UNIVERSITÄT ZU LÜBECK

From the Institute of Neurogenetics

of the University of Lübeck

Director: Prof. Dr. med. Christine Klein

Identification and characterization of novel genetic causes of dystonia

Dissertation

for Fulfilment of the Requirements

for the Doctoral Degree

at the University of Lübeck

from the Department of Natural Sciences

Submitted by

Hauke Baumann

from Celle

Lübeck, 2020

First referee: Prof. Dr. rer. nat. Katja Lohmann

Second referee: Prof. Dr. rer. nat. Henrik Oster

Date of oral examination: 24.07.2020

Approved for printing, Lübeck, 28.07.2020

To Bernd Baumann

LIST OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

1. Abstract

Dystonia is a rare neurological disorder characterized by excessive contraction of muscles leading to involuntary movements, postures, or both. Dystonic symptoms can either occur as the only disease manifestation (isolated dystonia) or may be part of a broader clinical syndrome (combined and complex dystonia). Over the past two decades, a number of genetically determined dystonia subtypes (DYT-[genename]) have been discovered, which improved our understanding of the pathophysiology of the disease to some extent. However, the molecular mechanisms underlying dystonia are still largely unknown. The main focus of this thesis was the molecular investigation of genetic variants in four genes recently linked to dystonia.

Pathogenic variants in the lysine methyltransferase 2B (*KMT2B*) gene have been discovered as a cause of childhood-onset dystonia (DYT-KMT2B). We identified a novel heterozygous *de-novo KMT2B* frameshift variant in a female dystonia patient. Gene expression analysis showed reduced *KMT2B* expression as a result of the frameshift variant and a non-frameshift variant from an unrelated DYT-KMT2B patient. Additionally, gene expression of dystonia genes THAP domain containing 1 (*THAP1*) and torsin family 1 member A (*TOR1A*) was reduced in blood compared to controls pointing towards a mechanistic link between these genetic dystonia subtypes (Klein, Baumann et al., 2019).

Previously, mutations in the guanine nucleotide-binding protein beta 1 (*GNB1*) gene were linked to a neurodevelopmental disorder accompanied by additional clinical features, including dystonia. The gene codes for the $G\beta_1$ subunit, which is part of a heterotrimeric G-protein complex. We identified sixteen carriers harbouring fourteen novel rare *GNB1* variants (two splice site, two frameshift, and ten missense changes). To assess the functional consequence of ten *GNB1* missense variants, dopamine-induced signal propagation was investigated using a cellular bioluminescence resonance energy transfer (BRET) assay. Aberrant functionality was detected in seven out of ten variants suggesting deficits in G-protein coupled receptor (GPCR) signalling. The presented data allow an assessment of the pathogenicity of a variant and provide valuable insights into a possible mechanism that may underly GNB1-related disorders (Lohmann, …, Baumann et al., 2017).

Recently, pathogenic variants in the VAC14 component of PIKFYVE complex (*VAC14*) gene have been identified as a cause of childhood-onset complex dystonia. VAC14 is a scaffold protein relevant for the regulation of phosphatidylinositol 3,5-bisphosphate ($PI(3,5)P_2$) and is hypothesized to form homodimers. In a dystonia patient with sudden-onset progressive dystonia harbouring we identified two novel rare compound-heterozygous variants in the *VAC14* gene. Both variants were located in the predicted homodimerization domain. To assess the influence of the variants and another recently reported variant on homodimerization, mutant VAC14 was overexpressed in a neuroblastoma cell line. Enhanced VAC14 homodimer formation was observed as a consequence of two variants compared to the wild-type. In contrast to previous reports, no enlarged vacuoles were detected in patient fibroblasts, possibly due to a compensatory effect of the milder second variant inherited from the mother. The functional results provide initial evidence of enhanced VAC14 homodimerization as a possible disease mechanism (Baumann et al., *in preparation*).

One of the most frequent monogenic subtypes of dystonia is caused by pathogenic variants in the THAP domain containing 1 (*THAP1*) gene (DYT-THAP1). Intriguingly, a large number (~50%) of pathogenic *THAP1* mutation carriers do not manifest dystonic symptoms, a phenomenon termed reduced penetrance. To date, little data is available on the genetic factors underlying reduced penetrance or the action of THAP1 during early human brain development. In the first part, a systematic literature search was conducted to review the published phenotypic and demographic literature of DYT-THAP1 patients

between December 2018 and March 2020 (Lange, Junker, Löns, Baumann et al., *in preparation*). The collected data was implemented in the Movement Disorder Society Genetic mutation database (MDSGene; https://www.mdsgene.org) as a resource for clinicians and geneticists. In the second part, a cellular model using human iPSC-derived cortical neurons of ten *THAP1* mutation carriers (five clinically affected and five clinically unaffected individuals) harbouring three different pathogenic variants was established to study the role of mutant THAP1 during early brain development. Twenty iPSC lines were successfully characterized confirming the clearance of Sendai virus, the genome integrity, the expression of pluripotency markers, and the potential to differentiate into all three germ layers (Baumann et al., 2018). Next, eighteen iPSC lines from THAP1 mutation carriers and eleven control iPSC lines were differentiated into cortical neurons. Global transcriptome analysis was conducted on day 44 of cortical differentiation to study the role of mutant THAP1 at early stages of development. Transcriptomic profiling revealed a penetrance-linked expression in which expressional differences were exclusively found in affected *THAP1* mutation carriers, but not in unaffected mutation carriers, when compared to controls. In addition, the gene expression of the disease gene *THAP1* in maturing cortical neurons was linked to the disease status as an upregulation was observed only in affected *THAP1* mutation carriers. An exploratory data analysis approach indicated an accumulation of genes related to perturbed cortical networks (*FOXP2*, *DCLK1*, *KIF26B*, *CNTN5*, *CNTNAP3B*, *PCDH19*, and *UNC5C*) which may influence DYT-THAP1 pathogenesis (Baumann et al., *in preparation*).

Collectively, the results presented in this work provide molecular insights into possible pathomechanisms underlying different forms of genetic dystonias. Further functional, genetic and clinical research is needed to fully resolve the pathophysiologic basis of dystonia in order to develop novel therapeutic treatments, and to provide the best care for patients.

2. Zusammenfassung

Dystonie ist eine seltene neurologische Erkrankung, welche durch eine übermäßige Kontraktion der Muskeln gekennzeichnet ist und zu unwillkürlichen Bewegungen, Körperhaltungen oder beidem führt. Dystone Symptome können entweder als einzige Krankheitsmanifestation auftreten (isolierte Dystonie) oder Teil eines umfassenderen klinischen Syndroms sein (kombinierte und komplexe Dystonie). In den letzten zwei Jahrzehnten wurde eine Reihe von genetisch bedingten Dystonie-Subtypen (DYT-[Genname]) entdeckt, die unser Verständnis der Pathophysiologie der Krankheit bis zu einem gewissen Grad verbessert haben. Die molekularen Mechanismen, die der Dystonie zugrunde liegen, sind jedoch noch weitgehend unbekannt. Der Schwerpunkt dieser Arbeit war die molekulare Untersuchung von genetischen Varianten in vier Genen, von denen seit Kurzem bekannt ist, dass sie Dystonien verursachen können.

Mutationen im *lysine methyltransferase 2B* (*KMT2B*)-Gen wurden als Ursache für die im Kindesalter einsetzende Dystonie (DYT-KMT2B) beschrieben. Wir identifizierten eine neue heterozygote de-novo *KMT2B frameshift*-Variante bei einer Dystonie-Patientin. Die Genexpressionsanalyse zeigte eine reduzierte KMT2B-Expression als Folge der *frameshift*-Variante und einer nicht-*frameshift*-Variante von einer anderen DYT-KMT2B-Patientin. Zusätzlich war die Genexpression der Dystoniegene *THAP domain containing 1* (*THAP1*) und *torsin family 1 member A* (*TOR1A*) in Blut im Vergleich zu Kontrollen reduziert, was auf eine mechanistische Verbindung zwischen diesen genetischen Dystonieformen hindeutet (Klein, Baumann et al., 2019).

Vor einiger Zeit wurden Mutationen im *G protein subunit beta 1* (*GNB1*)-Gen mit einer neurologischen Entwicklungsstörung in Verbindung gebracht, die von zusätzlichen klinischen Merkmalen einschließlich Dystonien begleitet ist. Das Gen kodiert für die Untereinheit Gβ1, die Teil eines heterotrimerischen G-Proteinkomplexes ist. Wir identifizierten sechzehn Träger, die vierzehn neue seltene *GNB1*-Varianten trugen (zwei Spleiß-, zwei *frameshift*- und zehn *missense*-Veränderungen). Um die funktionelle Konsequenz der *missense*-Varianten abzuschätzen, wurde die dopamininduzierte Signalweiterleitung mit Hilfe eines zellulären Biolumineszenz-Resonanz-Energietransfer-Assays (BRET) untersucht. Bei sieben von zehn Varianten wurde eine veränderte Funktionalität festgestellt, was auf ein Defizit bei der G-Protein-gekoppelten Rezeptor (GPCR)-Signalweiterleitung hindeutet. Die präsentierten Daten erlauben eine Beurteilung der Pathogenität einer Variante und liefern wertvolle Einsichten in einen möglichen Mechanismus, der den GNB1-bezogenen Störungen zugrunde liegen könnte (Lohmann, …, Baumann et al., 2017).

Pathogene Varianten im *VAC14 component of PIKFYVE complex* (*VAC14*)-Gen wurden als eine Ursache für eine in der Kindheit einsetzende komplexe Dystonie identifiziert. VAC14 ist ein Gerüstprotein, das für die Regulation von Phosphatidylinositol 3,5-Bisphosphat ($PI(3,5)P_2$) relevant ist und von dem angenommen wird, dass es Homodimere bildet. Wir haben bei einem Dystonie-Patienten mit plötzlich einsetzender progressiver Dystonie zwei neue, seltene, heterozygote Varianten des *VAC14*-Gens identifiziert, die in der vorhergesagten Homodimerisierungsdomäne liegen. Um den Einfluss der Varianten und einer weiteren kürzlich publizierten Variante auf die Homodimerisierung zu untersuchen, wurden die VAC14 Varianten ektopisch in einer Neuroblastomzelllinie überexprimiert. Eine verstärkte VAC14-Homodimerbildung wurde als Folge von zwei Varianten im Vergleich zum Wildtyp beobachtet. Im Gegensatz zu früheren Berichten wurden in den Fibroblasten des Patienten keine vergrößerten Vakuolen festgestellt, was möglicherweise auf einen kompensatorischen Effekt der von der Mutter vererbten milderen zweiten Variante zurückzuführen ist. Die funktionellen Ergebnisse liefern erste Hinweise auf eine verstärkte VAC14-Homodimerisierung als möglichen Krankheitsmechanismus (Baumann et al., *in Vorbereitung*).

Eine der prominentesten monogenen Subtypen der isolierten Dystonien wird durch pathogene Varianten im *THAP domain containing 1 THAP1*-Gen verursacht (DYT-THAP1). Interessanterweise zeigt eine große Anzahl (~50%) der Träger einer pathogenen *THAP1*-Mutation keine dystonen Symptome, ein Phänomen, das als reduzierte Penetranz bekannt ist. Bisher liegen nur wenige Daten über die genetischen Faktoren vor, die einer reduzierten Penetranz oder der Rolle von THAP1 während der frühen menschlichen Gehirnentwicklung zugrunde liegen. Im ersten Teil wurde eine systematische Literaturrecherche durchgeführt, um die publizierte phänotypische und demographische Literatur von DYT-THAP1-Patienten im Zeitraum zwischen Dezember 2018 bis März 2020 zusammenzutragen (Lange, Junker, Löns, Baumann et al., *in Vorbereitung*). Die gesammelten Daten wurden in die Datenbank der *Movement Disorder Society Genetic mutation database* (MDSGene; https://www.mdsgene.org) als Ressource für Kliniker und Genetiker aufgenommen. Im zweiten Teil wurde ein zelluläres Modell mit humanen iPSC-abgeleiteten kortikalen Neuronen von zehn *THAP1*-Mutationsträgern (fünf klinisch betroffene und fünf klinisch nicht betroffene Personen) erstellt, die drei verschiedene pathogene Varianten trugen, um die Rolle des mutierten THAP1-Proteins während der frühen Hirnentwicklung zu untersuchen. Zwanzig iPSC-Linien wurden erfolgreich in Bezug auf die Abwesenheit des Sendai-Virus, die Genomintegrität, die Expression von Pluripotenzmarkern und das Potenzial zur Differenzierung in alle drei Keimblätter charakterisiert (Baumann et al., 2018). Als nächstes wurden achtzehn iPSC-Linien von *THAP1*-Mutationsträgern und elf Kontroll-iPSC-Linien in kortikale Neuronen differenziert. Am Tag 44 der kortikalen Differenzierung wurde eine globale Transkriptomanalyse durchgeführt, um die Rolle von mutiertem THAP1 in frühen Entwicklungsstadien zu untersuchen. Die transkriptomweite Untersuchung ergab eine Penetranz-bezogene Expression, bei der sich im Vergleich zu den Kontrollen ausschließlich bei betroffenen, nicht aber bei gesunden, THAP1-Mutationsträgern Unterschiede in der Expression zeigten. Darüber hinaus war die Genexpression des Krankheitsgens *THAP1* in reifenden kortikalen Neuronen mit dem Krankheitsstatus verknüpft, da bei betroffenen Trägern im Vergleich zu nicht-betroffenen *THAP1*-Mutationsträgern eine Hochregulation beobachtet wurde. Eine explorative Datenanalyse deutete auf eine Akkumulation von Genen hin, die mit gestörten kortikalen Netzwerken (*FOXP2*, *DCLK1*, *KIF26B*, *CNTN5*, *CNTNAP3B*, *PCDH19* und *UNC5C*) zusammenhängen und somit die Pathogenese von DYT-THAP1 Patienten beeinflussen können (Baumann et al., *in Vorbereitung*).
Zusammengenommen liefern die in dieser Arbeit vorgestellten Ergebnisse molekulare Einblicke in mögliche Pathomechanismen, die verschiedenen Formen genetischer Dystonien zugrunde liegen. Weitere funktionelle, genetische und klinische Studien sind für die vollständige Aufklärung der pathophysiologischen Grundlagen von Dystonien erforderlich, um damit neue Therapieansätze zu entwickeln und die Patienten bestmöglich zu behandeln.

3. Introduction

3.1 Clinical characteristics of dystonia

Dystonia is the third most common movement disorder worldwide ranking behind Parkinson's disease and essential tremor. The term 'dystonia' was coined by the German neurologist Hermann Oppenheim in 1911 (Oppenheim, 1911). Dystonia is characterized by disabling intermittent or sustained contraction of muscles leading to abnormal postures, movements, or both (Albanese et al., 2013). Dystonic movements are often worsened by voluntary action associated with overflow muscle activation. The prevalence of dystonia is estimated to be approximately 16:100,000, while stratification based on dystonia subtype varies between epidemiological reports (Steeves et al., 2012). Considering that many dystonia patients remain undiagnosed, prevalence rates in the general population could be markedly higher (Das et al., 2007; Müller et al., 2002b). Dystonia is as yet incurable with a need for novel treatment options. To manage dystonia, symptomatic treatment of the affected muscle groups by botulinum toxin or deep brain stimulation surgery can alleviate symptoms; however, both treatment options are not always of lasting effect. Oral medications such as anticholinergic agents that block muscarinic acetylcholine receptors in the basal ganglia, dopaminergic drugs which suppress dopaminergic transmission in the basal ganglia, and gamma-aminobutyric acid (GABA) drugs which amplify GABA transmission have shown a transient benefit when tailored to the individual patient and dystonia subtype (Jinnah and Factor, 2015). For many years, the aetiology of dystonia was hypothesized to reside in the basal ganglia due to some cases that presented with discrete lesions in this brain area (Bhatia and Marsden, 1994). However, patients with isolated dystonia do not show neuropathological lesions or neurodegeneration (Prudente et al., 2013; Sharma, 2019). Voxel-based morphometry of more than 300 primary dystonia patient brains in several independent reports disclosed aberrant white and grey matter volumes in the basal ganglia, cortico-striato-pallido-thalamic and cerebello-thalamo-cortical axes (Ramdhani and Simonyan, 2013). Nowadays, dystonia is widely regarded as a network disorder comprising defects in neural inhibitory circuits, sensorimotor integration, and maladaptive plasticity (Gallea et al., 2018; Gilbertson et al., 2019; Jinnah et al., 2017; Schirinzi et al., 2018).

In general, dystonias encompass a broad clinical spectrum with heterogeneous causes and clinical manifestations (Balint et al., 2018). Currently, dystonia is categorised based on clinical and etiological characteristics (Albanese et al., 2013). Clinically, dystonia subtypes are stratified by the co-occurrence of other neurological or systemic manifestations (isolated, combined, or complex presentations), age at onset (infancy: birth - 2 years, childhood: 3 - 12 years, adolescence: 13 - 20 years, early adulthood: 21 - 40 years, and late adulthood: >40 years), body distribution (focal, segmental, multifocal, generalized with and without leg involvement, or hemidystonia), and temporal pattern (static or progressive disease course, persistent, action-specific, diurnal, or paroxysmal). Second, dystonia subtypes are categorised based on the aetiology such as nervous system pathology (degeneration, structural lesions, or no obvious pathology), or whether no cause was identified (idiopathic). Lastly, there is a subset of dystonia patients carrying pathogenic variants in genes causing dystonia (autosomal dominant, recessive, X-linked recessive, or mitochondrial) (Albanese et al., 2013). The genetic forms of dystonia will be described in the following section.

3.2 Genetics of dystonia

Historically, hereditary forms of dystonia (DYT) were linked to chromosomal regions and assigned DYT loci with ascending numerals in chronological order of discovery (DYT1, DYT2, and so forth). Due to accumulating inconsistencies in this nomenclature, such as erroneous and duplicate findings and unconfirmed genetic loci, a novel nomenclature was proposed by the International Parkinson's Disease and Movement Disorder Society Task Force on the Nomenclature of Genetic Movement Disorders. In fulfilment of its mandate, the task force generated recommendations for the naming system to address these issues of the old convention. The revised novel nomenclature assigns (co-)existing movement disorder prefixes based on the predominant manifestation coupled with the disease-causing gene (e.g., DYT-THAP1; Table 1).

Table 1 Summary of genes involved in monogenic forms of dystonia following the latest nomenclature recommendations (https://www.mdsgene.org). Modified from Klein et al., 2017 and Marras et al., 2016.

Dystonia subtype	MOI	Short clinical description
		Isolated dystonias
DYT-TOR1A	AD	Childhood or adolescence-onset, generalized
DYT-THAP1	AD	Adolescent-onset, cranial or generalized
DYT-ANO3	AD	Adult-onset, focal or segmental
DYT-GNAL	AD	Adult-onset, focal or segmental
DYT-KMT2B	AD	Early-onset, generalized, mild syndromic features
		Combined dystonias
DYT-PRKRA	AR	Dystonia, mild parkinsonism, inconsistent
DYT-HPCA	AR	Dystonia, infantile seizures
DYT/PARK-GCH1	AD/AR	Childhood-onset dystonia, parkinsonism, dopa-responsive
DYT/PARK-TH	AR	Infancy-childhood onset dystonia, parkinsonism, dopa-responsive
DYT/PARK-SPR	AR	Dystonia, parkinsonism
DYT/PARK-ATP1A3	A	Rapid-onset, chorea later in life
DYT/PARK-TAF1	XR	With parkinsonism, neurodegeneration
DYT-SGCE	AD	Dystonia, myoclonus
CHOR/DYT-ADCY5	AD	Facial dyskinesias, occasional myoclonus
		Complex dystonias
DYT/CHOR-HPRT	XR	Dystonia, chorea, occasionally ballism, Lesch-Nyhan syndrome
DYT/CHOR-ACAT1	AR	Dystonia and chorea
DYT/CHOR-GCDH	AR	Dystonia and chorea, later parkinsonism
DYT/CHOR-MUT	AR	Dystonia and chorea, occasionally ataxia
DYT/CHOR-PCCA/PCCB	AR	Dystonia, occasionally chorea
NBIA/DYT-DCAF17	AR	Woodhouse-Sakati syndrome Iron accumulation
DYT-PARK-SLC30A10	AR	Dystonia, polycythaemia, and liver cirrhosis
DYT/PARK-QDPR	AR	Dystonia, parkinsonism
DYT/PARK-SLC6A3	AR	Dystonia and parkinsonism (typically infantile-onset, atypical cases with juvenile-onset exist), occasionally chorea in infancy
NBIA/DYT-PANK2	AR	Pantothenate kinase-associated neurodegeneration
NBIA/DYT/PARK-PLA2G6	AR	Dystonia, parkinsonism, cognitive decline, pyramidal signs, psychiatric symptoms
DYT-ATP7B	AR	Wilson disease, dystonia, occasionally parkinsonism, and/or chorea
DYT-SLC19A3	AR	Dystonia, parkinsonism (mainly rigidity), occasionally ataxia, chorea
DYT-TIMM8A	X	Dystonia, Mohr-Tranebjaerg syndrome
DYT-mt-ND6	Mit	Dystonia, Leber's hereditary optic neuropathy
NBIA/DYT/PARK-CP	AR	Dystonia, ataxia, chorea, parkinsonism, tremor, iron accumulation
DYT-TUBB4A	AD	Spasmodic dysphonia is most common dystonic presentation, alternative phenotype hypomyelinating leukodystrophy

MOI = Mode of inheritance, AD = Autosomal dominant, AR = Autosomal recessive, X = X-linked, Mit = Mitochondrial, PARK = Parkinson's disease, CHOR = Chorea, NBIA = Neurodegeneration with brain iron accumulation

In 2016, the list of approved disease-causes for isolated, combined, and complex dystonias comprised 28 genetic loci (Marras et al., 2016). The genetic diagnosis of dystonia is established by single-gene testing, gene panels, and, if available, whole-exome or whole-genome sequencing. Multigene panels include a range of customizable genes and deletion/duplication analyses and are cost-effective with the limitation of a known set of genes that may not identify the underlying genetic cause of the dystonia in a given patient. If multigene panels fail to elucidate the genetic cause, comprehensive genomic testing (whole exome and whole genome sequencing) of the patient, or a parent-offspring trio, may identify genes or complex traits as soon as novel causative genes have been established. As a rough approximation, a recent study showed that the detection rate for diagnostic exome sequencing of young patients (average age at onset 13.7 years) with isolated and combined dystonia can be up to 20% (Powis et al., 2019). In addition to dominant, recessive, X-linked recessive, and mitochondrial causes of dystonia, several susceptibly genes have been reported that often do not reach the threshold to be disease-causing and therefore are not assigned a DYT designation (Marras et al., 2016).

This thesis comprises four functional projects of different genetic forms of dystonia. In the following sections, these dystonia subtypes will be introduced (sections 3.2.1 - 3.2.4). The main project of this thesis focused on DYT-THAP1 dystonia (section 3.2.4) and the question why some mutation carriers remain clinically unaffected despite carrying a pathogenic mutation. Therefore, an introduction into reduced penetrance in dystonia will be provided in section 3.3. A general introduction into the known cellular mechanisms of genetic dystonias will be given in section 3.4 to provide a basis for the interpretation of the candidate genes from the ribonucleic acid sequencing (RNA-seq) analysis.

3.2.1 DYT-KMT2B dystonia

Variants in the lysine methyltransferase 2B (*KMT2B*) gene were first described as a cause of monogenic complex childhood-onset progressive generalized dystonia in late 2016 (Zech et al., 2016) and early 2017 (Meyer et al., 2017) by two independent groups. The dystonia encompasses a wide variety of clinical presentations, often severely impacting the patients' quality of life. Clinical manifestations include limb dystonia, often leading to gait disturbances as well as a range of non-dystonia symptoms such as dysmorphic features, dysarthria, and short stature in some patients. To date, around 70% of individuals harbour a *de-novo* pathogenic variant. DYT-KMT2B is emerging as a prevalent genetic form of dystonia with frequencies first estimated to be 10% (Zech et al., 2016) (3/30 cases) in a group of generalized dystonia patients and higher rates in cases with childhood-onset dystonia from 21.5% (Carecchio et al., 2019) (14/65 cases) to 38% (Meyer et al., 2017) (13/34 cases). The majority of DYT-KMT2B patients respond well to deep brain stimulation with good to excellent scores (Lange et al., *in preparation*). *KMT2B* is located on chromosome 19q13.12 comprising 37 exons and encodes an epigenetic regulator (histone H3 lysine 4 specific N-methyltransferase), which plays an essential role in transcriptional control and posttranslational modification of histones during brain development (Shen et al., 2014).

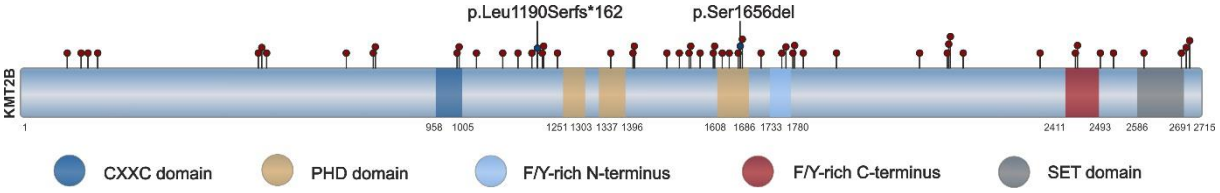

Figure 1 Schematic illustration of the lysine methyltransferase 2B (KMT2B) protein with predicted functional domains and published pathogenic variants (red pins) linked to dystonia. Pathogenic protein alterations investigated in this work are indicated by dark blue pins and labelled above. Numbers of amino acid residues are shown. Domain predictions were derived from the Pfam database (https://pfam.xfam.org) CXXC = CXXC-type zinc finger domain conferring DNA binding capability, PHD = plant homeodomain, F/Y-rich C/N-terminus = phenylalanine/ tyrosine-rich regions, SET = Su(var)3-9, enhancer-of-zeste, trithorax domain with methyltransferase activity.

KMT2B contains a CXXC-type zinc finger (CXXC) domain, three plant homeodomain (PHD) motifs and a Su(var)3-9, enhancer-of-zeste, trithorax (SET) domain, all of which are important for epigenetic regulation (Figure 1) (Qian and Zhou, 2006; Xu et al., 2018). The FY-rich N-terminus contains poorly characterized regions with a high content of phenylalanine and tyrosine that may form inter-domain interactions with the FY-rich C-terminus and SET domains such as in its paralog protein mixed-lineage leukaemia (MLL) (Hsieh et al., 2003). Most of the pathogenic variants have been reported in a single patient. In contrast, only two pathogenic variants were recurrently detected (p.Lys553Glnfs*46 and p.Arg1705Gln; Lange et al., *in preparation*). Intrafamilial variability, as well as reduced penetrance, have been reported (Carecchio et al., 2019; Meyer et al., 2017). Loss of KMT2B leads to dysregulated transcription program during the transdifferentiation of murine fibroblasts into neurons (Barbagiovanni et al., 2018). Meyer at al. demonstrated a link between dysfunctional KMT2B and a reduction THAP Domain Containing 1 (THAP1) and torsin family 1 member A (TOR1A) gene and protein expression, two genes known to cause dystonia, and provided a first mechanistic link between DYT-KMT2B and other forms of dystonia (Meyer et al., 2017). KMT2B has further been shown to coordinate early steps of differentiation in murine embryonic stem cells (Lubitz et al., 2007). More than 50 different pathogenic variants in *KMT2B* have been reported without apparent clustering in a specific predicted functional domain, however many are truncating variants. In keeping with the complex phenotype of the disorder, KMT2B is ubiquitously expressed (Uhlén et al., 2015). Intriguingly, several of the *KMT2* gene family members were already associated with numerous neurodevelopmental disorders listed in the Online Mendelian Inheritance in Man (OMIM) registry such as Wiedemann-Steiner syndrome (*KMT2A*; OMIM#605130) (Jones et al., 2012), Kleefstra syndrome 1 (*KMT2C*; OMIM#617768) (Kleefstra et al., 2012), and Kabuki syndrome 1 (*KMT2D*; OMIM#147920) (Ng et al., 2010).

3.2.2 GNB1-related disorders

Germline pathogenic variants in guanine nucleotide-binding protein beta 1 (*GNB1*) gene have been linked to a neurodevelopmental disorder accompanied by a heterogeneous group of additional clinical features such as hypotonia, epilepsy, dysmorphic features, dystonia and others (Endo et al., 2019; Petrovski et al., 2016). To date, a total of 50 cases have been published, of which eight patients exhibited signs of dystonia, including a patient from us (Steinrücke et al., 2016). The *GNB1* gene is located on the short arm of chromosome 1p.36.33 with 12 exons (NM_002074.5) and encodes the guanine nucleotide-binding protein subunit beta 1 ($G\beta_1$). Thus far, several pathogenic variants have been reported, many of which exhibited a specific clustering in exon 6 and 7 between the first and second WD40 repeats (Figure 2A). WD40 repeats typically form circularised beta-propeller structures to facilitate protein-protein or protein-deoxyribonucleic acid (DNA) interaction (Xu and Min, 2011). One recurrent substitution is located on position 80 of the $G\beta_1$ protein and occurred either as p.Ile80Thr, or p.Ile80Asn (Hemati et al.,

2018). Notably, molecular *in-silico* modelling of the tertiary protein structure of the Gαβγ-heterotrimer revealed that most pathogenic variants are located at the interface between Gα and Gβ-subunits raising the hypothesis of altered G-protein coupled receptor (GPCR) signalling as a mechanistic consequence (Lohmann et al., 2017; Petrovski et al., 2016). In fact, altered binding capabilities toward Gα-subunit and cytokine-independent growth have previously been established for the recurrent pathogenic p.Ile80Thr substitution (Yoda et al., 2015). Furthermore, five variants show an overlap with amino acid residues mutated across multiple human tumors (amino acid positions 57, 78, 80, 89, and 101), which cause cytokine-independent GPCR signal activation (Yoda et al., 2015).

