017/4318416> (2017).
- Liu, H., Duan, X., Yu, S. & Sun, X. Analysis of nucleosome positioning determined by DNA helix curvature in the human genome. *BMC Genomics* **12**, 72, <https://doi.org/10.1186/1471-2164-12-72> (2011).
- Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum Mol Genet* **24**, R102–110, <https://doi.org/10.1093/hmg/ddv259> (2015).
- Domanskyi, A., Alter, H., Vogt, M. A., Gass, P. & Vinnikov, I. A. Transcription factors Foxa1 and Foxa2 are required for adult dopamine neurons maintenance. *Front Cell Neurosci* **8**, 275, <https://doi.org/10.3389/fncel.2014.00275> (2014).
- Luo, J. *et al.* A calcineurin- and NFAT-dependent pathway is involved in alpha-synuclein-induced degeneration of midbrain dopaminergic neurons. *Hum Mol Genet* **23**, 6567–6574, <https://doi.org/10.1093/hmg/ddu377> (2014).
- Qu, D. *et al.* Role of Cdk5-mediated phosphorylation of Prx2 in MPTP toxicity and Parkinson's disease. *Neuron* **55**, 37–52, <https://doi.org/10.1016/j.neuron.2007.05.033> (2007).
- Yao, L. *et al.* Inhibition of transcription factor SP1 produces neuroprotective effects through decreasing MAO B activity in MPTP/MPP(+) Parkinson's disease models. *J Neurosci Res* **96**, 1663–1676, <https://doi.org/10.1002/jnr.24266> (2018).
- Cronin, K. D. *et al.* Expansion of the Parkinson disease-associated SNCA-Rep1 allele upregulates human alpha-synuclein in transgenic mouse brain. *Hum Mol Genet* **18**, 3274–3285, <https://doi.org/10.1093/hmg/ddp265> (2009).
- Chiba-Falek, O., Kowalak, J. A., Smulson, M. E. & Nussbaum, R. L. 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* **76**, 478–492, <https://doi.org/10.1086/428655> (2005).
- Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187, <https://doi.org/10.1038/nature09033> (2010).
- Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat Neurosci* **17**, 1330–1339, <https://doi.org/10.1038/nn.3808> (2014).
- Osato, N. Characteristics of functional enrichment and gene expression level of human putative transcriptional target genes. *BMC Genomics* **19**, 957, <https://doi.org/10.1186/s12864-017-4339-5> (2018).
- Pataskar, A. *et al.* NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J* **35**, 24–45, <https://doi.org/10.15252/embj.201591206> (2016).
- Lin, I. Y. *et al.* Suppression of the SOX2 neural effector gene by PRDM1 promotes human germ cell fate in embryonic stem cells. *Stem Cell Reports* **2**, 189–204, <https://doi.org/10.1016/j.stemcr.2013.12.009> (2014).
- Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–1363 (1995).
- Holzmann, C. *et al.* Polymorphisms of the alpha-synuclein promoter: expression analyses and association studies in Parkinson's disease. *J Neural Transm (Vienna)* **110**, 67–76, <https://doi.org/10.1007/s00702-002-0769-5> (2003).
- Xia, Y. *et al.* Characterization of the human alpha-synuclein gene: Genomic structure, transcription start site, promoter region and polymorphisms. *J Alzheimers Dis* **3**, 485–494 (2001).
- Mattapally, S., Singh, M., Murthy, K. S., Asthana, S. & Banerjee, S. K. Computational modeling suggests impaired interactions between NKX2.5 and GATA4 in individuals carrying a novel pathogenic D16N NKX2.5 mutation. *Oncotarget* **9**, 13713–13732, <https://doi.org/10.18632/oncotarget.24459> (2018).

