Received: 4 December 2018 | Revised: 16 November 2019 | Accepted: 24 December 2019

DOI: 10.1002/humu.23974

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

RESEARCH ARTICLE

A human minisatellite hosts an alternative transcription start site for *NPRL3* driving its expression in a repeat number-dependent manner

Maria Bertuzzi[1] | Dave Tang[2] | Raffaella Calligaris[1,3] | Christina Vlachouli[1] | Sara Finaurini[1,4] | Remo Sanges[1,5] | Stefano Goldwurm[6] | Mauro Catalan[3] | Lucia Antonutti[3] | Paolo Manganotti[3] | Gilberto Pizzolato[3] | Gianni Pezzoli[6] | Francesca Persichetti[4] | Piero Carninci[2,7] | Stefano Gustincich[1,5]

[1]Area of Neuroscience, SISSA, Trieste, Italy

[2]Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Japan

[3]Department of Medical Sciences, Neurology Unit, University of Trieste, Trieste, Italy

[4]Department of Health Sciences, Università del Piemonte Orientale and IRCAD, Novara, Italy

[5]Central RNA Laboratory, Istituto Italiano di Tecnologia, Genova, Italy

[6]Parkinson Institute, ASST G. Pini-CTO, ex ICP, Milan, Italy

[7]Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Japan

Correspondence
Piero Carninci, Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan.
Email: carninci@riken.jp

Stefano Gustincich, Central RNA Laboratory, Istituto Italiano di Tecnologia (IIT), via Melen 83, 16152 Genova, Italy.
Email: stefano.gustincich@iit.it

Funding information
Italian Ministry of Education, University and Research, Grant/Award Number: FIRB grant prot.RBAP11FRE9; IIT-Istituto Italiano di Tecnologia; Fondazione Telethon, Grant/Award Number: GGP10224; SISSA-International School for Advanced Studies; The Ministry of Education, Culture, Sports, Science and Technology, Japan

Abstract

Minisatellites, also called variable number of tandem repeats (VNTRs), are a class of repetitive elements that may affect gene expression at multiple levels and have been correlated to disease. Their identification and role as expression quantitative trait loci (eQTL) have been limited by their absence in comparative genomic hybridization and single nucleotide polymorphisms arrays. By taking advantage of cap analysis of gene expression (CAGE), we describe a new example of a minisatellite hosting a transcription start site (TSS) which expression is dependent on the repeat number. It is located in the third intron of the gene nitrogen permease regulator like protein 3 (*NPRL3*). NPRL3 is a component of the GAP activity toward rags 1 protein complex that inhibits mammalian target of rapamycin complex 1 (mTORC1) activity and it is found mutated in familial focal cortical dysplasia and familial focal epilepsy. CAGE tags represent an alternative TSS identifying *TAGNPRL3* messenger RNAs (mRNAs). *TAGNPRL3* is expressed in red blood cells both at mRNA and protein levels, it interacts with its protein partner NPRL2 and its overexpression inhibits cell proliferation. This study provides an example of a minisatellite that is both a TSS and an eQTL as well as identifies a new VNTR that may modify mTORC1 activity.

Maria Bertuzzi, Dave Tang, and Raffaella Calligaris are co-first authors.

KEYWORDS
blood transcriptomics, minisatellite, *NPRL3*, VNTR

1 | INTRODUCTION

The mammalian genome encodes the instructions to specify development from the zygote to the generation of full sets of organs necessary to become an adult, interact with the environment and reproduce (Abugessaisa et al., 2017; FANTOM Consortium and the RIKEN PMI and CLST DGT et al., 2014). The fine regulation of gene expression is extremely complex and requires several players for its correct tuning in time and space. Slight alterations could have dramatic cascade effects leading to loss of homeostasis and disease (Chen et al., 2008; Cookson, Liang, Abecasis, Moffatt, & Lathrop, 2009; Emilsson et al., 2008).

In these conditions, specific sets of transcription factors are induced or repressed. These factors provide proximal and distal regulatory inputs that are integrated at transcription start sites (TSSs) to control the expression of target genes. Most genes have more than one TSS, and the regulatory inputs that determine TSS choice and activity are diverse and complex (Abugessaisa et al., 2017; FANTOM Consortium and the RIKEN PMI and CLST DGT et al., 2014). A comprehensive overview of TSSs and promoters' usage across the human body has been recently assembled by taking advantage of cap analysis of gene expression (CAGE; Abugessaisa et al., 2017; FANTOM Consortium and the RIKEN PMI and CLST DGT et al., 2014), a technology based on the generation of short sequence tags from the 5′ end of full-length complementary DNAs (cDNAs) followed by high-throughput sequencing. When mapped to a reference genome, CAGE tags survey TSSs activity of specific promoters and measure expression levels on a massive scale (Carninci et al., 2006; Suzuki et al., 2009).

Genome-wide association studies have identified hundreds of genetic variants that affects human phenotypes or susceptibility to diseases. In this context, a long-standing challenge is the unraveling of the chain of molecular events linking genetic variation to gene expression. To this purpose, a large effort has been devoted to identify expression quantitative trait loci (eQTL) correlating genetic polymorphisms to expression profiles (Jansen & Nap, 2001). Although eQTL identification has helped to unveil the molecular basis of some complex disorders, genomic analyses have been mainly focused on single nucleotide polymorphisms (SNPs). More recently, by taking advantage of the 1000 Genome Project, gene expression differences have been correlated to selected structural variants (SVs; 1000 Genomes Project Consortium et al., 2015). A substantial portion of SVs involves repetitive elements (REs), sequences that occur multiple times in the genome. These present a large variety of DNA elements of diverse structure and origin. REs can be grouped into two classes: tandem repeats (TRs) are small nucleotide stretches repeated in a head-to-tail orientation, while transposable elements are DNA sequences with the ability to move from one place of the genome to another. They represent almost 50% of the human genome (Babatz & Burns, 2013).

Minisatellites, also called variable number of tandem repeats (VNTRs), are a class of RE that have a unit length of 10–60 bp with a conserved core sequence of 10–15 bp, spanning 0.1–15 kb. VNTRs are frequently found in functional genic regions (such as coding sequence, promoter, untranslated region [UTR]) and may affect gene transcription, messenger RNA (mRNA) splicing, posttranscriptional modification or translational efficiency (Brookes, 2013). In selected cases, a causative link between their copy number and gene expression has been proved (Manca et al., 2018; van Dyck et al., 2005; Warburton, Breen, Bubb, & Quinn, 2016). Importantly, VNTRs have been correlated to human diseases (Brummett et al., 2007; Faraone, Doyle, Mick, & Biederman, 2001; Kennedy, Weed, Forget, & Morrow, 1994; Ogilvie et al., 1996).

Unfortunately, our understanding of their role as eQTLs has been limited by their absence in comparative genomic hybridization (CGH) and SNPs arrays.

The nitrogen permease regulator-like 3 (*NPRL3*) gene is adjacent to the α globin cluster on chromosome 16p13.3 and gives rise to five *NPRL3* mRNAs isoforms differing in their 5′-UTR regions including the translational starting codon (Vyas, Vickers, Picketts, & Higgs, 1995). *NPRL3* expression is ubiquitous and high in human and mouse erythroid cells (Kowalczyk et al., 2012a; Lower et al., 2009). NPRL3 protein dimerizes with its homologous NPRL2 and together with DEPDC5 they form GAP activity toward rags 1 (GATOR1), a protein complex responsible of mammalian target of rapamycin complex 1 (mTORC1) inhibition. Recently, *NPRL3* gene mutations have been found in patients with familial focal cortical dysplasia (FCD) and familial focal epilepsy, leading to pathological mTOR activation (Sim et al., 2015).

Our laboratories have been interested in the etiology and molecular diagnosis of Parkinson's disease (PD), a slowly progressive degenerative disorder of the central nervous system. No pharmacological treatment is currently available to slow or arrest the neurodegenerative process. Furthermore, accurate early diagnosis suffers from the lack of reliable biomarkers. In the quest of a gene expression signature for PD, we have previously shown that gene expression profiling of peripheral blood samples discriminates PD patients from healthy controls identifying differentially expressed genes by the use of the Affymetrix platform (Calligaris et al., 2015).

In a follow-up study, we have taken advantage of CAGE applied to tiny amounts of RNAs (Grison et al., 2014; Pascarella et al., 2014; Plessy et al., 2010, 2012), to describe TSS usage in the peripheral blood of 20 drug-naïve de novo PD patients and 20 healthy age and sex matched controls (HC), a subset of those previously analysed (Calligaris et al., 2015).

While a complete analysis of differential gene expression is not the topic of this study and it will be described elsewhere, here we report the identification of a new alternative TSS for *NPRL3* (*TAGNPRL3*) that lies within a 29 nt long minisatellite. High expression is associated to the presence of 13-repeats allele while the 16-repeats allele is the most common. Although neither expression or allelic frequency is related to PD, *TAGNPRL3* from

16/13 genotypes is highly expressed in red blood cells (RBCs). It encodes for a protein lacking the N-terminal part but able to bind NPRL2 and inhibit cell proliferation as its full-length isoform.

This study provides a novel example of a TSS located in a minisatellite, which usage depends on the number of TRs increasing the repertory of sequence architectures for *cis*-acting eQTLs.

2 | MATERIAL AND METHODS

2.1 | Blood collection and RNA purification

The study was approved by the local institutional Ethical Committee at the Movement Disorders Center of the Neurologic Clinic of Trieste (Italy). Study participants gave written informed consent. We enrolled 20 patients with a first clinical diagnosis of PD, according to the UK Parkinson's Disease Society Brain Bank criteria. Twenty healthy age- and ethnicity-matched control subjects (HC) traveling with the patients were also included in the study.

Blood was collected from study subjects after a fasting period. Samples were harvested directly and sequentially into eight PAXgene Blood RNA tubes (PreAnalytiX, Hombrechtikon, CH) via a 21G butterfly needle and then frozen and kept at −80°C. Total RNA was purified using PAXgene™ Blood RNA Kit (PreAnalytiX GmbH, Qiagen, Hilden, Germany) and DNase I treatment was performed by "on-column" treatment as recommended by manufacturer's instructions plus a second treatment subsequent to elution. RNA was then purified using RNase column (Qiagen, Hilden, Germany) and quantified by Nanodrop ND-100 Spectrophotometer (NanoDrop Technologies, Wilmington, DE). RNA integrity was determined with 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and exclusively samples with RNA integrity number ≥ 8 were included in the subsequent investigations.

2.2 | Cap analysis of gene expression

The CAGE libraries were prepared and analyzed as previously described (Hasegawa, Daub, Carninci, Hayashizaki, & Lassmann, 2014; Takahashi, Kato, Murata, & Carninci, 2012). First, cDNA was reverse-transcribed by the reverse transcriptase using a random primer including an Eco P15I sequence. First-strand cDNA reaction was then cleaned up and Cap-trapping was performed. Chemical oxidation with $NaIO_4$ was used to open the RNA diol groups. The derived oxidized dialdehyde was incubated with a long arm biotin hydrazide, resulting in biotinylation of the cap-site and the 3′ ends of RNAs. Before capturing the biotinylated cap, samples were treated with RNase I to cleave single stranded RNA (ssRNA) regions followed by a short incubation at 65°C. The cDNAs including the biotinylated cap site were finally captured with streptavidin-coated magnetic beads.

By the single-strand linker ligation method, a primer sequence for second strand cDNA synthesis was added. The second strand cDNA synthesis was then initiated by a biotin modified primer and carried out with a thermostable DNA polymerase. cDNA was then treated with Antarctic phosphatase and cleaved by Eco P15I, a Type III restriction enzyme that cleaves 27 nt downstream of the enzyme recognition site. After cleavage, the 3′ linker, ending with two NN protruding nucleotides, was ligated at the Eco P15I cleavage site providing a priming site for the subsequent polymerase chain reaction (PCR) amplification carried out with Phusion High-Fidelity DNA Polymerase (Thermo Scientific).

2.3 | Rapid amplification of cDNA ends (RACE)

RACE technique allows the amplification of full-length 5′ and 3′ ends of cDNA starting from a known partial sequence obtained, for example, from library screening such as CAGE.

RACE was performed using the GeneRacer cDNA Amplification Kit (Invitrogen).

One microgram of total RNA has been retrotranscribed into cDNA modified with the GeneRacer 5′ and 3′ adaptors following the manufacturer's protocol, using poly-T primer in the 20 μl reverse transcription reaction.

All RACE reactions were performed by nested PCR. For the first amplification, 1 μl cDNA obtained from the retrotranscription reaction was used as a template together with GeneRacer™ 5′ Primer or the GeneRacer™ 3′ Primer, when performing 5′ or 3′ RACE reactions, respectively, and gene-specific primers: REV_first_5RACE (CCCAGAGCATGCTGTAGCAGTGTT) for 5′RACE and FW_3RACE (GAGTGTGTGATCCTGTTTCTCAGCGTG) for 3′RACE.