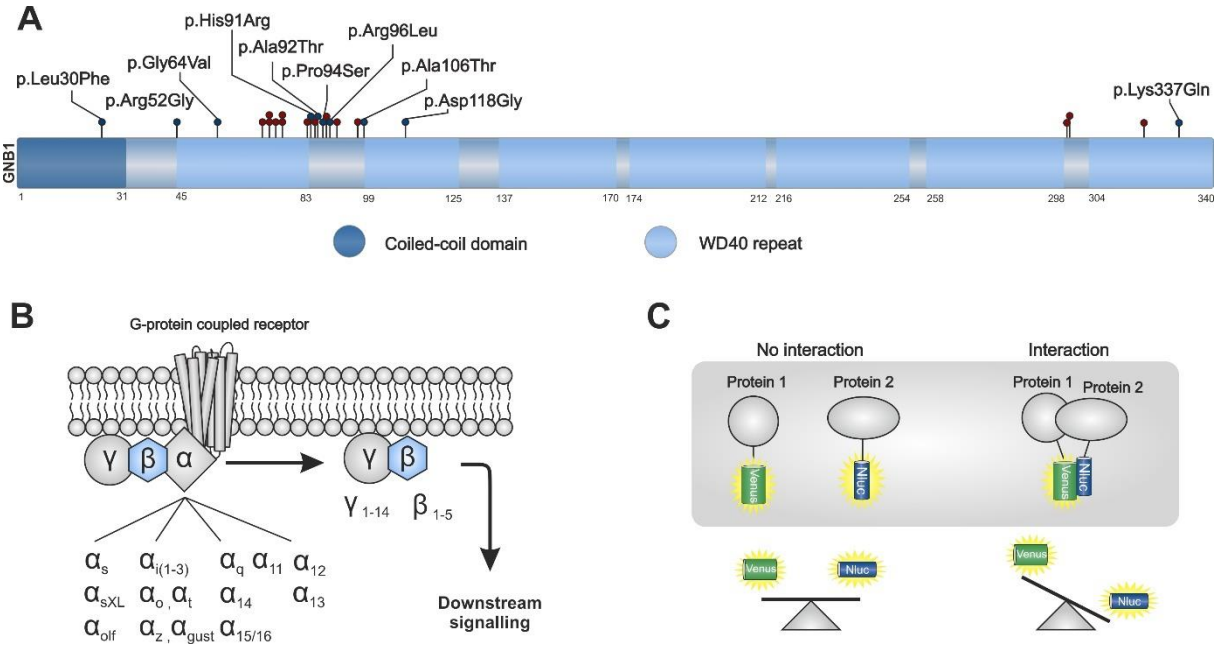

Figure 2 Characteristics of the Guanine nucleotide-binding protein beta 1 (GNB1; $G\beta_1$) protein and G-protein coupled receptor signalling. (A) Schematic illustration of the $G\beta_1$ protein with predicted functional domains and published pathogenic germline variants (red pins). Protein alterations investigated in this work are indicated by dark blue pins and labelled above. Numbers of amino acid residues are shown. Domain predictions were derived from the Pfam database (https://pfam.xfam.org). (B) Diversity of G-protein coupled receptors subunits modified from Wootten et al., 2018. (C) Principle of bioluminescence resonance energy transfer (BRET) in two proteins fused with bioluminescent Venus (acceptor) or Nluc (donor). Coiled-coil domain = structural protein motif of two or more coiled α-helices, WD40 repeat = tryptophan-aspartic-acid domains that can form circular protein structures.

The $G\beta_1$ protein forms heterotrimers with a multitude of Gα and Gγ proteins for GPCR signal propagation, depending on the tissue (Figure 2B) (Wootten et al., 2018). For instance, GPCR signalling via $G\alpha_{olf}$ and $G\gamma_7$ is crucial in striatal dopaminergic neurons - subtypes of neurons that are relevant to a broad spectrum of movement disorders (Corvol et al., 2004; Girault, 2012; Schwindinger et al., 2010). In this thesis, GPCR signalling of $G\alpha_{olf}\beta_1\gamma_7$ was investigated by the bioluminescent resonance energy transfer (BRET) technique, which is similar to Förster resonance energy transfer (FRET) (Figure 2C). Instead of chromophores in FRET, BRET employs two bioluminescent luciferases (Dimri et al., 2016). Co-expression of two fusion proteins that interact can result in the transfer of energy from the donor (Nluc) to the acceptor (Venus), which indicates an interaction of these proteins within a proximity of 10 - 100 Å (Dimri et al., 2016; Lohmann et al., 2017). This technique is widely used to study GPCR signalling pathway components (Salahpour et al., 2012).

3.2.3 VAC14-related disorders

VAC14 appeared as a novel recessive disease-causing gene for a neurological disease characterized by sudden childhood-onset progressive dystonia in the year 2016 (Lenk et al., 2016). The first report described two siblings of a family with compound-heterozygous *VAC14* variants with a progressive movement disorder and striatal abnormalities (Lenk et al., 2016). Since then, a total of eleven patients with *VAC14* variants from eight families have been reported. Notably, the initial psychomotor development of the reported patients is normal followed by an abrupt onset, which, in severe cases, can lead to an acute neurological deterioration with vomiting followed by death (Stutterd et al., 2017). Neuropathological findings of post-mortem brains revealed neuronal vacuolation and neurodegeneration in the putamen, caudate nucleus, and globus pallidus (Stutterd et al., 2017). Moreover, the occurrence of enlarged vacuoles in a subset of fibroblasts has been reported, which was rescued by transient transfection with VAC14 wild type protein (de Gusmao et al., 2019; Lenk et al., 2016). The age at onset is early ranging from 1.5 to 13 years with initial symptoms occurring in the limbs or as slurred speech (Lenk et al., 2016; Lyon et al., 2019; Stutterd et al., 2017; Taghavi et al., 2018). The disorder is disabling and often accompanied by a wide spectrum of clinical presentations, including hypersalivation, impaired speech, and a dystonic gait (see Table 8). So far, four compound-heterozygous and four homozygous variants were reported with one recurrent pathogenic variant (c.1271G>T; p.Trp424Leu) in two families (Lenk et al., 2016; Stutterd et al., 2017). Many pathogenic variants cluster in the FIG4 binding domain with only a few deviating from this signature.

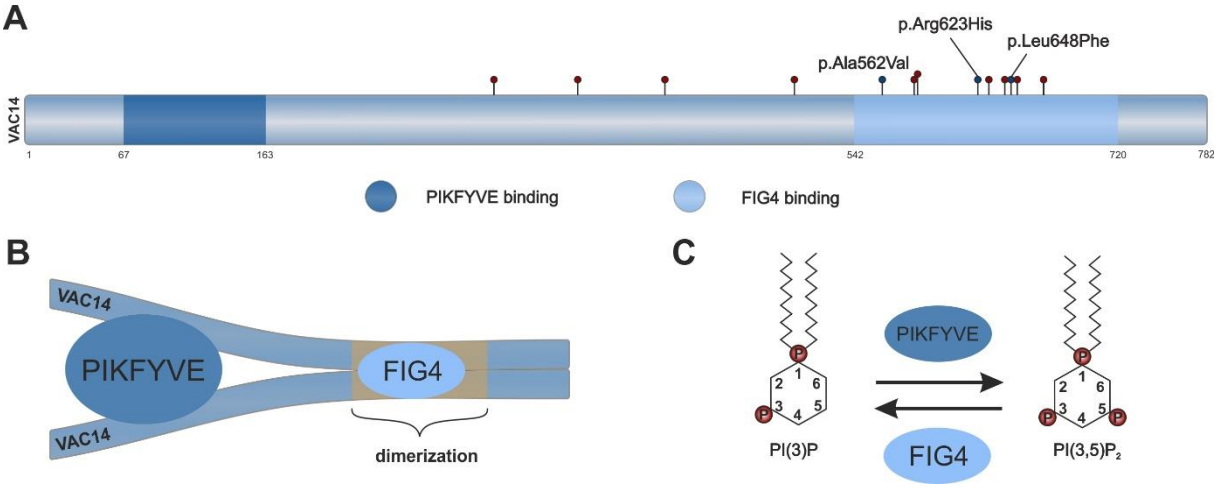

Figure 3 Characteristics of the VAC14 protein and its interacting proteins. (A) Schematic illustration of the VAC14 protein with predicted functional domains and published pathogenic variants (red pins) linked to dystonia. Protein alterations investigated in this work are indicated by dark blue pins and labelled above. Numbers of amino acid residues are shown. Domain predictions were derived from the Pfam database (https://pfam.xfam.org). (B) Model of the VAC14 protein complex. (C) Regulation of phosphatidylinositol 3,5-bisphosphate ($PI(3,5)P_2$) turnover. PIKFYVE binding = element that binds phosphoinositide kinase, FYVE-type zinc finger containing (PIKFYVE), FIG4 binding = domain that binds FIG4 phosphoinositide 5-phosphatase (FIG4).

The *VAC14* gene is located on chromosome 16q22 with 19 exons encoding the scaffold protein VAC14, which forms a ternary complex with FYVE-type zinc finger containing (PIKFYVE) and FIG4 phosphoinositide 5-phosphatase (FIG4) (Figure 3B). The trimolecular complex regulates the synthesis and turnover of the signalling lipid phosphatidylinositol 3,5-bisphosphate ($PI(3,5)P_2$), in which FIG4 and PIKFYVE have opposing functions (Figure 3 C) (Jin et al., 2008). Of note, upon physical interaction of FIG4 and PIKFYVE in the ternary complex, the kinase activity of PIKFYVE is enhanced resulting in the synthesis of $PI(3,5)P_2$. Deficiency in $PI(3,5)P_2$ can cause impaired endosomal membrane properties and result in enlarged cytoplasmic vesicles (Lines et al., 2017). Pathogenic variants in FIG4 are known to cause several

neurological disorders such as Yunis-Varon syndrome (OMIM#216340) (Yunis and Varón, 1980), Charcot-Marie-Tooth disease type 4J (OMIM#611228) (Lenk et al., 2011), and amyotrophic lateral sclerosis 11 (OMIM#612577)(Chow et al., 2009). Haploinsufficiency of *PIKFYVE*, the second member of the ternary protein complex, is known to cause corneal fleck dystrophy (OMIM#121850)(Li et al., 2005). While all pathogenic variants causative for Yunis-Varon syndrome thus far were located in the *FIG4* gene, one first report has associated mutations in *VAC14* with the clinical spectrum of the Yunis-Varon syndrome (Lines et al., 2017). Loss of *Vac14* in mice resulted in accumulation of enlarged cytoplasmic vacuoles and neurodegeneration recapitulating pathological findings in humans (Stutterd et al., 2017; Zhang et al., 2007).

3.2.4 DYT-THAP1 dystonia

A dystonia locus on chromosome 8p21 - 8q22 was first described in a genetically isolated Amish-Mennonite population with autosomal dominantly inherited dystonia with reduced penetrance in 1997 (Almasy et al., 1997). After narrowing down the locus in other affected Mennonite families (Saunders-Pullman et al., 2007), variants *THAP1* were identified as a cause of autosomal-dominant isolated adolescence-onset dystonia (Bressman et al., 2009; Fuchs et al., 2009). Since this discovery, more than 230 dystonia patients from various ethnic groups were identified to harbour pathogenic *THAP1* variants. Cranial involvement is much more common than in other hereditary forms of dystonia, especially in muscles of the face, neck, larynx, and vocal cords. The mean age at onset in DYT-THAP1 patients is 15 years, and the first symptoms typically occur in the arms, the neck, or cranial muscles. Over time, contractions often spread into other body regions. Co-morbidities such as global developmental delay, cognitive impairment, or short stature are rare in DYT-THAP1. Reduced penetrance is observed in ~50% in *THAP1* mutation carriers (Saunders-Pullman et al., 2007; Lange et al., *in preparation*) with no molecular explanation yet reported. Mutations in *THAP1* underlie a small proportion of dystonia patients with reported frequencies of 0.87%, 1%, and 2.5% investigated in 231 Chinese (Song et al., 2011), 160 Caucasian (Djarmati et al., 2009), or 362 (Houlden et al., 2010) Caucasian primary dystonia patients, respectively. The *THAP1* gene is located close to the centromere and encodes the transcription factor THAP1 with 213 amino acid residues (Figure 4). Pathogenic variants are dispersed across the coding region of the gene, and more than half of the variants are located in the DNA-binding domain.

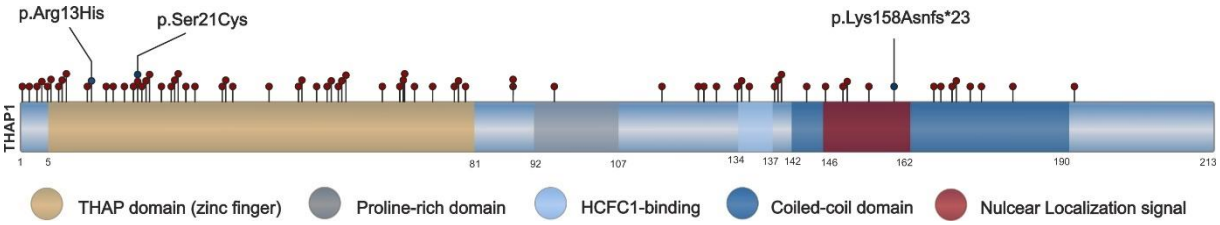

Figure 4 Schematic illustration of the THAP1 protein with predicted functional domains and published pathogenic variants (red pins) linked to dystonia. Pathogenic protein alterations investigated in this work are indicated by dark blue pins and labelled above. Numbers of amino acid residues are shown. Domain predictions were derived from the Pfam database (https://pfam.xfam.org) and the nuclear localization mapper database (http://nls-mapper.iab.keio.ac.jp/). THAP domain (zinc finger) = zinc-dependent domain that binds DNA in a sequence-specific manner, Proline-rich domain = relevant for spliceosomal assembly, HCFC1-binding = domain binding transcription factor HCFC1, Coiled-coil domain = structural protein motif of two or more coiled α-helices, Nuclear localization signal = nuclear transport signal.

The protein is characterized by a conserved DNA- binding domain (family-designating THAP domain) at its N-terminus (amino acids 5 - 81), a central proline-rich region (amino acids 92 - 107), a HCFC1-binding domain (amino acids 134-137; (Mazars et al., 2010)), and a coiled-coil domain (amino acids 142 - 190) including a short proposed nuclear localization signal (amino acids 146 - 162) at its C-terminus. As a

transcription factor, THAP1 gains entry into the nucleus through binding importin α. Together, these proteins move through the nuclear pore complex that spans the nuclear envelope (Kosugi et al., 2009). The three-dimensional structure of THAP1 has been resolved by nuclear magnetic resonance analysis showing that the two antiparallel ß-sheets of the THAP domain are inserted into the DNA major groove (Campagne et al., 2010). The protein contains an evolutionary conserved THAP1 domain which confers DNA binding via a zinc-coordination C2CH signature ($DysX_{2\text{-}4}CysX_{35\text{-}53}CysX_2His$), binds an 11-nucleotide sequence within the promotor region of target genes, and is characteristic for all twelve human THAP proteins (THAP0 - THAP11) (Clouaire et al., 2005; Roussigne et al., 2003). THAP1 has been demonstrated to autoregulate its own expression via binding of its own promotor, and pathogenic variants in the THAP domain (e.g., Arg13His) are associated with decreased DNA binding and elevated expression (Erogullari et al., 2014). Moreover, THAP1 was shown to repress the expression of the dystonia gene *TOR1A* (Gavarini et al., 2010; Kaiser et al., 2010). Proteomic and transcriptomic analysis revealed that THAP1 protein expression is widespread and highest in brain, testis, and blood and RNA expression in bone marrow, blood, endocrine tissues, and brain (Uhlén et al., 2015; http://www.proteinatlas.org). Transgenic mice carrying a *Thap1* null allele are early embryonic lethal, indicating an important role of THAP1 during development (Ruiz et al., 2015). Furthermore, $Thap1^{C54Y/+}$ and $Thap1^{+/-}$ mice exhibited a reduced number of glutamatergic and GABAergic projection neurons in the cerebellum compared to control mice (Ruiz et al., 2015). Immunoblot analysis of mouse brain tissue showed the highest THAP1 protein expression in prenatal whole brain lysates (embryonic day 13) and declined with age according to different brain regions (Gavarini et al., 2010). These results are consistent with *Thap1* gene expression analysis in rats indicating the highest levels during early development (Zhao et al., 2013). Expanding the role of THAP1 during development, Aguilo et al. investigated mouse embryonic stem cells in which Thap1 was shown to regulate pluripotency-associated genes and led to an enhanced rate of cell death and partial failure of ectodermal lineage commitment (Aguilo et al., 2017). Moreover, in the same cell model, a possible connection to the epigenetic regulator O-GlcNAc transferase (OGT) was established (Mazars et al., 2010). Furthermore, retroviral-mediated gene transfer in human endothelial cells inhibited G1/S cell-cycle progression (Cayrol et al., 2007). This finding is in line with *in-vitro* experiments that linked THAP1 to cell-cycle regulator Host cell factor 1 (HCF-1) with possible transcriptional regulation of S-phase regulatory ribonucleotide reductase catalytic subunit M1 (RRM1) in primary human endothelial cells (Mazars et al., 2010). Neuropathological assessment of two DYT-THAP1 post-mortem brains confirmed that carriers did not show discrete anatomically abnormal regions, neurodegeneration, or lesions (Paudel et al., 2016). In conclusion, despite tremendous effort, the role of THAP1 during DYT-THAP1 disease progression remains to be elucidated and highlights the need to investigate pathogenic *THAP1* variants in a neuronal cellular model.

3.3 Reduced penetrance in genetic dystonias

In genetics, the penetrance of a disease-causing variant defines the proportion of mutation carriers that show clinical symptoms. Thus, 'reduced penetrance' refers to the fact that a certain proportion of individuals is clinically healthy despite carrying a disease-causing variant. In addition, the severity of clinical symptoms in manifesting mutation carriers can vary, showing 'variable expressivity'. Reduced penetrance and variable expressivity are important when interpreting a genetic variant, because disease-causing variants in unaffected individuals may occur. Intriguingly, all monogenic autosomal dominantly inherited dystonias (currently except for DYT-TUBB4A; Tubulin beta 4A class IVA) exhibit reduced penetrance and all inherited subtypes of dystonia show some degree of variable expressivity. For example, the most common inherited form of dystonia is caused by pathogenic variants in *TOR1A*, for which penetrance is reduced to 30% (Risch et al., 2007). These circumstances greatly impact on genetic counselling as offspring of an affected or unaffected *TOR1A* mutation carrier have a 50% chance of inheriting the pathogenic variant and, if inherited, a 30% of developing symptoms during their

lifespan. Clinical symptoms in DYT-TOR1A can range from mild focal dystonia to severe dystonic storm even within families (Opal et al., 2002). Elucidating the causes of reduced penetrance may reveal mechanisms of endogenous disease protection and has important implications in genetic counselling and personalized medicine. In DYT-TOR1A it has been successfully shown that a common polymorphism D216H (rs1801968) acts as a genetic modifier of pathogenic GAG deletion by reducing disease penetrance from ~30% to ~3% (Kamm et al., 2008; Risch et al., 2007). Furthermore, pathogenic variants in a number of dystonia genes usually cause disease-onset during childhood to adolescence (e.g., *TOR1A, THAP1, and KMT2B;* https://www.mdsgene.org), indicating a time-window of susceptibility. Another mode of disease protection was discovered in the epsilon-sarcoglycan (SGCE) gene causing myoclonus-dystonia (DYT-SGCE). The gene is transcribed only from the paternal allele while the maternal allele is imprinted, and thus, mutations on the maternal allele are not penetrant (Müller et al., 2002a). Intriguingly, neuroimaging of DYT-THAP1 patients revealed that cortico-striato-pallido-thalamo brain circuits are disrupted in manifesting carriers compared to non-manifesting mutation carriers, suggesting that DYT-THAP1 is a neurodevelopmental circuit disorder (Niethammer et al., 2011). However, the underlying molecular basis of this observation is yet to be identified. Recent progress in genetic dystonia indicates that multiple risk factors such as genetic triggers, brain network organisation, and the presence of environmental factors may collectively lead to disease manifestation (Siokas et al., 2019).

3.4 Converging pathways of genetic dystonias

The failure to detect neuropathological lesions or neurodegeneration in dystonia patients makes it very difficult to identify a single brain region underlying the pathogenesis of dystonia (Prudente et al., 2013; Sharma, 2019). However, multiple lines of evidence suggest abnormal connectivity between the cortex, striatum, thalamus, and cerebellum as well as altered plasticity in dystonia (Gilbertson et al., 2019; Jinnah et al., 2017; Schirinzi et al., 2018). In general, information of movement in the human brain is sent to the striatum by cortical layer V. In the striatum, information is processed by the substantia nigra and the globus pallidus. These structures provide feedback to the striatum and pass the information to the thalamus, which sends the signals back to the cortex. This feedback loop is governed by glutamatergic neurons, dopaminergic neurons, cholinergic neurons, and GABAergic neurons (Balint et al., 2018; Breakefield et al., 2008). Looking at the protein function of hereditary causes of dystonia, it is apparent that dopaminergic dysfunction is a common theme. For instance, pathogenic variants in dopa-responsive dystonia genes GTP cyclohydrolase 1 (*GCH1*), tyrosine hydroxylase (*TH*) and sepiapterin reductase (*SPR*) are implicated in reduced levels of dopamine and related metabolites at the presynapse (Bonafé et al., 2001; Ichinose et al., 1994; Lüdecke et al., 1995). In cellular models, G protein subunit alpha L (GNAL) and adenylate cyclase 5 (ADCY5) exert functions at the postsynapse of dopaminergic neurons of the striatum (Balint et al., 2018). $G\alpha_{olf}$, encoded by *GNAL*, is hypothesized to act as a regulator of both adenosine and dopamine signalling via dopamine type 1 receptors (D1R) and adenosine A_{2A} receptors, respectively (Hervé, 2011; Hervé et al., 2001). Both receptors activate ADCY5, which regulates cyclic adenosine monophosphate synthesis in the striatum (Chen, 2012). Furthermore, $G\beta_1$, encoded by *GNB1* (see section 3.2.2), is a binding partner of $G\alpha_{olf}$ implicated in similar signalling cascades. Consistent with these findings, animal knockout models of SGCE revealed dysfunctional dopaminergic plasticity in the striatum, which was reversed by inhibition of adenosine A2A receptor signalling (Maltese et al., 2017). Furthermore, impaired plasticity of the striatum has emerged as a key concept common to all forms of dystonia (Peterson et al., 2010). Changes in cortico-striatal plasticity were also found in DYT-THAP1 and DYT-TOR1A mouse models (Martella et al., 2009; Pisani et al., 2006; Zakirova et al., 2018). Deficits in dopamine type 2 receptor (D2R) binding activity of striatal neurons

caused by a TOR1A-related dystonia mouse model further strengthens the hypothesis of a basal ganglia circuit dysfunction (Napolitano et al., 2010; Yokoi et al., 2011). One hypothesis is that TOR1A in its role as a molecular chaperone inhibits the dopamine transporter via the chaperone calnexin (Free et al., 2007). Consistent with these results, animal models show reduced endoplasmic reticulum (ER) volume, increased ER stress and aberrant trafficking of ER proteins (Chen et al., 2010; Liang et al., 2014; Nery et al., 2011). Along this line, elongation factor 2α signalling as part of ER-stress response is implicated in dysfunctional neuronal long term potentiation and synaptic plasticity linked to dystonia (Di Prisco et al., 2014). Less functional evidence is given for the group of ATPase Na+/K+ transporting subunit alpha 3 (ATP1A3), anoctamin 3 (ANO3) and hippocalcin (HPCA), all linked disturbed ion channel activities often referred to as channelopathies. While both ANO3 and HPCA regulate calcium signalling, their physiological role during pathogenesis in the relevant cell types remains to be elucidated (Charlesworth et al., 2012; Helassa et al., 2017). In summary, cellular studies of genes causative for dystonia are pivotal to elucidate shared molecular pathways. A substantial body of evidence links hereditary dystonias to maladaptive plasticity and sensorimotor integration in the cortico-striato-cerebello axis. This network hypothesis may help to explain why it is difficult to identify a single anatomical region as a pathological locus. However, despite marked advances in the understanding of the genetic basis of dystonia, many facets of the pathogenesis are still poorly understood.

3.5 Aims

The aims of this work were to investigate novel (*KMT2B*, *GNB1,* and *VAC14*) and established (*THAP1*) dystonia-causing genes on the molecular level. Specifically, the following sub-aims were addressed:

KMT2B

Premise: Pathogenic variants in *KMT2B* have recently emerged as a frequent genetic cause of dystonia. KMT2B protein acts as an epigenetic regulator that may alter the gene expression of various genes. We identified two novel variants in *KMT2B* in two severely affected dystonia patients. Available biomaterials were blood and fibroblasts.
Aim: To explore the functional impact of two *KMT2B* variants in blood and fibroblasts in two unrelated dystonia patients

GNB1

Premise: Mutations in *GNB1* can cause a clinically heterogeneous neurodevelopmental disorder, including dystonia in a subset of patients. *GNB1* encodes $G\beta_1$, a protein involved in G-protein coupled receptor signalling. We identified novel missense variants of uncertain significance in *GNB1* in patients with global developmental delay, dystonia, and other symptoms. It was unknown whether these variants influence the functionality of the G-protein receptor signalling.
Aim: To perform molecular cloning of ten *GNB1* missense variants to assess their functionality in a cell-based assay (BRET) and draw conclusions about their pathogenicity

VAC14

Premise: Mutations in *VAC14* have recently been identified as a cause of a form of childhood-onset complex dystonia with a progressive course. VAC14 homodimers are suggested to regulate the turnover of specific phosphoinositide phosphate lipids. We identified a dystonia patient with novel compound heterozygous variants in the predicted homodimerization domain of the *VAC14* gene.
Aim: To assess the impact of *VAC14* variants on the capacity of VAC14 to form homodimers

THAP1

Premise: Pathogenic variants in *THAP1* cause early-onset isolated dystonia (DYT-THAP1). THAP1 acts as a zinc finger transcription factor. Despite the autosomal dominant pattern of inheritance, DYT-THAP1 manifests in only ~50% of all cases, which is termed reduced penetrance. To date, little data is available on the precise action of THAP1 during early human brain development. Moreover, the underlying pathomechanisms of reduced penetrance in DYT-THAP1 remain elusive.

Aims:

-To systematically review published literature of the dystonia gene *THAP1* and other isolated dystonia genes from December 2018 to March 2020 to update the Movement Disorder Society Genetic mutation (MDSGene) database

-To generate and characterize twenty *THAP1* mutant induced pluripotent stem cell (iPSC)-lines derived from affected and unaffected members of three families harbouring three different mutations

-To investigate iPSC-derived cortical neurons as an endogenous cellular disease model to study the expressional consequences of pathogenic *THAP1* mutations by global transcriptome analysis

-To identify penetrance-modifying genes in developing neurons that may lead to reduced penetrance in DYT-THAP1 by global transcriptome analysis

4. Patients, materials, and methods

4.1 Patients

In the present thesis, blood, primary dermal fibroblasts, iPSCs, or iPSC-derived cortical neurons of *KMT2B*, *VAC14*, and *THAP1* mutation carriers and neurologically healthy individuals were investigated. Written informed consent was obtained from all individuals, and the study was approved by the Ethics Committee at the University of Lübeck (ethics approval 04-180, 11-021, and 16-039). All patients were diagnosed based on accepted clinical criteria by movement disorder specialists (Norbert Brüggemann, Alexander Münchau, Christine Klein; Clinic of Neurology, University Hospital Schleswig Holstein, Campus Lübeck).

KMT2B patient

A 31-year-old female patient of Afghan origin was seen in the Clinic of Neurology due to severe progressive dystonia. She was a daughter of consanguineous parents with a negative family history. She developed gait problems at the age of 7 years. These were attributed to two car accidents with unspecified operations of the feet in Afghanistan as well as by exhaustive walks that the patient underwent during her time as a refugee. Neurological examination in Germany at the age of 15 years showed generalized dystonia with prominent involvement of the lower extremities and trunk, initially largely sparing the face and neck. Speech was limited due to dysarthria. Furthermore, dystonic symptoms of the patient continuously worsened, including the development of severe cervical dystonia until the age of 31 years. The remainder of the neurological examinations were unremarkable (Klein et al., 2019).

VAC14 patient

A 26-year old male Caucasian dystonia patient was seen in the Clinic of Neurology due to generalized dystonia, anarthria, incomplete vertical supranuclear palsy, inner ear deafness, and drumstick fingers. After an uneventful pregnancy and delivery, his initial motor development appeared normal, although the patient described frequent falls during childhood. His speech development was delayed, which may be attributed to his inner ear deafness. At the age of 13 years, the patient developed a writer's cramp. Blepharospasm was noted at the age of 23 years. One year later, the symptoms worsened and the patient presented with poor fine motor skills and loss of strength in the hands, anarthria, dysphagia, and hypoglossal palsy with absent swallowing reflexes and hypersalivation. At the age of 26 years, he developed a progressive gait disturbance (toe walking), an overstretching to the back, and cervical dystonia with an intermittent retrocollis. His mother reported periods of increased irritability, accompanied by a weight loss of 10 kg. Brain MRI of the index patient at the age of 29 years showed T2-weighted hypointensities in the globus pallidus and T2-weighted hyperintensities in the striatum.

THAP1 patients

Detailed phenotypic information on members of Family A and Family B were previously published (Djarmati et al., 2009; Zittel et al., 2010). An overview of the clinical status of *THAP1* mutation carriers and controls is provided in Figure 5 and Table 2.

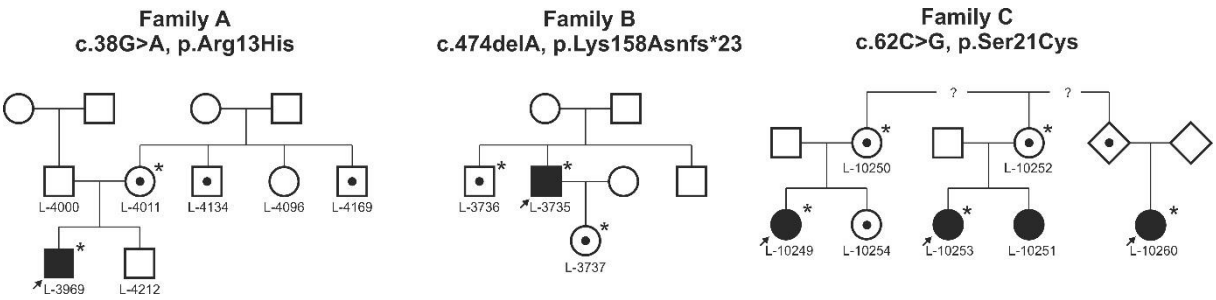

Figure 5 Pedigrees of Families A, B, and C harbouring pathogenic *THAP1* variants. Squares represent male subjects; circles represent female subjects; diamonds represent disguised genders; filled symbols indicate affected *THAP1* mutation carriers; a black dot in the middle of a symbol represents clinically unaffected *THAP1* mutation carriers; the index patient is identified by an arrow pointing towards the symbol; the question mark indicates missing information; the L-identifier is located below a symbol; asterisks indicate that fibroblasts were collected.

Table 2 Demographic and phenotypic information on DYT-THAP1 patients and matched controls. All individuals are of Caucasian ancestry.