24. Mehta, J., Asthana, S., Mandal, C. C. & Saxena, S. A molecular analysis provides novel insights into androgen receptor signalling in breast cancer. *PLoS One* **10**, e0120622, <https://doi.org/10.1371/journal.pone.0120622> (2015).
25. Chowdhury, K. *et al.* Presence of a consensus DNA motif at nearby DNA sequence of the mutation susceptible CG nucleotides. *Gene* **639**, 85–95, <https://doi.org/10.1016/j.gene.2017.10.001> (2018).
26. Fernandez-Santiago, R. *et al.* Aberrant epigenome in iPSC-derived dopaminergic neurons from Parkinson's disease patients. *EMBO Mol Med* **7**, 1529–1546, <https://doi.org/10.15252/emmm.201505439> (2015).
27. Wullner, U., Kaut, O., deBoni, L., Piston, D. & Schmitt, I. DNA methylation in Parkinson's disease. *J Neurochem* **139**(Suppl 1), 108–120, <https://doi.org/10.1111/jnc.13646> (2016).
28. de Boni, L. & Wüllner, U. Epigenetic Analysis in Human Neurons: Considerations for Disease Modeling in PD. *Frontiers in Neuroscience* **13**, <https://doi.org/10.3389/fnins.2019.00276> (2019).
29. Sharma, A. *et al.* Skewed X-chromosome inactivation and XIST locus methylation levels do not contribute to the lower prevalence of Parkinson's disease in females. *Neurobiol Aging* **57**, 248 e241–248 e245, <https://doi.org/10.1016/j.neurobiolaging.2017.05.021> (2017).
30. Sharma, A. *et al.* Detailed methylation map of LINE-1 5'-promoter region reveals hypomethylated CpG hotspots associated with tumor tissue specificity. *Mol Genet Genomic Med* **7**, e601, <https://doi.org/10.1002/mgg3.601> (2019).
31. Sharma, A. *et al.* Classic bladder exstrophy and adenocarcinoma of the bladder: Methylome analysis provide no evidence for underlying disease-mechanisms of this association. *Cancer Genet* **235–236**, 18–20, <https://doi.org/10.1016/j.cancergen.2019.05.004> (2019).
32. McVicker, G. *et al.* Identification of genetic variants that affect histone modifications in human cells. *Science* **342**, 747–749, <https://doi.org/10.1126/science.1242429> (2013).
33. Guhathakurta, S., Bok, E., Evangelista, B. A. & Kim, Y. S. Deregulation of alpha-synuclein in Parkinson's disease: Insight from epigenetic structure and transcriptional regulation of SNCA. *Prog Neurobiol* **154**, 21–36, <https://doi.org/10.1016/j.pneurobio.2017.04.004> (2017).
34. Xie, J. *et al.* DNMMHMM: An approach to identify the differential nucleosome regions in multiple cell types based on a Hidden Markov Model. *Biosystems* **185**, 104033, <https://doi.org/10.1016/j.biosystems.2019.104033> (2019).
35. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897–903, <https://doi.org/10.1038/ng.154> (2008).
36. Sharma, A. *et al.* Mutational Landscape of the BAP1 Locus Reveals an Intrinsic Control to Regulate the miRNA Network and the Binding of Protein Complexes in Uveal Melanoma. *Cancers (Basel)* **11**, <https://doi.org/10.3390/cancers11101600> (2019).
37. Jamil, M. A. *et al.* F8 Inversions at Xq28 Causing Hemophilia A Are Associated With Specific Methylation Changes: Implication for Molecular Epigenetic Diagnosis. *Front Genet* **10**, 508, <https://doi.org/10.3389/fgene.2019.00508> (2019).
38. Sharma, A. *et al.* Towards a Central Role of ISL1 in the Bladder Exstrophy(-)Epispadias Complex (BEEC): Computational Characterization of Genetic Variants and Structural Modelling. *Genes (Basel)* **9**, <https://doi.org/10.3390/genes912060> (2018).
39. Hrdlickova, B., de Almeida, R. C., Borek, Z. & Withoff, S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim Biophys Acta* **1842**, 1910–1922, <https://doi.org/10.1016/j.bbdis.2014.03.011> (2014).
40. Brenner, S., Wersinger, C. & Gasser, T. Transcriptional regulation of the alpha-synuclein gene in human brain tissue. *Neurosci Lett* **599**, 140–145, <https://doi.org/10.1016/j.neulet.2015.05.029> (2015).
41. Marchese, D. *et al.* Discovering the 3' UTR-mediated regulation of alpha-synuclein. *Nucleic Acids Res* **45**, 12888–12903, <https://doi.org/10.1093/nar/gkx1048> (2017).
42. Afek, A. *et al.* Toward deciphering the mechanistic role of variations in the Rep1 repeat site in the transcription regulation of SNCA gene. *Neurogenetics* **19**, 135–144, <https://doi.org/10.1007/s10048-018-0546-8> (2018).
43. Gamez-Valero, A. & Beyer, K. Alternative Splicing of Alpha- and Beta-Synuclein Genes Plays Differential Roles in Synucleinopathies. *Genes (Basel)* **9**, <https://doi.org/10.3390/genes9020063> (2018).
44. Kantor, B. *et al.* Downregulation of SNCA Expression by Targeted Editing of DNA Methylation: A Potential Strategy for Precision Therapy in PD. *Mol Ther* **26**, 2638–2649, <https://doi.org/10.1016/j.yymthe.2018.08.019> (2018).
45. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* **533**, 95–99, <https://doi.org/10.1038/nature17939> (2016).
46. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324> (2009).
47. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, <https://doi.org/10.1093/bioinformatics/btp352> (2009).
48. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958, <https://doi.org/10.1093/bioinformatics/btp340> (2009).
49. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, <https://doi.org/10.1093/bioinformatics/btq033> (2010).
50. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, <https://doi.org/10.1186/gb-2009-10-3-r25> (2009).

Acknowledgements

The study was supported by the German Federal Ministry of Education and Research (BMBF/ANR) through the EpiPD (Epigenomics of Parkinson's disease) project, under the auspices of the bilateral Epigenomics of Common and Age-related Diseases Programme (grant # 01KU1403B; UW). Hongde Liu was supported by grants from the National Natural Science Foundation of China (No. 61972084; No. 81830053).

Author contributions

A.S. and U.W. conceived and designed the study; A.S., N.O., H.L., S.A., T.C.D., G.A., P.B. and I.S. contributed to data analysis; A.S. and U.W. wrote the manuscript, all authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-54865-w>.

Correspondence and requests for materials should be addressed to U.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019