PCR was performed with Platinum Taq High Fidelity (Invitrogen) under the following conditions: 94°C for 2 min, followed by five cycles at 94°C for 30 s, 72°C for 1 min; five cycles at 94°C for 30 s, 70°C for 1 min, 25 cycles at 94°C for 30 s, 68°C for 30 s, 68°C for 1 min, and a final extension at 68°C for 10 min.

One microliter of this PCR product was used as template for the nested PCR together with GeneRacer™ 5′ Nested Primer or the GeneRacer™ 3′ Nested Primer, when performing 5′ or 3′ reactions, respectively, and a second nested gene-specific primer: REV_nested_5RACE (ATAACATCTGAAAACCAGACAAGAAACA) for 5′-RACE and FW_3RACE (the same of the first reaction) for 3′-RACE.

Nested PCR was performed under the following conditions: 94°C for 2 min, followed by 40 cycles at 94°C for 30 s, 65°C for 30 s, 68°C for 2 min, and a final extension at 68°C for 10 min. PCR products were resolved on 1% agarose gel. The amplicons were cloned in pGEM-T Easy vector (Promega) and sequenced at the Eurofins MWG Operon Inc. facility.

2.4 | Identification of minisatellite polymorphism

The validation set of PD and HC samples was obtained from the PD DNA Bank at the Centro Studi Parkinson (Milan, Italy) and additional HC from the Associazione Donatori di Sangue (Trieste, Italy).

The independent set of PD and HC samples was obtained from "Clinica Neurologica di Trieste" and "Telethon Biobank" for PD, and from "Associazione Donatori Sangue Trieste" for HC.

Genomic DNA has been extracted from blood samples taking advantage of QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's protocol.

PCRs were performed in 96-well plates, on a total of 20 ng/5 μl of human genomic DNA per well, by using ExTaq TaKaRa on a total of 50 μl reaction volume, as specified by the manufacturer. Primers are VNTR_FW (GCAGAAGTGCCACCATTAAGCA), VNTR_REV (AAACCCCATGGTAAGCGTTGA) and they were used at a final concentration of 200 nM. The two primers were designed to map externally to the repeated region to obtain the amplification of the entire minisatellite. The cycling conditions were: 94°C for 3 min, followed by 35 cycles at 94°C for 30 s, 62.4°C for 30 s, 72°C for 1 min, and a final extension at 72°C for 5 min. Reaction products were visualized on a 1.3% agarose gel.

2.5 | Reverse transcription PCR (RT-PCR) and real-time TaqMan PCR

To validate the full-length *TAGNPRL3* transcript, 1 μg of total RNA was reverse-transcribed in a final volume of 20 μl using Superscript III Reverse Transcriptase (Invitrogen), 25 ng random hexamers and 2.5 μM oligo(dT) 20 primers according to the manufacturer's recommendations.

PCR mix was prepared by adding 1 μl of the so prepared cDNA to 5 units of TaKaRa Ex Taq, 10× Ex Taq Buffer, dNTP Mixture (200 μM final concentration each) (TaKaRa), 200 nM of primer FW_3RACE (see Table 2), 200 nM of primer REV (CCTTTCCAAACCTGCGCACC), and water to a final volume of 50 μl.

The FW_3RACE is the same forward primer used to amplify the full-length transcript in 3′RACE assay, while the REV primer was designed in the 3′-UTR region of *TAGNPRL3*. The primers used to amplify *NPRL3* isoform 1 are NPRL3_FW (CCCCACGGCGGG ATGCGGGA) and NPRL3_REV (TCGCCCGTGTTGCTGGCAGCGT).

The PCR protocol is the following: 95°C for 5 min, followed by 40 cycles at 95°C for 15 s, 60°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 5 min. PCR products were resolved on 1% agarose gel.

Real-time PCR was performed in the presence of 1 μl of cDNA template, TaqMan gene expression master mix, commercially available TaqMan gene expression assay for *PGK1* (Hs99999906_m1), 250 nM specific primer mix, 150 nM specific TaqMan MGB probe (6-carboxyfluorescein dye-labelled) and run on an iCycler IQ (Bio-Rad) in a 20 μl reaction volume according to the manufacturer's instructions. Each sample has three technical replicates. *PGK1* has been used as a reference gene for this study because it was among the most reliable reference genes for peripheral blood gene expression analyses (Calligaris et al., 2015).

NPRL3_FW and NPRL3_REV primers have been used together with the following specific probe on Nprl3 isoform 1: AGCCAGGA GCACCCGGCGTCC.

The specific primers and probes designed on *TAGNPRL3* are TAGNPRL3_FW (GTGATCCTGTTTCTTGTCTGGT), TAGNPRL3_REV (CAGAGCATGCTGTAGCAGTGT), and the probe GTTATTCTGGC AACAATTTTGGCA.

Thermal cycler conditions were as follows: 95°C for 10 min, followed by 40 cycles of amplification at 95°C for 15 s and 60°C for 1 min.

Standard curves of cDNA ranging from 120 to 0.2 ng were used to verify that the 50 ng dilution tested was within the linear range of reaction. Primer efficiency and multiplexing efficacy was verified by linear regression to the standard curve with a slope near −3.32 representing acceptable amplification efficiency.

The amplified products were separated on a 2% agarose gel. Results were normalized to PGK1 and the initial amount of the template of each sample was determined as relative expression versus a pool of healthy control samples used as calibrator. The relative expression of each sample was calculated by the formula $2^{-\Delta\Delta Ct}$ (User Bulletin 2 of the ABI Prism 7700 Sequence Detection System).

2.6 | RNA and protein preparation from blood fractions

Blood was collected into ethylenediaminetetraacetic acid (EDTA) tubes and the plasma, the RBCs and the peripheral blood mononuclear cells (PBMCs) have been separated starting from the whole blood samples by using Ficoll Histopaque gradient following the manufacturer's protocol (Sigma-Aldrich).

To extract total RNA from the blood fractions each sample was added with a proper amount of TRIzol reagent (Invitrogen) followed by manufacturer's protocol.

A fraction of the total RNA sample was treated with DNase I (Ambion) at 37°C for 1 hr, and the sample was then purified on RNAeasy Mini Kit columns (Qiagen). The final quality of RNA sample was tested on the Agilent 2100 bioanalyzer using the Eukaryote Total RNA Nano assay.

The PBMC protein lysate has been prepared adding a proper amount of 2× sodium dodecyl sulfate (SDS) sample buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM β-mercaptoethanol). The protein lysates from whole blood, plasma, and RBCs have been prepared according to Lin et al. (2012) ("Whole blood sample preparation" chapter) followed by an albumin depletion using trichloroacetic acid/acetone (Chen et al., 2005).

The obtained precipitates were diluted in a proper amount of 2× SDS sample buffer.

2.7 | Expression plasmids

NPRL3: pcDNA3 vector containing the full-length DNA sequence of human *NPRL3* isoform 1 and pcDNA3 vector containing coding DNA sequence of human *NPRL3* isoform 2 were kindly provided by Licio

Collavin's laboratory (Consorzio Interuniversitario per le Biotecnologie, Trieste).

TAGNPRL3: the full-length DNA sequence of *TAGNPRL3* was amplified from whole blood cDNA modified with the GeneRacer 5′ and 3′ adaptors via PCR with the GeneRacer™ 5′ Primer (see Table 2) and a gene specific reverse primer: TTTTGCTTGGCCTGG CTTTTATCTTGAA.

The pGEM T-easy-*TAGNPRL3* plasmid was generated by subcloning the cDNA sequence into the *Eco*RI cloning site of the pGEM T-easy cloning vector from Promega for sequencing. This cassette was finally cloned into the pcDNA3.1 (−) expression vector using the same restriction enzymes.

*NPRL2*_Myc: pcDNA3.1-vector containing the coding region of human *NPRL2* amplified from whole blood cDNA. The Myc tag was fused to the C-terminus of the protein.

2.8 | Cell culture, transfections, and immunoblotting

HEK 293T (human embryonic kidney) cells were grown in DMEM (Gibco) supplemented with 10% fetal bovine serum (Sigma-Aldrich), 100 U/ml penicillin and 100 μm/ml streptomycin (Sigma-Aldrich) at 37°C in a humidified CO_2 incubator. HEK cells were transfected with FuGENE HD Transfection Reagent (Promega) following manufacturer instructions. After 48 hr from transfection, cells were either collected with 2× SDS sample buffer or analysed via immunofluorescence.

All the protein lysates have been boiled 5 min at 95°C and analysed by western blot.

For western blot, samples were resolved on SDS/polyacrylamide gel electrophoresis, and proteins were transferred to nitrocellulose membrane (Amersham, GE Healthcare). Membrane was blocked with 5% nonfat milk in Tris buffer saline solution (TBST), then incubated overnight at 4°C with the following primary antibodies: anti-NPRL3 (1:500; Cat #ab121346; RRID:AB_11129281; Abcam), anti-β Actin (1:5,000; Cat #A5441; RRID:AB_476744; Sigma-Aldrich).

Proteins were detected by horseradish peroxidase-conjugated secondary antibodies (DakoCytomation, Glostrup, Denmark) and ECL Western Blotting Detection Reagents (GE Healthcare).

2.9 | Small interfering RNA (siRNA) experiment

By taking advantage of siTran 2.0 siRNA transfection reagent (Cat. no.: TT320001; Origene), SH-SY5Y cells were transfected with 10 nM siRNA for human NPRL3 (Human) (Cat. no.: SR305374; Origene) and with a Universal scrambled negative control siRNA. Fluorescent labeled siRNA has been used to optimize transfection efficiency (Trilencer-27 Fluorescent-labeled transfection control siRNA duplex; 1 nmol; Cat. no.: SR30002; Origene). Forty-eight later samples were harvested for western blot and qRT-PCR analysis.

2.10 | Immunofluorescence

For immunocytochemistry experiments, cells were fixed in 4% paraformaldehyde (Sigma-Aldrich) for 10 min, then washed with phosphate-buffered saline solution (PBS) two times, treated with 0.1 M glycine for 4 min in PBS and permeabilized with 0.1% Triton X-100 in PBS for another 4 min. After washing with PBS and blocking with 0.2% bovine serum albumin (BSA), 1% normal goat serum, 0.1% Triton X-100 in PBS (blocking solution), cells were incubated with the indicated antibodies diluted in blocking solution for 90 min at room temperature. After washes in PBS, cells were incubated with labelled secondary antibodies for 60 min. For nuclear staining, cells were incubated with 1 μg/ml 4′,6-diamidino-2-phenylindole (DAPI) for 5 min. Cells were washed and mounted with antifade fluorescent mounting medium (Vectashield Hard Set, VectorLabs). All images were collected using a Leica DM6000 fluorescence microscope.

2.11 | Coimmunoprecipitation

HEK cells were transfected with FuGENE HD Transfection Reagent (Promega) following manufacturer instructions. After 24 hr from transfection, cells were harvested and lysed in the following buffer supplemented with complete EDTA-free Protease Inhibitor Tablets (Roche): 150 mM NaCl, 50 mM Tris-HCl pH 7.4, 0.5% NP-40, 1 mM EDTA.

After 30 min of lysis at 4°C samples were sonicated twice (7 μm amplitude, 10 s) and then centrifuged 30 min, 4°C at maximum speed. Supernatants were collected and a fraction was saved as input. Lysates were incubated with 1 μl of Myc-Tag (9B11) antibody from Cell Signaling Technology (Cat# 2233S, RRID:AB_10693328). After 3 hr of incubation, 40 μl of Protein G Sepharose beads (GE Healthcare) were added to each sample and incubated 1 hr. After washing, beads were dried and eluted in 2× SDS sample buffer, boiled 5 min at 95°C and analysed by western blot.

2.12 | Bromodeoxyuridine (BrdU) assay

HEK cells were transfected with FuGENE HD Transfection Reagent (Promega) following manufacturer instructions. As a control, a pcDNA3.1-vector expressing green fluorescent protein (GFP) was used. After 48 hr from transfection, BrdU was added to each plate at a final concentration of 15 μg/ml for 1 hr.