Identifier (L-Code/SFC-Code)	Gender	Age at biopsy [years]	Clinical status	Disease-causing *THAP1* mutation (protein alteration)
L-3969	Male	30	Manifesting (Generalized dystonia)	c.38G>A (p.Arg13His)
L-4011	Female	56	Asymptomatic but subtle writer's cramp and mild cervical dystonia	c.38G>A (p.Arg13His)
L-3736	Male	70	Non-manifesting	c.474delA (p.Lys158Asnfs*23)
L-3735*	Male	69	Manifesting (Writer's dystonia followed by laryngeal dystonia)	c.474delA (p.Lys158Asnfs*23)
L-3737	Female	40	Non-manifesting	c.474delA (p.Lys158Asnfs*23)
L-10249	Female	29	Manifesting	c.62C>G (p.Ser21Cys)
L-10250	Female	52	Non-manifesting	c.62C>G (p.Ser21Cys)
L-10252	Female	57	Non-manifesting	c.62C>G (p.Ser21Cys)
L-10253	Female	32	Manifesting	c.62C>G (p.Ser21Cys)
L-10260	Female	44	Manifesting	c.62C>G (p.Ser21Cys)
	30% Male	Mean: 47.9 Range: 29 - 70		
L-5991 (SFC062)	Female	67	Neurologically healthy control	No mutation
L-6534 (SFC065)	Male	65	Neurologically healthy control	No mutation
L-5872 (SFC067)	Male	72	Neurologically healthy control	No mutation
L-6353 (SFC068)	Female	68	Neurologically healthy control	No mutation
L-4993 (SFC084)	Female	63	Neurologically healthy control	No mutation
L-6396 (SFC086)	Female	57	Neurologically healthy control	No mutation
L-6004 (SFC089)	Female	64	Neurologically healthy control	No mutation
L-6069 (SFC156)	Male	75	Neurologically healthy control	No mutation
L-4646 (SFC163)	Male	66	Neurologically healthy control	No mutation
SFC841	Male	36	Neurologically healthy control	No mutation
SFC856	Female	78	Neurologically healthy control	No mutation
	40% Male	Mean: 64.6 Range: 36 - 78		

*also known as L-2257 in Djamarti et al., 2009

4.2 Materials

4.2.1 Solutions

Specification	Components
10x Loading buffer for agarose gels	30% (v/v) Glycerine 70% (v/v) Water 0.25% (w/v) Bromophenol blue
10x TBE buffer	890 mM Tris 890 mM Boric acid 20 mM EDTA Solved in water
10x Tris-buffered saline (TBS)	500 mM Tris-HCl pH =7.5 1.5 M Sodium chloride Solved in water
EDTA-PBS	0.5 µM EDTA solved in 1x PBS
Fibroblast culture medium	DMEM 10% (v/v) Fetal bovine serum 1% (v/v) PenStrep
iPS medium	400 ml Knockout-DMEM 100 ml KnockOut™ Serum Replacement 7.5 ml 1 M HEPES 5 ml Glutamine 5 ml MEM-NEAA 1 ml 50 mM 2-Mercaptoethanol 10 ng/ml human FGF2, freshly added
LB agar plates	LB broth 2% (w/v) Bacto™ Agar Solved in water
Lysis buffer	150 mM Sodium chloride 1 % Nonidet P40 0.1% SDS 50 mM Tris-HCl pH = 7.4 1 tablet cOmplete™ protease Inhibitor per 10 ml
Lysogeny broth (LB) medium	1% (w/v) Bacto™ Tryptone 0.5% (w/v) Bacto™ Yeast Extract 1% Sodium chloride Adjust pH to 7.0 with sodium hydroxide (NaOH) Solved in water
Neural induction medium (KSR)	212.5 ml Knockout DMEM Gibco® 37.5 ml KnockOut™ Serum Replacement 2.5 ml L-glutamine 2.5 ml MEM-NEAA 455 µl 2-Mercaptoethanol 2.5 ml PenStrep
Neural maintenance medium (NMM)	400 ml DMEM/F12 2.5 ml N2 supplement 125 µl Insulin 5 ml L-glutamine 2.5 ml MEM-NEAA 450 µl 2-Mercaptoethanol 240 ml Neurobasal™ 5 ml NeuroCult™

TBS with Tween 20 (TBS-T)	1x TBS 0.1 (v/v) Tween 20

4.2.2 Factors for differentiation

Specification	Stock conc.	Final conc.	Supplier
Ascorbic acid	100 mM in water	0.2 mM	Sigma-Aldrich, St. Louis, United States
Dorsomorphine	1 mM in water	1 µM	Tocris, Bristol, United Kingdom
Recombinant human BDNF	40 ng/µl in 1x PBS	20 ng/ml	Peprotech
Recombinant human FGF2	10 ng/µl in	20 ng/ml	Merck Millipore, Darmstadt, Germany
Recombinant human GDNF	10 ng/µl in 1x PBS	20 ng/ml	Peprotech, Hamburg, Germany
SB-431542	10 mM in DMSO	10 µM	Tocris, Bristol, United Kingdom
Y-27632	10 mM in water	10 µM	Merck Millipore, Darmstadt, Germany

4.2.3 Technical equipment

Specification	Supplier
2100 Bioanalyzer	Agilent Technologies, Santa Clara, United States
3130xL Genetic Analyzer	Applied Biosystems, Waltham, United States
3500xL Genetic Analyzer	Applied Biosystems, Waltham, United States
Bacterial Incubator ST 5042	Heraeus, Hanau, Germany
Biometra T Advanced Thermocycler	Analytik Jena, Jena, Germany
Canon EOS 700D agarose gel visualization	Canon, Ota, Japan
Centrifuge Fresco 21	Thermo Fisher Scientific, Waltham, United States
ChemiDoc MP Imaging System	Bio-Rad Laboratories, Hercules, United States
Confocal Microscope LSM, LSM710	Zeiss, Oberkochen, Germany
Digital Scale 572	Tovatech, Maplewood, United States
Electroporator 2510	Eppendorf, Hamburg, Germany
Eppendorf Plate Centrifuge 5804	Eppendorf, Hamburg, Germany
EPS 500/400 Electrophoresis power supply	GE Healthcare, Chicago, United States
FlexCycler Thermocycler	Analytik Jena, Jena, Germany
FVL-2400N Combi-Spin Vortexer	Biosan, Riga, Latvia
Heracell 150i CO_2 incubator	Thermo Fisher Scientific, Waltham, United States
Keyence microscope BZ-9000	Keyence, Osaka, Japan
LightCycler®480	Roche, Basel, Switzerland
Mastercycler™ Pro	Eppendorf AG, Hamburg, Germany
Micro scale ACS 80-4	Kern & Sohn, Balingen, Germany
Mini Plate Spinner MPS1000	Labnet International, Edison, United States
MJ Mini Personal Thermocycler	Bio-Rad Laboratories, Hercules, United States
Multifuge 1 S-R	Heraeus, Hanau, Germany
NanoDrop™ Spectrophotometer ND-1000	VWR International, Radnor, Pennsylvania
NUC6I7KYK mini-PC (termed Seqbox)	Intel, Santa Clara, California, United States
pH-Meter pH7110	Xylem, Rye Brook, United States
Plate Reader Synergy™ HT and HTX	BioTek, Winooski, United States
Rotina420 centrifuge	Hettich, Tuttlingen, Germany
Safety Cabinet Biowizard SilverLine BlueSeries	Kojair, Mänttä-Vilppula, Finland
TS1 ThermoShaker	Analytik Jena, Jena, Germany
Tube rotator	VWR International, Radnor, Pennsylvania
Ultracentrifuge Optima™L-80 XP	Beckmann Coulter, Brea, United States
UV visualization DarkHood DH-50	Biostep, Burkhardtsdorf, Germany
Zeiss Axiovert200M Microscope	Zeiss, Oberkochen, Germany

4.2.4 Kits

Specification	Application	Supplier
CytoTune™-iPS 2.0 Sendai Reprogramming Kit	Reprogramming of fibroblasts into iPSCs	Thermo Fisher Scientific, Waltham, United States
DC Protein Assay	Protein measurement	Bio-Rad Laboratories, Hercules, United States
DNeasy Blood & Tissue Kit	DNA extraction	Qiagen, Hilden, Germany
Endo-free® Maxiprep Kit	Bacterial DNA isolation	Qiagen, Hilden, Germany
Platinum™ Pfx Kit	PCR	Thermo Fisher Scientific, Waltham, United States
Q5® High-Fidelity Mix	PCR	New England Biolabs, Ipswich, United States
QIAamp® DNA Mini Kit	Bacterial DNA isolation	Qiagen, Hilden, Germany
QIAquick® Gel Extraction Kit	DNA isolation from agarose gels	Qiagen, Hilden, Germany
Quant-iT™ PicoGreen™ dsDNA Assay	DNA concentration measurement	Thermo Fisher Scientific, Waltham, United States
QuikChange II Site-Directed Mutagenesis Kit	Introduction of point mutations into vectors	Agilent Technologies, Santa Clara, United States
RNeasy® Mini Kit	RNA extraction	Qiagen, Hilden, Germany
RNeasy® Plus Mini Kit	RNA extraction	Qiagen, Hilden, Germany
SYBR Green I Master	qRT-PCR	Roche, Basel, Switzerland
Taq Core Kit	PCR	MP Biomedicals, Santa Ana, United States

4.2.5 Antibodies

Primary Antibodies	Source		Application (ICC/WB)*	Dilution	Supplier
BRN2	Goat		ICC	1:500	Santa Cruz, Dallas, United States
CTIP2	Rat		ICC	1:500	Abcam, Cambridge, United Kingdom
GAPDH 14C10	Rabbit		WB	1:30000	Cell Signaling, Danvers, United States
MAP2	Mouse		ICC	1:200	Milllipore, Burlington, United States
NANOG	Rabbit		ICC	1:200	Stemgent, Cambridge, United States
OCT4	Rabbit		ICC	1:500	Stemgent, Cambridge, United States
SSEA4	Mouse		ICC	1:200	Stemgent, Cambridge, United States
TRA-1-60	Mouse		ICC	1:200	Stemgent, Cambridge, United States
TUJ1	Mouse		ICC	1:1000	Covance, Princeton, United States
VAC14 C-10	Mouse		WB	1:10000	Santa Cruz, Dallas, United States
vGLUT1	Rabbit		ICC	1:500	Abcam, Cambridge, United States
Secondary Antibodies	**Source**	**Against**		**Dilution**	**Supplier**
AlexaFluor® 488	Goat	Rabbit	ICC	1:500	Thermo Fisher Scientific, Waltham, United States
AlexaFluor® 488	Donkey	Goat	ICC	1:400	Thermo Fisher Scientific, Waltham, United States
AlexaFluor® 568	Goat	Mouse	ICC	1:500	Thermo Fisher Scientific, Waltham, United States
AlexaFluor® 568	Donkey	Rabbit	ICC	1:400	Thermo Fisher Scientific, Waltham, United States
AlexaFluor® 568	Goat	Rat	ICC	1:400	Thermo Fisher Scientific, Waltham, United States
AlexaFluor® 635	Goat	Mouse	ICC	1:400	Thermo Fisher Scientific, Waltham, United States
AlexaFluor® 647	Donkey	Mouse	ICC	1:400	Thermo Fisher Scientific, Waltham, United States
HRP-linked	Horse	Mouse	WB	1:5000	Cell Signaling, Danvers, United States
HRP-linked	Goat	Rabbit	WB	1:5000	Cell Signaling, Danvers, United States

ICC = Immunocytochemistry, WB = Western Blot, HRP = Horseradish peroxidase

4.2.6 Laboratory supplies and reagents

Specification	Supplier
1 kB DNA ladder	New England Biolabs, Ipswich, United States
1 kB DNA ladder Quick-Load® Extend	New England Biolabs, Ipswich, United States
100 bp DNA ladder	New England Biolabs, Ipswich, United States
10x Antarctic phosphatase reaction buffer	New England Biolabs, Ipswich, United States
10x CutSmart® buffer	New England Biolabs, Ipswich, United States
10x Incubation Mix T. Pol with $MgCl_2$	MP Biomedicals, Santa Ana, United States
10x Incubation Mix T. Pol without $MgCl_2$	MP Biomedicals, Santa Ana, United States
12-well cell culture plate with lid	Greiner Bio-One, Kremsmünster, Austria
1x Dulbecco's Phosphate Buffered Saline (PBS)	Thermo Fisher Scientific, Waltham, United States
2-Mercaptoethanol, 50 mM	Thermo Fisher Scientific, Waltham, United States
6-well cell culture plate with lid	Greiner Bio-One, Kremsmünster, Austria
96 PCR plate half skirt (for sequencing)	Sarstedt, Nümbrecht, Germany
96-well Lightcycler plate (for LightCycler)	Sarstedt, Nümbrecht, Germany
96-well PCR plates without skirt (for PCR)	Sarstedt, Nümbrecht, Germany
Acetic acid, 3 M, 0.2 µm filtered	Sigma-Aldrich, St. Louis, United States
Agarose	Biozym, Hessisch Oldendorf, Germany
Ampicillin sodium salt	Sigma-Aldrich, St. Louis, United States
Antarctic phosphatase 5000 U/ml	New England Biolabs, Ipswich, United States
Assistent® Neubauer counting chamber	Karl Hecht, Sondheim vor der Rhön, Germany
*Ava*I 50,000 U/ml	New England Biolabs, Ipswich, United States
Bacto™ Agar	Becton Dickinson, Franklin Lakes, United States
Bacto™ Tryptone	Becton Dickinson, Franklin Lakes, United States
Bacto™ Yeast Extract	Becton Dickinson, Franklin Lakes, United States
*Bam*HI 50,000 U/ml	New England Biolabs, Ipswich, United States
Betaine, 5 M	Sigma-Aldrich, St. Louis, United States
BigDye® Terminator 5x sequencing buffer	Applied Biosystems, Foster City, United States
BigDye® Terminator v3.1	Applied Biosystems, Foster City, United States
Biosphere® filter tips 20, 100, 1000	Sarstedt, Nümbrecht, Germany
Boric acid	Merck Millipore, Darmstadt, Germany
Bromphenol blue	Serva Electrophoresis, Heidelberg, Germany
Cell culture dishes T25, T75, and T175	Sarstedt, Nümbrecht, Germany
Cell spatula 20 mm	TPP, Trasadingen, Switzerland
cOmplete™, Mini Protease Inhibitor	Sigma-Aldrich, St. Louis, United States
Corning® Matrigel® hESC-Qualified Matrix	Corning, New York, United States
Costar® Ultra-Low Attachment 6-well	Corning, New York, United States
CRYO.S cryo tubes	Greiner Bio-One, Kremsmünster, Austria
DAPI-Fluoromount G®	Southern Biotech, Birmingham, United States
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, St. Louis, United States
DMEM/F12	Thermo Fisher Scientific, Waltham, United States
Dulbecco's Modified Eagle Medium (DMEM) + 4.5 g/l D-glucose + L-glutamine	Thermo Fisher Scientific, Waltham, United States
EDTA, UltraPure™ 0.5 M, pH = 8.0	Invitrogen, Carlsbad, United States
Essential 6™ Medium	Thermo Fisher Scientific, Waltham, United States
Exonuclease 10 U/µl	Thermo Fisher Scientific, Waltham, United States
FastAP, thermosensitive, 1 U/µl	Thermo Fisher Scientific, Waltham, United States
Fetal Bovine Serum, Performance Plus certified	Thermo Fisher Scientific, Waltham, United States
Formamide HiDi™	Applied Biosystems, Foster City, United States
FuGENE®HD	Promega, Madison, United States
GeneScan™-600 LIZ® Size Standard v2.0	Thermo Fisher Scientific, Waltham, United States
Gibco® 100X L-glutamine 200 mM	Thermo Fisher Scientific, Waltham, United States
Gibco® 100X MEM-Non-Essential Amino Acids (NEAA)	Thermo Fisher Scientific, Waltham, United States
Gibco® 100X N-2 Supplement	Thermo Fisher Scientific, Waltham, United States
Gibco® KnockOut™ DMEM/F-12 Medium	Thermo Fisher Scientific, Waltham, United States
Gibco® Opti-MEM™ Reduced Serum	Thermo Fisher Scientific, Waltham, United States

Glass Pasteur pipets 230 mm	Th. Geyer, Höxter-Stahle, Germany
Glycine	Sigma-Aldrich, St. Louis, United States
Human Genomic DNA (human blood donors)	Roche, Basel, Switzerland
HyClone Fetal Bovine Serum, defined	GE Healthcare, Chicago, United States
Insulin solution, human, from *S. cerevisae*	Sigma-Aldrich, St. Louis, United States
J.T. Baker™ Ethanol, absolut	Thermo Fisher Scientific, Waltham, United States
J.T. Baker™ HPLC-grade water	Thermo Fisher Scientific, Waltham, United States
Kanamycin	Sigma-Aldrich, St. Louis, United States
KnockOut™ Serum Replacement (KOSR)	Thermo Fisher Scientific, Waltham, United States
LE Agarose	Biozym Scientific, Hessisch Oldendorf, Germany
Menzel Microscope Slides 76 x 26 mm	Thermo Fisher Scientific, Waltham, United States
Microscope cover glasses 12 mm diameter	Karl Hecht, Sondheim vor der Rhön, Germany
MIDORIGreen Advance	Nippon Genetics, Dueren, Germany
Mouse embryonic fibroblasts CF-1, irradiated	Thermo Fisher Scientific, Waltham, United States
mTeSR™1	Stemcell Technolgies, Vancouver, Canada
Neurobasal™	Thermo Fisher Scientific, Waltham, United States
NeuroCult™	Stemcell Technolgies, Vancouver, Canada
*Nhe*I-HF 20,000 U/ml	New England Biolabs, Ipswich, United States
Nitrocellulose 0.2 µm Protran® Amersham™	Carl Roth, Karlsruhe, Germany
Nonidet P40	Sigma-Aldrich, St. Louis, United States
Normal Goat Serum, 0.1% Sodium Azide	Thermo Fisher Scientific, Waltham, United States
NuPAGE™ 4-12% Bis-Tris Protein Gels	Thermo Fisher Scientific, Waltham, United States
NuPAGE™ LDS Sample Buffer	Thermo Fisher Scientific, Waltham, United States
NuPAGE™ SDS Running Buffer	Thermo Fisher Scientific, Waltham, United States
NuPAGE™ Transfer Buffer	Thermo Fisher Scientific, Waltham, United States
One Shot™ TOP10 Chem. Comp. *E.coli*	Thermo Fisher Scientific, Waltham, United States
One Shot™ TOP10 Electrocomp. *E.coli*	Thermo Fisher Scientific, Waltham, United States
Paraformaldehyde, 37% (v/v) in H_2O	Sigma-Aldrich, St. Louis, United States
Peha-soft® nitrile gloves	Paul Hartmann, Heidenhein, Germany
PenStrep, 10,000 Units/ml Pen, 10,000 µg/ml Strep	Thermo Fisher Scientific, Waltham, United States
PhosSTOP™	Roche, Basel, Switzerland
Plastic Pasteur pipets	Ratiolab, Dreieich, Germany
Poly D-lysine hyrdobromide	Sigma-Aldrich, St. Louis, United States
Poly-L-ornithine, 0.01% (w/v)	Sigma-Aldrich, St. Louis, United States
Precision Plus Protein™ All Blue	Bio-Rad Laboratories, Hercules, United States
Proteinase K, 30 mAnson-U/mg	Carl Roth, Karlsruhe, Germany
Puromycin	Sigma-Aldrich, St. Louis, United States
Research Plus pipets	Eppendorf, Hamburg, Germany
RNase™ AWAY	Thermo Fisher Scientific, Waltham, United States
S.O.C. Medium	Invitrogen, Carlsbad, United States
Serological pipet tips	Sarstedt, Nümbrecht, Germany
Serological pipets 5, 10,25, and 50 ml	Sarstedt, Nümbrecht, Germany
Sodium acetate, 3 M, pH = 5.2	Sigma-Aldrich, St. Louis, United States
Sodium chloride, 99.7%	TH. Geyer, Renningen, Germany
StemPro® Accutase®	Thermo Fisher Scientific, Waltham, United States
SuperSignal™ Chemiluminescent Substrate	Thermo Fisher Scientific, Waltham, United States
T4 10x buffer with 10 mM ATP	New England Biolabs, Ipswich, United States
T4 DNA Ligase, 400,000 U/ml	New England Biolabs, Ipswich, United States
TeSR™-E8™ Essential 8	Stemcell Technolgies, Vancouver, Canada
Tip StackPack 200 µl	Sarstedt, Nümbrecht, Germany
Triton X-100	Sigma-Aldrich, St. Louis, United States
VSV-G, ΔR8.2 plasmids	Addgene, Watertown, United States
*Xho*I 20,000 U/ml	New England Biolabs, Ipswich, United States

4.2.7 Primer

Primer name	Sequence forward	Sequence reverse	Ann. [°C]
Genomic primer			
Mycoplasma rDNA	GGGAGCAAACAGGATTAGATACCCT	TGCACCATCTGTCACTCTGTTAACCTC	58
THAP1_Ex1	AAACACCTGGCCTCAGCCAATA	TGTTCCAGGAGCGCGAGAAA	59
THAP1_Ex2	AAGGTTCCAGGCACATTTATTC	ACAACAGAGCAAGACGGCATC	59
THAP1_Ex3	TATTTTTAGCAGAGACAGGGTT	TTTAAATGAAACTCCTTTACAGG	59
VAC14_Ex13	GAGCAACAACCCAAAGGAC	GCCCCATTCACTGTAGTCTG	59
VAC14_Ex14	GTGCGGAAAAACATCAACACTG	TAACACCCCCCACACACATC	59
VAC14_Ex15	AGTGGACCAGGGGAAGTCTC	CACACTGGCCTCAGGAAAG	59
VAC14_Ex16	CTCTTTGCTGCCAATCTCG	GCTGTCAGCATTTCCTTTGTG	59
cDNA primer			
ACTB	TGAAGTGTGACGTGGACATC	GGAGGAGCAATGATCTTGAT	58
cMYC_CytoTune	TAACTGACTAGCAGGCTTGTCG	TCCACATACAGTCCTGGATGATGATG	55
FLAD1	AAGGTGTTGGTGTCCTGAGTG	AGGGCCTCTGAACTTTCTC	58
FPGS	CGAAGCTATGGCCTGAAGAC	AAGAGCTCAGGACTGATGGG	58
GAPDH	GTCAGCCGCATCTTCTTTTG	GCGCCCAATACGACCAAATC	58
GATA4_Endoderm	CTAGACCGTGGGTTTTGCAT	TGGGTTAAGTGCCCCTGTAG	58
GDF3	AAATGTTTGTGTTGCGGTCA	TCTGGCACAGGTGTCTTCAG	58
GNB1	ACCCAGTGGGAAGAATC	TGGCTGAAGCATCACAA	58
KLF4_CytoTune	TTCCTGCATGCCAGAGGAGCCC	AATGTATCGAAGGTGCTCAA	55
KMT2B	AGTCGGTGGAAGCTAAGAGAGA	ACTGAGGCGAGGGACATC	58
KOS_CytoTune	ATGCACCGCTACGACGTGAGCGC	ACCTTGACAATCCTGATGTGG	55
MSX1_Mesoderm	CGAGAGGACCCCGTGGATGCAGAG	GGCGGCCATCTTCAGCTTCTCCAG	64
NANOG	TGAACCTCAGCTACAAACAG	TGGTGGTAGGAAGAGTAAAG	58
NCAM_Ectoderm	ATGGAAACTCTATTAAAGTGAACCTG	TAGACCTCATACTCAGCATTCCAGT	58
OCT4	CCTCACTTCACTGCACTGTA	CAGGTTTTCTTTCCCTAGCT	58
PAX6_Ectoderm	GTCCATCTTTGCTTGGGAAA	TAGCCAGGTTGCGAAGAACT	60
RUNX1_Mesoderm	CCCTAGGGGATGTTCCAGAT	TGAAGCTTTTCCCTCTTCCA	58
SeV_CytoTune	GGATCACTAGGTGATATCGAGC	ACCAGACAAGAGTTTAAGAGATATGTATC	55
SOX17_Endoderm	CTCTGCCTCCTCCACGAA	CAGAATCCAGACCTGCACAA	58
SOX2	CCCAGCAGACTTCACATGT	CCTCCCATTTCCCTCGTTTT	58
THAP1	AAACTTTAAACCCACCAAGTATAG	GGAAGCTGTTCCTGTGGCTC	58
TOR1A	TGTGCGAGGTCCATCTTCATA	CTTGATGTCTTCCCTCTGCTTT	58
VAC14_1	TGTGTGAGGTTGTTCTTGG	GGATGAGGTCATCTGTTGTC	58
VAC14_2	AGCCGGCCTACTGAACAC	TCATGGTGGGAGTTGAAGG	58
VAC14_5	AGAGCAGCGCCTTCCAGCT	AGAGCAGCGCCTTCCAGCT	58
YWHAZ	GTAGGTCATCTTGGAGGGTCGT	GGTATGCTTGTTGTGACTGATCG	58
Molecular cloning primer			
M13 vector	GTAAAACGACGGCCAG	CAGGAAACAGCTATGAC	55
T7/BGHR vector	TAATACGACTCACTATAGGG	TAGAAGGCACAGTCGAGG	58
VAC14_NheI_XhoI_NK57	GGTTCCCCCAGCTAGCGCCATGAACCCCGAGAAG GATTTC	CCTCCGTGCCTCGAGTCAGAGGACAACCC TCCGGTC	59
Site directed Mutagenesis primer			
GNB1_c.88C>T	TGATCTGAGAGAAAGTTGCATCTGCACA TGCTTTCC	GGAAAGCATGTGCAGATGCAACTTTCTC TCAGATCA	55
GNB1_c.154C>G	CAGGTGCCCCCCCAGTGTCCTCC	GGAGGACACTGGGGGGGCACCTG	55
GNB1_c.191G>T	CCTGGAGTCTGTGACCCAGTGCATGGC	GCCATGCACTGGGTCACAGACTCCAGG	55
GNB1_c.272A>G	GCAGAGGGATGGCGCGGACCTTGTTGGTG	CACCAACAAGGTCCGCGCCATCCCTCTGC	55
GNB1_c.274G>A	GCGCAGAGGGATGGTGTGGACCTTGTTGG	CCAACAAGGTCCACACCATCCCTCTGCGC	55
GNB1_c.280C>T	GAGGAGCGCAGAGAGATGGCGTGGACC	GGTCCACGCCATCTCTCTGCGCTCCTC	55
GNB1_c.287G>T	GACCCAGGAGGAGAGCAGAGGGATGGC	GCCATCCCTCTGCTCTCCTCCTGGGTC	55
GNB1_c.316G>A	CCCAGAAGGGGTATATGCACAGGTCATGACCCA	TGGGTCATGACCTGTGCATATACCCCTTCTGGG	55
GNB1_c.353A>G	ATGGAGCAAATGTTACCCAGGCCACCGCAGG	CCTGCGGTGGCCTGGGTAACATTTGCTCCAT	55
GNB1_c.1009A>C	GAGTTAGTTCCAGATCTGGAGGAAGCTATCCCAGG	CCTGGGATAGCTTCCTCCAGATCTGGAA CTAACTC	55
VAC14_c.1685C>T	GTGGAAGATGTTCTCCACATTCAGCAGGAGGCA	TGCCTCCTGCTGAATGTGGAGAACATCTTCCAC	55
VAC14_c.1868G>A	GTGGCACCAGGAGTGGTACAGGCAGCA	TGCTGCCTGTACCACTCCTGGTGCCAC	55
VAC14_c.1942C>T	CCAAACTTCTGGATGAAGTCATAGGCGTGCCGG	CCGGCACGCCTATGACTTCATCCAGAAGTTTGG	55

Short tandem repeat (STR) primer			
D1S2663	FAM-M13-ATGACTGTTCTCTGTTTTAGGAACC	CAGTTGTTGGGAAGCG	55
D1S2845	FAM-M13-CCAAAGGGTGCTTCTC	GTGGCATTCCAACCTC	55
D1S2870	FAM-M13-CGATCATGCCAATGCACT	CCAGGGTGACACAGCA	55
D3S3647	FAM-M13-CGGCTCAGAGCAGGCATAC	CCTAACTAATAAAATGTGTTCCCAG	55
D5S2090	FAM-M13-CCATGGGCATGTTTCAAA	AGTACCTCCTTAGTAACTCTGGGC	55
D6S305	FAM-M13-CACCAGCGTTAGAGACTGC	GCAAATGGAGCATGTCACT	55
D14S306	FAM-M13-AAAGCTACATCCAAATTAGGTAGG	TGACAAAGAAACTAAAATGTCCC	55
D19S878	FAM-M13-CGCCTGGGCGACAGAGAAT	GGTTGCCCGCAGAGTG	55
D20S888	FAM-M13-GGACTTGCTAAGCCTCCAC	GTCAGGGCTCCCTAGAGAA	55
D21S1909	FAM-M13-CCTGTGATTGTGTTTTCCATTTAGCA	TTCCACACTGAGTCAAGAGCAGG	55
DYS390	FAM-M13-CTATATTTTACACATTTTTGGGCC	TGACAGTAAAATGAACACATTGC	55
FAM-M13	CACGACGTTGTAAAACGAC		55

Ann. = Annealing temperature

4.3 Methods

The following table provides an overview of laboratory techniques used in the four subprojects. The scope of methods in each subproject reflects the experimental approach. Nucleic acid-based methods were mainly used for the KMT2B project to compare gene expression levels in patients harbouring pathogenic *KMT2B* variants and controls. Molecular cloning methods were particularly crucial for the two projects related to GNB1 and VAC14 for different functional assays. The main project (THAP1) involved reprogramming of ten fibroblast lines, characterization of twenty iPSC lines, and subsequent differentiation into neurons to investigate gene expression levels by transcriptome analysis (Table 3).

Table 3 Overview of methods related to the four subprojects in this thesis. Dark grey boxes indicate the application of a method in a project.

Methods / Projects	4.3.1 Mammalian cell culture							4.3.2 Nucleic acid-based methods									4.3.3 Molecular cloning							Other				
	4.3.1.1 Fibroblasts	4.3.1.2 Reprogramming of skin fibroblasts into iPSCs	4.3.1.3Coating	4.3.1.4 iPSC culture	4.3.1.5 Embryoid body formation	4.3.1.6 Differentiation of iPSCs into cortical neurons	4.3.1.7Mycoplasma detection	4.3.2.1 Extraction of nucleic acids	4.3.2.2 First-strand cDNA synthesis	4.3.2.3 Quantitative real-time PCR (qRT-PCR)	4.3.2.4 Polymerase chain reaction (PCR)	4.3.2.5 Sanger sequencing	4.3.2.6 Short tandem repeat (STR) analysis	4.3.2.7 RNA integrity	4.3.2.8SNP karyotype analysis	4.3.2.9Whole transcriptome analysis	4.3.3.1 Insert generation and ligation	4.3.3.2 Site-directed mutagenesis	4.3.3.3 Transformation	4.3.3.4 Gel extraction	4.3.3.5Plasmid extraction	4.3.3.6 Subcloning of genomic DNA fragments	4.3.3.7 Lentiviral transduction	4.3.4 Immunofluorescence	4.3.5 Western blotting	4.3.6 Cytometry	4.3.7 Systematic literature search	4.3.8 Statistical analysis
KMT2B	■						■	■	■	■	■	■																■
GNB1								■	■		■	■					■	■	■	■	■							
VAC14	■						■	■	■		■	■					■	■	■	■	■	■	■		■	■		■
THAP1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■								■			■	■

4.3.1 Mammalian cell culture

A range of human cellular models was used throughout this work, including primary human dermal fibroblasts, iPSCs, iPSC-derived neurons, neuroblastoma cells (SH-SY5Y), and human embryonic kidney cells (HEK293T). All mammalian cells were maintained under humidified conditions at 37 °C and 5% CO_2.

4.3.1.1 Fibroblasts

A skin biopsy of a few millimetres in width and around one centimetre in length was collected by a clinician (e.g., Norbert Brüggemann; Clinic of Neurology) under local anaesthetics. The piece of skin was immediately transferred into Dulbecco's Modified Eagle Medium (DMEM) supplemented with +10% (v/v) fetal calf serum and 1% Penicillin-Streptomycin (PenStrep), hereafter called 'culture medium'. All following steps were conducted under sterile working conditions. The biopsy was transferred into a 6-well dish. Small pieces were cut off and placed in the centre of a dry 6-well. A sterile glass coverslip was gently applied onto the skin pieces, and culture medium was filled into the spacing between coverslip and bottom of the dish. Next, 2 ml of culture medium was added, and the primary cells were incubated at 37 °C and 5% CO_2 for several weeks. Cells were passaged and cryopreserved at first passage when reaching 80 - 90% confluency.

To freeze cells, the supernatant was aspirated, and cells were rinsed with 1x Phosphate buffered saline (PBS) at room temperature. Cells were detached using Accutase for 10 min at 37 °C and 5% CO_2. The suspension was centrifuged at 200 x g for 5 min. Next, the pellet was resuspended in 1 ml culture medium supplemented with 10 % dimethylsulfoxide. The cryo-vials were placed into a freezing container with isopropyl alcohol and stored at -80 °C. For the purpose of long-term storage vials were transferred into a tank filled with liquid nitrogen (-180 °C).

To thaw cells, cryo-vials were gently heated in a water bath at 37 °C until a small icicle was left. The cells were then carefully transferred into the culture medium. The tube was gently inverted centrifuged at 200 x g for 5 min at room temperature. The supernatant was aspirated, and the pellet was resuspended in culture medium and incubated on plastic dishes at 37 °C and 5% CO_2. The culture medium composition, culture conditions, as well as cryopreservation and thawing procedures outlined above also apply to HEK293T and SH-SY5Y cells.

4.3.1.2 Reprogramming of skin fibroblasts into iPSCs

Reprogramming in biology describes the remodelling of epigenetic marks either during development or artificially in cell culture. In the present thesis, human pluripotency factors POU class 5 homeobox 1 (POU5F1, also OCT4), SRY-box transcription factor 2 (SOX2), MYC proto-oncogene (MYC), and Kruppel like factor 4 (KLF4) were used to reprogram fibroblasts into iPSCs via non-integrative Sendai virus. During this work, the iPSC facility changed the reprogramming protocol from a feeder culture with irradiated mouse embryonic fibroblasts to a feeder-free culture, which guarantees more defined culture conditions (Figure 6B). Before this transition, lines L-3737 and L-3736 were cultivated on MEF feeder cells during the reprogramming phase (Figure 6A).

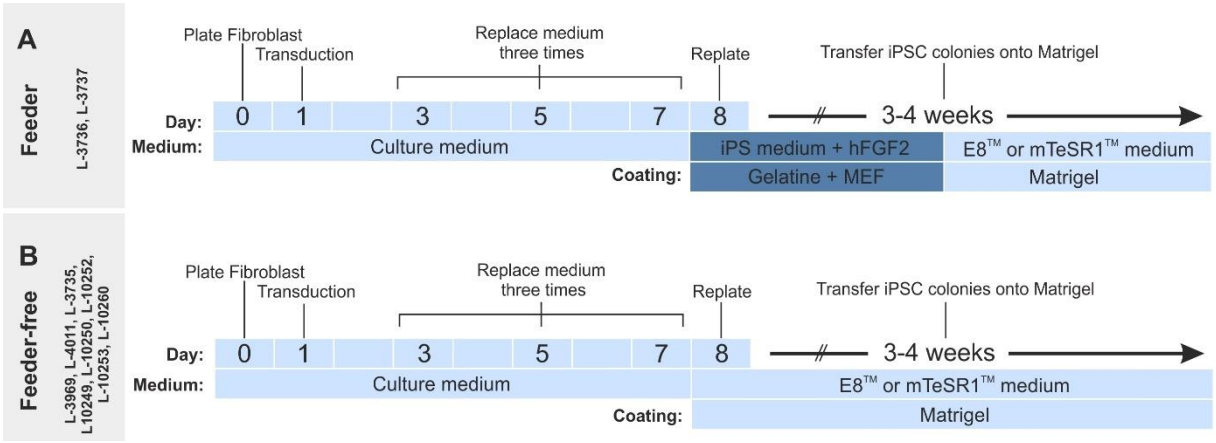

Figure 6 Timeline of reprogramming of human dermal fibroblasts into iPSCs using the CytoTune-iPS 2.0 Sendai Reprogramming Kit. Comparison of (A) feeder and (B) feeder-free protocols with the corresponding cell lines (total = ten fibroblast lines). Differences are highlighted in dark blue. E8 = Essential 8™ medium, iPS medium = induced pluripotent stem cell medium, hFGF2 = human fibroblast growth factor 2, MEF = mouse embryonic fibroblasts.

Early passages of patient fibroblasts were thawed and cultivated in fibroblast medium for several days. Fibroblasts were seeded at a density of 1 x 10^5 cells in one well of a 6-well dish and incubated at 37°C and 5% CO_2 for 24 h. Sendai virus particles from the CytoTune-iPS 2.0 Sendai Reprogramming Kit with a multiplicity of infection (MOI) of polycistronic KOS virus (acronym for Klf4, Oct4, and Sox2) MOI = 5, Myc virus MOI = 5, and Klf4 virus MOI = 3 were collected in 2 ml prewarmed culture medium (without PenStrep). Subsequently, the fibroblast medium was aspirated, and 2 ml virus-containing medium was carefully applied to the cells. After 24 h of incubation, the virus-containing medium was removed. Cells were washed twice with 1x PBS and replaced with 2 ml culture medium. After 48 h, the culture medium was aspirated and cells were detached using Accutase for 10 min at 37 °C and 5% CO_2. Cells were collected in culture medium and centrifuged at 200 x g for 5 min. The pellet was resuspended in culture medium, and cells were seeded on 4 - 6 Matrigel-coated wells of a 6-well dish (except for L-3736 and L-3737 that were seeded on a 1% (w/v) gelatine-coated 10 cm^2 dish; Figure 6). Cells were evenly distributed and incubated for 24 h. Essential 8 (E8) or mTeSR1 was used as culture medium for daily medium changes on the following days. For L-3736 and L-3737, iPS medium supplemented with human fibroblast growth factor 2 (FGF2) (10 ng/ml) was used until iPSC colonies were picked and transferred onto Matrigel. The culture medium was switched to E8 or mTeSR1, as shown in Figure 6B. Colonies were passaged when reaching 80 % confluency.

Starting three weeks post transduction, cells were monitored daily to assess the formation of single iPSC colonies within the fibroblast monolayer. Up to 10 iPSC clones per cell line picked and transferred on Matrigel-coated 4-well dishes with E8 or mTeSR1 supplemented with 10 µM Rho kinase inhibitor Y-27632 to enhance survival of single stem cells. The iPSC colonies were incubated at 37 °C and 5% CO_2. iPSC colonies were expanded on Matrigel-coated plastic dishes either with E8 or mTeSR1 medium. Up to 10 clones per fibroblast line were expanded and banked. Two iPSC clones of each parental line were characterized in-depth as described in section 5.4.2.

4.3.1.3 Coating

Matrigel as well as poly-L-ornithine/laminin (PLO/LA) and poly-D-lysine/laminin (PDL/LA) coatings promote adhesion of cultured cells onto plastic or glass surfaces. While iPSCs were maintained on Matrigel-coated plates, PLO/LA and PDL/LA adhesive substrates were used during cortical differentiations.

To generate Matrigel coating solution, a 300 µl Matrigel aliquot was solved in 24 ml ice-cold DMEM Knockout F12. Next, the Matrigel coating solution was distributed on precooled (-20 °C) cell culture plates using precooled (-20 °C) serological pipets (1 ml per 6-well; 0.5 ml per 12-well, and 0.5 ml per 24-well). The plates were sealed with Parafilm and incubated for a minimum of 1 h at room temperature. Matrigel-coated plates were stored at 4 °C for up to several weeks. Prior to usage, Matrigel-coated plates were placed at room temperature for a minimum of 30 min or at 37 °C and 5% CO_2 for 10 min and the solution was aspirated.

During neuronal differentiations, PLO/LA coating was used for early steps (day 12 and 18), while PDL/LA coatings were experimentally shown to provide improved adhesion for neurons at later stages (day 24, 30, 44, and 75). PLO/LA coating: Culture dishes were incubated with a solution of 0.01% (w/v) poly-L-ornithine diluted in sterile water at 37 °C and 5% CO_2 overnight. The coating was aspirated and rinsed with 1x PBS for three times with 30 min of incubation each. Next, the supernatant was aspirated, and 2 ml laminin (20 µg/ml (w/v) in 1x PBS) was applied and incubated at 37 °C and 5% CO_2 overnight. PDL/LA coating: Culture dishes were incubated with a solution of 0.077 mg/ml (w/v) poly-D-lysine solved in 0.1 M borate buffer, pH = 8.5 and 0.2 µm filtered. Upon incubation overnight at 37 °C and 5% CO_2, the solution was aspirated and washed twice with sterile water and incubated overnight at room temperature protected from exposure to daylight. On the next day, 2 ml 20 µg/ml (w/v) laminin solved in 1x PBS was applied and incubated at 37 °C and 5% CO_2 overnight. The coating was aspirated prior to usage.

4.3.1.4 iPSC culture

iPSC cultures require careful handling techniques to maintain high-quality pluripotent cultures at each passage. Undifferentiated multicellular iPSC colonies grow as compact adherent clusters characterized by distinct borders with a homogeneous morphology when viewed under phase-contrast microscope. To passage iPSCs, colonies were purified from differentiated areas that are characterized by loss of border integrity, or regions of irregular cell morphology within a colony. The medium was aspirated, and colonies were detached by incubation in EDTA-PBS for 1 - 4 min at room temperature. Subsequently, colonies were dispersed with EDTA-PBS using a 5 ml serological pipet. The cell aggregates were carefully transferred to Matrigel-coated wells using a plastic Pasteur pipet. The cells were evenly distributed and incubated at 37 °C and 5% CO_2. In the case of profoundly differentiated iPSC cultures, cells were passaged mechanically. In brief, undifferentiated colonies were selected under a stereomicroscope and transferred by cutting horizontal and vertical lines using a pipet tip to obtain ~1 mm squares of cell patches. Next, these squares were scraped from the surface and transferred to a Matrigel-coated well with a plastic Pasteur pipet. The cells were evenly distributed and incubated at 37 °C and 5% CO_2. In order to freeze iPSCs, colonies were cleaned from differentiated areas under a stereomicroscope and the medium was aspirated. Next, EDTA-PBS solution was added for 1.5 minutes at room temperature. After incubation, cells were dispersed with 1 ml Freeze Medium (60% (v/v) HyClone Fetal Bovine Serum, 20% (v/v) DMEM Knockout F12, and 20% (v/v) DMSO mixed with E8 or mTeSR1 in a 1:1 ratio). Subsequently, the iPSCs were scraped from the surface and transferred into a cryovial using a plastic Pasteur pipet. The cryovial was placed into a -80°C freezer in a container filled with propan-2-ol, which allowed a cooling rate of -1 °C/min. For long term storage, cryovials were transferred into a liquid nitrogen tank. To thaw cells, cryo-vials were gently heated in a water bath at 37 °C until a small icicle was left. Subsequently, cells were transferred into E8 or mTeSR1 using a plastic Pasteur pipet. The tube

was gently inverted and centrifuged at 58 x g for 2 min at room temperature. The supernatant was aspirated, and the pellet was carefully resuspended in E8 or mTeSR1. The cells were distributed on Matrigel-coated wells and incubated at 37 °C and 5% CO_2 overnight.

4.3.1.5 Embryoid body formation

Embryoid body (EB) formation is a commonly used method to assess the potential of iPSC lines to differentiate spontaneously into all three germ layers. In order to generate EBs, 1-2 confluent 6-well plates of iPSCs were treated with EDTA-PBS for 2 min and then mechanically detached using a cell scraper. The cell suspension was resuspended in 3 ml E8 supplement with 4 mg/ml polyvinyl alcohol and transferred into a Costar Ultra-Low-Attachment 6-well plate. Next, the cells were incubated at 37 °C and 5% CO_2. Culture medium (Essential 6 Medium) was exchanged every other day. Cells were harvested after 14 days of culture using a cell scraper. The cells were centrifuged at 800 x g for 5 min. After removing the supernatant, the pellet was stored at -80 °C. RNA extraction and first-strand cDNA synthesis were carried out as described in section 4.3.2.1 and 4.3.2.2. Relative messenger ribonucleic acid (mRNA) levels of ectodermal marker (neural cell adhesion molecule 1 (*NCAM*), paired box 6 (*PAX6*)), endodermal marker (GATA binding protein 4 (*GATA4*), SRY-box transcription factor 17 (*SOX17)*), and mesodermal marker (msh homeobox 1 (*MSX1*), RUNX family transcription factor 1 (*RUNX1*)) were analysed by quantitative real-time PCR as described in section 4.3.2.3. Primer sequences are listed in section 4.2.7.

4.3.1.6 Differentiation of iPSCs into cortical neurons

Human iPSC-derived neuronal models can be utilized to yield insights into the role of pathogenic mutations and complex neurodevelopmental processes. A graphical scheme illustrates the relevant time points of cortical differentiation (Figure 7).

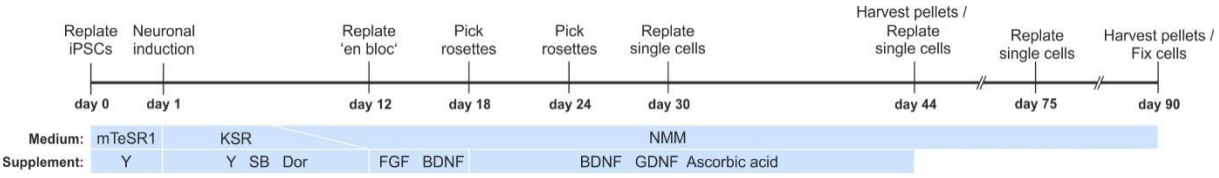

Figure 7 Timeline of the cortical differentiation protocol over a period of 90 days. The protocol was adapted from Shi et al., 2012 as described in Grütz et al., 2017. KSR = Neural induction medium, NMM = Neural maintenance medium, Y = Y-27632, SB = SB-431542, Dor = Dorsomorphine, FGF = human fibroblast growth factor 2, BDNF = Brain-derived neurotrophic factor, GDNF = Glial cell-derived neurotrophic factor.

Day 0 A densely populated 6-well of undifferentiated iPSCs was rinsed with EDTA-PBS and detached using Accutase for 5 - 10 min at 37 °C and 5% CO_2. To generate a single-cell suspension, iPSCs were resuspended with a 1000 µl pipet tip and centrifuged at 200 x g for 5 min. The supernatant was discarded, and the cell pellet was resuspended mTeSR1 supplemented with 10 µM Y-27632. The cell suspension was plated at a density of 1.2 - 1.6 x 10^6 cells on a Matrigel-coated 6-well with mTeSR1 supplemented with 10 µM Y-27632. Subsequently, the cells were incubated at 37 °C and 5% CO_2.

Day 1 When the cells reached confluency, the medium was switched to a neural induction medium (KSR; Medium composition is described in section 4.2.1), which marks the first day of this protocol. The neural induction medium was supplemented with 10 µM Y-27632 + 10 µM SB-431642 (SB) + 1 µM Dorsomorphine (Dor) and was applied from first to fourth day. The culture medium was gradually replaced from KSR to NMM (neural maintenance

medium; composition described in 4.2.1). The volume of KSR was gradually decreased at KSR:NMM ratios of 1:3, 1:1 and 3:1 on the fifth, seventh, and ninth day, respectively. The supplementary factors 10 µM Y-27632 + 10 µM SB + 1 µM Dor were maintained.

Day 12 On the twelfth day, the neuroepithelial sheet was cut into small squares of a few hundred cells by cutting vertical and horizontal lines with a 100 µl pipet tip. Then, the supernatant was discarded, and 4 ml prewarmed NMM medium was applied. The aggregates were scraped from the surface and transferred into a PLO/LA-coated 6 cm^2 dish with a 1000 µl pipet tip. The cells were incubated at 37 °C at 5% CO_2. On the next day, the culture medium was changed to NMM supplemented with 20 ng/µl human FGF2 + 20 ng/ml = Brain-derived neurotrophic factor (BDNF) with 3 ml daily changes of medium until day 17.

Day 18 On day 18, the medium was replaced by 4 ml NMM supplemented with 20 ng/ml BDNF, 20 ng/ml Glial cell-derived neurotrophic factor (GDNF), and 0.2 mM ascorbic acid. The rosettes were detached manually using a cell scraper and transferred into two PLO/LA-coated 6-well dishes using a 1000 µl pipet tip. An exchange of medium with 2 ml fresh NMM + BDNF + GDNF per 6-well was performed every other day. The medium was applied until day 44 (Figure 7).

Day 24 On day 24, the same procedure as described on day 18 was carried out.

Day 30 On day 30, cells were rinsed with 1x PBS and detached using Accutase for 10 min at 37 °C and 5% CO_2. Next, the detached layer of cells was mechanically dissociated by pipetting up and down three times using a 1000 µl tip and incubated at 37 °C and 5% CO_2 for 10 min. The mechanical dissociation step was repeated two times. Next, 1 ml NMM was added and cells were centrifuged at 200 x g for 5 min at room temperature. The single-cell suspension was replated onto PDL/LA-coated wells at a density of 1 - 1.5 Mio cells per 6-well. All remaining cells were cryopreserved in NMM medium supplemented with 10% (v/v) DMSO.

Day 44 On day 44, the same procedure as on day 30 was repeated. Additionally, three wells were pelleted for bulk RNA-Seq analysis and two wells were used for single-cell replating as described above. From day 45 onwards, the medium was exchanged every second or third day with 2 - 4 ml NMM.

Day 75 On day 75, cells were detached as described on day 30. A total of 120,000 cells were seeded onto a PDL/LA-coated glass coverslips in a 12-well cavity.

Day 90 Two 6-wells were cultured and harvested on day 90 post induction. The coverslips were rinsed with 1x PBS, fixed with 4% (w/v) paraformaldehyde solved in 1x PBS for 15 min at room temperature followed by three washing steps with 1x PBS and stored in 1x PBS at 4 °C. Subsequent immunostaining was performed as described in section 4.3.4.

4.3.1.7 Mycoplasma detection

A specialized Mycoplasma-PCR was performed to detect low levels of mycobacterial contamination in cell cultures using primers flanking mycobacterial ribosomal DNA conserved in various mycobacteria species (for primer sequences see section 4.2.7). First, 100 µl supernatant from overnight cultures was centrifuged at 200 x g for 5 min. The supernatant was mixed with 10 mg Proteinase K, 1 µl 10x Incubation Mix without $MgCl_2$, and incubated at 55 °C for 50 min. This was followed by incubation at 95 °C for 10 min. As a next step, a PCR with Taq polymerase (for conditions see section 4.3.2.4) with following cycling conditions: 95 °C 5 min, 40x [95 °C 30s, 58 °C 30s, 72 °C 30s], 72 °C 10 min, 4 °C ∞ was performed. The PCR product was visualized on an agarose gel, as described in section 4.3.2.4.

4.3.2 Nucleic acid-based methods

The following section summarizes all nucleic acids-related methods related to this work. Several techniques, such as RNA extraction, first-strand cDNA synthesis, PCR, and Sanger sequencing, were furthermore applied in molecular cloning experiments (section 4.3.3).

4.3.2.1 Extraction of nucleic acids

The isolation of DNA refers to the purification of DNA from a sample. DNA extraction from cell cultures was carried out using the DNeasy Blood & Tissue Kit following the manufacturer's instructions. DNA and RNA from blood were already available at the Institute of Neurogenetics. If not otherwise indicated, all centrifugation steps were performed at 8000 x g at room temperature. The cell pellet was resuspended in 200 µl 1x PBS supplemented with 20 µl Proteinase K (600 U/ml). Next, 4 µl of RNase (100 mg/µl) was added and the sample was pulse-vortexed for 15 s. This was followed by an incubation of 56 °C for 10 min. Subsequently, 200 µl of pure ethanol was added and samples were pulse-vortexed for 15 s. The volume was transferred to a DNeasy spin column and centrifuged for 1 min. This was followed by a washing step with 500 µl AW1 and centrifugation for 1 min. Next, 500 µl of AW2 was applied to the column. A centrifugation step at 20,000 x g for 3 min was performed. The column was placed in a DNase-free reaction tube and eluted with AE buffer for 1 min twice. The quantity and quality of DNA samples were assessed by measuring the absorbance of the samples at 260/280 nm and 260/230 nm with the NanoDrop1000 spectrometer using the NanoDrop software v3.5.

The isolation of RNA from a sample was performed using the RNeasy Mini Kit and RNeasy Mini Plus Kit according to the manufacturer's instructions. In brief, a whole-cell pellet was lysed in 350 µl RLT buffer by ten times of pulse vortexing. All following centrifugation steps were carried out at 10,000 x g at room temperature. Upon centrifugation for 3 min, the supernatant was transferred into a new reaction tube. Next, 350 µl of 70% (v/v) ethanol was added. The solution was mixed by pipetting up and down. The volume was transferred to the RNeasy spin column, which was centrifuged for 15 s. The flow-through was discarded and 350 µl of RW1 buffer was applied to the column. Next, the column was centrifuged for 15 s, and the flow-through was discarded. In order to hydrolyse DNA impurities, a mixture of 10 µl DNase I and 70 µl RDD was added to the centre of the column and incubated for 15 min. Subsequently, 350 µl of RW1 buffer was applied to the column, centrifuged for 15 s, and the flow-through was discarded. This was followed by two washing steps with 500 µl RPE buffer for 15 s. The column was centrifuged for 1 min in an empty collection tube to ensure the removal of all washing reagents. Next, the RNA was eluted with 30 µl RNase free water into a fresh RNase free reaction tube by centrifugation for 1 min. In contrast to the RNeasy Mini Kit, the RNeasy Mini Plus Kit utilizes dedicated gDNA Eliminator columns to obtain high-quality RNA without DNA impurifications. Furthermore, a special RLT-Plus extraction buffer was supplemented with 0.01% (v/v) 2-mercaptoethanol. All other steps were identical between both protocols. The RNA quantity and quality were routinely assessed by measuring the absorbance at 260/280 nm and 260/230 nm using the NanoDrop1000 spectrometer with the Nucleic acid v3.8.1 software. The RNA integrity number (RIN) was assessed only for RNA samples that were subjected to RNA-Seq (section 4.3.2.7).

4.3.2.2 First-strand cDNA synthesis

The reverse transcription of DNA from an RNA template produces complementary DNA (cDNA). The cDNA can be used as a template for DNA-dependent DNA polymerases such as the ones described in section 4.3.2.3 and 4.3.2.4. The RNA was reversely transcribed with desoxythymidin oligo$(dT)_{18}$ primer and random hexamer primer following the manufacturer's instructions (Maxima First-Strand cDNA

Synthesis Kit). In brief, the equivalent volume of 500 ng RNA was mixed with 1 µl of 10x dsDNase buffer, 1 µl of dsDNase, and nuclease-free water to a total volume of 10 µl. This mix was incubated at 37 °C for 2 min and then immediately cooled to 4 °C. Then, 4 µl of 5x Reaction Mix, 2 µl Maxima Enzyme Mix, and 4 µl of nuclease-free water was added and mixed by pipetting up and down. Subsequently, the sample was incubated at 25 °C for 10 min, 50 °C for 15 min, 85 °C for 5 min, and cooled to 4 °C. The samples were stored at -20 °C.

4.3.2.3 Quantitative real-time PCR (qRT-PCR)

Quantitative RT-PCR is a technique to determine the expression level of a gene in a given biospecimen. A serial dilution of pooled cDNAs from all samples served as a standard curve. Each reaction contained 0.3 µM of forward and reverse primer, 2.4 µl ddH_2O, and 5 µl Maxima SYBR Green I dye. A list of target and reference genes used in this work is shown in section 4.2.7. The qRT-PCR was performed according to the following protocol (Table 4). The results were processed with the software Lightcycler 480 v1.5.1.62. The relative amplification of a target gene compared to the reference gene was either calculated by the ΔΔCt method or by the efficiency-corrected E-method (Livak and Schmittgen, 2001; Tellmann, 2006). The ΔΔCt method assumes a duplication of PCR product after each cycle, while the E-method accounts for changes in the gene amplification by means of a standard curve. Different reference genes were used in this work; these include actin beta (*ACTB*), tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein (*YWHAZ*), flavin adenine dinucleotide synthetase 1 (*FLAD1*), folylpolyglutamate synthase (*FPGS*) and Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*).

Table 4 Cycling parameters of the LightCycler480 instrument. The annealing temperature (X) and amplification time (Y) were adapted according to the primer and amplicon size.

Program	Temperature [°C]	Cycles	Analysis mode	Time [mm:ss]	Ramp rate [°C/s]
Pre-incubation	95	1	None	10:00	4.4
Denaturation	95	38	None	00:15	4.4
Annealing	X	38	None	00:30	2.2
Amplification	72	38	Quantification	Y	4.4
Melting curve	95	1		00:05	4.4
	65	1		01:00	2.2
5 Acquisitions per °C	97	1	Continuous	--	0.11

4.3.2.4 Polymerase chain reaction (PCR)

The PCR is a technique to amplify a particular region of the genome. Several polymerases were used for the amplification of DNA (or cDNA) fragments prior to Sanger sequencing or cloning. The respective polymerase for each reaction was selected based on the application of the PCR product. The non-proofreading Taq-polymerase was used for routine PCR, Mycoplasma PCR, and creating A-overhangs for topoisomerase-mediated subcloning (TOPO-TA). In addition, the Q5 High-Fidelity polymerase was used for molecular cloning due to its high reliability and proofreading capability. The Platinum Pfx kit was used for difficult amplifications such as exon 3 of the *THAP1* gene. All steps were performed on ice, reagents were mixed thoroughly, and spun down for subsequent PCR amplification. Next, the PCR product was subjected to agarose gel electrophoresis. An overview of PCR reagents and cycling conditions is displayed in Table 5.

Table 5 PCR kit components and corresponding cycling conditions of Taq, Q5, and Pfx polymerases. Volumes are provided for a single reaction. The annealing temperature (X) and amplification time (Y) were adapted according to the primer and amplicon size.

Taq-polymerase	Stock concentration	Volume [µl]	Temperature [°C]	Time [mm:ss]	Cycles
ddH_2O	~	4.23	95	05:00	1x
dNTPs	1 mM	3.00	95	00:30	
10x Taq Buffer	10x	1.50	X	00:30	35x
Primer F	10 µM	0.60	72	Y	
Primer R	10 µM	0.60	72	00:10	1x
Taq	5 U/µl	0.07	4	∞	
Template	10 ng/µl	5.00			
Total:		15.00			
Q5 High-Fidelity					
ddH_2O	~	5.00	98	02:00	1x
Primer F	10 µM	1.25	98	00:10	
Primer R	10 µM	1.25	X	00:30	35x
Q5 Master Mix	2x	12.50	72	Y	
Template	10 ng/µl	5.00	72	05:00	1x
Total:		25.00	4	∞	
Platinum Pfx					
ddH_2O	~	4.00	95	05:00	1x
Amplification Buffer	10x	1.50	95	00:30	35x
dNTPs	10 mM	0.75	X	00:30	
Primer F	10 µM	0.75	68	Y	
Primer R	10 µM	0.75	68	10:00	1x
$MgSO_4$	50 mM	0.60	4	∞	
PCR_x Enhancer	10x	1.50			
Pfx Taq	2.5 U/µl	0.15			
Template	10 ng/µl	5.00			
Total:		15.00			

A 1 – 2% (w/v) agarose solution was prepared in 1x Tris-borate-EDTA (TBE) buffer containing 0.05% (v/v) DNA-intercalating dye MIDORIGreen. The solution was heated in the microwave, poured into a gel casting chamber, and cured at room temperature. A volume of 3 µl PCR product was mixed with 3 µl 10x loading buffer (composition see section 4.2.1) and loaded into cavities of the agarose gel. Subsequently, negatively charged nucleic acids were separated according to size by applying a current of 120 V and 400 mA for 20 - 30 min. Nucleic acids were visualized by UV-light in the DarkHood DH-50. Images were processed by the software Argus X1 v7.14.6.

4.3.2.5 Sanger sequencing

Sanger sequencing is an analytical method to determine the DNA sequence by selective incorporation of a chain-terminating fluorescent dideoxynucleotide. In order to remove residual PCR primers and dephosphorylate deoxyribose nucleoside triphosphates (dNTPs) after PCR amplification, a mixture of 10 U Exonuclease I and 1 U alkaline phosphatase (FastAP) was added to each PCR product with volumes ranging from 15 - 25 µl (see section 4.3.2.4). This was followed by incubation at 37 °C for 15 min, 85 °C for 15 min, and cooling to 4 °C. A reaction mix for a single reaction contained 6.5 µl ddH_2O, 1.5 µl 5x

Sequencing buffer, 0.5 µl BigDye Terminator v3.1, 0.5 µl (10 pmol/µl) of either forward or reverse primer, and 1 µl PCR product. Next, the reaction mix was vortexed, spun down, and DNA fragments were amplified according to the following conditions using a thermocycler: 96 °C for 1 min, 25x [96 °C 10s, 60 °C 80s], 4 °C ∞. In order to precipitate the DNA, 10 µl PCR-grade water, 2 µl 3 M sodium acetate, and 50 µl 100% pure ethanol was added to the sequencing reaction and incubated for 15 min at room temperature in the dark. All following centrifugation steps were carried out at room temperature. Next, the reaction mix was centrifuged for 15 min at 13,000 x g. The supernatant was removed and 100 µl 70% (v/v) ethanol was applied to each well. This was followed by centrifugation for 10 min at 13,000 x g. The supernatant was fully removed, and any remaining supernatant was dried off at 55 °C for several minutes. Next, 15 µl Formamide was added to each reaction and incubated at 37 °C for 20 min or at 4 °C overnight. The samples were sequenced on an ABI3130xL or ABI3500xL Genetic Analyzer and interpreted using the software Mutation Surveyor v3.25 or Chromas v2.6.

4.3.2.6 Short tandem repeat (STR) analysis

A STR analysis can be used as an approach to compare two samples from an individual based on highly polymorphic repetitive regions in the genome. The respective primer sequences were derived from the Marshfield Comprehensive Human Genetic Maps. Individual fragment lengths of the polymorphic markers were compared to two fully characterized control individuals (1331.1 and 1331.2) from the *Centre d'Etude du Polymorphisme Humain individuum 1* (CEPH1) and CEPH2 (Broman et al., 1998). The following PCR ingredients were mixed for a single reaction: 2.08 µl PCR-grade water, 1 µl dNTPs (1 mM), 0.5 µl 10x Taq buffer, 0.06 µl forward primer (10 µM), 0.06 µl reverse primer (10 µM), 0.05 µl 6-fluorescein amidite labelled M13 Primer (10 µM), 0.05 µl Taq polymerase (5 U/µl), 1 µl DNA template (10 ng/µl). The PCR was performed according to following conditions: 95 °C 5 min, 35x [95 °C 30s, 55 °C 30s, 72 °C 30s], 72 °C 10 min, 4 °C ∞. Next, the PCR product was diluted with water at a ratio 1:6, and 1 µl was added to 10.7 µl formamide and 0.3 µl 600 LiZ Size Standard. Subsequently, the reaction mix was denatured at 95 °C for 5 min and then immediately cooled on ice. Fragment length analysis was carried out on an ABI 3500xL Genetic Analyzer and visualized using the software GeneMapper v4.1.