Cells were then fixed in 4% paraformaldehyde (Sigma-Aldrich) for 10 min, then washed with PBS two times, treated with 0.1 M glycine for 4 min in PBS and permeabilized with cold acetone for 30 s. After washing with PBS cells are treated with HCl 3 N for 15 min to separate DNA into single strands so the primary antibody can access the incorporated BrdU. After several washes with PBS cells were incubated at 37°C with a blocking solution of 1% BSA for 30 min, followed by 37°C incubation with an anti-BrdU antibody (1:50; Cat #B2531; RRID:AB_476793; Sigma-Aldrich) and an anti-NPRL3

antibody (1:500; Cat #ab121346; RRID:AB_11129281; Abcam) in blocking solution for 90 min. After washes in PBS, cells were incubated with labelled secondary antibodies for 60 min. For nuclear staining, cells were incubated with 1 μg/ml DAPI for 5 min. Cells were washed and mounted with antifade fluorescent mounting medium (Vectashield Hard Set, VectorLabs). All images were collected using a Leica DM6000 fluorescence microscope.

The significance of differences between the means in experimental groups and conditions was analyzed using one-way ANOVA followed by *post hoc* Tukey test or Dunnet's multiple comparison test, or nonparametric tests (for small data samples) Wilcoxon test, using Prism (GraphPad Software Inc.). Significance levels indicated in all figures are as follows: $^{**}p < .01$, $^{***}p < .001$.

3 | RESULTS

3.1 | CAGE analysis of peripheral blood identifies a highly expressed tag in the intronic region of the *NPRL3* gene

In a follow up of Calligaris et al. (2015), we have taken advantage of CAGE to map TSSs in the peripheral blood of 20 drug-naïve de novo PD patients and 20 age and sex matched healthy controls (HC). On average for each CAGE library, 7,158,414 tags were sequenced and 1,799,298 were uniquely mapped to the reference genome (Table S1). A complete analysis of differential gene expression between PD and HC will be described elsewhere and it is not the topic of this study.

In this study, we focus our attention on a CAGE tag (27 nt) that was found upregulated by more than 50-fold in six out of 20 (30%) PD individuals respect to the remaining patients and HC. This highly expressed tag maps to the third intron of the *NPRL3* gene, on the telomeric region of chromosome 16, with coordinates 173660-174239 (genomic assembly hg19). The multimapping nature of this tag is appreciable on Figure 1, suggesting it represents a repetitive region. By Genome Browser sequence analysis, this tag maps on a minisatellite sequence of 29 nucleotides repeated 16 times. The mapping is on the (−) strand of genomic DNA, in the same direction of *NPRL3* transcription. This minisatellite is present only in this genomic locus and it is human-specific. The most common tag sequence is GAGTGTGTGATCCTGTTTCTCAGCGTG. Table 1 reports the tag per million (tpm) values of the 40 CAGE samples. Samples presenting the highest tag levels are PD2 (30.08), PD4 (23.12), PD7 (19.42), PD11 (46.05), PD13 (50.88), PD18 (45.57), while the average tpm for the remaining PD are 1.89 and for HC are 1.43.

3.2 | Identification of tag-containing transcripts

To identify tag-containing transcripts we took advantage of 3′RACE applying this technique to the PD sample that presented the highest amount of tags on the CAGE assay (PD13; Table 1). PD13 RNA was retrotranscribed into cDNA modified with the GeneRacer 3′ adaptor

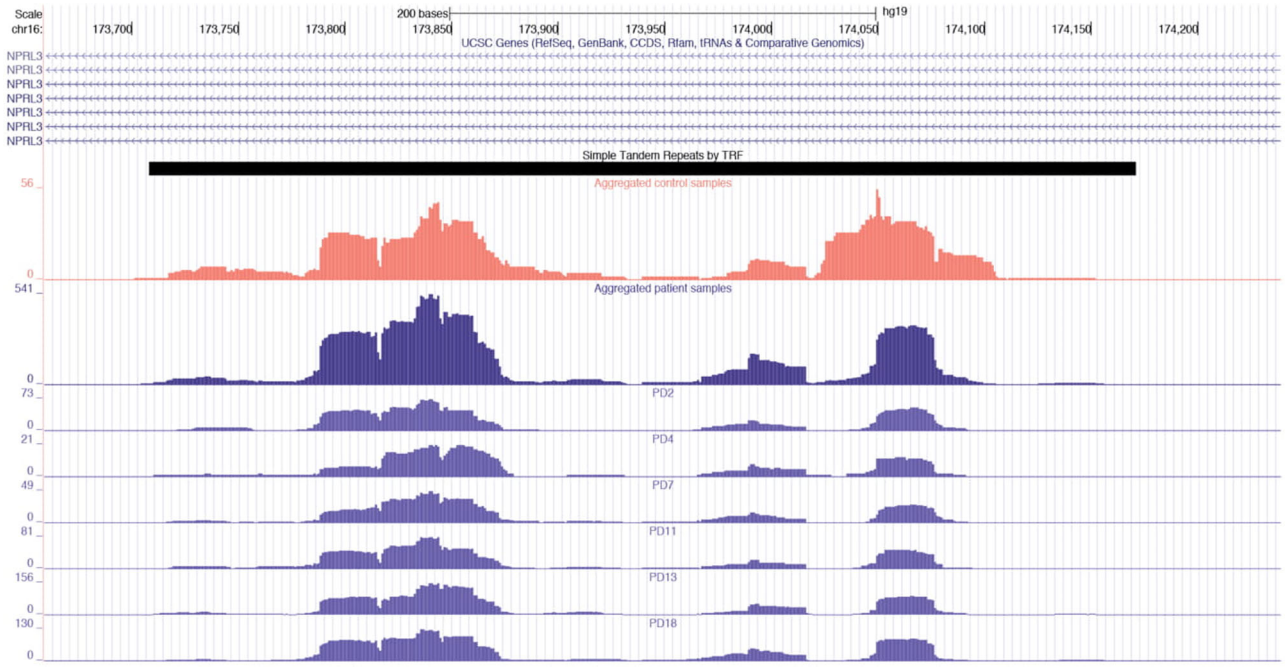

FIGURE 1 CAGE tags upregulated in PD samples in the intronic region of *NPRL3* (chr16:173660-174239 on hg19). The UCSC Genome Browser screenshot shows the intronic region of *NPRL3*, a minisatellite sequence detected by the Tandem Repeats Finder, and several raw coverage plots of CAGE tags. The first two coverage plots are an aggregation of all control (red) and patient (blue) samples. The following six coverage plots shows the patient samples with the highest number of CAGE tags: PD2, PD4, PD7, PD11, PD13, and PD18. CAGE, cap analysis of gene expression; NPRL3, nitrogen permease regulator like protein 3; PD, Parkinson's disease

TABLE 1 Cap analysis of gene expression upregulated tag in PD: tpm values for each HC and PD sample, mapping on chr16:173660-174239 region (hg19)

Sample	tpm
HC1	1.48
HC2	1.19
HC3	2.15
HC4	0.27
HC5	1.84
HC6	1.18
HC7	1.76
HC8	0.90
HC9	0.19
HC10	0.69
HC11	2.94
HC12	1.43
HC13	0.92
HC14	1.84
HC15	3.92
HC16	0.68
HC17	1.30
HC18	0.88
HC19	2.42
HC20	0.62
PD1	0.62
PD2	30.08
PD3	4.16
PD4	23.12
PD5	0.55
PD6	0.15
PD7	19.42
PD8	0.81
PD9	5.21
PD10	2.67
PD11	46.05
PD12	1.78
PD13	50.88
PD14	1.61
PD15	1.89
PD16	0.65
PD17	1.06
PD18	45.57
PD19	3.19
PD20	2.14

Abbreviations: HC, healthy control; PD, Parkinson's disease; tpm, tag per million.

using poly-T primer and the first-step RT-PCR was carried out using the GeneRacer 3′ Primer and a forward primer corresponding to the tag sequence (Figure 2a). Nested RT-PCR, performed using the GeneRacer 3′ Nested Primer and the same forward primer, allowed amplification of a major product (about 2,000 bp; Figure 2e). Subcloning and sequencing of this fragment revealed a new *NPRL3* isoform with the TSS in the third intron of the *NPRL3* gene (Figure 2b).

Together with this major product of amplification, we identified four *NPRL3*-hemoglobin A1 chimeric transcripts that we called transalpha1–4. The breakpoints of these transcripts involve the intronic minisatellite sequence (2–4 repeats are transcribed) and the third exon of hemoglobin 1A (HBA1). Only in one case (transalpha4) we could observe a more complex transplicing RNA composed of minisatellite intronic sequences spliced to the subsequent two exons of *NPRL3* interrupted by part of the second and all the third exons of HBA1 (Figure 3). Unfortunately, we were not able to validate these chimeric transcripts due to the high abundance of HBA1 mRNA in blood samples.

We then completed our study of the tag-containing transcript with a 5′RACE experiment. To this purpose, we took advantage of another high tag-expressing PD blood sample (PD11; Table 1). After confirming that this PD patient expressed the same tag-containing transcript of PD13 with a 3′RACE experiment (data not shown), the first-step RT-PCR for the 5′RACE was carried out using the GeneRacer 5′ Primer and a reverse primer located to the fourth exon of the *NPRL3* gene. 5′RACE nested RT-PCR was performed using the GeneRacer 5′ Nested Primer and a nested reverse primer spanning the fourth *NPRL3* exon and the transcribed intronic sequence found with 3′RACE (Figure 2c). The 5′ RACE gel run revealed the amplification of a major product (about 200 bp) (Figure 2f) that mapped the TSS of this new *NPRL3* isoform (*TAGNPRL3*) to the minisatellite sequence of *NPRL3* third intron (Figure 2d). Interestingly, *TAGNPRL3* TSS sequence is TC+1TTTCT which strongly resemble a TCT motif (YC+1TYTYY), a key component of an RNA polymerase II system that is directed toward the expression of ribosomal protein genes as well as other genes encoding factors involved in protein synthesis (Parry et al., 2010). Despite the high sequence homology between the different repeats, the consensus TC+1TTTCT is present only in the eleventh repeat providing TSS specificity.

To validate the *TAGNPRL3* transcript identified with RACE, we carried out an RT-PCR experiment with a reverse primer in the 3′-UTR region and the same forward primer of the 3′RACE assay mapping to the CAGE tag sequence (Figure 4a). We thus confirmed that *TAGNPRL3* is the RNA isoform detected as differentially expressed in the CAGE experiment since PD13 and PD2, high-expressing tag samples, showed a higher level of *TAGNPRL3* compared with the low-expressing tag sample HC14 (Figure 4b). The identity of the amplicon was confirmed by sequencing.

On the contrary, by performing an RT-PCR with primers specific for the *NPRL3* isoform 1, we showed that its expression is quite comparable among the three samples (Figure 4c).