4.3.2.7 RNA integrity number (RIN)

The RIN is an algorithm for assessing the quality of RNA and assigning quality values in a numeric system from 1 to 10 (in which 10 represents intact RNA)(Schroeder et al., 2006). The method is based on electrophoretic separation of stained RNA by size in a microfabricated chip. The analysis using the Agilent 6000 Nano Kit was performed according to manufacturer's instruction. In brief, a gel matrix was filtered by centrifugation in a spin filter for 10 min at 1500 x g and 1 µl of supplied dye concentrate was mixed with 65 µl filtered gel. The mixture was spun down for 10 min at 13,000 x g. The electrodes of the 2100 Bioanalyzer were cleaned with RNase Away and RNase free water using the electrode cleaner chip. After setting up the priming station, 9 µl of the gel-dye mix was distributed in the capillary system by applying air pressure with a syringe. Next, 5 µl marker and 1 µl sample or 1 µl heat-denatured RNA ladder were added to the remaining 12 wells. The chip was vortexed horizontally at 2400 rpm for 1 min. The chip was run on the 2100 Bioanalyzer and electropherograms were evaluated using the 2100 Expert software vB022.08.SI648.

4.3.2.8 SNP karyotype analysis

Single-nucleotide polymorphism (SNP) array karyotyping allows the quantitative detection of allelic imbalances on a genome-wide level. The comparative analysis of parental fibroblast lines and derived iPSC clones was used to confirm the parental identity and the genomic integrity of the iPSC clones.

Genomic DNA was extracted from whole-cell pellets (fibroblasts and iPSCs) using the DNeasy Blood & Tissue Kit as described in section 4.3.2.1. The DNA concentration was measured using the Quant-iT PicoGreen dsDNA Assay and diluted to 50 ng/µl in ddH_2O. Genotyping of 200 ng DNA was performed using a Human Omniexpress-24 Beadchip v1.3 (~710,000 markers) following the Infinium HTS assay protocol and analysed with KaryoStudio v1.4.3.0 by our collaborator Professor Sally Cowley (Sir William Dunn School of Pathology, Oxford).

4.3.2.9 Whole transcriptome analysis

Whole transcriptome analysis is a technology to analyse the quantity and sequence of RNA molecules in a biological sample based on next-generation sequencing. In this work, 32 iPSC-derived neuronal samples after 44 days of differentiation were analysed. Cells were collected using a spatula and rinsed with 1x PBS. Subsequently, the supernatant was aspirated, and pellets were frozen at -80 °C. The RNA was extracted using the RNeasy Plus kit (section 4.3.2.1). The quality was assessed by RIN analysis (section 4.3.2.7). An equivalent volume of 1200 ng of RNA was aliquoted and shipped to a company (Novogene) for paired-end bulk RNA sequencing. The company's workflow included mRNA enrichment by oligo(dT) beads and rRNA removal using the Ribo-Zero kit, and random RNA fragmentation using fragmentation buffer. The double-stranded cDNA library was synthesized using random hexamer primer and second strand synthesis was initiated using dNTPs, RNase H and DNA polymerase I. Upon end repair, A-tailing and adapter ligation (5' - AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG-3'), the double-stranded cDNA library was subjected to size selection (150 bp) and PCR enrichment. Subsequently, the samples were sequenced with Illumina's method 'sequencing by synthesis' (SBS). The company sent the raw data (738.6 GB for all 32 samples) to us via an external hard drive. The paired-end FASTQ-formatted raw read files were then processed using the open-source 4SeqGUIv2 pipeline that runs software packages in docker containers v1.13.1 (Beccuti et al., 2018) to decrease dependencies and to achieve a high degree of reproducibility. This pipeline includes the removal of sequence adapters by skewer (docker2017.01). The tool Multi-QC (docker2018.01) compiled crucial quality control measures of the dataset, such as the number of uniquely mapped reads, mean Phred quality scores, duplication levels, adapter content, overrepresented sequences, and per sequence guanine-cytosine (GC)-content. Reads were mapped to reference Ensembl GRCh38 using Spliced Transcripts Alignment to a Reference (STAR; docker251.2017.01) (Dobin et al., 2013). Accurate transcript abundance (Transcripts per kilobase million, TPM values) were obtained using RNA-Seq by Expectation-Maximization (RSEM; docker2017.01) (Li and Dewey, 2011). Differential expression analysis of the raw counts was carried out using the tool DESeq2 (Love et al., 2014). In addition to the pipeline, several R-packages were used for visualization. Volcano plots were visualized using the R-package EnhancedVolcano v1.2.0. Heatmaps were rendered using the R-package heatmap.2 by gplots v3.0.1.1. Principal component analysis was performed using the R-package Glimma v1.12.0. Gene identifiers were generated using the R-package BiomaRt v2.40.5.

4.3.3 Molecular cloning

Molecular cloning refers to a set of experimental methods for the replication of recombinant DNA molecules within a living host (Figure 8A). In this work, DNA fragments from human origin were combined with bacterial vectors to generate recombinant molecules. The methods described below were used for molecular cloning. The plasmid encoding Venus-156-239-Gβ1 was provided by Kirill A. Martemyanov (The Scripps Research Institute, La Jolla, United States) and contained a split Venus

protein (GenBank accession DQ092360) tagged to the *GNB1* wild-type sequence via a GGSGGG linker (Hollins et al., 2009). Two vectors used in this work for VAC14 cloning are shown in Figure 8B and C.

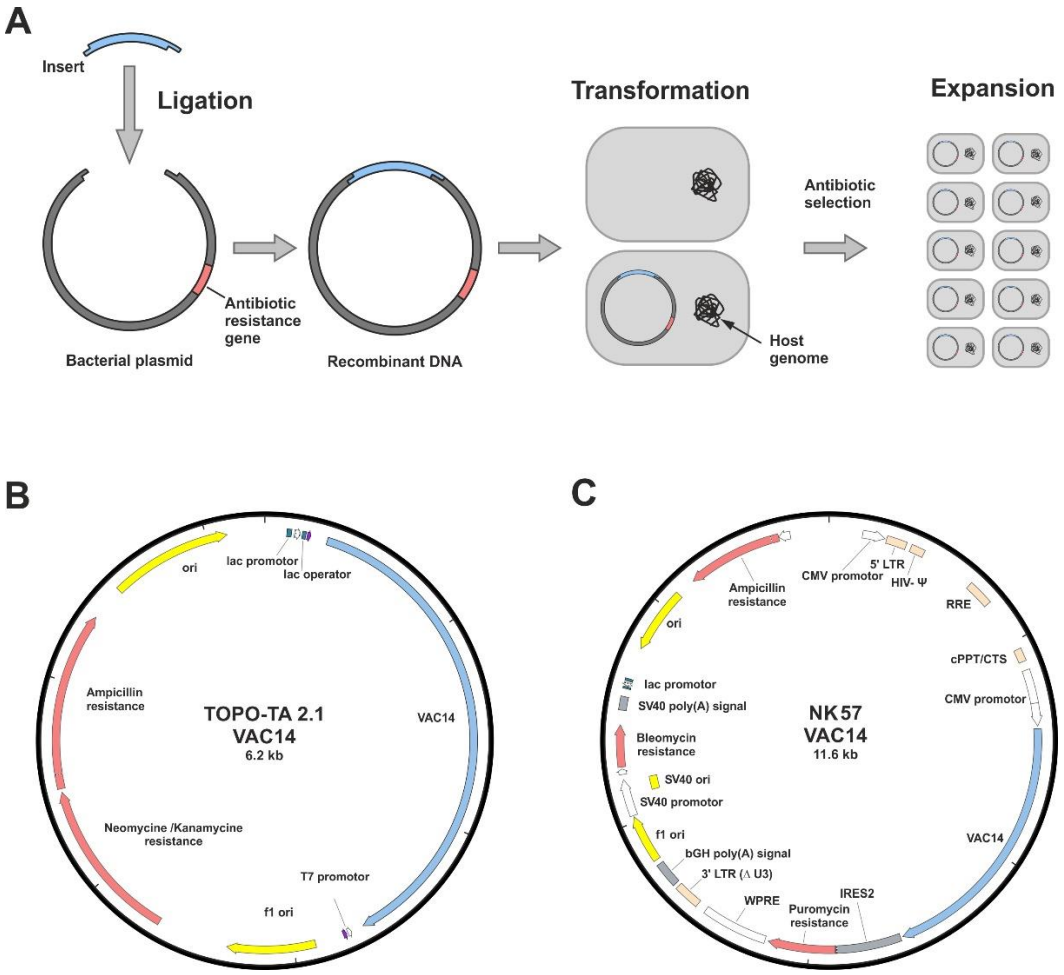

Figure 8 Schematic illustration of the molecular cloning workflow. (A) First, a DNA fragment of human origin, termed insert, is joined with a bacterial plasmid through a technique called ligation. The recombinant DNA is then introduced into competent *E.coli* by a process called transformation. Bacteria are then exposed to antibiotics and selected based on the presence of an antibiotic resistance gene within the vector. (B) TOPO-TA™-VAC14 vectors were used for subcloning and (C) NK57-VAC14 vector used for lentiviral transduction.

4.3.3.1 Insert generation and ligation

One of the first steps in molecular cloning is the generation of an insert sequence. The primers were designed to carry overhangs with defined recognition sites for restriction enzymes. Upon PCR of a control cDNA with a proofreading polymerase, the amplicon was applied to an agarose gel, extracted from the gel (section 4.3.3.4), and solved in 60 µl EB buffer. Next, 30 µl of the gel-extracted amplicon was mixed with 10 U of each restriction enzyme (*Ava*I and *Bam*HI for the Venus-156-239-Gβ1 vector and *Nhe*I and *Xho*I for the NK57 VAC14) and 1x CutSmart buffer. The DNA was hydrolysed for 1 h at 37 °C. Likewise, this procedure was done for 5 µg of the empty vector with the same restriction enzymes. To prevent religation of the vector, 5'-ends were dephosphorylated using 5000 units Antarctic Phosphatase with 1x Antarctic Phosphatase buffer during an incubation for 30 min at 37 °C followed by inactivation at 70 °C for 5 min. The ligation was performed at a ratio of 1:3 (vector to insert) with regard to the length of both educts according to the following formula:

$$ng\ insert\ required = 3 \cdot 50\ ng\ vector\ \frac{size\ of\ insert\ (bp)}{size\ of\ vector\ (bp)}$$

The reaction mix (1x T4 buffer, 50 ng vector, calculated amount of insert, 400 Units T4 ligase, and water added to a final volume of 20 µl) was incubated for 10 min at room temperature. This was followed by an incubation for 10 min at 65 °C to inactivate the heat-labile alkaline phosphatase. The ligated vector was then used as a template for transformation into competent *Escherichia coli* (*E.coli*).

4.3.3.2 Site-directed mutagenesis

Site-directed mutagenesis is a method that is used to introduce a specific change to the DNA sequence. This was carried out using the QuikChange II Site-Directed Mutagenesis Kit according to manufacturer's instructions. The primers were designed online (www.agilent.com/store/primerDesignProgram.jsp). A bacterial plasmid harbouring a wild-type insert served as a template. A single PCR reaction contained 1x Reaction buffer, 130 pmol forward and reverse primer, 0.6 µl dNTP mix, 50 ng plasmid DNA, 6.25 Units Pfu Ultra high-fidelity DNA polymerase, and water to a total of 25 µl. A PCR was performed according to following conditions: 95 °C 1 min, 12x [95 °C 30s, 55 °C 1 min, 68 °C Y min], 68 °C 10 min, and 4 °C ∞ with variable Y adjusted according to plasmid length (1 min/kb). A list of primers relevant to this method can be found in section 4.2.7. To cleave the parental methylated and hemimethylated bacterial template, 9 Units *Dpn*I were added to the reaction mix, and incubated at 37 °C for 1 h. Subsequently, 1.5 µl of hydrolysed template was used for transformation into competent *E.coli*.

4.3.3.3 Transformation

Transformation in molecular biology describes the procedure to transfer exogenous DNA into bacterial cells. For this work, the laboratory strains One Shot TOP10 Electrocompetent *E.coli* and One Shot TOP10 Chemically Competent *E.coli* were used according to manufacturer's instructions. For electroporation, one vial of electrocompetent cells was thawed on ice, mixed carefully with 1.5 µl of ligated vector, and transferred into an electroporation cuvette on ice. Cells were subjected to a pulse of 2500 volts, and gently mixed with 400 µl provided S.O.C medium. In contrast to that, one vial of chemically competent cells was incubated with 1 - 5 µl ligated vector for 30 min on ice and heated to 42 °C for 45 s in a water bath. Subsequently, the mixture was incubated for 2 min on ice and then mixed with 250 µl S.O.C. medium. In both procedures, this was followed by an incubation on a horizontally shaking incubator at 200 rpm for 1 h. Next, 20 - 100 µl of bacterial suspension was spread on a LB-agar plate containing 50 µg/ml of antibiotics according to the resistance gene of the vector. The plates were incubated at 37 °C overnight, and single-cell colonies were picked for further analysis.

4.3.3.4 Gel extraction

Nucleic acids were extracted from an agarose gel by use of the QIAquick Gel Extraction Kit according to manufacturer's instructions. All following centrifugation steps were carried out at 16,000 x g at room temperature. In short, the desired piece of agarose was excised using a scalpel and incubated in 400 µl QG buffer at 50 °C for 10 min. Next, 150 µl of isopropyl alcohol was added. The mixture was subjected to a QIAquick Spin Column. After centrifugation for 1 min, the column was rinsed with 500 µl QG buffer. This was followed by a centrifugation step for 1 min. Next, 750 µl of PE buffer was added to the column. Upon incubation at room temperature for 2 min, the column was centrifuged for 1 min. To eliminate residual washing buffer, the column was centrifuged for 1 min. The DNA was eluted with 60 µl EB buffer by centrifugation for 1 min in a new collection tube.

4.3.3.5 Plasmid extraction

Small amounts of bacterial plasmid DNA were isolated and purified using the QIAprep Spin Miniprep Kit. The protocol was performed according to manufacturer's instructions. First, 9.5 ml of overnight bacterial culture in LB-medium was centrifuged at 4500 x g for 5 min at room temperature. All following centrifugation steps in this protocol were carried out at 16,000 x g at room temperature. The pellet was resuspended in 250 µl P1 and transferred into a 1.5 ml reaction tube. Next, 250 µl of P2 was added, the tube was inverted six times, and 350 µl of N3 was added immediately, followed by inverting the tube six times. Subsequently, the samples were centrifuged for 10 min and 900 µl of supernatant was carefully applied to the QIAprep 2.0 spin column. After centrifugation for 1 min the flow-through was discarded and 700 µl of PE was added to the column. The flow-through was discarded and tubes were centrifuged for 1 min to remove residual washing buffer. Then, the columns were placed on 1.5 ml collection tubes and incubated in 35 µl EB buffer for 1 min, and plasmid DNA was eluted by centrifugation for 1 min. For long-term storage, 500 µl of overnight culture was mixed with 80% (v/v) glycerol in water (ratio 1:1) and frozen at -80 °C.

Large amounts of bacterial plasmid DNA were isolated and purified using the EndoFree QIAprep Maxiprep Kit according to manufacturer's instructions. A total of 100 ml of overnight bacterial culture was centrifuged at 4500 x g for 15 min at 4 °C. The bacterial pellet was resuspended in 10 ml P1. Next, 10 ml of P2 was added to the tube followed by inverting six times. Subsequently, 10 ml of pre-cooled P3 buffer was added and thoroughly mixed. The tube was centrifuged at 2000 x g for 5 s. The supernatant was filtered through a QIAfilter cartridge and mixed with 2.5 ml ER buffer. The tube was inverted several times and incubated for 30 min on ice. In the meantime, a QIAGEN-tip was equilibrated with 10 ml QBT buffer for 10 min. The cell lysate was then applied onto the QIAGEN-tip and emptied by gravity flow. Next, 60 ml of QN buffer were applied to the QIAGEN-tip in a stepwise fashion. The sample was eluted with 15 ml QN buffer and DNA was precipitated with 10.5 ml isopropyl alcohol. After centrifugation for 30 min at 15,000 x g at 4 °C the supernatant was removed, and the pellet was solved in 70% (v/v) ethanol. Subsequently, the solution was centrifuged at 13,000 x g for 10 min, and the air-dried pellet was resuspended in 150 - 250 µl TE buffer.

4.3.3.6 Subcloning of genomic DNA fragments

Recessive human diseases are caused by pathogenic variants affecting both alleles. These two pathogenic variants can be located on the same allele (in cis) or on different alleles (in trans). In the absence of DNA from family members, subcloning of the alleles into separate bacterial vectors can be performed to clarify on which allele each variant is located. In this study, subcloning was used to set the phase for two variants in the *VAC14* gene found in a dystonia patient. In the first step, the DNA region of interest (VAC14 exon 16) was amplified using a non-proofreading Taq-polymerase with a final extension at 72 °C for 10 min to produce A-overhangs. The PCR was conducted as described in section 4.3.2.4. The PCR amplicon was introduced into a TOPO-vector using the TOPO-TA Cloning Kit for Subcloning by mixing 2 µl of the PCR product, 1 µl Salt solution, 2 µl PCR-grade water, and 1 µl TOPO-vector. This mixture was incubated for 5 min at room temperature. This was followed by a 1:4 dilution with ddH_2O, of which 1.5 µl were mixed gently with 50 µl One Shot TOP10 electrocompetent *E.coli* for transformation (as described in section 4.3.3.3). Up to ten single recombinant colonies were picked, solved in 50 µl PCR-grade water, and incubated at 95 °C for 5 min. The mutation was verified by Sanger sequencing in which 100 ng of the vector was used as a template (as described in section 4.3.2.5).

4.3.3.7 Lentiviral transduction

Lentiviruses have the ability to integrate DNA fragments into the genome of a host. Therefore, lentiviral transduction can be used to generate stable cell lines that express a gene of interest. In this work, lentiviruses encoding mutant VAC14 vectors were used to transduce SH-SY5Y neuroblastoma cells (Figure 8). All following steps were performed under biosafety level 2 conditions. For lentiviral production, 300 µl Opti- Minimum Essential Medium (MEM) was mixed with 2.5 µg purified NK57-VAC14 vector, 2.5 µg ΔR8.2 plasmid, 0.3 µg VSV-G plasmid, and 20 µl FuGENE-HD transfection reagent and vortexed fifteen times. After incubation for 25 min at room temperature, the mixture was applied dropwise to a 10 cm^2 plate with 60% confluent HEK293T cells. The cells were incubated at 37 °C and 5% CO_2. After 24 h, the culture medium was exchanged, and plates were incubated for another 24 h at 37 °C and 5 % CO_2. On the next day, the supernatant containing lentiviral particles was passed through a 0.45 µm filter and centrifuged at ~140,000 x g for 2 h at 4 °C. The lentiviral pellet was resuspended in 150 µl Opti-MEM, aliquoted and stored at -80 °C. For lentiviral transduction, 12 µl of the virus were supplemented to the culture medium of SH-SY5Y cells growing in a 12-well dish. Cells were selected for the presence of a puromycin resistance gene by supplementing 2 µg/ml puromycin to the culture medium for 14 days. Medium exchange with fresh antibiotics-containing medium was performed every other day. After ten days, stably transfected lines could be handled according to biosafety level 1 condition. The successful overexpression of VAC14 was confirmed by Western blotting (Figure 15).

4.3.4 Immunofluorescence

Immunofluorescence relies on the use of antibodies coupled to fluorescent dyes to visualize the distribution of biomolecules. In this work, immunostaining of cells was carried out on glass coverslips. At first, coverslips were sterilized at 180 °C for 4 h, placed in 4-well plates, and coated with Matrigel or PDL (as described in section 4.3.1.3). Neurons were seeded on the coverslips as described in section 4.3.1.6 and iPSCs were cultured for around one week. Cells were fixed with 4% (w/v) formaldehyde solved in 1x PBS for 15 min at room temperature, washed three times with 1x PBS and stored at 4 °C in 1x PBS. The coverslips were transferred onto a hydrophobic Parafilm layer and washed once with 75 µl 1x PBS. As a next step, 75 µl of blocking buffer (10% (v/v) normal goat serum and 0.1% (v/v) Triton X-100 solved in 1x PBS) was applied for 45 min at room temperature. The blocking buffer was aspirated and the coverslips were washed with 1x PBS. The primary antibody or antibodies (for concentrations see 4.2.5) was diluted in 1x PBS containing 1% (v/v) normal goat serum. After overnight incubation, coverslips were washed three times with 75 µl 1x PBS with each 5 min of incubation time. Then, 75 µl of the secondary antibody solved in 1x PBS was applied to each coverslip. Each coverslip was inverted and placed on one drop of Fluoromount-G pre-cleaned glass microscope slides. The excess of mounting medium was removed, and the slide was incubated overnight at room temperature and stored at 4 °C. The samples were either visualized with the Keyence microscope and the BZ-II Viewer v1.0 software or the confocal microscope LSM710 with the Zen 2012 software.

4.3.5 Western blotting

Western blotting is a technique to visualize specific proteins of a biological sample, which are separated by molecular weight. In this work, the technique has been used in conjunction with formaldehyde crosslinking of lentivirus-transfected SH-SY5Y cell lines overexpressing VAC14 (results shown in section 5.3). First, the cells were washed with 1x PBS and detached using Accutase for 10 min at 37 °C and 5 °C CO_2. All following steps were carried out at room temperature. The cell suspensions were centrifuged at 200 x g for 5 min at. The supernatant was aspirated, and the pellet was resuspended in

0.5% formaldehyde solved in 1x PBS. Subsequently, the cell suspensions were rotated for 10 min at 20 rpm. This was followed by a centrifugation step at 1500 x g for 5 min. As a next step, the sample was quenched with 500 mM glycine solved in 1x PBS for 10 min and then washed with 1x PBS. The cells were centrifuged at 1500 x g for 5 min and solved in the appropriate amount of Lysis buffer. Subsequently, the protein concentration in each sample was set to 10 µg using the DC Protein Assay. The samples were solved in 1x NuPAGE LDS Sample Buffer and incubated at 95 °C for 4 min. Next, the samples were loaded into a NuPAGE 4 - 12% Bis-Tris polyacrylamide gel. The molecular weight standard Precision Plus Protein was applied on the same gel. Proteins were separated at 100 V for 5 min followed by 150 V for 60 min. The proteins were then transferred onto a 0.2 µm nitrocellulose membrane by using 32 V for 1 h in a XCell SureLock Blot Module. The membrane was blocked with 1% (w/v) milk-TBS-T for 1 h while shaking on an incubator. Next, the membrane was incubated with the primary antibody (for concentrations see section 4.2.5) diluted in 1% (w/v) milk-TBS-T at 4 °C overnight while rotating at 20 rpm. On the next day, the membrane was washed three times with 1x TBS-T for 5 min. The horseradish peroxidase-conjugated secondary antibody (1:5,000 in 1% (w/v) milk-TBS-T) was applied to the membrane. This was followed by an incubation for 60 min on a rotator and three washing steps with TBS-T. The proteins were visualized by means of the Super Signal West Pico PLUS Chemiluminescent Substrate with the ChemiDoc MP Imaging system using the software Image Lab v4.0.1.

4.3.6 Cytometry

Flow cytometry refers to a technique based on the absorbance of light or emission of fluorescence light of single cells at high speed. Fibroblast culture from the patient with compound heterozygous *VAC14* variants (L-7964) and two controls (L-6410 and L-2152) were rinsed with 1x PBS and cells were detached using Accutase for 10 min at 37 °C. The cells were collected in culture medium and centrifuged at 200 x g for 5 min at room temperature. Next, the supernatant was aspirated, and the cell pellet was resuspended in 1x PBS. Subsequently, the cell suspensions were passed through a 70 µm filter and measured using the LSRII cytometer. The area of the side scatter (SSC-A, respectively) was quantified using FlowJo v10.

4.3.7 Systematic literature search

The MDSgene database (https://www.mdsgene.org) aims to provide a comprehensive and curated overview of the published literature of patients with a genetically determined movement disorder. In this work, a systematic literature search was conducted to identify publications on dystonia patients harbouring variants in *TOR1A*, *THAP1*, *ANO3*, *GNAL*, *PRKRA*, or *HPCA* from December 2018 to March 2020 using the National Center for Biotechnology database (www.ncbi.nlm.nih.gov/pubmed) with standardized search terms (Table S1). A mutation carrier was included in the database if a genetic variant was reported, and the clinical data indicated a disease manifestation of dystonia. Patients with a contiguous deletion, multiple genetic diagnoses, or a minor allele frequency ≥ 1% in the Genome Aggregation Database (gnomAD; https://gnomad.broadinstitute.org) were excluded. Furthermore, a Pathogenicity Scoring system was applied to rank a genetic variant according to segregation, population frequency, Combined Annotation Dependent Depletion (CADD) score, published *in-vivo* and *in-vitro* evidence of pathogenicity or functional alterations (https://mdsgene.org/methods). Patients with variants that were interpreted as “benign” using the Pathogenicity Scoring system were excluded from the database.

4.3.8 Statistical analysis

The data sets were statistically analysed using GraphPad Prism 5 and IBM Statistics SPSS v24 software. For gene expression data, a one-way analysis of variance (ANOVA) with the Bonferroni post-hoc correction was performed to account for multiple comparisons using GraphPad Prism v5.02. Phenotypic data in section 5.4 were displayed as median ± interquartile range (IQR). Gene expression analyses were displayed as mean ± standard error of the mean (SEM). The bioinformatic analysis of the whole transcriptome dataset is described in section 4.3.2.9.

5. Results

5.1 *KMT2B* gene expression analysis

The following section will cover the genetic analysis of biospecimens from a female 31-year old Afghan patient harbouring a heterozygous pathogenic frameshift variant, and a previously reported carrier of a pathogenic non-frameshift three bp-deletion (p.Ser1656del), in the *KMT2B* gene (Klein et al., 2019; Lange et al., 2017). A detailed case report of the patient with the frameshift variant is provided in section 4.1. Genetic panel sequencing was negative for variants in *TOR1A*, *THAP1*, *GNAL*, *GCH1,* and *SGCE*, but positive for a novel heterozygous frameshift variant in *KMT2B* (p.Leu1190fs161X; c.3568_3577delCTGAGTGTGC) resulting in a premature stop-codon on protein level (Figure 9). The electropherogram of the *KMT2B* gene shows the absence of a ten base-pair sequence on one chromosome in the index patient L-10377. Both parents did not harbour the variant, which therefore occurred *de novo* in the patient. The variant was absent in the gnomAD database, including over 125,000 exomes and 15,000 genomes from individuals of various ethnic backgrounds.

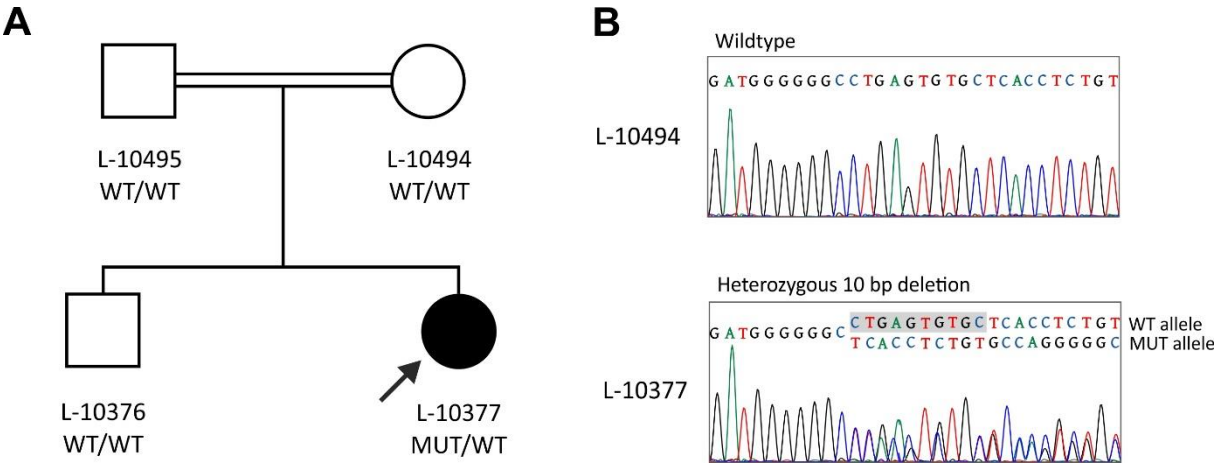

Figure 9 Novel *KMT2B* variant in a consanguineous family in a (A) pedigree of second-degree cousins indicated by the double line with the affected female index patient (black, arrow). (B) Mutational analysis by Sanger sequencing of amplified genomic DNA from the index patient confirmed the heterozygous c.3568_3577delCTGAGTGTGC (NM_014727.2) frameshift variant (marked in grey). Her mother shows the *KMT2B* wildtype sequence. Modified from Klein et al., 2019.

In order to gain insights into the pathophysiological effect of the novel frameshift variant, and the previously reported non-frameshift variant, gene expression levels of *KMT2B*, *THAP1*, and *TOR1A* in blood and fibroblasts were investigated, as alterations in these genes have recently been linked to *KMT2B* mutations (Meyer et al., 2017). Experiments in blood were conducted using four unrelated controls, while two unrelated controls were investigated in fibroblasts. Four different reference genes (*ACTB, YWHAZ*, *FLAD1*, *FPGS*) were analysed. The *KMT2B* expression in both carriers of pathogenic *KMT2B* variants was significantly reduced in fibroblasts compared to controls (Figure 10). In blood, however, no reduction in *KMT2B* expression was observed in the patient harbouring the frameshift variant (p.Leu1190fs161X). Intriguingly, significantly reduced *THAP1* and *TOR1A* expression levels were detected in the blood of both *KMT2B* mutation carriers, but not in fibroblasts (data not shown), compared to controls.

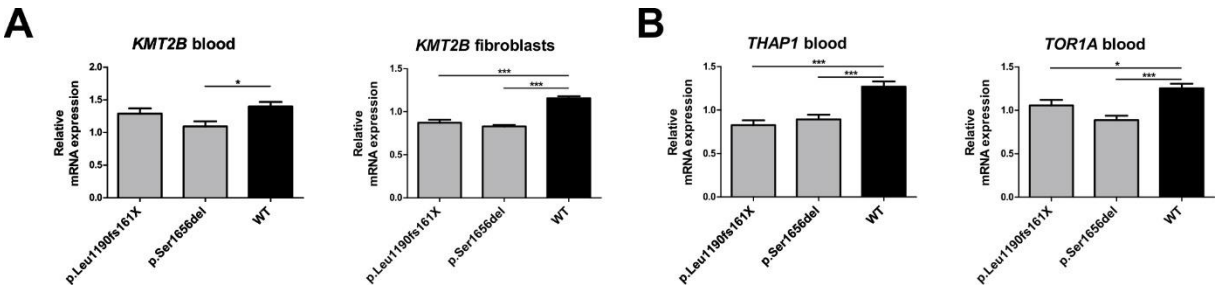

Figure 10 Gene expression analysis of two DYT-KMT2B patients compared to controls. The mean ± standard error of the mean (SEM) expression normalized to four different reference genes is displayed. Four independent experiments were performed in duplicates. For statistical analysis, one-way ANOVA followed by Bonferroni post-hoc test was carried out. Modified from Klein et al., 2019. *P < 0.05; ***P < 0.001

In summary, gene expression analysis of a female dystonia patient with a novel *KMT2B* frameshift variant revealed reduced *KMT2B* mRNA levels in fibroblasts compared to controls. Interestingly, putative target genes *THAP1* and *TOR1A* were reduced in blood compared to controls suggesting a functional link between these three dystonia genes.