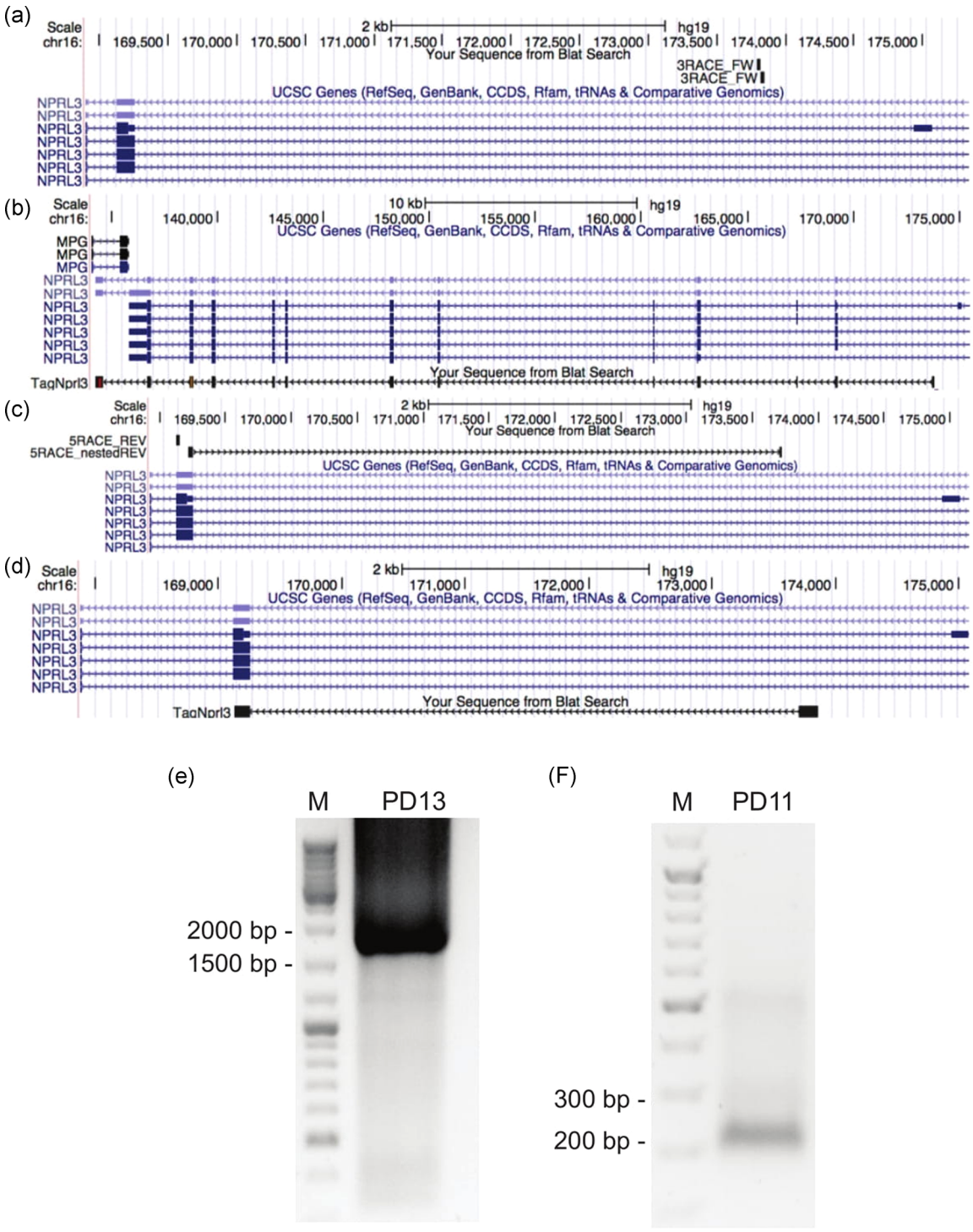

FIGURE 2 Identification of the tag containing transcripts by 3′ RACE-PCR analysis and validation by 5′ RACE-PCR analysis. (a) UCSC Genome Browser snapshot showing the position of 3′RACE primer. The PCR product run on a 1% agarose gel is shown in (e)). (b) 3′RACE product sequence blat on UCSC Genome Browser. (c) UCSC Genome Browser snapshot showing the position of 5′RACE primers. The PCR product run on a 1% agarose gel is shown in (f)). (d) 5′RACE product sequence blat on UCSC Genome Browser. PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends

3.3 | *TAGNPRL3* expression levels are associated to the minisatellite genotype

Given the high variability in the number of repeats of minisatellite sequences, we investigated whether the minisatellite containing *TAGNPRL3* TSS is polymorphic. To this purpose, we designed two primers that map externally to the repeated region to obtain the amplification of the entire minisatellite sequence. PCR experiments on genomic DNA revealed the presence of two amplification products (Figure S1). By sequencing, we found that these PCR

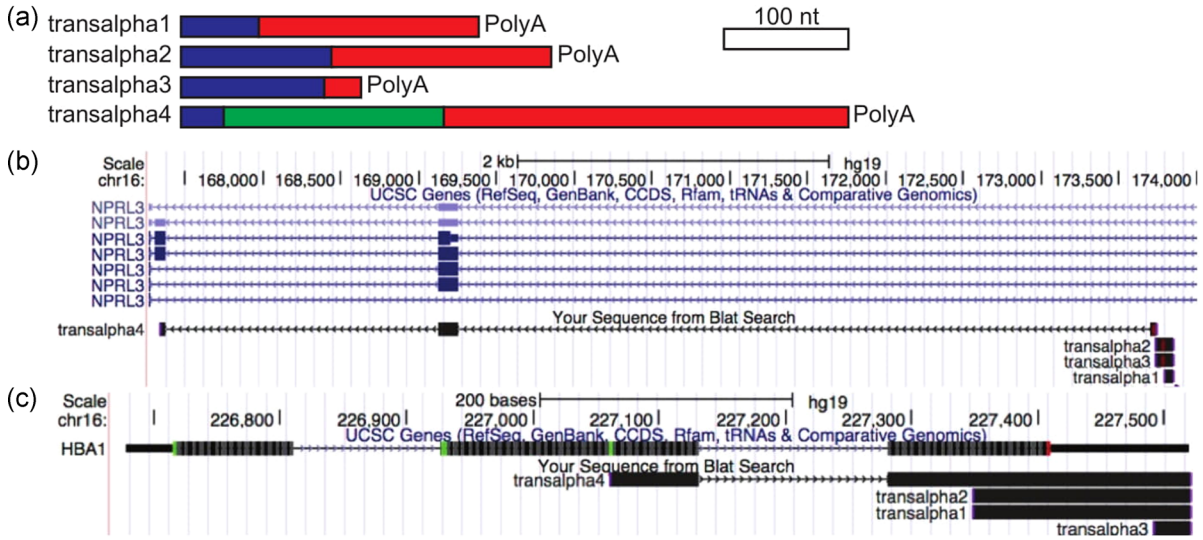

FIGURE 3 Identification of chimeric products by 3′ RACE-PCR analysis. (a) Schematic representation of chimeric products found with 3′RACE analysis. The white block indicates the scale. In blue, *NPRL3* intronic part; in green, *NPRL3* exonic part; in red, *HBA1* exonic part. (b) UCSC Genome Browser representation of 5′ ends of the chimeric transcripts on *NPRL3* gene. (c) UCSC Genome Browser representation of 3′ ends of the chimeric transcripts on *HBA1* gene. NPRL3, nitrogen permease regulator like protein 3; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends

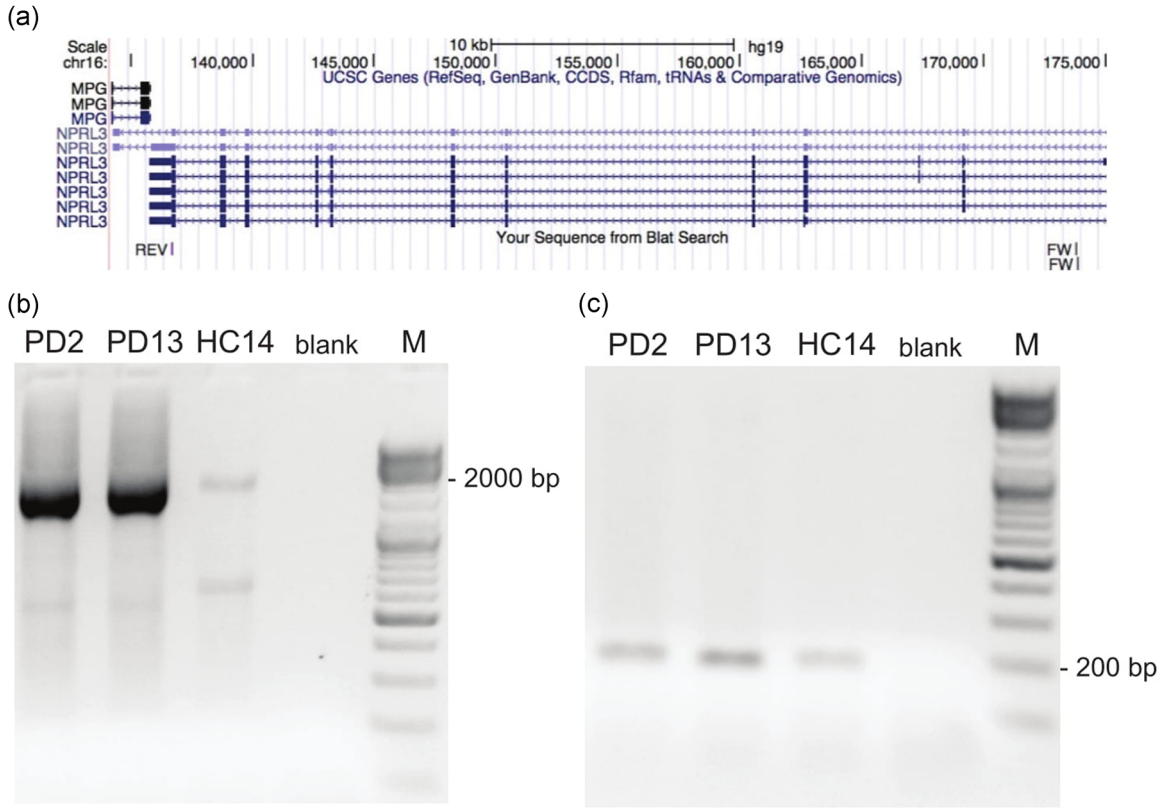

FIGURE 4 Validation of *TAGNPRL3* transcript. (a) UCSC Genome Browser representation of the FW primer (double mapping on the minisatellite sequence) and the REV primer used for *TAGNPRL3* validation. (b) *TAGNPRL3* RT-PCR performed on first-strand cDNA prepared from total RNA from PD2, PD13, and HC14 whole blood. PCR products were analysed on a 1% agarose gel with EtBr staining. The main 2,000 bp product corresponds to *TAGNPRL3* and its expression is barely detectable in HC14 individual (16/16, low-tag expressing) whereas is high in PD2 and PD13 (both 16/13, high-tag expressing). (c) *NPRL3* RT-PCR performed on the same samples (PD2, PD13, and HC14). PCR products were analysed on a 1% agarose gel with EtBr staining. The 200 bp product corresponds to *NPRL3* and its expression is comparable among the three individuals. cDNA, complementary DNA; NPRL3, nitrogen permease regulator like protein 3

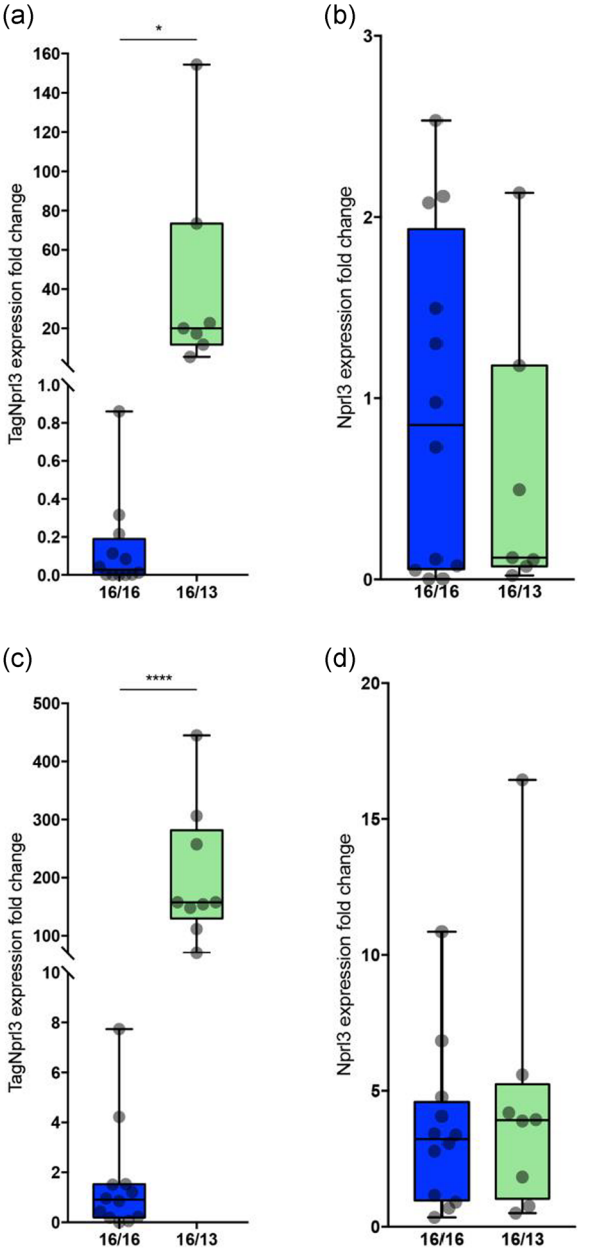

FIGURE 5 *TAGNPRL3* upregulation in blood is associated to 16/13 minisatellite genotype. Real-Time TaqMan PCR normalized with *PGK1* assay. Fold change is indicated on the *y* axis. 16/16 are indicated in blue/dark and 16/13 in green/light. (a) *TAGNPRL3* assay on selected CAGE samples reveals that its high expression is linked to 16/13 minisatellite genotype. (b) *NPRL3* assay on selected CAGE samples. Canonical *NPRL3* isoform expression is variable among individuals but it is independent from minisatellite genotype. (c) *TAGNPRL3* assay on an independent sample set (PD and HC, 16/13 and 16/16, see Table 2) confirmed that all 16/13 individuals (both PD and HC) express high levels of *TAGNPRL3*. (d) *NPRL3* assay on an independent sample set (PD and HC, 16/13 and 16/16, see Table 2). *NPRL3* expression is independent from minisatellite genotype. Data presented as min to max with median. $^{*}p < .05$; $^{****}p < .0001$. CAGE, cap analysis of gene expression; HC, healthy control; NPRL3, nitrogen permease regulator like protein 3; PCR, polymerase chain reaction; PD, Parkinson's disease

fragments contained two repeats variants of the same 29 mer unit: the larger one consists of 16 repeats while the shorter amplicon consists of 13 repeats. In details, six PD patients (30%) and one HC individual (5%) (sample HC15) were 16/13 heterozygous. Interestingly, we noted that the individuals with high-tag counts, and therefore highly expressing *TAGNPRL3* mRNA, were 16/13 heterozygous.

On the basis of their genotypes we selected 18 RNA samples among those used for the CAGE analysis: six 16/13 (PD samples) and 12 16/16 (six PD and six HC). We then carried out Real-Time TaqMan RT-PCR with the assays for *TAGNPRL3* and *NPRL3* mRNAs. We took advantage of *PGK1* as invariant gene for normalization. *TAGNPRL3* high expression was strongly correlated with the 16/13 genotype while it was low in homozygous 16/16 individuals (Figure 5a). On the other hand, *NPRL3* expression levels were quite variable among different samples but showed no correlation with genotype (Figure 5b). The expression of both transcripts was not correlated with PD.

We then assembled an independent set of samples (see Section 2 for details) composed of eight 16/13 and 12 16/16 subjects (details in Table 2) and performed a Real-Time TaqMan RT-PCR in the same conditions as above (*TAGNPRL3*, *NPRL3*, and *PGK1* for normalization). While high *TAGNPRL3* expression was linked to the 16/13 minisatellite genotype (Figure 5c) in a statistically significant manner, *NPRL3* expression did not correlate with the genotype (Figure 5d). Again, no correlation was found between mRNA levels and PD.

In summary, *TAGNPRL3* is a new *NPRL3* isoform with a TSS within a minisatellite region in its third intron. This minisatellite presents two variants with 16- and 13-repeats. The 13-repeats allele is associated to high expression of *TAGNPRL3* mRNA.

3.4 | Genomic analysis of VNTR locus

We then carried out an analysis of the VNTR locus for genomic features and expression data interrogating large consortium datasets. The VNTR overlaps two eQTLs identified in the version 6 of the GTEx (GTEx Consortium, 2013) data analysis. These are associated to the rs17146023 SNP and result to impact the expression of the *NPRL3* gene in the lung and of RNA polymerase III subunit K (*POLR3K*) in the cerebellum. The VNTR also overlaps a DNAse I hypersensitivity cluster identified within the ENCODE (ENCODE Project Consortium, 2012) version 3 analysis. Out of 125 cell types, the DNAse I hypersensitivity site from this VNTR has been identified only in the T helper 1 (Th1) cell lineage.

To look for potential regulatory networks, we carried out a binding site analysis using Jaspar (Khan et al., 2018) and identified myeloid zinc finger 1 (MZF1; Morris, Hromas, & Rauscher, 1994) as the transcription factor with the highest probability of binding the VNTR (Figure 6). MZF1 is a zinc finger protein with an established role in hemopoiesis. In agreement with a potential overlap with an enhancer region, the VNTR also overlaps a peak of H3K27

TABLE 2 List of real-time TaqMan polymerase chain reaction blood samples and their minisatellite genotype

Experiment	Minisatellite genotype	Sample name
Validation	16/16	2C
		5C
		7C
		8C
		10C
		14C
	16/13	15C
	16/16	1PD
		6PD
		8PD
		9PD
		10PD
		19PD
	16/13	2PD
		4PD
		7PD
		11PD
		13PD
		18PD
Independent	16/16	PD_C2616
		PD_M2371
		PD_Z0411
		PD_S1613
		PD_T0770
		PD_C2766
	16/13	PD_025
		PD_022
		PD_Z0397
		PD_G1593
	16/16	C_49AB
		C_51PR
		C_52FC
		C_54MK
		C_47IC
		C_48MK
	16/13	**C_12CD**
		C_28CD
		C_43SB
		C_43CS

Note: The two samples in bold have been used for the fractionation experiment.

acetylation and falls few hundred nucleotides downstream of a peak of conservation in 100 vertebrate species.

Notably, a long noncoding RNA is transcribed in antisense to the *NPRL3* gene starting at about 1,500 nucleotides downstream the VNTR. It results expressed at very low levels (from 0.04 to 0.45 median transcript per million) in all the tissues sampled in the GTEx Project, with its higher expression in cervix and cerebellum.

3.5 | 13-repeat variant prevalence

We then analysed the prevalence of the 13-repeats variant. To this purpose we took advantage of a validation set of PD and HC samples (see Section 2 for details) and we carried out a PCR-based genotype assay. Among a total of 1,124 subjects (both PD patients and HC), 212 were 16/13 (18.86%); 11 subjects were 13/13 (0.98%), and 901 subjects were 16/16 (80.16%). The allelic frequency of the 13-repeats was 10.41%. This analysis showed that the 13-repeat variant is not linked to PD.

The VNTR sequence in the *NPLR3* gene has been included into the dbVAR, the NCBI's database of human genomic structural variation, as nstd179.

3.6 | *TAGNPRL3* mRNA is expressed in RBC in a 13-repeats-dependent manner

We then assessed the main source of *TAGNPRL3* expression in the blood. To this purpose we took advantage of two 16/16 HC (MB and RC, from our laboratory stock) and two 16/13 HC (from the Real-Time TaqMan PCR validation set; C_43_CS and C_12_CD, see Table 2) individuals. Blood was collected into EDTA tubes and PBMC; RBCs and plasma were separated. After RNA extraction, we performed a Real-Time TaqMan RT-PCR with the assays for *TAGNPRL3*, *NPRL3*, and *PGK1* for normalization.

NPRL3 and *TAGNPRL3* expression was quite exclusive of RBC fraction while *TAGNPRL3* mRNA was highly expressed in heterozygous individuals. PBMC expressed low levels of *NPRL3* while *TAGNPRL3* was undetectable except for individual C_43_CS where we cannot exclude a minor contamination due to the high level of *TAGNPRL3* mRNA expression (Figure 7). Both *NPRL3* and *TAGNPRL3* mRNAs were undetectable in plasma (data not shown).

3.7 | TAGNPRL3 protein is highly expressed in RBC of 16/13 individuals

NPRL3 gene has five annotated different alternative splicing transcripts. *NPRL3* isoform 1 (NM_001077350) encodes for a protein of 64 kDa. Isoforms 2–5 are predicted to translate a shorter NPRL3 protein, starting from an alternative AUG, but they were undetectable in several cell types (Lunardi et al., 2009). The coding sequence of TAGNPRL3 corresponds to the coding sequence of NPRL3 isoform 3 (NM_001243247) but with different UTRs. Respect to isoform 1, TAGNPRL3 and NPRL3 isoform 3 share a N-terminal truncated (Δ1–78 amino acids) NPRL3 protein.

After confirming with an siRNA experiment that the anti-NPRL3 antibody (1:500; Cat #ab121346; Abcam) specifically detects endogenous NPRL3 in SH-SY5Y cells (Figure S2A), we cloned the full-length cDNA sequence of *TAGNPRL3* into the pcDNA3.1-expression vector and transiently transfected it into HEK 293T cells. As shown in Figure S2B, TAGNPRL3 was detected as a 60 kDa migrating band while the NPRL3 isoform 1 gave rise to a 64 kDa protein. Both proteins migrated at an apparent molecular weight higher than their conceptual translation (55 kDa for TAGNPRL3 and 60 kDa for NPRL3 isoform 1).

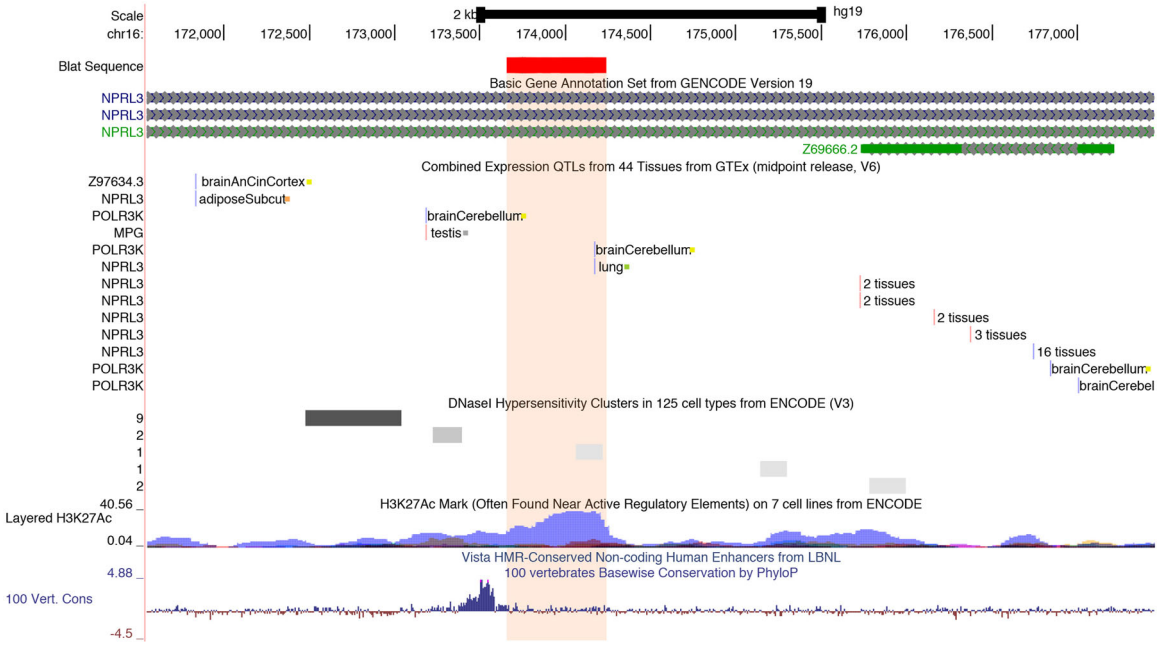

FIGURE 6 Analysis of genomic features of VNTR locus. Image modified from the UCSC genome browser (Casper et al., 2018) describing the genomic context of the VNTR (in red). The region represents a portion of the intron of the *NPRL3* gene containing the VNTR. The genomic features overlapped by the VNTR are highlighted in the light orange box. The antisense long noncoding RNA transcribed within the intron is the green transcript at the end of the figure. NPRL3, nitrogen permease regulator like protein 3; VNTR, variable number of tandem repeat

Immunofluorescence analysis showed that both isoforms present a similar nucleocytoplasmic distribution (Figure S2C). We then assessed the main source of TAGNPRL3 protein expression in the blood in vivo and whether TAGNPRL3 protein levels were dependent on the presence of the 13-repeats variant. To this purpose we carried out whole blood fractionation of sample MB, a 16/16 HC, collecting PBMC and RBC fractions. After preparing protein lysates, by western blot analysis we detected a band of about 60 kDa in whole blood and RBCs, as expected from RNA expression analysis. Endogenous bands were of similar size of the one detected ectopically expressing *TAGNPRL3* cDNA in HEK 293T cells (Figure 8a). We then compared the expression in RBC of MB with two 16/13 HC individuals (C_43_CS and C_12_CD), as previously carried out for *TAGNPRL3* mRNA levels. As expected, TAGNPRL3 protein was highly expressed in the two 16/13 HC compared with the 16/16 individual proving that high mRNA levels lead to higher protein expression in RBC (Figure 8b). Although the protein band in this experiment does not correspond exactly in migration size to the overexpressed protein in HEK 293T cells, we hypothesize this may be due to the different lysate preparations and total protein amount loaded on the gel. Furthermore, we cannot exclude the presence of cell-type specific posttranslational modifications.

3.8 | TAGNPRL3 protein maintains properties of NPRL3 isoform 1

NPRL3 isoform 1 dimerizes with its partner NPRL2 and, together with DEPCDC5, they form a protein complex called GATOR1 in mammals, which in turns is involved in many cellular processes such as stimulation of protein synthesis, growth, metabolism, and the inhibition of autophagy (Bar-Peled et al., 2013). Furthermore, NPRL3 is also able to inhibit cell proliferation. It is therefore important to carry out a preliminary assessment of the ability of the TAGNPRL3 protein to maintain some of the biological function of NPRL3 isoform 1.