5.2 Functional evaluation of *GNB1* mutations

Pathogenic variants in *GNB1,* a gene that codes for the $G\beta_1$ subunit, were identified to cause a neurodevelopmental disorder with hypotonia, seizures, and dystonia (Petrovski et al., 2016; Steinrücke et al., 2016). To expand the spectrum of *GNB1* mutations, we searched for rare variants in the *GNB1* gene in a database containing >4300 whole-exome sequencing datasets of undiagnosed patients of various disorders (Centogene AG). In total, we identified 16 carriers harbouring 14 different rare variants (two frameshift, two splice site, and ten missense changes, one of which was recurrently found in three patients)(Lohmann et al., 2017). Within this thesis, the missense changes were investigated by a cellular model utilizing transient overexpression of components of the GPCR signalling pathway. Therefore, sufficient amounts of purified expression vectors harbouring one of the ten single-nucleotide mutations were generated. The molecular cloning experiments were performed as described in section 4.3.3 in cooperation with Sofia Steinrücke (cand. med.) with equal contributions (Lohmann et al., 2017). PCR-based mutagenesis of ten missense variants was conducted with *GNB1* mutagenesis primers listed in section 4.2.7. The vector template (Venus-156-239-Gβ1-wt-GNB1 vector) was provided by Kirill A. Martemyanov (The Scripps Research Institute; La Jolla, United States). The *GNB1* mutagenesis PCR product was transformed into electrocompetent *E.coli* and selected on Ampicillin-containing plates. Subsequently, several colonies for each point mutation were picked, amplified, and DNA was purified. The constructs were hydrolysed using *Bam*HI and *Ava*I restriction enzymes, which flanked the insert. Subsequently, the hydrolysed vectors were applied on an agarose gel and separated by size (Figure 11A). Visualization with UV-light revealed bands at ~1 kb and ~5 kb, the matching sizes of the *GNB1* insert (1023 bp) and the vector backbone (~5.6 kb). Each construct harboured the targeted point-mutation as depicted in the sequencing electropherograms (Figure 11B). No further genetic alterations of the *GNB1* reference sequence were detected. Upon confirmation of the point mutations by Sanger sequencing, vector DNA was purified in large quantities (Table 6). A total amount of 200 µg purified vector of each point mutation was sent to our collaborators (Kirill A. Martemyanov; The Scripps Research Institute; La Jolla, United States) for the functional evaluation.

Table 6 Venus-156-239-Gβ1-wt-GNB1 vector DNA concentrations and absorbance ratios.

	L30F	R52G	G64V	H91R	A92T	P94S	R96L	A106T	D118G	K337Q
Conc.* [ng/µl] in 150 µl	2767	2617	2789	1839	3345	2807	3082	2290	2445	2692
260/280 nm	1.85	1.85	1.82	1.89	1.78	1.84	1.81	1.88	1.84	1.85
260/230 nm	2.29	2.29	1.78	2.36	1.81	2.27	2.17	2.33	2.29	2.29

* Conc. = Concentration

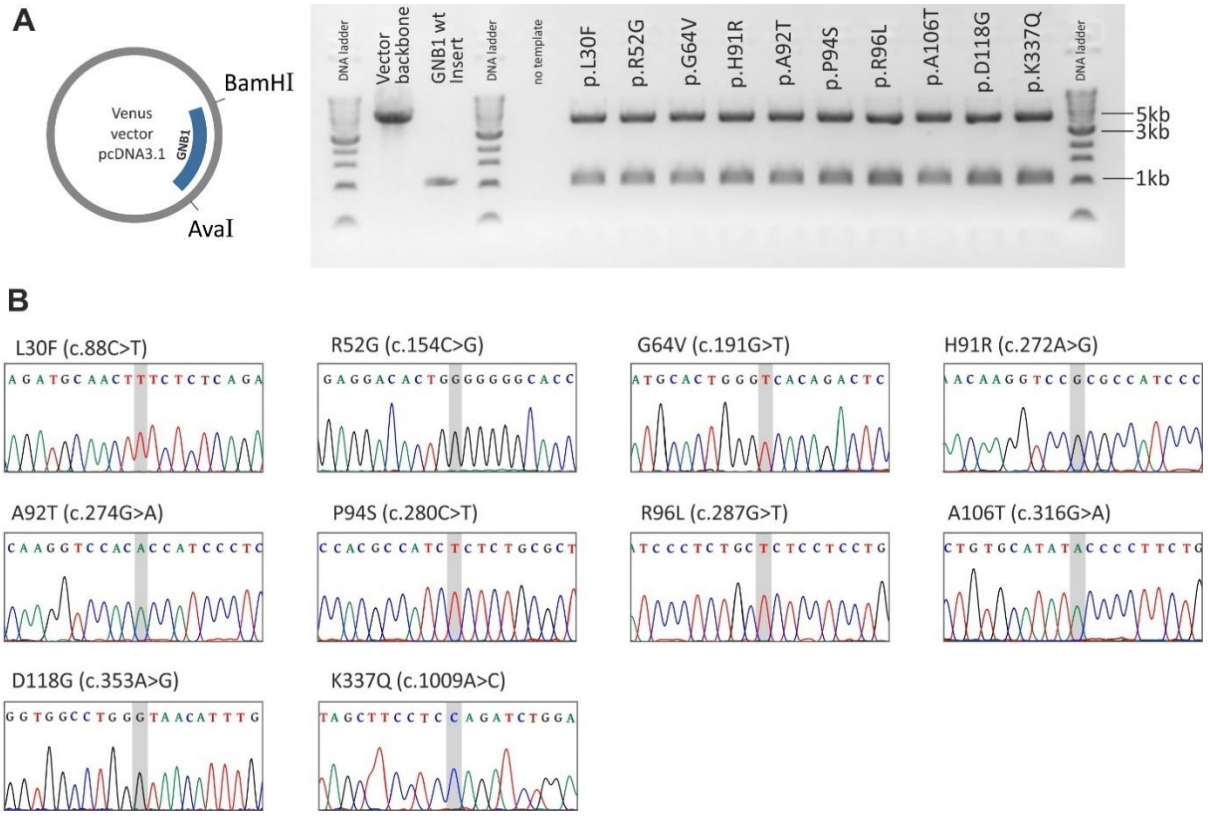

Figure 11 Site directed mutagenesis of the Venus-156-239-Gβ1-wt-GNB1 vector. (A) The mutant Venus-156-239-Gβ1 vectors were hydrolysed using the restriction enzymes *Bam*HI and *Ava*I and separated by size on an agarose gel. (B) Sequencing electropherograms of the single point mutations (grey box) that were introduced into the vectors.

Next, the missense variants were functionally investigated using a cell-based BRET assay by our collaborators (Ikuo Masuho; The Scripps Research Institute; La Jolla, United States). The changes in bioluminescence were used to quantify changes in the interaction of $G\beta_1\gamma_7$ and GRK3 proteins. The complete analysis involved three steps in which different aspects of the GPCR signalling pathway were monitored. First, this involved the ability of $G\beta_1$ to form constitutive dimers with $G\gamma_7$. Second, the influence of missense changes of the $G\beta_1\gamma_7$ dimer with Gα was assessed. Third, the response to dopamine stimulation on signal propagation was investigated (Lohmann et al., 2017). For the illustration of the results, the third assay is depicted in Figure 12. Stimulation of the dopamine 1 receptor (D1R) with the agonist dopamine rapidly leads to a dissociation of the α-subunit releasing β_1/γ_7-dimers tagged with the reporter venus. The G protein-coupled receptor kinase 3 (GRK3) tagged with Nluc luciferase becomes available. Binding of GRK3 and $G\beta_1\gamma_7$ results in the transfer of bioluminescence from Nluc to Venus. The rapid increase in Venus/Nluc bioluminescence, as displayed in Figure 12B, serves as a read-out for the functional activity of the $G\beta_1\gamma_7$ subunit. The assay revealed significantly reduced dopamine-induced GPCR signalling compared to the wild-type in all variants except for the $G\beta_1$ missense changes L30F, H91R, and K337Q (Lohmann et al., 2017).

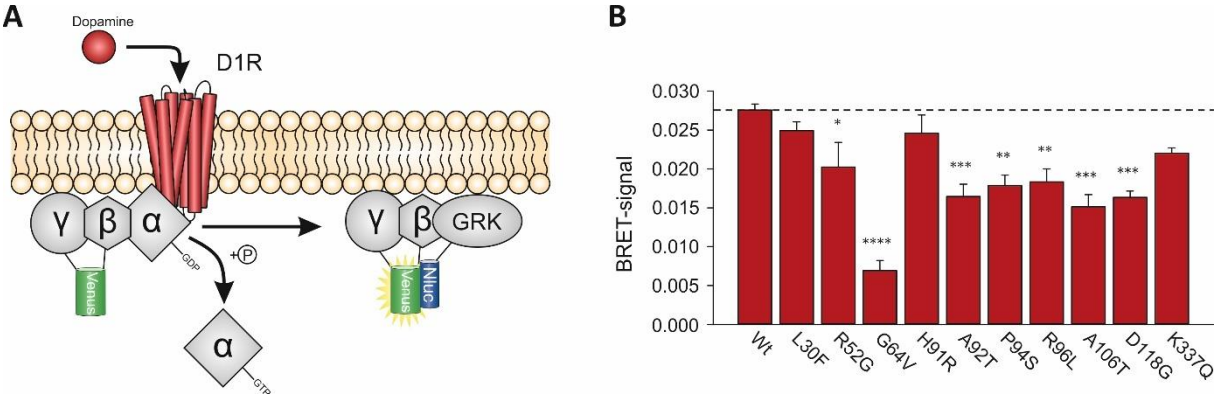

Figure 12 Characterization of ten missense *GNB1* variants upon dopamine stimulation. (A) Assay design and (B) functional effect of *GNB1* variants in a Bioluminescence Energy Transfer (BRET) assay after transient transfection of human embryonic kidney (HEK293) cultures with G-protein coupled signalling pathway components. Three independent experiments were carried out using six replicates. Values represent mean ± SEM. For statistical analysis, one-way ANOVA followed by Dunett's post-hoc test was conducted. D1R = Dopamine 1 receptor; GRK = G-protein coupled receptor kinase 3. Modified from Lohmann et al., 2017. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001

In summary, ten single-nucleotide variants resulting in missense changes in $G\beta_1$ were successfully introduced into Venus-156-239-Gβ1-wt-GNB1 vectors. These vectors were sequenced, amplified, and purified for subsequent use in transient transfection experiments. By assaying the functionality of the $G\beta_1$ protein in an *in-vitro* cell model, aberrant signal propagation was detected in seven out of ten variants. This indicates that these three variants are either benign or exhibit other deficits that are not monitored by this assay (Lohmann et al., 2017).

5.3 Molecular characterization of VAC14

In 2016, mutations in the dimeric scaffold protein VAC14 were established as a novel cause of dystonia (Lenk et al., 2016). A detailed case report can be found in section 4.1. By exome sequencing, we identified a 30-year-old male Caucasian dystonia patient with compound heterozygous missense variants located in exon 16 of the *VAC14* gene (p.L648F and p.R623H; Figure 13). Both of the variants were verified by Sanger sequencing in the index patient. The variant c.1868G>A (p.R623H) was present in the unaffected mother. As DNA of the father was not available, it was unclear if the other variant occurred *de novo* on the maternal allele or was located on the other allele and thus in the compound heterozygous state. In order to clarify zygosity, genomic DNA (*VAC14* exon 16) of the index patient was amplified and subcloned into TOPO-TA vectors (see section 4.3.3.6). The *VAC14* variants were present in different clones (four clones with c.1868G>A and four clones with c.1942C>T) indicating that the variants are compound heterozygous. A deleterious biological effect of the missense variants was predicted by several *in-silico* prediction tools such as Polymorphism Phenotyping v2 (PolyPhen-2), Sorting Tolerant From Intolerant (SIFT), MutationTaster2, and CADD (Table 7).

Table 7 *In-silico* predicted pathogenicity of VAC14 (NM_018052.5) missense variants by PolyPhen-2, SIFT, MutationTaster2, and CADD.

	PolyPhen-2	SIFT	MutationTaster2	CADD v1.4
p.Arg623His (c.1868G>A)	Probably damaging (score = 0.99)	Deleterious (score = 0.02)	Disease causing (score = 0.99)	Top 0.1% most deleterious (score = 33)
p.Leu648Phe (c.1942C>T)	Probably damaging (score = 1.00)	Deleterious (score = 0.00)	Disease causing (score = 0.99)	Top 0.1% most deleterious (score = 32)

Consistent with these findings, multiple sequence alignment of VAC14 orthologs demonstrated that both variants are highly conserved across species (Figure 13). Based on the data from gnomAD, VAC14 missense variant p.Leu648Phe was not observed among approximately 140,000 individuals. Furthermore, VAC14 missense variant p.Arg623His (rs567126350) was rare in the gnomAD database accounting for 0.001995% of the global population (five heterozygous and no homozygous individuals). All five individuals were of European (Non-Finnish) ancestry in which the population allele frequency was highest with 0.001126%.

In order to evaluate the role of VAC14 in dystonia, Sanger sequencing in 160 unresolved dystonia patients in the predicted homodimerization domain of VAC14 (coding exons 13 – 16 and adjacent exon-intron boundaries) was performed. However, we did not detect rare novel mutations.

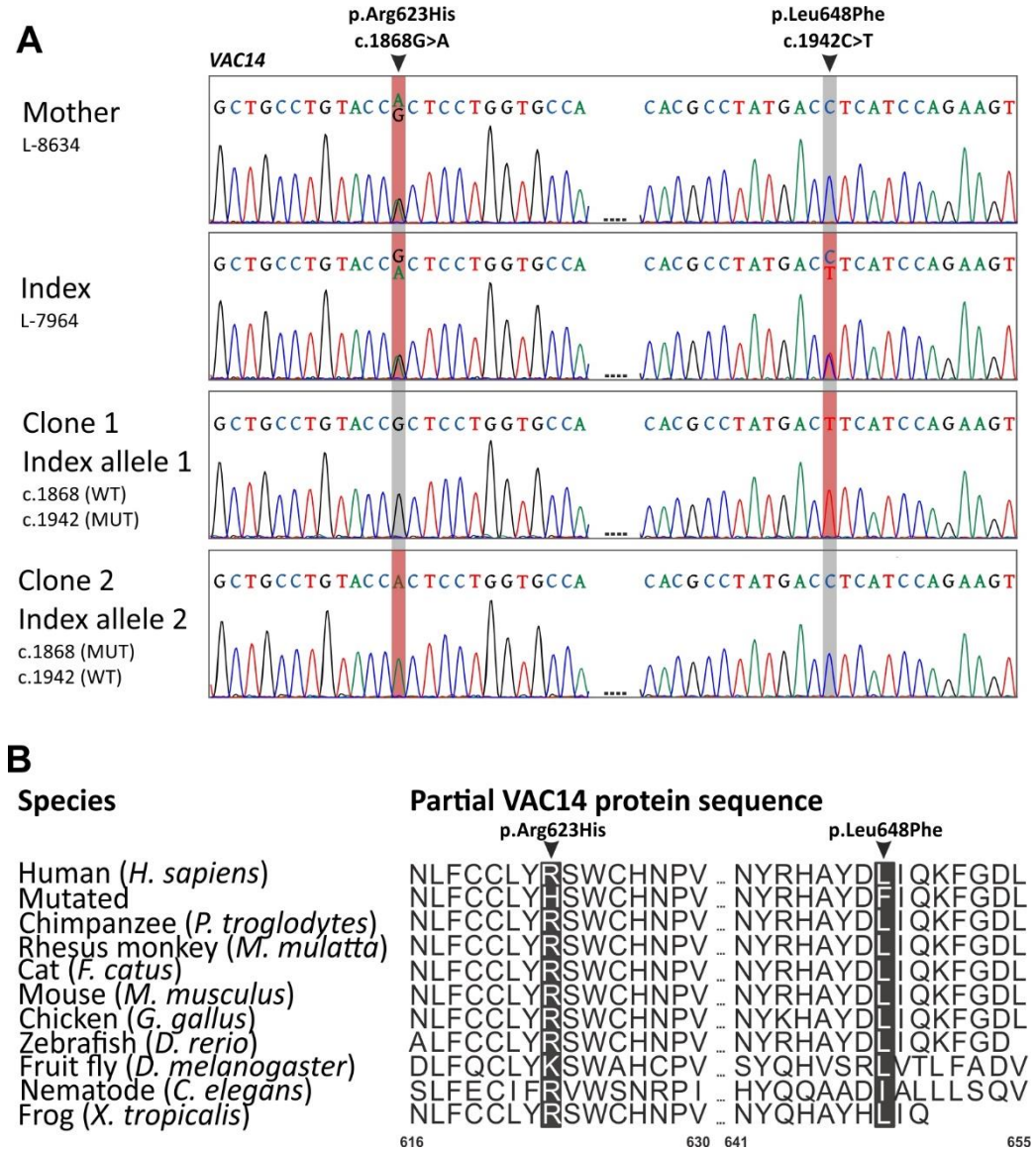

Figure 13 Mutation analysis and protein conservation of VAC14. (A) Sequencing electropherogram of genomic DNA showing two heterozygous variants in the index patient (c.1868G>A and c.1942C>T; NM_018052.5) and one heterozygous variant in the mother (c.1868G>A). Due to the lack of paternal DNA, *VAC14* exon 16 of the index patient was amplified and subcloned into vectors to separate the two alleles. The two variants were found in different clones and thus located on different alleles, i.e., in the compound state. The presence of a mutated nucleobase is indicated by a red background. (B) Protein conservation of VAC14 protein residues across a broad spectrum of animal species using MutationTaster. Arrows indicate VAC14 missense mutations at amino acid positions 623 and 648.

Three reports have demonstrated the occurrence of enlarged cytoplasmic vacuoles in fibroblasts of patients with pathogenic variants in the homodimerization domain of VAC14 (de Gusmao et al., 2019; Kaur et al., 2019; Lenk et al., 2016). In our patient, no enlarged round cytoplasmic structures were observed in fibroblasts of the index patient by brightfield analysis, as shown in Figure 14A. Since we could not detect vacuoles by microscopy, we sought to test for its presence by another method. It was hypothesized that smaller, not microscopically visible vacuoles might be detected and quantified by cytometry analysis. As a read-out, we selected signals of the side scatter (SSC-A), which represent the granularity of a particle and, therefore, are indicative of vacuoles with a high refractive index. In a preliminary experiment, the SSC-A was not elevated in fibroblasts derived from the index patient compared to two control lines (Figure 14B).

Furthermore, we investigated whether starving of fibroblasts in serum-reduced Opti-MEM medium over a period of 72 h induced the formation of intracellular round structures. However, we did not detect the appearance of vacuoles (data not shown), raising doubts about the pathogenicity of the variants.

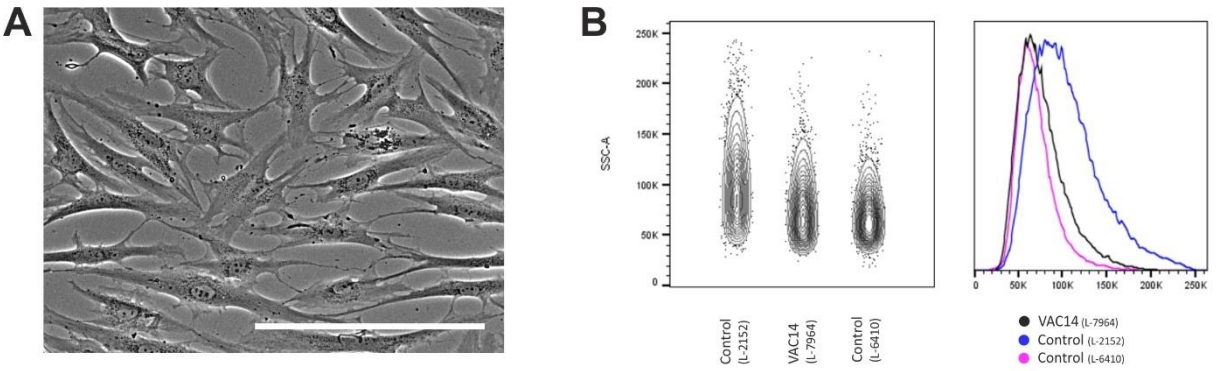

Figure 14 Assessment of vacuoles in VAC14 mutant fibroblasts. (A) Representative brightfield micrograph of patient's fibroblasts (L-7964). Scale bar represents 50 µm. (B) Quantification of the area of the side scatter (SSC-A) by flow cytometry analysis of patient's fibroblasts (10,000 events) versus two controls.

To further test for functional changes mediated by the two missense variants in our patient, we next hypothesized that pathogenic variants in the VAC14 homodimerization domain may alter the capacity of the proteins to form homodimers. To explore the capacity of VAC14 to form homodimers, we generated a disease-associated overexpression model of VAC14 in SH-SY5Y neuroblastoma cells. Transgenic lines of the above-mentioned missense variants, and a recently published missense mutation in a patient with an overlapping clinical phenotype (Taghavi et al., 2018), were generated using lentiviral transfection of SH-SY5Y cells (section 4.3.3.7). Cell suspensions of transgenic cell lines were crosslinked with formaldehyde to retain VAC14 subunit interaction, and subjected to Western blot analysis (section 4.3.5). VAC14-specific chemiluminescent signals were exclusively found in the VAC14 transgenic overexpression lines. Furthermore, the detected signals correlated with the expected molecular weights of VAC14 monomers, dimers, and possibly trimers. A trend of enhanced VAC14 dimer formation was observed in two transgenic missense lines (p.Ala562Val and p.Leu648Phe; dimer/monomer normalized to GAPDH signals) compared to wild-type VAC14 (Figure 15B). Due to the variation of the VAC14 wild-type signal, statistical significance was not reached. Interestingly, the VAC14 missense change inherited by the mother (p.Arg623His) did not alter VAC14 dimerization compared to the wild-type (Figure 15B).

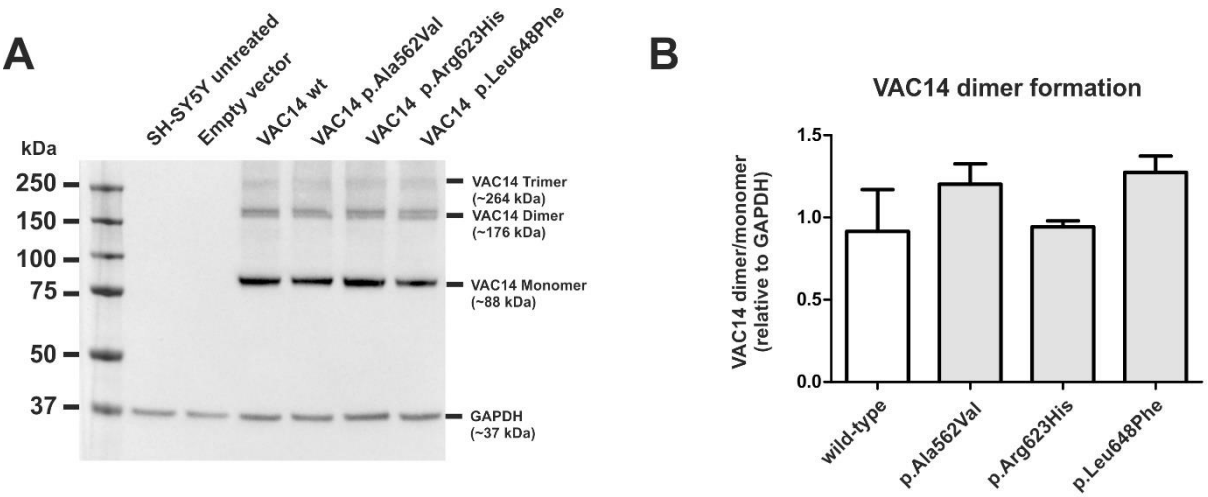

Figure 15 Western blot analysis showing the impact of VAC14 missense variants on VAC14 multimer formation. Transgenic VAC14-SH-SY5Y and untreated control cells were detached, lysed, treated with 0.5% formaldehyde and 10 µg of protein was analysed by Western blot analysis. Images were visualized using chemiluminescence. (A) Representative raw Western blot of combined VAC14 C-10 (dilution 1:10,000) and GAPDH (1:30,000) signals are shown. VAC14 signal (at expected sizes for monomer, dimer, and trimer) was specific to transgenic VAC14 overexpression cell lines. (B) Western blot quantification of lentiviral-transduced SHY-SY5Y VAC14 overexpression lines showing VAC14 dimer formation (dimer/monomer) normalized to the reference protein GAPDH. The mean ± SEM of three experiments is displayed.

Taken together, two novel rare missense variants in the *VAC14* gene were identified in a severely affected dystonia patient. The variants are predicted to have deleterious biological effects by several *in-silico* programs and are highly conserved across species. Molecular characterization of three missense changes showed an accumulation of VAC14 dimers for two missense changes compared to the wild-type in a cellular VAC14 overexpression model. This enhanced dimer formation suggests a functional impact of two missense mutations, which may influence interaction with other proteins, such as FIG4 or PIKFYVE. Peculiarly, no enlarged vacuoles were detected in the patient's fibroblasts, neither by light microscopy nor by quantitative flow cytometry analysis.

5.4 Clinical spectrum and genetic analysis of DYT-THAP1 patients

5.4.1 Literature review DYT-THAP1

The International Parkinson's disease and Movement Disorder Society Genetic online mutation database (MDSGene; www.mdsgene.org) aims to provide a comprehensive and curated overview of published peer-reviewed articles of genetically determined movement disorders. The contribution of the doctoral applicant consisted of a systematic 15-month update regarding the isolated dystonia genes *TOR1A*, *THAP1*, *ANO3*, *GNAL*, *PRKRA*, and *HPCA* from December 2018 to March 2020 and the calculation of patient statistics. Demographic, genetic as well as clinical data was extracted adhering to a standardized protocol (around 200 variables per patient). The clinical characteristics were extracted in collaboration with Professor Christine Klein. A systematic literature search for *TOR1A*, *THAP1*, *ANO3*, *GNAL*, *PRKRA*, and *HPCA* identified 51 original articles. Of these, 19 were eligible articles (36.4%) that reported 36 patients fulfilling the established inclusion criteria (section 4.3.7). Other articles were excluded based on the title, abstract, and full text in a multistep fashion (Figure 16).

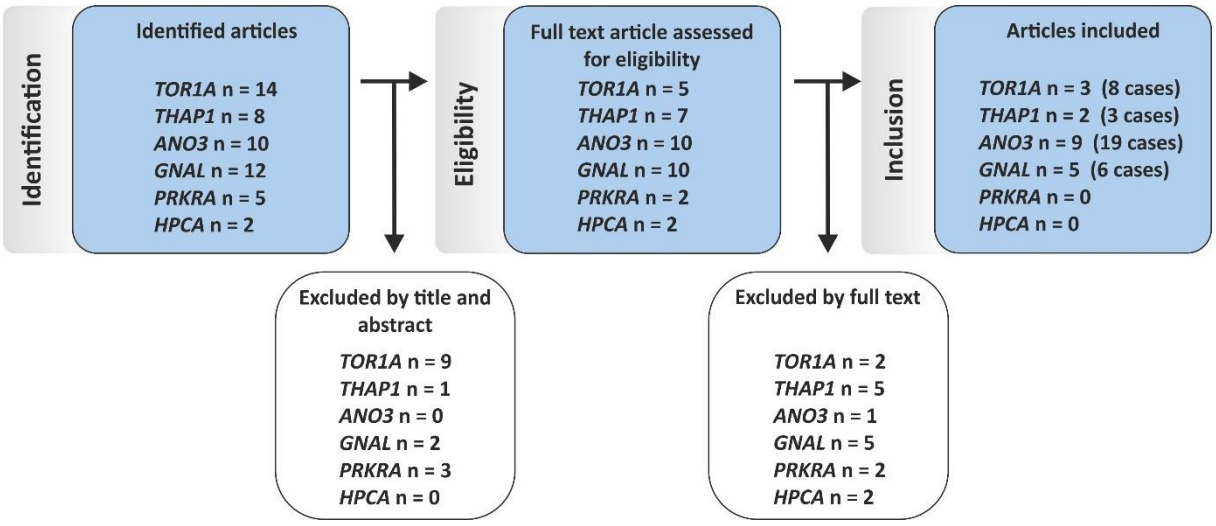

Figure 16 Outline of the PubMed literature search of selected dystonia genes from December 2018 until March 2020. The process involved a three-step assessment of which the number of included and excluded articles after each checkpoint is shown.

As the main project of this thesis was focused on the functional analysis of pathogenic variants of *THAP1*, the following section will summarize the clinical and demographic results of DYT-THAP1 patients. While I contributed to the dataset, the full body of data has been jointly collected by the MDSGene dystonia team (https://www.mdsgene.org; Lange et al., *in preparation*).

In the following, a comprehensive and up-to-date (as of March 2020) overview of the demographic and phenotypic spectrum underlying DYT-THAP1 dystonia is provided. In total, 241 *THAP1* mutation carriers originating from 163 families were identified, 49.8% of which were male (12.4% missing data). The family history was reported to be positive in 64.4%, negative in 35.6% of the cases. Among the reported mutation carriers, Caucasian ethnicity was most common (57.1%, n = 101), followed by Mixed/other (26.9%, n = 47), Asian (14.3%, n = 25), Indian (1.1%, n = 2), and Jewish (0.6%, n = 1) ethnicities (Figure 17). However, 27.4% of all reports did not provide any information regarding the ethnic background. The majority (62.4%) of DYT-THAP1 patients manifested the disease during childhood (age range 2 – 12 years) and adolescence (13 - 20 years) with a median age at onset (AAO) of 15.0 years (IQR 9.00 - 28.25 years; Figure 17B). As a hallmark of DYT-THAP1 dystonia, first symptoms were prominently reported in the upper half of the body (33.0% upper limb, 26.9% neck, 14.7% larynx, and 10.2% craniofacial). Only 9.6% of the cases presented with first symptoms in the lower limbs, 5.6% of cases at ≥two different sites, and none with first symptoms in the trunk (Figure 17C).