To assess if TAGNPRL3-encoded protein could interact with NPRL2, a coimmunoprecipitation assay has been set up by cotransfecting HEK cells with NPRL2_MYC and TAGNPRL3. As control, we used NPRL3 isoform 1 and NPRL3 isoform 2: the former is the canonical NPRL3 isoform involved in GATOR1 complex, the latter is a N-terminal truncated (Δ1–179 amino acids) protein-coding isoform. According to Levine et al. (2013), NPRL3 isoform 2 should not coimmunoprecipitate because it lacks the N-terminal longin domain that should dimerize with the N-terminal longin domain of NPRL2. As shown in Figure 9a, NPRL2 interacted with all the three different NPRL3 isoforms. This result suggests that there could be at least one additional domain involved in dimerization. To investigate whether TAGNPRL3-encoded protein inhibits proliferation as NPRL3 (Lunardi et al., 2009), we transfected plasmids expressing TAGNPRL3 or GFP and assayed cells proliferation after 48 hr with BrdU staining. By immunofluorescence with anti-BrdU antibody, we found that *TAGNPRL3*-transfected cells showed a significant decrease in the rate of proliferation compared with controls: an average of 40% of GFP-positive cells were growing in control experiment while only about 20% of TAGNPRL3-overexpressing cells incorporated BrdU. We thus proved that TAGNPRL3 retains NPRL3 isoform ability to inhibit cell proliferation (Figure 9b).

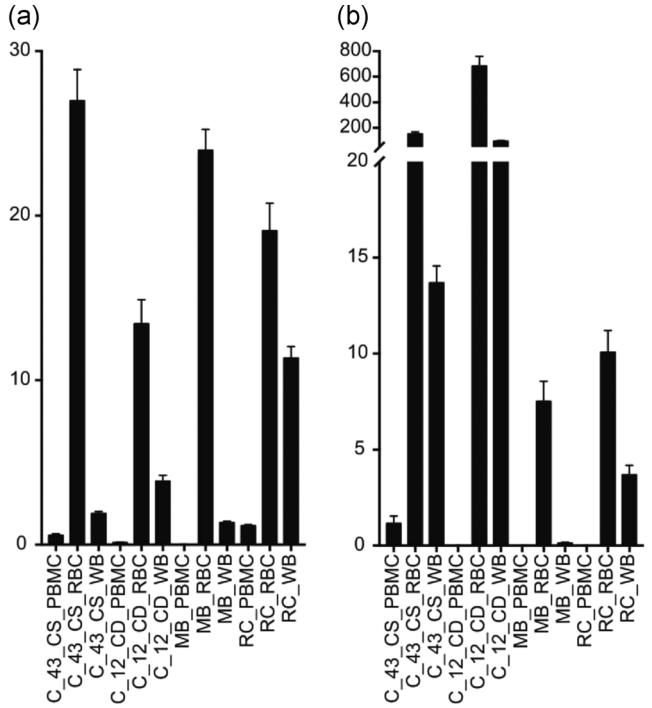

FIGURE 7 Red blood cells are the main source of *TAGNPRL3* expression in blood and the 16/13 minisatellite genotype induces *TAGNPRL3* transcript. Real-Time TaqMan PCR on blood fractions of four HC: two heterozygous (C_43_CS and C_12_CD) and two homozygous (MB and RC). Whole blood samples have been fractionated with Histopaque®-1077 reagent (Sigma-Aldrich) and total RNAs have been extracted with TRIzol reagent (Invitrogen). After cDNA preparation, Real-Time TaqMan assays have been performed. (a) *NPRL3* assay shows that its expression is variable among the four individuals and ubiquitous in the cell fractions, although prevalent in RBC. (b) *TAGNPRL3* assay shows that its expression seems to be exclusive of RBC (except from C_43_CS PBMC) and it is highly induced in the RBC of heterozygous individuals. Data presented as mean ± SEM of three technical replicates. NPRL3, nitrogen permease regulator like protein 3; PBMC, peripheral blood mononuclear cell; RBC, red blood cell

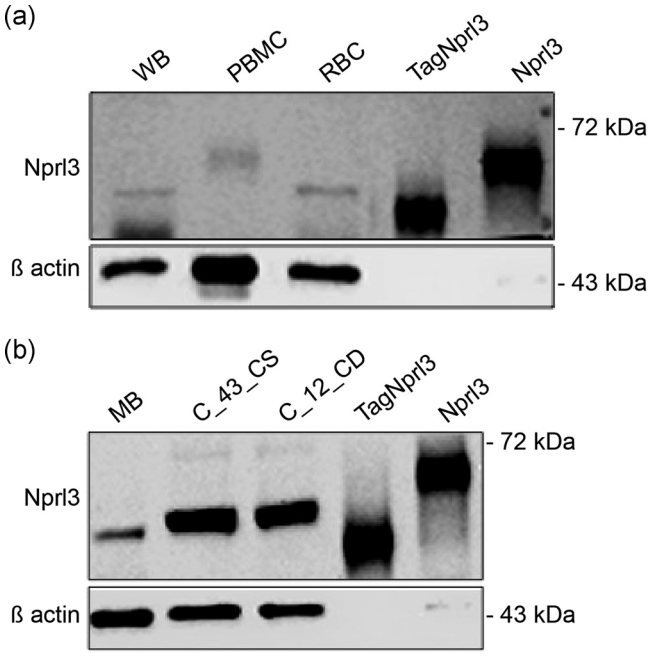

FIGURE 8 TAGNPRL3 protein expression analysis. (a) Western blot analysis with anti-NPRL3 and anti-β actin antibodies of MB, a 16/16 HC individual. WB, PBMC, and RBC are compared with cell lysates of HEK 293T cells upon *TAGNPRL3* and *NPRL3* isoform 1 transfection. As expected from analysis of *TAGNPRL3* mRNA levels in the same individual, TAGNPRL3 protein is expressed mainly in RBC. Endogenous bands were of similar size of the one detected ectopically expressing *TAGNPRL3* cDNA. (b) Western blot analysis with anti-NPRL3 and anti-β ACTIN antibodies of RBC of MB and two 16/13 HC individuals (C_43_CS and C_12_CD). TAGNPRL3 protein was highly expressed in the two 16/13 HC compared with the 16/16 individual proving that high mRNA levels lead to higher protein expression in RBC. mRNA, messenger RNA; NPRL3, nitrogen permease regulator like protein 3; PBMC, peripheral blood mononuclear cell; RBC, red blood cell; WB, whole blood

4 | DISCUSSION

The FANTOM projects have profoundly changed the view of the mammalian transcriptome. First, in conjunction with the isolation of a very large collection of full length cDNAs, many additional types of RNAs were identified leading to the discovery of long noncoding RNAs. By the use of CAGE in combination with a paired-end tag sequencing method (Wei et al., 2004), it was unveiled that the genome is pervasively transcribed—with more than 63% of the genome producing transcripts (Carninci et al., 2005), and more than 73% of the genes showing some form of antisense transcription (Katayama et al., 2005). A first comprehensive promoter map both for human and mouse (Carninci et al., 2006) was then assembled leading to the discovery of different classes of promoter architectures (Carninci et al., 2006; Lenhard, Sandelin, & Carninci, 2012). Furthermore, REs were shown to be frequently transcribed in a cell-type specific and regulated manner (Faulkner et al., 2009). This provided evidence that REs contained promoters driving the transcription of coding and noncoding RNAs in various tissues, especially in embryonic stages where long interspersed nuclear element and long terminal repeat elements are particularly active.

Here we show that CAGE tags identify a TSS host by minisatellite sequences and that its expression is influenced by genomic VNTRs.

Since their first description in the 80s, minisatellites have been intensely studied for their high degree of polymorphism and effects on the expression of the adjacent gene. While most of them are quite stable, selected loci, called hypermutable minisatellites, show hypervariability leading to the establishment of human DNA fingerprinting assays (Denoeud, Vergnaud, & Benson, 2003).

Minisatellites polymorphisms in the coding region of a gene are the best candidates for functional effects. The most studied example is the minisatellite present in the third exon of the dopamine receptor D4 (DRD4) gene that has been associated with attention deficit hyperactive disorder, response to clozapine in schizophrenia treatment (Shaikh et al., 1993) and other neurological disorders. The 48 bp minisatellite ranges from 2 to 11 repeats and, once translated,

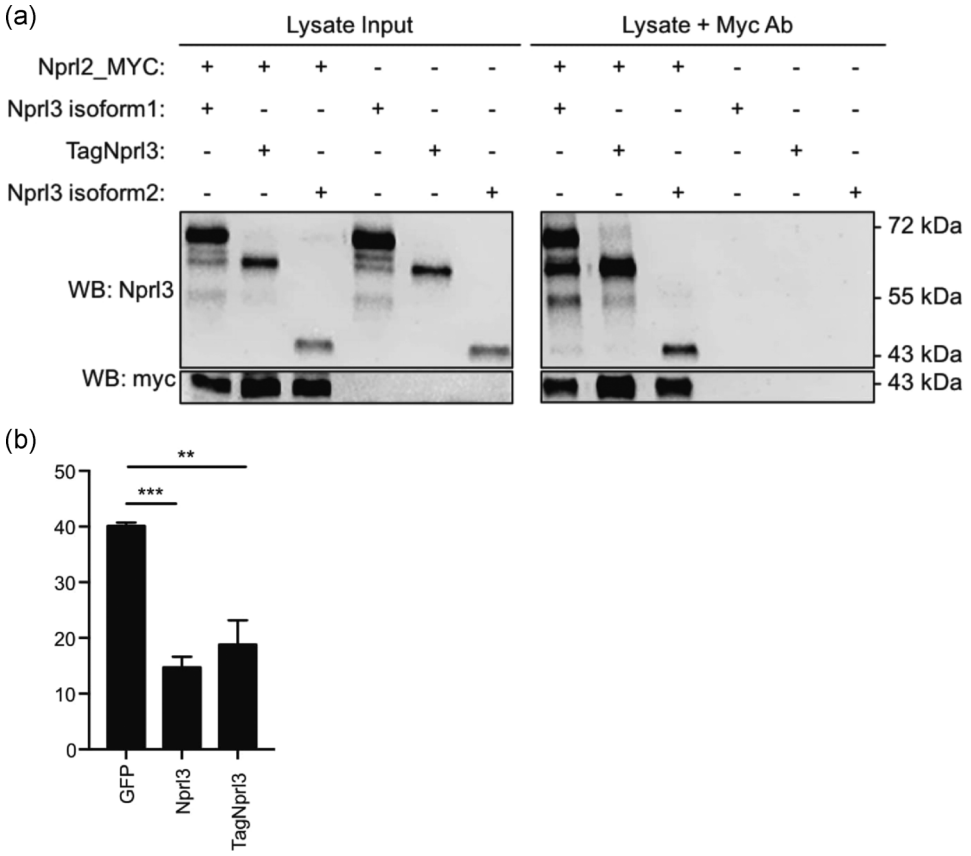

FIGURE 9 Coimmunoprecipitation of NPRL2 with NPRL3 isoforms. (a) Western blot showing the coimmunoprecipitation assay in HEK 293T cells transiently transfected with *NPRL2* and one of the three *NPRL3* isoforms. After 24 hr from transfection, cell lysates have been incubated with Myc antibody that recognize the C-terminal tag of NPRL2. Protein G Sepharose beads have been used to precipitate the protein complexes bound by Myc antibody that have been tested for NPRL3 isoforms presence by western blot. NPRL2 coimmunoprecipitates with all the three NPRL3 isoforms. (b) TAGNPRL3 retains the ability of the full-length protein to inhibit proliferation, since it inhibits BrdU incorporation. A vector expressing GFP was transfected as a control. The *y* axis indicates the % of BrdU positive cells in the transfected population. $N = 4$. BrdU, bromodeoxyuridine; GFP, green fluorescent protein; NPRL3, nitrogen permease regulator like protein 3

each TR adds 16 amino acids to the third cytoplasmic loop of the protein (Chio et al., 1994; Van Tol et al., 1992). Albeit the function of the receptor is only partially influenced by the number of repeats and only in certain conditions, different DRD4 variants display differential sensitivity to dopamine chaperone effect in the cell (Van Craenenbroeck, De Bosscher, Vanden Berghe, Vanhoenacker, & Haegeman, 2005).

Minisatellite polymorphisms are also found in promoters where they may influence gene expression levels. As representative examples, the 20–23 bp of the 5HTT promoter are repeated 14 or 16 times and its short variant reduces transcriptional efficiency. It is associated to anxiety-related traits (Lesch et al., 1996), increased risk of depression (Lotrich & Pollock, 2004), and reduced gray matter volume and connectivity in limbic regions (Pezawas et al., 2005). MAOA gene has a 30 bp repeat sequence in the promoter region which variants are associated to its mRNA levels. These may influence bipolar disorder, impulsivity, and antisocial behaviour (Brookes, 2013). Finally, minisatellite polymorphisms can be present in intronic and UTRs of genes influencing RNA splicing, localization, stability, and translational efficiency. A 3′-UTR minisatellite of 40 bp repeated 7–11 times influences the expression of the dopamine transporter gene (DAT1/SLC6A3), which in turn affect dopamine re-uptake processes and behaviour (Brookes, 2013).

Nevertheless, since they are not represented in CGH and SNPs arrays, the genome-wide identification of minisatellite polymorphisms present in noncoding regions and acting as eQTL remains elusive.

The *NPRL3*-associated VNTR is a 29 nt long minisatellite with two variants of 16 and 13 repeats. By mapping CAGE tags and 5′RACE sequencing data, we found that TSS of *TAGNPRL3* mRNA localizes in the 11th repeat, the one that exclusively presents the tag sequence despite the high homology between repeats. Transcription starts from a relatively uncommon core promoter sequence called TCT motif (YC+1TYTYY), which has been found to be specific for genes of ribosomal and translation-related proteins (Parry et al., 2010). This is intriguing since *NPRL3* ablation in mice causes the perturbation of protein synthesis (Kowalczyk et al., 2012a).

While our initial observation suggested that high levels of *TAGNPRL3* mRNA expression in the blood can be associated to PD in a portion of cases, the analysis of a larger number of individuals

proved that high *TAGNPRL3* were occurring in 16/13 and not in 16/16 minisatellite genotypes, irrespective of the pathology. This may suggest that a selective group of individuals within the local population shows a higher allelic frequency of the 13-repeats variant independently of PD. Future population studies may address this important point.

The molecular basis of *TAGNPRL3* differential expression and its dependency on repeat number remains unclear as well as whether the 13-repeat variant influences the level of *TAGNPRL3* expression in other tissues including the brain. As expected, minisatellite variants do not influence *NPRL3* canonical isoform 1 mRNA expression. The VNTR locus overlaps two eQTLs that are associated to a SNP modulating the expression of the NPRL3 gene in the lung. It also overlaps a DNAse I hypersensitivity cluster that has been previously identified in the Th1 cell lineage. In agreement with a potential overlap with an enhancer region, the VNTR also overlaps a peak of H3K27 acetylation. By transcription factor binding site analysis we identified MZF1 as the transcription factor with the highest probability of binding the VNTR. Since MZF1 is involved in hemopoiesis, future work will investigate a potential causal relationship between the expression of this zinc finger protein and transcription at this locus. Since a long noncoding RNA is transcribed in antisense to the NPRL3 gene starting at about 1,500 nucleotides downstream the VNTR, it will be also worth investigating whether its expression is dependent on the VNTR alleles.

TAGNPRL3 mRNAs expression occurs in RBCs of 16/13 individuals and is increased in RBCs of 16/13 genotypes. This pattern seems mirrored at the protein levels since we found that a band migrating at the same apparent size of transfected *TAGNPRL3* was present in RBC of the very same 16/16 HC. Furthermore, when comparing RBC of this individual with those of two 16/13 we detected a much higher expression in heterozygotes. While in our experimental conditions the migration of the TAGNPRL3 protein from RBC lysates of 16/13 individuals does not perfectly mimic the one from overexpressing *TAGNPRL3* cDNA in HEK 293T cells, it is very likely that they are indeed the same protein and this difference may be due to different lysate preparations (RBC vs. a cultured cell line) and total protein amounts loaded on gel (much less in the case of overexpressing HEK 293T cells). Importantly, we cannot exclude the presence of cell-type specific posttranslational modifications since TAGNPRL3 presents consensus sequences for ubiquitination and phosphorylation.

Interestingly, a SNP in NPRL3 first intron (rs7203560) has been found correlated to haemolysis in sickle cell anemia (Milton et al., 2013). Furthermore, the minisatellite maps only 3 kb away of MCS-R3, one of the major α-globin enhancer (Higgs, 2013). Future work will assess whether minisatellite alleles may influence hemoglobin expression.

A central question concerns the biological significance of TAGNPRL3. NPRL3 dimerizes with its homologous NPRL2 and was initially identified as mediating an amino acid starvation signal to TORC1 in yeast. Together with DEPDC5, they form the GATOR1 protein complex, an evolutionary conserved inhibitor of mTORC1 activity (Wei, Reveal, Cai, & Lilly, 2016). Recently, *NPRL3* gene mutations have been found in FCD and in familial focal epilepsy (Sim et al., 2015) where the loss of NPRL3 function led to mTOR dysregulation. In eukaryotes, TOR is the major sensor of nutrients, energy, and stress. Alterations in its pathway have been correlated with diseases and conditions where growth and homeostasis are compromised like cancer, metabolic diseases, and aging (Bjedov et al., 2010; Hansen et al., 2008; Johnson, Rabinovitch, & Kaeberlein, 2013; Ravikumar et al., 2004). It is therefore tempting to speculate that TAGNPRL3 expression levels may influence mTOR activity and regulate metabolic homeostasis and cellular response to stress. In the blood, imbalances in mTORC1 function lead to different types of anemias with changes in RBC size, number, and hemoglobin content while its pharmacological activation can ameliorate some types of hereditary anemias (Bayeva et al., 2012; Jaako et al., 2012; Knight, Schmidt, Birsoy, Tan, & Friedman, 2014; La, Yang, & Dennery, 2013; Ohyashiki et al., 2009; Payne et al., 2012). Interestingly, *NPRL2* absence in mice severely compromises hematopoiesis (Dutchak et al., 2015).

The functional effects of TAGNPRL3 should be discussed in terms of isoform protein structure and level of expression. TAGNPRL3 encodes for a nucleocytoplasmic protein lacking the first 78 amino acids at the N-terminal of isoform 1. We experimentally showed that TAGNPRL3 maintains two features of the full-length protein. First, TAGNPRL3 is able to bind its partner NPRL2 and therefore potentially be part of the GATOR1 complex. Interestingly, our data shows that LD at the N-terminal (Levine et al., 2013) is not required for this interaction since NPRL3 isoform 2, lacking this domain, maintains its ability to dimerize with NPRL2. Furthermore, by performing BrdU assays, we demonstrated that TAGNPRL3 is able to inhibit proliferation. Since this function seems to be dependent on the binding of the NPRL3 isoform 1 to all p53-family proteins (Lunardi et al., 2009), further experiments should assess whether TAGNPRL3 is able to preserve these protein interactions.

A crucial open question concerns which fraction of the increased quantity of TAGNPRL3, driven by the 13-repeat variant, is included in the GATOR1 complex and whether an increased TAGNPRL3 expression influences mTORC1 activity. While in our cohort the 13-repeat variant presents an allelic frequency of 10.41%, genetic analysis of *TAGNPRL3* VNTRs should be carried out for all those pathological contexts dependent on mTORC1 activity. It will be also interesting to address whether the lack of the N-terminal portion can influence selected NPRL3 protein interactions or biochemical activities as well as gives rise to a dominant negative isoform.

In summary, this study provides a new example of a VNTR that is both a TSS and an eQTL of an important gene crucial for mTORC1 regulation and in a genomic location enriched in regulatory elements for α globins.

ACKNOWLEDGMENTS

We express our gratitude to patients, control subjects who participated in this study and to "Associazione Donatori di

Sangue," Trieste. We are indebted to all the members of the SG and PC labs for thought-provoking discussions, to Dr. Helena Krmac, Cristina Leonesi (SISSA) and to Omar Peruzzi and Eva Ferri (IIT) for technical help, to Monica Sirk, Annalisa Sulli (SISSA) and Alessandra Sanna (IIT) for administrative work. This study was supported by SISSA intramural grants, by the Italian Ministry of Education, University and Research (FIRB grant prot.: RBAP11FRE9) to SG and GP, by the grant (GGP10224) of the Telethon Foundation to SG and GP, by IIT intramural grants to SG and by MEXT (The Ministry of Education, Culture, Sports, Science and Technology) to PC.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

M. B. designed and performed the experiments, wrote the manuscript. D. T. performed the experiments and data analyses, revised the manuscript. R. C. designed and performed the experiments, wrote the manuscript. C. V. and S. F. performed the experiments and revised the manuscript. R. S. and F. P. designed the experiments, discussed the results, and revised the manuscript. S. Goldwurm, P. M., G. B., and G. P. perform patients' diagnosis, provided their samples, and clinical information. S. G. and P. C. conceived the project, designed the experiments, analyzed the data, and write the manuscript. All authors read and approved the final version of the manuscript.

ORCID

Maria Bertuzzi http://orcid.org/0000-0001-9444-6166

REFERENCES

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., … Kasukawa, T. (2017). FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Scientific Data*, *4*, 170107. https://doi.org/10.1038/sdata.2017.107

Babatz, T. D., & Burns, K. H. (2013). Functional impact of the human mobilome. *Current Opinion in Genetics and Development*, *23*(3), 264–270. https://doi.org/10.1016/j.gde.2013.02.007

Bar-Peled, L., Chantranupong, L., Cherniack, A. D., Chen, W. W., Ottina, K. A., Grabiner, B. C., … Sabatini, D. M. (2013). A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. *Science*, *340*(6136), 1100–1106. https://doi.org/10.1126/science.1232044

Bayeva, M., Khechaduri, A., Puig, S., Chang, H.-C., Patial, S., Blackshear, P. J., & Ardehali, H. (2012). mTOR regulates cellular iron homeostasis through tristetraprolin. *Cell Metabolism*, *16*(5), 645–657. https://doi.org/10.1016/j.cmet.2012.10.001

Bjedov, I., Toivonen, J. M., Kerr, F., Slack, C., Jacobson, J., Foley, A., & Partridge, L. (2010). Mechanisms of life span extension by rapamycin in the fruit fly *Drosophila melanogaster*. *Cell Metabolism*, *11*(1), 35–46. https://doi.org/10.1016/j.cmet.2009.11.010

Brookes, K. J. (2013). The VNTR in complex disorders: The forgotten polymorphisms? A functional way forward? *Genomics*, *101*(5), 273–281. https://doi.org/10.1016/j.ygeno.2013.03.003

Brummett, B. H., Krystal, A. D., Siegler, I. C., Kuhn, C., Surwit, R. S., Züchner, S., … Williams, R. B. (2007). Associations of a regulatory polymorphism of monoamine oxidase-A gene promoter (MAOA-uVNTR) with symptoms of depression and sleep quality. *Psychosomatic Medicine*, *69*(5), 396–401. https://doi.org/10.1097/PSY.0b013e31806d040b

Calligaris, R., Banica, M., Roncaglia, P., Robotti, E., Finaurini, S., Vlachouli, C., … Gustincich, S. (2015). Blood transcriptomics of drug-naïve sporadic Parkinson's disease patients. *BMC Genomics*, *16*, 876. https://doi.org/10.1186/s12864-015-2058-3

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., & Maeda, N., … RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). (2005). The transcriptional landscape of the mammalian genome. *Science*, *309*(5740), 1559–1563. https://doi.org/10.1126/science.1112014

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., … Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, *38*(6), 626–635. https://doi.org/10.1038/ng1789

Casper, J., Zweig, A. S., Villarreal, C., Tyner, C., Speir, M. L., Rosenbloom, K. R., … Kent, W. J. (2018). The UCSC Genome Browser database: 2018 update. *Nucleic Acids Research*, *46*(D1), D762–D769. https://doi.org/10.1093/nar/gkx1020

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., … Ng, H.-H. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, *133*(6), 1106–1117. https://doi.org/10.1016/j.cell.2008.04.043

Chen, Y.-Y., Lin, S.-Y., Yeh, Y.-Y., Hsiao, H.-H., Wu, C.-Y., Chen, S.-T., & Wang, A. H.-J. (2005). A modified protein precipitation procedure for efficient removal of albumin from serum. *Electrophoresis*, *26*(11), 2117–2127. https://doi.org/10.1002/elps.200410381

Chio, C. L., Drong, R. F., Riley, D. T., Gill, G. S., Slightom, J. L., & Huff, R. M. (1994). D4 dopamine receptor-mediated signaling events determined in transfected Chinese hamster ovary cells. *The Journal of Biological Chemistry*, *269*(16), 11813–11819.

Cookson, W., Liang, L., Abecasis, G., Moffatt, M., & Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nature Reviews Genetics*, *10*(3), 184–194. https://doi.org/10.1038/nrg2537

Van Craenenbroeck, K., De Bosscher, K., Vanden Berghe, W., Vanhoenacker, P., & Haegeman, G. (2005). Role of glucocorticoids in dopamine-related neuropsychiatric disorders. *Molecular and Cellular Endocrinology*, *245*(1–2), 10–22. https://doi.org/10.1016/j.mce.2005.10.007

van Dyck, C. H., Malison, R. T., Jacobsen, L. K., Seibyl, J. P., Staley, J. K., Laruelle, M., … Gelernter, J. (2005). Increased dopamine transporter availability associated with the 9-repeat allele of the SLC6A3 gene. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine*, *46*(5), 745–751.