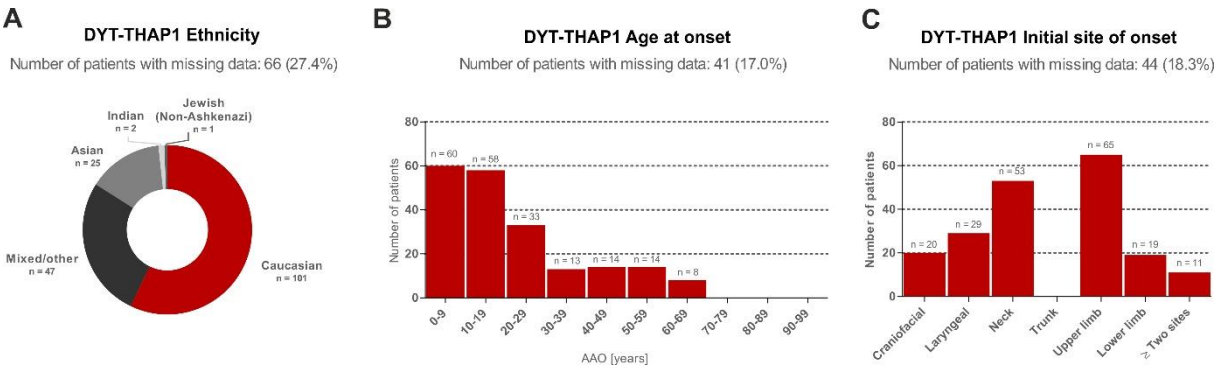

Figure 17 Demographic and clinical characteristics of all published DYT-THAP1 patients regarding (A) ethnicity, (B) age at onset, and (C) initial site of onset (https://www.mdsgene.org).

We observed that generalization of symptoms only occurred in around one-third of patients (34.8%, n = 72; Figure 18). Symptoms spread to multiple regions in 40.6% (n = 84) and remained focal in 22.6% (n = 51). On examination, the most commonly affected body part by dystonia was the neck (73.6%, n = 156). DYT-THAP1 patients were often affected by limb dystonia (n = 145). The limb involvement can be decomposed into arm/hand dystonia (n = 126 and n = 64, respectively) and less frequent leg/foot dystonia (n = 77 and n = 33, respectively). Moreover, dystonic symptoms were reported in the craniofacial region (n = 103), as well as in the larynx (n = 94), the trunk (n = 55), and the vocal cords (n = 38). In addition to dystonia, several patients presented with tremor (n = 40), dysarthria (n = 23), or myoclonus (n = 5). Other motor and non-motor symptoms such as dysmorphic features, short stature, delayed speech and language development, global developmental delay, microcephaly, motor delay, or signs of parkinsonism were not reported for any patient. Cognitive impairment was present in five individuals (Figure 18A). Deep brain stimulation of 33 patients resulted in 36.4% of patients in good to excellent, 42.4% in moderate, and 12.2% in none to minimal/intermittent improvement and remained unreported for 9.0% (Figure 18B).

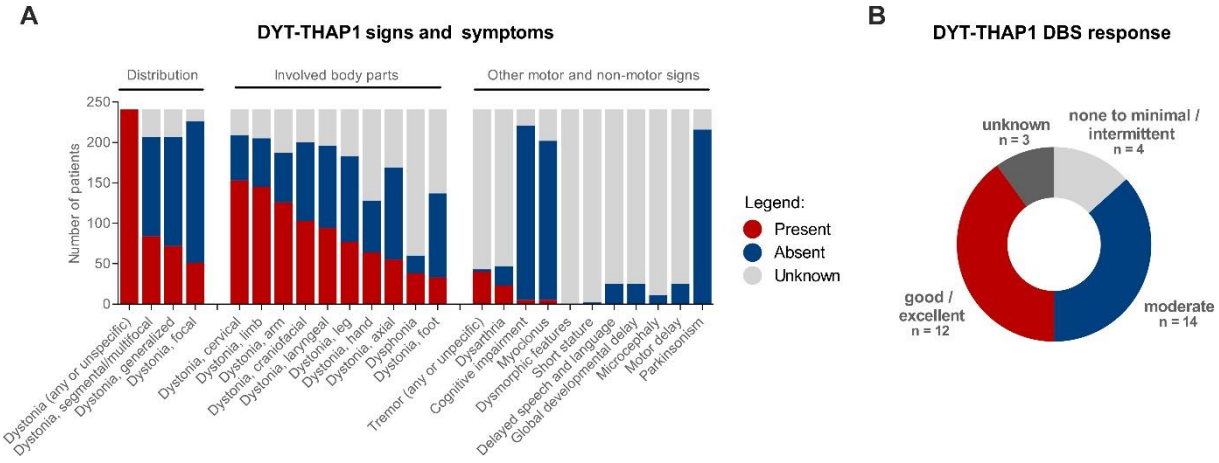

Figure 18 Data from the literature on clinical signs and deep brain stimulation (DBS) outcome in DYT-THAP1 patients. (A) Stacked histogram ranked across various categories in descending order of appearance of signs and symptoms and (B) response quantification chart after DBS.

In summary, DYT-THAP1 is a distinctive genetically determined progressive monogenic dystonia, usually starting during childhood or adolescence. DYT-THAP1 patients often presented with neck and arm dystonia and showed a small range of co-morbidities. Furthermore, DBS stimulation resulted in a marked or moderate improvement in the majority of patients.

5.4.2 Generation and characterization of twenty *THAP1* mutant iPSC lines

Induced pluripotent stem cells (iPSCs) exhibit an excellent potential for neurological disease modelling as they exhibit the potential to be differentiated into neurons. In this project, the role of the transcription factor THAP1 in developing iPSC-derived human cortical neurons was explored by global gene expression analysis. Thus, a set of twenty iPSC lines derived from three THAP1 families (see Figure 5) harbouring three different *THAP1* mutations was generated. Reprogramming of fibroblasts into iPSCs, expansion, banking, and characterization of iPSC-lines was conducted between 01/2017 and 04/2018 (Baumann et al., 2018). Up to eight clones per patient were banked, and of these, two clones were comprehensively characterized. Furthermore, 14 control fibroblast-derived iPSC lines were provided by our collaborators from the stem cells for Biological Assays of Novel drugs and prediCtive toxiCology (stemBANCC) project. The control lines were matched for sex (patients: 30% male vs. controls: 40%

male) and ethnicity (all of the Caucasian ancestry; Table 2). The iPSC control lines were reprogrammed and cultivated in the same way as patient-derived iPSC lines.

Morphologic and molecular characterization of iPSC lines is a crucial quality control step to ensure high-quality iPSC lines for further experiments. The entire set of results is shown in Supplementary Figure S1-Figure S5. In this section, members of Family A will serve as an example throughout the characterization process. First, the pathogenic THAP1 variant of all iPSC-derived lines and fibroblasts was validated by Sanger sequencing. An exemplary set of sequencing electropherograms from individual L-3969 is displayed in Figure 19A. In order to demonstrate the genomic integrity, DNA from parental fibroblasts, as well as iPSC lines, was subjected to quantitative SNP array genotyping (Figure 19B). Insertions or deletions are marked as a bar in green and red colour, respectively. Due to KaryoStudio software incompatibilities with the raw data, the X chromosome in male individuals is recognized as missing as indicated by the red bar. Manual screening confirmed no abnormalities within these regions. All lines displayed a normal karyotype (46 XY) with no large insertions or deletions (defined as >1 Mb). The quantitative SNP arrays and *THAP1* mutational analyses of all lines are displayed in Figure S1 and S2.

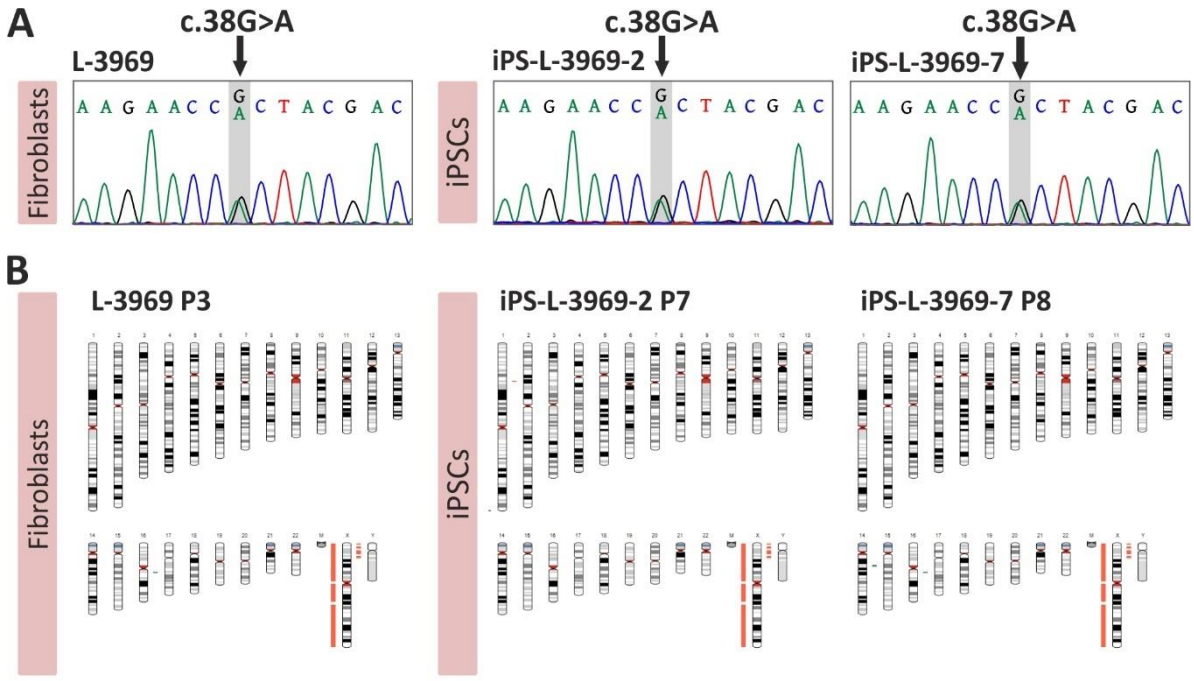

Figure 19 Mutation verification and SNP array analysis of *THAP1*-mutant fibroblast and iPSC lines. (A) Exemplary sequencing electropherograms of the heterozygous pathogenic *THAP1* c.38G>A variant (NM_018105.3) from individual L-3969 marked in grey and iPSC clone 2 and 7 (iPS-L-3969-2 and iPS-L-3969-7, respectively). (B) SNP array results confirm genetically stable iPSC clone (visualized by KaryoStudio).

Subsequently, the complete clearance of Sendai virus in patient iPSC lines was tested by PCR (Figure 20). Specific amplification of the Sendai vector sequence by PCR revealed the absence of Sendai virus (expected band at 181 bp) in all iPSC clones. The house-keeping gene *ACTB* (expected band at 151 bp) was used to show the integrity of the cDNA. The absence of CytoTune™ 2.0 Sendai virus components *MYC*, *KLF4*, *SeV*, and *KOS* of all generated iPSC lines was confirmed by PCR (Figure S4).

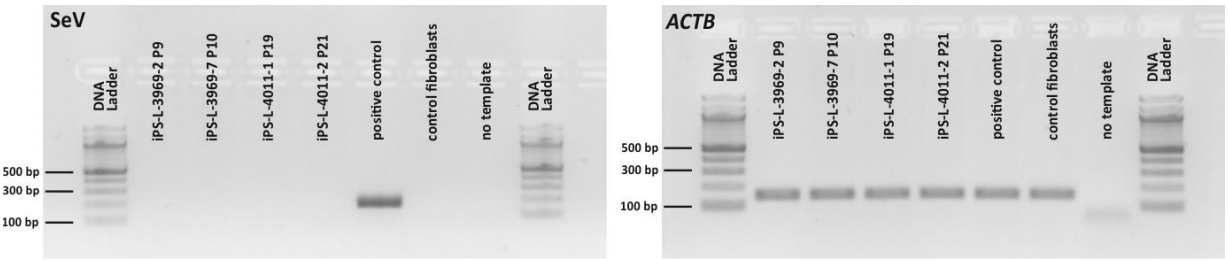

Figure 20 Clearance of Sendai virus in iPSC lines with passages ≥9 using a Sendai virus positive control (fibroblasts seven days post Sendai virus infection), a negative control (fibroblasts), and *ACTB* as a reference gene. SeV = Sendai virus

Next, the pluripotency of the iPSC lines was assessed by immunochemistry of stem cell markers and quantitative RT-PCR. In addition, the potential of all iPSC lines to differentiate into all three germ layers was investigated. All iPSC lines displayed the expected colony-forming morphology with small cells and large nuclei (Figure 21A). Immunofluorescence analysis of patient iPSC lines revealed high levels of endogenous pluripotency marker proteins nanog homeobox (NANOG), stage-specific embryonic antigen 4 (SSEA4), OCT4, and TRA-1-60. In accordance with these results, mRNA expression of *NANOG*, growth differentiation factor 3 (*GDF3*), *OCT4*, and SRY-box transcription factor 2 (*SOX2*) was highly abundant compared to control human dermal fibroblasts (Figure 21B). The capacity of pluripotent stem cells to form cells of all three germ layers was shown by quantitative PCR of spontaneously differentiated embryoid bodies with markers specific for all three germ layers (ectoderm: neural cell adhesion molecule 1 (*NCAM*) and paired box 6 (*PAX6*); endoderm: GATA binding protein 4 (*GATA4*) and SRY-box transcription factor 17 (*SOX17*); mesoderm: msh homeobox 1 (*MSX1*) and RUNX family transcription factor 1 (*RUNX1*)). The variable gene expression levels of germ layer markers are a consequence of spontaneous differentiation of iPSC aggregates into embryoid bodies (Figure 21C and Figure S3).

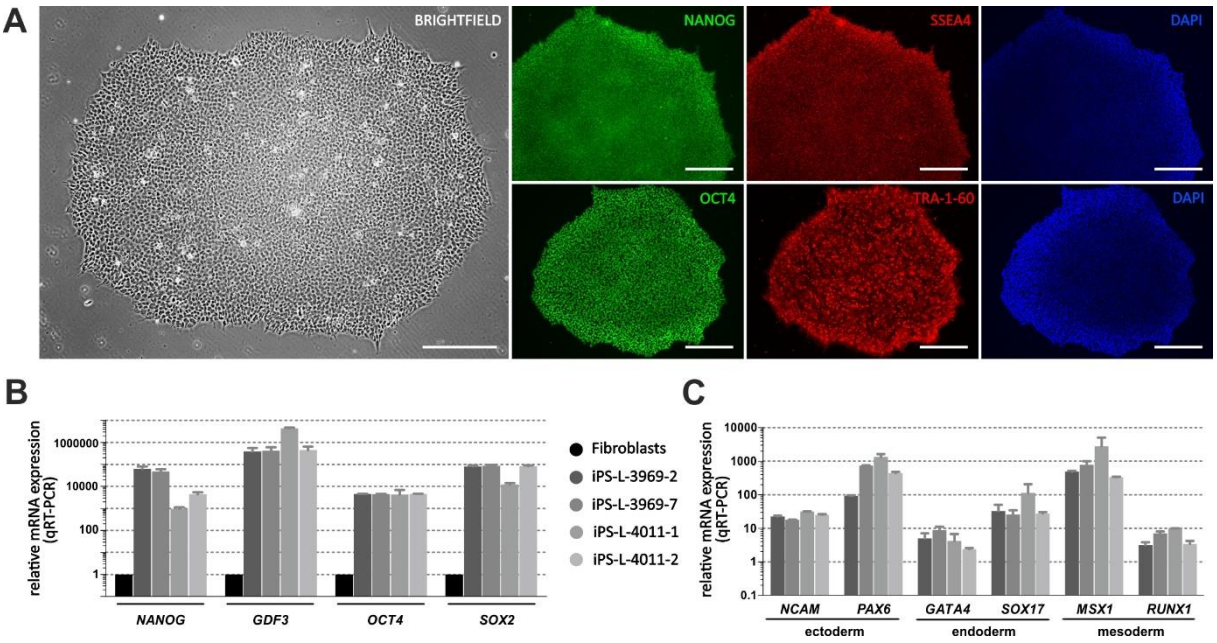

Figure 21 Functional characterization of iPSC lines. (A) Cellular morphology and expression of pluripotency markers in iPS-L-3969-2. Scale bar represents 300 µm. Relative mRNA expression of (B) pluripotency markers in iPSC lines compared to control fibroblasts and normalized to *ACTB*. (C) Gene expression of specific germ layer markers upon spontaneously differentiated iPSC lines into embryoid bodies compared to the respective iPSC line and normalized to *ACTB*.

In summary, twenty human dermal fibroblast-derived iPSC lines of ten *THAP1* mutation carriers harbouring three different *THAP1* mutations passed the comprehensive quality control. The characterization encompassed analyses regarding the assessment of genome integrity, the absence of Sendai virus, the verification of the pathogenic *THAP1* variants, and the confirmation of pluripotency.

5.4.3 Neural differentiation of human iPSCs into neuron

The lack of overt histopathological changes in human brains of dystonia patients creates a conceptual challenge with regard to which cell type at what time point is relevant to disease pathogenesis. We assumed that THAP1, as a transcription factor, impacts gene expression also in cortical neurons during the early development of the brain. Under this assumption, an endogenous iPSC-derived neuronal *in-vitro* model was generated to be able to harvest precursor neurons at an early stage of cortical differentiation. An established differentiation protocol was used, which was adapted from Dr. Karen Grütz (Institute of Neurogenetics, Lübeck; Grütz et al., 2017), as shown in Figure 22A. Sendai virus-free iPSC lines were subjected to neural differentiation over a period of 90 days and cell pellets were harvested at different time points during differentiation (Table S3). Neural differentiations were experimentally conducted between 02/2018 and 12/2018.

The patient-specific identity of iPSC-derived cortical neurons was proven by STR analysis of five polymorphic loci (together with the assistance of Joachim Weber, Master student). The neuronal cultures exhibited matching DNA profiles with blood from the donor except for one marker in L-3736 (Table S4). To exclude a sample mix-up, STR analysis of six additional polymorphic markers in L-3736 was performed with matching profiles in neuronal cultures and blood, indicating that the repeat expansion likely occurred during cell culture.

To illustrate the cellular morphology during differentiation, brightfield microscopic images of different stages during differentiation are shown in Figure 22B. While iPSC colonies prior to neuronal induction exhibited a flat colony-forming character (day 0), a timely coordinated supplementation of factors necessary for human corticogenesis resulted in a morphologic change. Neuralization of human iPSCs was achieved by dual SMAD inhibition, followed by replating of spheres (day 30) with the appearance of round structures indicative of the formation of columnar neuroepithelial rosettes. After single-cell replating on day 44, neural precursor cells were morphologically characterized by short cellular extensions. After several weeks of maturation, cells appeared as interconnected cell clusters with elongated cellular projections resembling axons and dendrites (day 90).

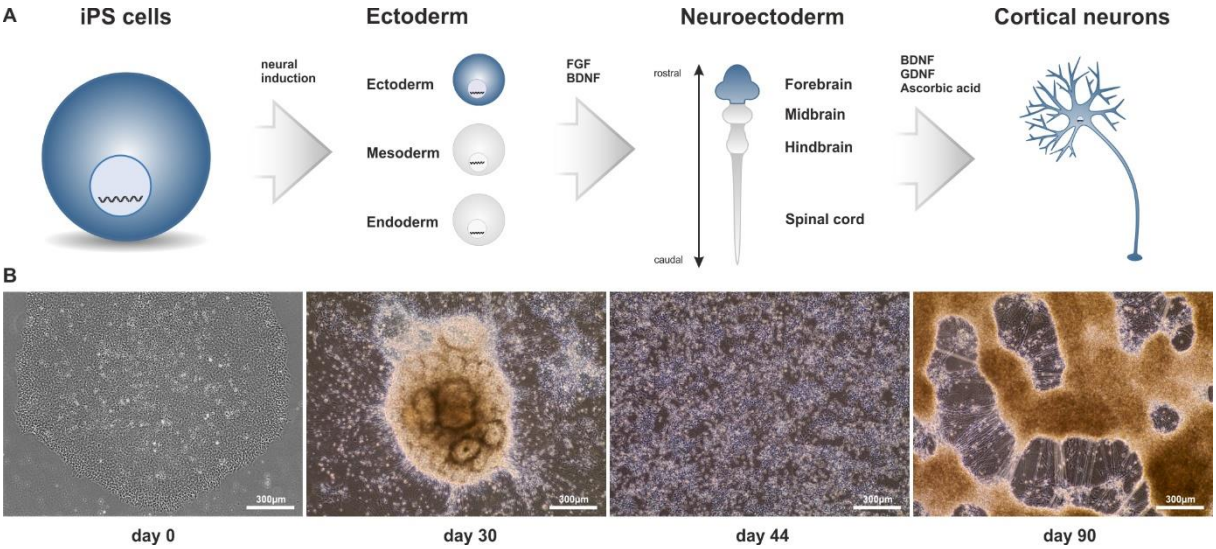

Figure 22 Stages of cortical differentiation. (A) Schematic representation of the cortical differentiation protocol used in this work and corresponding (B) morphologic transition from iPSCs (day 0), rosette stage (day 30), single-cell suspension (day 44), to mature neurons (day 90) captured by brightfield microscopy.

As a next step, using immunochemistry, it was focused on protein markers specific to deep POU class 3 homeobox 2 (POU3F2, also BRN2) and upper-layer COUP-TF-interacting protein 2 (CTIP2) cortical neurons to assess the adoption of glutamatergic identity after 90 days of differentiation. For this purpose, cells were fixed and permeabilized on day 90 of differentiation, immunoblotted against neuronal markers, and visualized by confocal microscopy (Figure 23). Immunostaining revealed that iPSC-derived neurons were positive for the pan-neuronal markers microtubule associate protein 2 (MAP2) and class III β-tubulin (TUJ1) signals, which appeared as thin elongated extensions resembling axons and dendrites (Figure 23). To further demonstrate the cortical identity, the presence of the transporter necessary and sufficient for the uptake of glutamate, the vesicular glutamatergic transporter 1 (vGLUT1), was investigated. Upon immunostaining with vGLUT1 antibodies, signals co-localized with MAP2 signals and were mainly found around 4′,6-Diamidin-2-phenylindol (DAPI) signal positive nuclei. Moreover, mature cortical neurons were immunoreactive for cortical deep-layer marker (BRN2), and upper layer marker CTIP2.

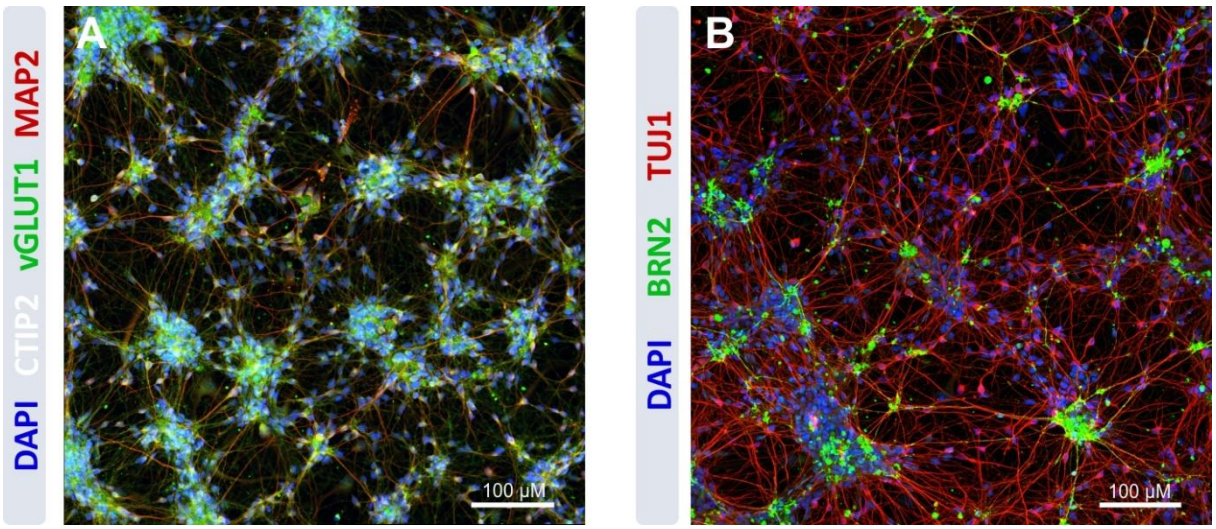

Figure 23 Immunofluorescence micrographs of 93 day-old iPSC-derived glutamatergic neurons (Co_9_086-03-03-02A). Neurons were stained with antibodies against neuron-specific (MAP2 and TUJ1) and cortical markers (vGLUT1, CTIP2, and TUJ1) and mounted with a DAPI-containing medium. Scale bar represents 100 µm.

In conclusion, after differentiation of human iPSCs-derived cells, an expected neuron-like morphology of neural precursors and mature neurons was observed. It was demonstrated that the cells derived from human iPSCs acquired cortical identity after 90 days of differentiation as measured by immunostaining of specific cortical marker proteins.

5.4.4 Whole transcriptome analysis of 44-day-old cortical neurons

To elucidate the dysregulated transcriptional signature caused by mutant THAP1 during early neuronal development, an unbiased global transcriptome analysis with patient iPSC-derived developing neurons was conducted. It is hypothesized that pathogenic mutations in *THAP1* may cause transcriptional dysregulation of THAP1 downstream targets, which are involved in DYT-THAP1 pathogenesis. The pellets of 44-day-old neurons were harvested as described in section 4.3.2.9. The RNA quality was assessed by the RNA integrity number (RIN) with a mean of 9.9 (range 9.5 – 10; Table S5). The theoretical basis to process whole transcriptome datasets was acquired during the European Molecular Biology Laboratory (EMBL) course "Whole Transcriptome Data Analysis" in Heidelberg and later established in the Institute of Neurogenetics. The bioinformatic workflow used an open-source, reproducible, and efficient software pipeline (docker4seq; Beccuti et al., 2018) that was installed on a high-performance NUC6I7KYK mini-computer (Seqbox) in a Linux environment.

Evaluating the quality of the raw data is of central importance for the subsequent biological interpretation. In order to characterize the quality of the raw reads, the median read qualities across the full read length were investigated (Figure 24A). The Phred score is a measure to characterize nucleobases by automated sequencing. Phred categories range from good (28-40), reasonable (20-27), to poor (0-19). Usually, single-end sequencing results in a noticeable decline of read qualities along the length of the reads. In contrast, paired-end sequencing typically shows a homogeneous mean quality along the length of the reads. The median Phred score of the paired-end reads in this study showed consistent values with a global median of 35.5 and did not fall below 33 at any position. The degree of duplications of the first 100,000 sequences was unremarkable and passed the quality control (Figure 24B). Specifically, a median of 84.9% was constituted by unique reads (range 81.0% - 89.6%), 15.1% by duplicate reads (range 10.4% - 19.0%), 3.0% by reads found in >10 times (range 0.4% - 5.8%), and 11/32 samples exhibited at least one read that was measured more than one hundred times in the dataset.

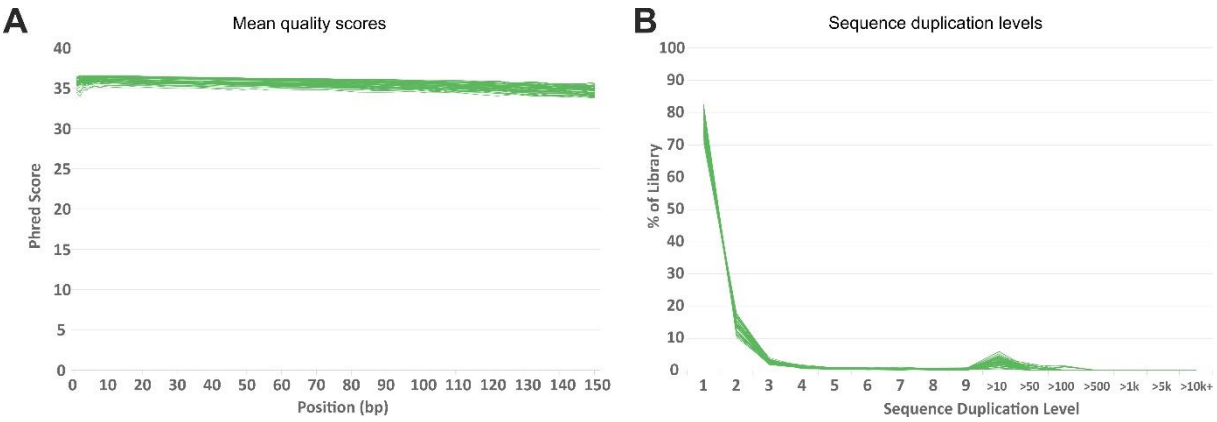

Figure 24 Quality metrics of raw sequence reads used for transcriptome analysis. (A) Line plot showing mean Phred quality scores for each base in the read and (B) relative proportion of duplicate counts.

To detect a possible enrichment bias of the independent cortical differentiations in the eleven samples with highly duplicated reads, a quality assessment of the forward and reverse strand sorted by ascending cortical differentiations is displayed in Figure 25 (eleven samples are underlined). As a result, the sum of top overrepresented sequences (defined to constitute more than 0.1% of the dataset) represented 2.5% (range 0.1% - 5.7%) of the total amount of reads in all samples. Notably, the R1 sequencing direction was more prone to read multiplication, indicating a technical issue rather than a biological origin (Figure 25A). The overrepresented sequences did not show any apparent batch effects between the differentiations. The GC-content was roughly Gaussian distributed with a median of 57% GC-content with two major peaks at 52% and 60% GC-content. The elevated GC-content at 52% was present only in the R1 forward sequencing direction (median group difference to R2 sequencing direction 0.8%), likely reflecting the same technical bias as seen in Figure 25A. The second peak at 60% GC-content was shared by all samples.