Denoeud, F., Vergnaud, G., & Benson, G. (2003). Predicting human minisatellite polymorphism. *Genome Research*, *13*(5), 856–867. https://doi.org/10.1101/gr.574403

Dutchak, P. A., Laxman, S., Estill, S. J., Wang, C., Wang, Y., Wang, Y., … Tu, B. P. (2015). Regulation of hematopoiesis and methionine homeostasis by mTORC1 inhibitor NPRL2. *Cell Reports*, *12*(3), 371–379. https://doi.org/10.1016/j.celrep.2015.06.042

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., … Stefansson, K. (2008). Genetics of gene expression and its effect on disease. *Nature*, *452*(7186), 423–428. https://doi.org/10.1038/nature06758

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A. R. R., Kawaji, H., Rehli, M., Baillie, J. K., de Hoon, M. J. L., … Hayashizaki, Y. (2014). A promoter-level mammalian expression atlas. *Nature*, *507*(7493), 462–470. https://doi.org/10.1038/nature13182

Faraone, S. V., Doyle, A. E., Mick, E., & Biederman, J. (2001). Meta-analysis of the association between the 7-repeat allele of the dopamine D(4) receptor gene and attention deficit hyperactivity disorder. *The American Journal of Psychiatry*, *158*(7), 1052–1057. https://doi.org/10.1176/appi.ajp.158.7.1052

Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., … Carninci, P. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nature Genetics*, *41*(5), 563–571. https://doi.org/10.1038/ng.368

Grison, A., Zucchelli, S., Urzì, A., Zamparo, I., Lazarevic, D., Pascarella, G., … Gustincich, S. (2014). Mesencephalic dopaminergic neurons express a repertoire of olfactory receptors and respond to odorant-like molecules. *BMC Genomics*, *15*, 729. https://doi.org/10.1186/1471-2164-15-729

GTEx Consortium. (2013). The genotype-tissue expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Hansen, M., Chandra, A., Mitic, L. L., Onken, B., Driscoll, M., & Kenyon, C. (2008). A role for autophagy in the extension of lifespan by dietary restriction in C. elegans. *PLOS Genetics*, *4*(2), e24. https://doi.org/10.1371/journal.pgen.0040024

Hasegawa, A., Daub, C., Carninci, P., Hayashizaki, Y., & Lassmann, T. (2014). MOIRAI: A compact workflow system for CAGE analysis. *BMC Bioinformatics*, *15*, 144. https://doi.org/10.1186/1471-2105-15-144

Higgs, D. R. (2013). The molecular basis of α-thalassemia. *Cold Spring Harbor Perspectives in Medicine*, *3*(1), a011718. https://doi.org/10.1101/cshperspect.a011718.

Jaako, P., Debnath, S., Olsson, K., Bryder, D., Flygare, J., & Karlsson, S. (2012). Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. *Blood*, *120*(11), 2225–2228. https://doi.org/10.1182/blood-2012-05-431437

Jansen, R. C., & Nap, J. P. (2001). Genetical genomics: The added value from segregation. *Trends in Genetics*, *17*(7), 388–391.

Johnson, S. C., Rabinovitch, P. S., & Kaeberlein, M. (2013). mTOR is a key modulator of ageing and age-related disease. *Nature*, *493*(7432), 338–345. https://doi.org/10.1038/nature11861

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., & Nakamura, M., … FANTOM Consortium. (2005). Antisense transcription in the mammalian transcriptome. *Science*, *309*(5740), 1564–1566. https://doi.org/10.1126/science.1112009

Kennedy, S. P., Weed, S. A., Forget, B. G., & Morrow, J. S. (1994). A partial structural repeat forms the heterodimer self-association site of all beta-spectrins. *The Journal of Biological Chemistry*, *269*(15), 11400–11408.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., … Mathelier, A. (2018). JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research*, *46*(D1), D1284. https://doi.org/10.1093/nar/gkx1188

Knight, Z. A., Schmidt, S. F., Birsoy, K., Tan, K., & Friedman, J. M. (2014). A critical role for mTORC1 in erythropoiesis and anemia. *eLife*, *3*, e01913. https://doi.org/10.7554/eLife.01913

Kowalczyk, M. S., Hughes, J. R., Babbs, C., Sanchez-Pulido, L., Szumska, D., Sharpe, J. A., … Higgs, D. R. (2012). Nprl3 is required for normal development of the cardiovascular system. *Mammalian Genome*, *23*(7–8), 404–415. https://doi.org/10.1007/s00335-012-9398-y

La, P., Yang, G., & Dennery, P. A. (2013). Mammalian target of rapamycin complex 1 (mTORC1)-mediated phosphorylation stabilizes ISCU protein: Implications for iron metabolism. *The Journal of Biological Chemistry*, *288*(18), 12901–12909. https://doi.org/10.1074/jbc.M112.424499

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, *13*(4), 233–245. https://doi.org/10.1038/nrg3163

Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., … Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, *274*(5292), 1527–1531.

Levine, T. P., Daniels, R. D., Wong, L. H., Gatta, A. T., Gerondopoulos, A., & Barr, F. A. (2013). Discovery of new Longin and Roadblock domains that form platforms for small GTPases in Ragulator and TRAPP-II. *Small GTPases*, *4*(2), 62–69. https://doi.org/10.4161/sgtp.24262

Lin, X., Cook, T. J., Zabetian, C. P., Leverenz, J. B., Peskind, E. R., Hu, S.-C., … Zhang, J. (2012). DJ-1 isoforms in whole blood as potential biomarkers of Parkinson disease. *Scientific Reports*, *2*, 954. https://doi.org/10.1038/srep00954

Lotrich, F. E., & Pollock, B. G. (2004). Meta-analysis of serotonin transporter polymorphisms and affective disorders. *Psychiatric Genetics*, *14*(3), 121–129.

Lower, K. M., Hughes, J. R., De Gobbi, M., Henderson, S., Viprakasit, V., Fisher, C., … Higgs, D. R. (2009). Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21771–21776. https://doi.org/10.1073/pnas.0909331106

Lunardi, A., Chiacchiera, F., D'Este, E., Carotti, M., Dal Ferro, M., Di Minin, G., … Collavin, L. (2009). The evolutionary conserved gene C16orf35 encodes a nucleo-cytoplasmic protein that interacts with p73. *Biochemical and Biophysical Research Communications*, *388*(2), 428–433. https://doi.org/10.1016/j.bbrc.2009.08.027

Manca, M., Pessoa, V., Lopez, A. I., Harrison, P. T., Miyajima, F., Sharp, H., … Quinn, J. P. (2018). The regulation of monoamine oxidase A gene expression by distinct variable number tandem repeats. *Journal of Molecular Neuroscience: MN*, *64*(3), 459–470. https://doi.org/10.1007/s12031-018-1044-z

Milton, J. N., Rooks, H., Drasar, E., McCabe, E. L., Baldwin, C. T., Melista, E., … Steinberg, M. H. (2013). Genetic determinants of haemolysis in sickle cell anaemia. *British Journal of Haematology*, *161*(2), 270–278. https://doi.org/10.1111/bjh.12245

Morris, J. F., Hromas, R., & Rauscher, F. J. (1994). Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: Two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Molecular and Cellular Biology*, *14*(3), 1786–1795. https://doi.org/10.1128/mcb.14.3.1786

Ogilvie, A. D., Battersby, S., Bubb, V. J., Fink, G., Harmar, A. J., Goodwim, G. M., & Smith, C. A. (1996). Polymorphism in serotonin transporter gene associated with susceptibility to major depression. *Lancet (London, England)*, *347*(9003), 731–733.

Ohyashiki, J. H., Kobayashi, C., Hamamura, R., Okabe, S., Tauchi, T., & Ohyashiki, K. (2009). The oral iron chelator deferasirox represses signaling through the mTOR in myeloid leukemia cells by enhancing expression of REDD1. *Cancer Science*, *100*(5), 970–977. https://doi.org/10.1111/j.1349-7006.2009.01131.x

Parry, T. J., Theisen, J. W. M., Hsu, J.-Y., Wang, Y.-L., Corcoran, D. L., Eustice, M., … Kadonaga, J. T. (2010). The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes and Development*, *24*(18), 2013–2018. https://doi.org/10.1101/gad.1951110

Pascarella, G., Lazarevic, D., Plessy, C., Bertin, N., Akalin, A., Vlachouli, C., … Gustincich, S. (2014). NanoCAGE analysis of the mouse olfactory epithelium identifies the expression of vomeronasal receptors and of proximal LINE elements. *Frontiers in Cellular Neuroscience*, *8*, 41. https://doi.org/10.3389/fncel.2014.00041

Payne, E. M., Virgilio, M., Narla, A., Sun, H., Levine, M., Paw, B. H., … Khanna-Gupta, A. (2012). L-Leucine improves the anemia and developmental defects associated with Diamond-Blackfan anemia and del(5q) MDS by activating the mTOR pathway. *Blood*, *120*(11), 2214–2224. https://doi.org/10.1182/blood-2011-10-382986

Pezawas, L., Meyer-Lindenberg, A., Drabant, E. M., Verchinski, B. A., Munoz, K. E., Kolachana, B. S., … Weinberger, D. R. (2005). 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: A genetic susceptibility mechanism for depression. *Nature Neuroscience*, *8*(6), 828–834. https://doi.org/10.1038/nn1463

Plessy, C., Bertin, N., Takahashi, H., Simone, R., Salimullah, M., Lassmann, T., … Carninci, P. (2010). Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. *Nature Methods*, *7*(7), 528–534. https://doi.org/10.1038/nmeth.1470

Plessy, C., Pascarella, G., Bertin, N., Akalin, A., Carrieri, C., Vassalli, A., … Carninci, P. (2012). Promoter architecture of mouse olfactory receptor genes. *Genome Research*, *22*(3), 486–497. https://doi.org/10.1101/gr.126201.111

Ravikumar, B., Vacher, C., Berger, Z., Davies, J. E., Luo, S., Oroz, L. G., … Rubinsztein, D. C. (2004). Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nature Genetics*, *36*(6), 585–595. https://doi.org/10.1038/ng1362

Shaikh, S., Collier, D., Kerwin, R. W., Pilowsky, L. S., Gill, M., Xu, W. M., & Thornton, A. (1993). Dopamine D4 receptor subtypes and response to clozapine. *Lancet*, *341*(8837), 116.

Sim, J. C., Scerri, T., Fanjul-Fernández, M., Riseley, J. R., Gillies, G., Pope, K., … Leventer, R. J. (2015). Familial cortical dysplasia caused by mutation in the mTOR regulator NPRL3. *Annals of Neurology*, *79*(1), 132–137. https://doi.org/10.1002/ana.24502

Suzuki, H., FANTOM Consortium, Forrest, A. R. R., van Nimwegen, E., Daub, C. O., Balwierz, P. J., … Hayashizaki, Y. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genetics*, *41*(5), 553–562. https://doi.org/10.1038/ng.375

Takahashi, H., Kato, S., Murata, M., & Carninci, P. (2012). CAGE (cap analysis of gene expression): A protocol for the detection of promoter and transcriptional networks. *Methods in Molecular Biology*, *786*, 181–200. https://doi.org/10.1007/978-1-61779-292-2_11

Van Tol, H. H., Wu, C. M., Guan, H. C., Ohara, K., Bunzow, J. R., Civelli, O., … Jovanovic, V. (1992). Multiple dopamine D4 receptor variants in the human population. *Nature*, *358*(6382), 149–152. https://doi.org/10.1038/358149a0

Vyas, P., Vickers, M. A., Picketts, D. J., & Higgs, D. R. (1995). Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. *Genomics*, *29*(3), 679–689. https://doi.org/10.1006/geno.1995.9951

Warburton, A., Breen, G., Bubb, V. J., & Quinn, J. P. (2016). A GWAS SNP for schizophrenia is linked to the internal MIR137 promoter and supports differential allele-specific expression. *Schizophrenia Bulletin*, *42*(4), 1003–1008. https://doi.org/10.1093/schbul/sbv144

Wei, C.-L., Ng, P., Chiu, K. P., Wong, C. H., Ang, C. C., Lipovich, L., … Ruan, Y. (2004). 5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(32), 11701–11706. https://doi.org/10.1073/pnas.0403514101

Wei, Y., Reveal, B., Cai, W., & Lilly, M. A. (2016). The GATOR1 complex regulates metabolic homeostasis and the response to nutrient stress in *Drosophila melanogaster*. *G3*, *6*(12), 3859–3867. https://doi.org/10.1534/g3.116.035337

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Bertuzzi M, Tang D, Calligaris R, et al. A human minisatellite hosts an alternative transcription start site for *NPRL3* driving its expression in a repeat number-dependent manner. *Human Mutation*. 2020;41:807–824. https://doi.org/10.1002/humu.23974