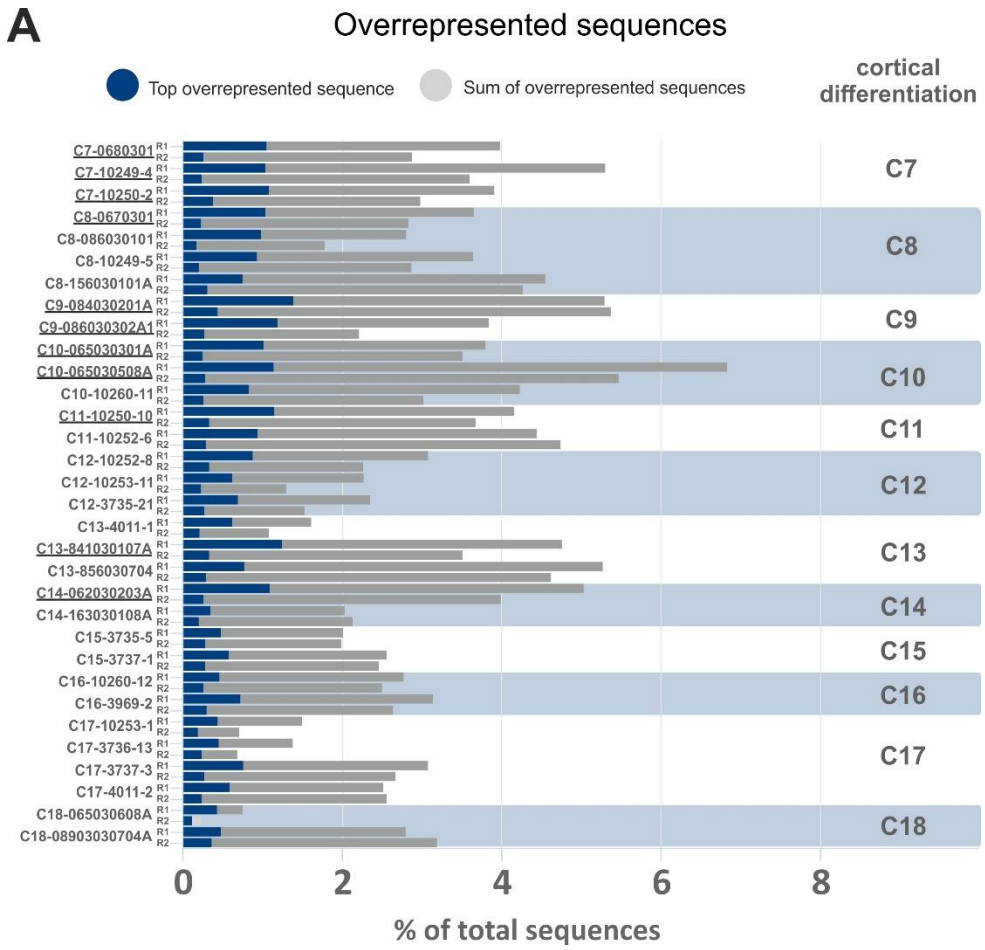

Figure 25 continued

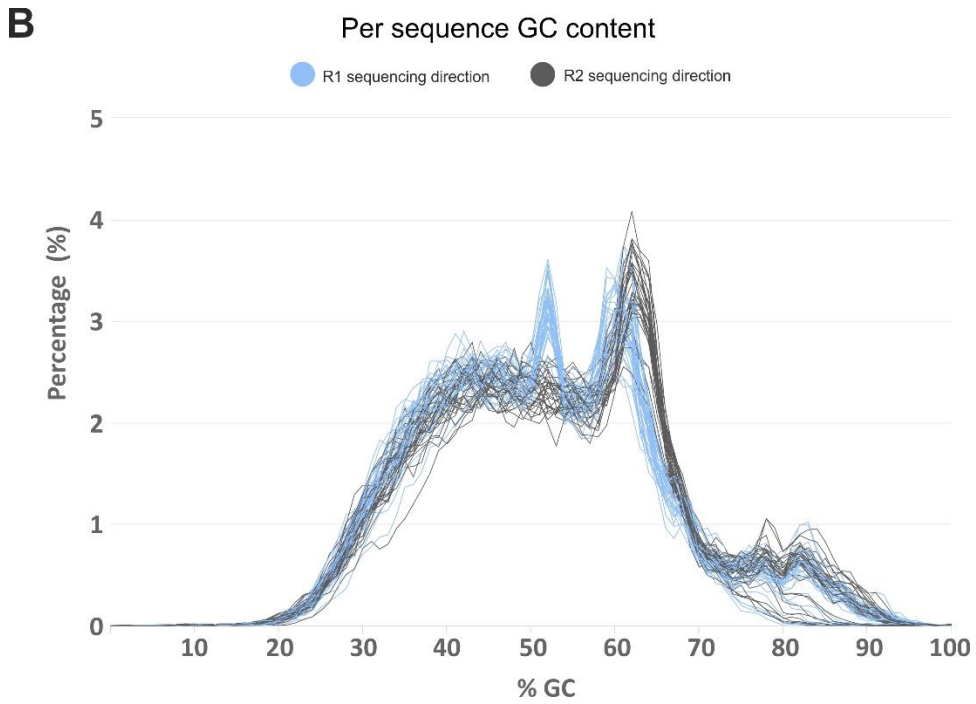

Figure 25 Raw sequence reads assessment sorted by cortical differentiations. (A) Relative amount of top- and the sum of overrepresented sequences in the library sorted by cortical differentiations (C7 - C18) and sequencing direction. Samples with >100 read duplication are underlined. (B) Per sequence GC content graph sorted by sequencing direction.

The removal of adapter sequences of the reads was performed efficiently using the Skewer software (Figure S6). The trimmed reads were mapped using STAR to the human reference genome GRCh38 and subsequently normalized to the gene length (Dobin et al., 2013). Mapping statistics are a simple and intuitive way to assess the quality of RNA sequencing. Upon STAR alignment, an average of 87.5% reads was uniquely mapped (mean of 67.2 Mio; range 49.7 - 85.0 Mio), indicating a high data quality. The percentages of reads aligned to the reference genome of each sample is shown in Figure 26.

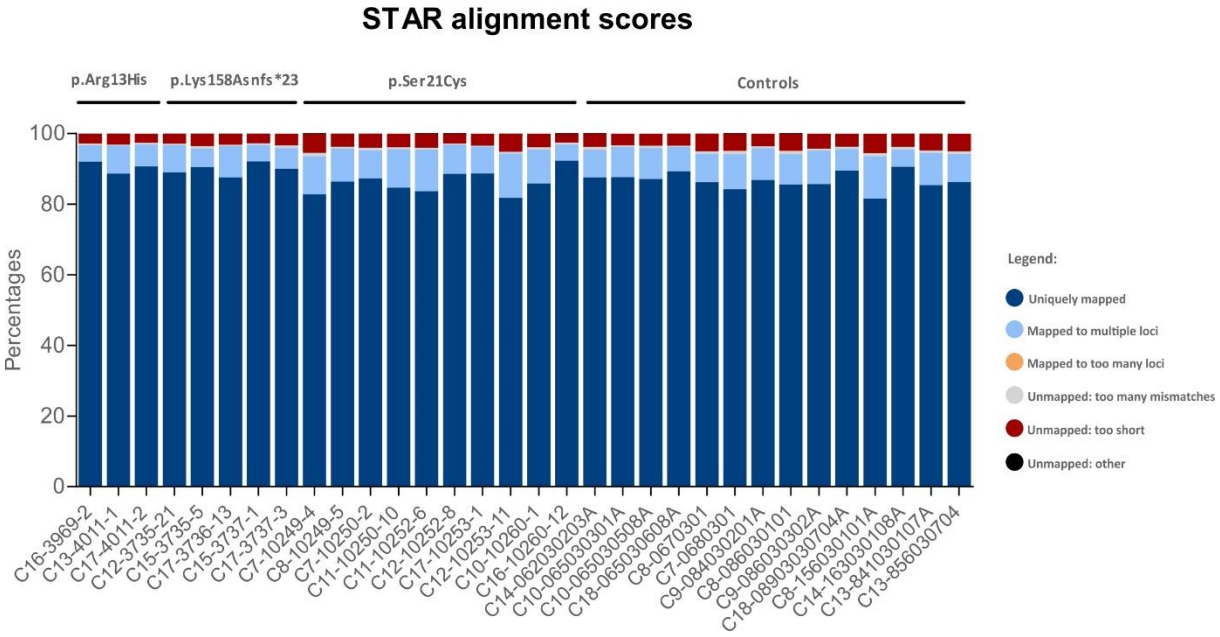

Figure 26 STAR alignment scores sorted by *THAP1* mutations. Relative mapping percentage of 32 neuronal cultures is displayed in a stacked bar plot in which most reads were mapped to a single locus to the reference genome (uniquely mapped).

In summary, the library construction, the quality control of the raw reads, and mapping of the reads to the reference genome were successful. All samples were used for further analysis.

5.4.5 Exploratory data analysis

The exploratory data analysis is an approach to summarize the main characteristics of a dataset by visual methods. A principal component analysis (PCA) is often used to decompose a complex dataset into the most variable components. Here, the gene length-normalized (TPM, transcripts per kilobase million) expression matrix was used to associate global gene expression with clinical status and gender by principal component analysis. During the course of the exploratory data analysis, three control lines (of 32 samples) were excluded from further analysis, based on recurrent outlier status in the PCA as well as unsupervised heatmap clustering (Figure 27).

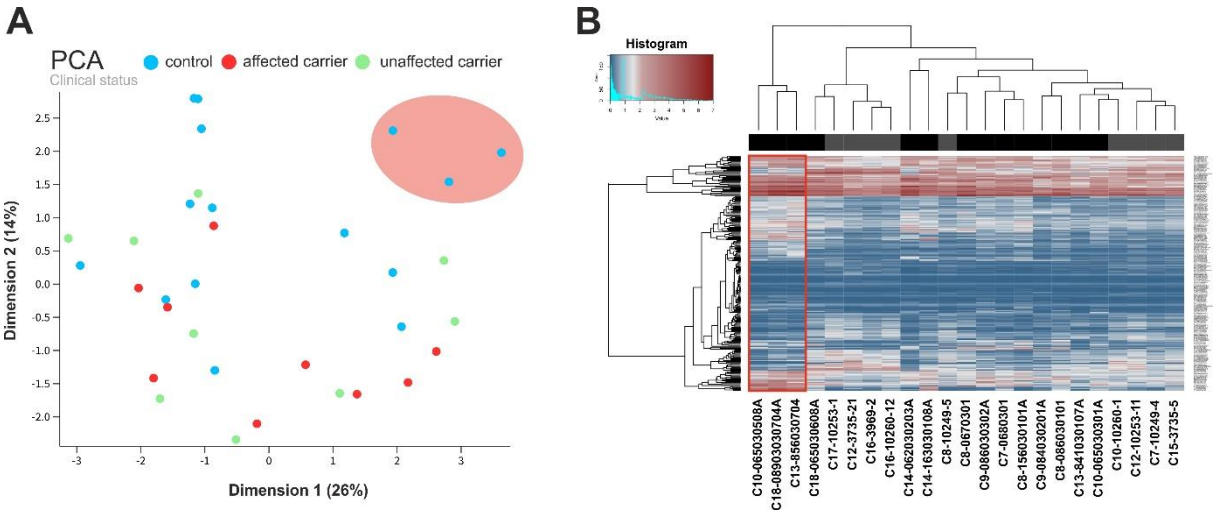

Figure 27 Exclusion of three control samples (C10-065030508A, C18-08903030704A, and C13-856030704) from further analysis. (A) PCA analysis revealed that all three samples are clustered together (light red circle) (B) The heatmap shows 245 differentially expressed genes (DEGs) between affected *THAP1* mutation carriers and controls. By unsupervised clustering three samples were detected as outliers as indicated by the leftmost dendrogram branch (red rectangle).

A second PCA analysis with 29 samples was conducted. Principal component coordinate 1, 2, and 3 accounted for 26%, 13%, and 9% of the total variance, respectively (Figure 28). As seen in Figure 28A, a distinction between the group of controls, affected and unaffected individuals was not apparent. In order to identify biological subgroups or technical factors, various covariates were taken into account, such as cortical differentiation batches, age of the probands, and different *THAP1* mutations (data not shown). These analyses did not show remarkable clusters. As shown in Figure 28B, a marked distinction in the principal coordinates 2 and 3 was related to gender.

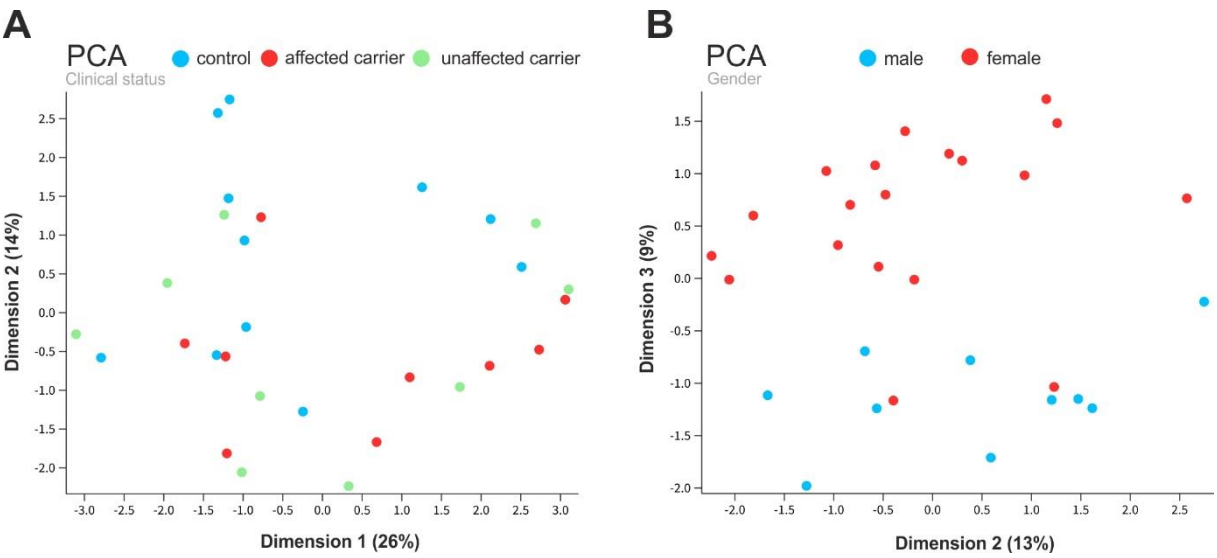

Figure 28 Principal component scatterplot of 29 cultures *in-vitro* grown neurons harvested at day44. (A) Principal component 1 and 2 of the clinical status (affected *THAP1* mutation carriers n = 9, red; unaffected *THAP1* mutation carriers n = 9, green; and control samples n= 11, blue). (B) Gender-specific clustering in principal component 2 and 3.

Assessment of global gene expression by PCA was indistinctive with regard to clinical status, age at biopsy, batches of differentiations, or different *THAP1* mutations. An association between global gene expression and gender was observed by PCA analysis.

Differential expression analysis can be performed to statistically analyse gene expression differences between sample groups. To address the question of altered gene expression in *THAP1* mutation carriers and to assess gene networks regulated by mutant *THAP1*, differential gene expression abundance estimates were performed between "all THAP1 carriers", "affected carriers", and "unaffected carriers" each versus the group of controls using DESeq2 (Love et al., 2014). The difference in expression was visualized as effect score ($\log_2$ fold change) versus confidence score ($-\log_{10}$ adjusted p-value) for ~20,000 genes of the human genome in a volcano plot (Figure 29).

The threshold to define a candidate gene is dependent on each analysis. The volcano plot was used to define boundaries (regarding log2 fold change and adj. P-value) to define candidate genes. Overall, fold changes were modest, ranging to a maximum of -2.3 and +3.9 and adjusted p-values up to 3.4×10^{-8}. Differentially expressed genes (DEGs) were considered as candidate genes with a $\log_2 fc > 0.8$ (equals >1.7-fold change) and adj. p-value < 0.1 (displayed in red). To reflect the fact that several genes had high confidence scores with low fold changes, genes with $\log_2 fc < 0.8$ and for higher specificity an increased adj. p-value of < 0.05 were also defined as candidate genes. The number of resulting genes is displayed in Figure 29C. By comparing gene expression abundance of maturing neurons (day 44) of affected mutation *THAP1* mutation carriers (n = 9 samples) with controls (n = 11 samples), 104 genes were called differentially expressed with 74 genes exclusively found in affected individuals (Figure 29A). Strikingly, the group of unaffected carriers exhibited only five DEGs of >1.7-fold with $\log_{10} p < 0.1$ compared to controls, while four genes were shared with the other two groups. All *THAP1* mutation carriers (n = 18) showed 63 DEGs when compared to controls (n = 11), while around half (n = 30) were identical with DEGs in affected individuals. The number of unique candidate genes resulting from these three analyses represented 0.00687% (n = 137/ 19919) of the total number of analysed transcripts and was compiled as a list of candidate genes in Supplementary Table S2. The number of significantly up- or downregulated genes within the merged list was comparable (68 and 69 genes, respectively).

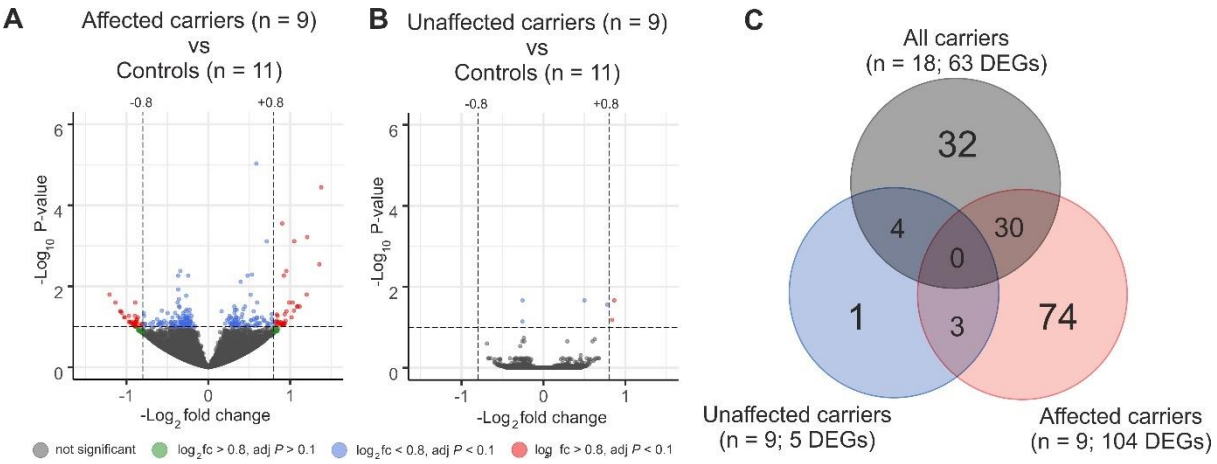

Figure 29 Global transcriptomic changes in developing neurons of THAP1 mutation carriers. (A) Volcano plot of DEGs in iPSC-derived cortical neurons harvested at day 44 between affected or (B) unaffected *THAP1* mutation carriers versus control. Red dots represent DEGs (with log_2fold change > 0.8 and adj P < 0.1) meeting the requirements for candidate genes (C) Venn diagram illustrating the number of DEGs of selected subgroups versus controls.

Intriguingly, differential gene expression in developing neurons revealed a signature that was unique to affected individuals and not present in unaffected individuals demonstrating altered gene expression that correlates to disease manifestation of DYT-THAP1 patients.

To address the question of whether the set of candidate genes is clustered in specific genomic regions, such as in close vicinity of THAP1, a chromosomal map was constructed based on the combined candidate genes (Figure 30). None of the candidate genes was located in the topologically associated domain around the mutated gene *THAP1*. Six candidate genes were particularly clustered on the q-arm of chromosome 19. Furthermore, twelve clusters of genes in close vicinity were both up- or downregulated (marked with a rectangle in Figure 30). To correlate our candidate genes to known THAP1 binding sites discovered by the encyclopaedia of DNA elements (ENCODE) project, all gene symbols were converted into unique Ensembl gene identifiers (ENSG) using the R-package BioMart. Six genes predicted to be regulated by binding of the transcription factor *THAP1*, exhibited a random distribution across the genome (marked in blue in Figure 30). These genes include the transmembrane protein 79 (*TMEM79)*, ubiquitin specific peptidase 21 (*USP21*), transporter 2, ATP binding cassette subfamily B member (*TAP2*), GTP binding protein 3, mitochondrial (*GTPBP3*), plakophilin 3 (*PKP3*), and PC-esterase domain containing 1A (*PCED1A*)).

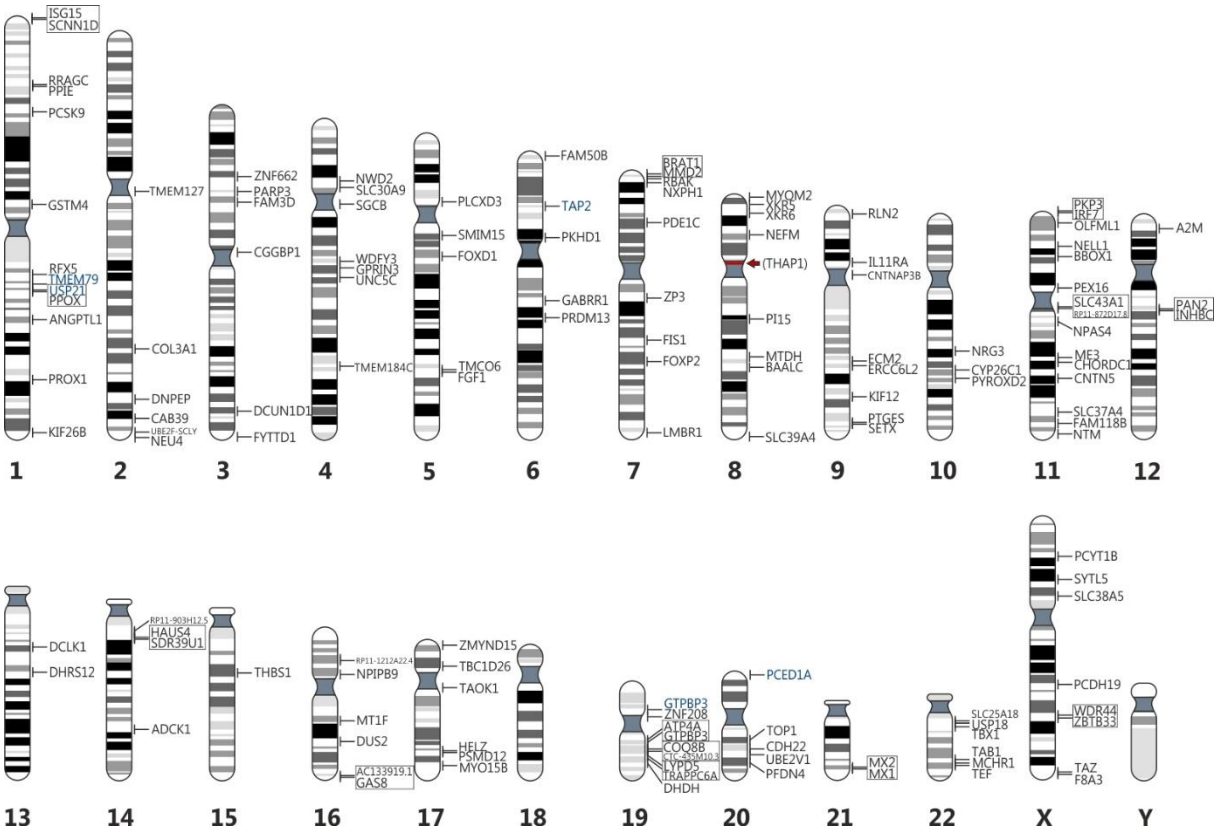

Figure 30 Genome-wide localization of possible disease-contributing genes derived from differential expression in developing neurons. The combined candidate list (137 genes) of *THAP1* mutation carriers and controls in maturing neurons (day 44) is displayed. Adjacent gene with a shared up- or downregulation were marked with a rectangle. Putative THAP1-regulated genes discovered by the ENCODE project are coloured in blue. The full gene names are listed Table S2.

In conclusion, a genome-wide distribution of candidate genes identified 26/137 genes to be regulated in pairs suggesting a shared transcriptional regulation by the transcription factor THAP1.

Subsequently, enriched biological processes of the DEGs were explored by gene ontology (GO) analysis. Typically, GO analysis lists the enriched GO-terms in numeric order of the p-value without accounting for the fact that GO-terms are redundant in different levels of hierarchy and therefore, could be grouped in a semantic fashion. The Database for Annotation, Visualization, and Integrated Discovery (DAVID; https://david.ncifcrf.gov) allows to group functionally enriched biological GO-term annotations based on shared molecular functions thereby allowing for a more meaningful biological representation of gene sets (Hosack et al., 2003). Among the highest enriched biological processes within the set of upregulated genes were oxidoreductase activity, immune response, GTPase activity, transmembrane protein, ATP-binding, and signal peptide each represented by a cluster of genes in Figure 31A. Within the set of downregulated genes the most enriched biological processes were related to circadian rhythm (prospero homeobox 1 (*PROX1*), senataxin (*SETX*) and topoisomerase 1 (*TOP1*)), integrin binding and cell adhesion, transcription factor activity and zinc finger binding, calcium ion binding, epidermal growth factors, and WD1-5/40 domain (Figure 31B). Noteworthy, many GO-terms were not assigned to a cluster of biologic processes. Among those, axon guidance (forkhead box D1 (*FOXD1*), kinesin family member 26B (*KIF26B* and unc-5 netrin receptor C (*UNC5C*)), and neurogenesis (doublecortin like kinase 1 (*DCLK1*), neuronal PAS domain protein 4 (*NPAS4*), and *SETX*) were context-specific categories with putatively relevant implications in DYT-THAP1 disease pathogenesis.

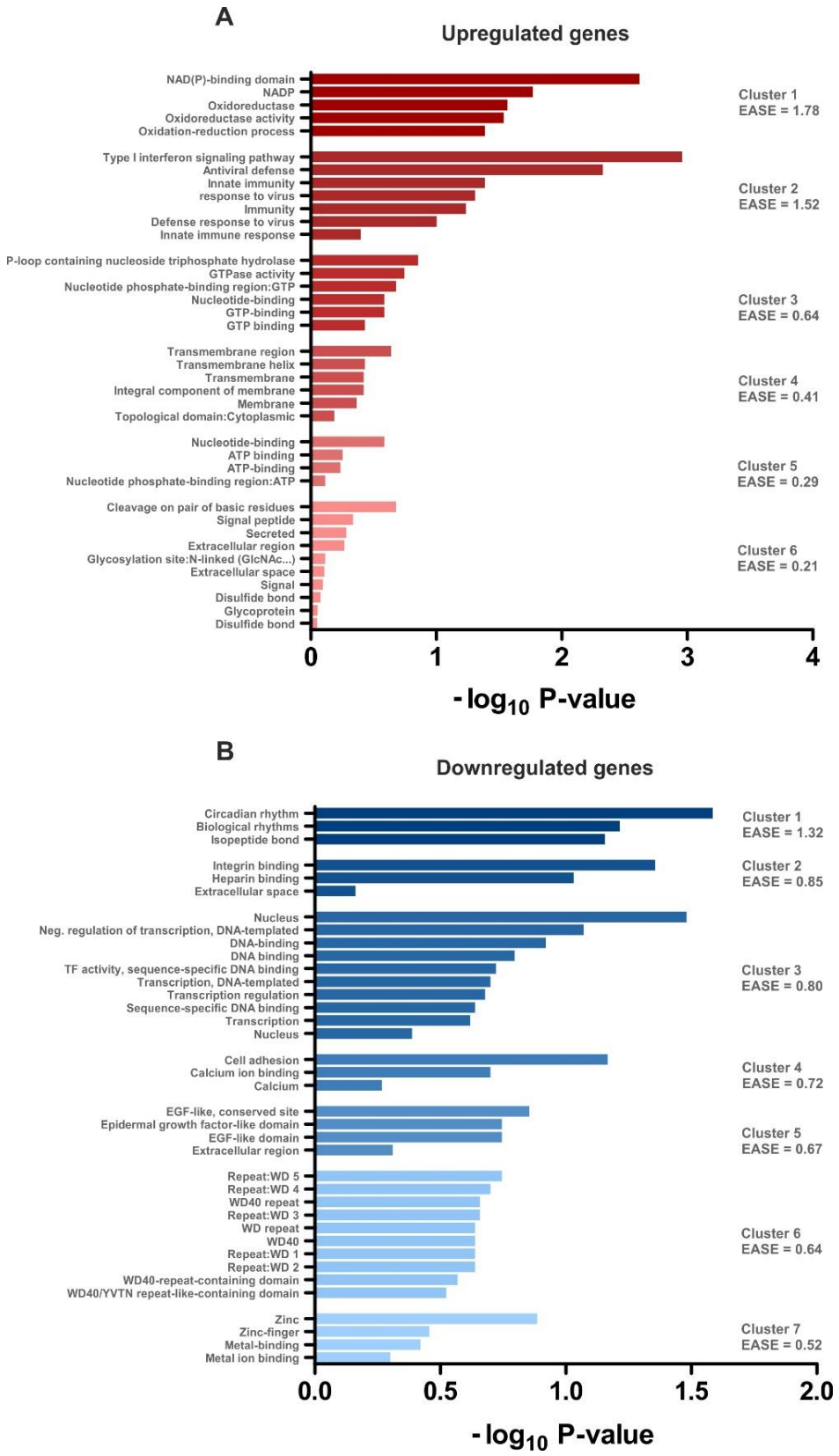

Figure 31 Gene ontology enrichment of DEGs in maturing cortical neurons. Significantly enriched biological processes of the combined list of candidate genes sorted by (A) up- and (B) downregulated genes and displayed in enriched clusters of similar GO-terms ordered by the enrichment Expression Analysis Systematic Explorer (EASE) score.

To account for biological networks, a similarity-based GO enrichment analysis was carried out. The GO-term analysis provides a link that THAP1-sensitive candidate genes were associated with defective nucleus localization and cell adhesion with upregulation of immune response genes and oxidoreductases.

A major aspect of omics approaches is the interpretation of candidate genes. Thus, candidate genes were selected based on putative function, relevance to dystonia, animal models of neurological diseases, disease association in the OMIM registry, and annotated in a neurological GO-category in human and mouse databases. To focus on relevant neurological genes, information various databases including Mouse Genome Informatics (MGI) and the Database for Annotation, Visualization, and Integrated Discovery (DAVID) regarding neurological phenotypes in mouse (31 hits) and human neurological association (29 hits) were retrieved. Candidate genes present in at least one category were selected for visualization and sorted by differential fold expression (Figure 32). A total of 78 of the 137 DEGs were listed in at least one of the six categories extracted from different databases. A confirmed hereditary cause of disease in the OMIM database was found in 26/137 genes of which four genes were linked to movement disorders in a broader sense: Forkhead box P2 (*FOXP2*) to dyspraxia, a motor-speech disorder with primary difficulties in the planning and execution of movements (OMIM#602081), *SETX* to amyotrophic lateral sclerosis (OMIM#606002), protocadherin 19 (*PCDH19*) to epileptic encephalopathy (OMIM#300088), and BRCA1 associated ATM activator 1 (*BRAT1*) to rigidity and multifocal seizure syndrome (OMIM#614498). Moreover, the TAO kinase 1 (*TAOK1*) gene has recently been implicated in neurodevelopmental delay (Dulovic-Mahlow et al., 2019).

To provide added power in assessing the pathogenic impact of the candidate genes, analyses regarding tolerance to Loss of Function (LoF) variants based on expected and observed number of respective variants in a particular gene were carried out. Genes with a high probability of LoF variation encompass the majority of known haploinsufficient genes and therefore may be more vulnerable to modulated gene expression. Overall, 22/137 genes exhibited a probability of being intolerant to a Loss of Function variant with a (pLI) score 0.9 or higher. Interestingly, 21/22 of LoF-intolerant candidate genes were found amongst the downregulated genes, possibly with insufficient protein being produced in *THAP1* mutation carriers.

Additionally, it was particularly looked for genes previously linked to dystonia, using published causative or risk factor genes as a reference (Siokas et al., 2019), but no overlap was detected. Notably, WD repeat domain 44 (*WDR44*) and beta-sarcoglycan (*SGCB*), which were among the DEGs, belong to protein families known to cause neurological disorders (such as WDR45 causing Neurodegeneration with brain iron accumulation 5 (OMIM#300894) and SGCE causing myoclonus dystonia DYT-SGCE OMIM#159900)). However, *WDR44* and *SGCB* have not been associated with a neurological disease so far. A total number of six solute carrier (SLC) genes was found among the dysregulated genes (*SLC25A18*, *SLC30A9*, *SLC37A4*, *SLC38A5*, *SLC39A4*, and *SLC43A1*) performing various functions ranging from glutamate, zinc, glucose-6-phosphate, glutamine and neutral protein transport.

As depicted in Figure 32, two genes were listed in five categories (*PCDH19 and* forkhead box 2 (*FOXP2)*) and four genes appeared in four categories (contactin 5 (*CNTN5*), neuregulin 3 (*NRG3*), ras related GTP binding C (*RRAGC*), and *UNC5C*), which were all found to be downregulated by mutant THAP1.

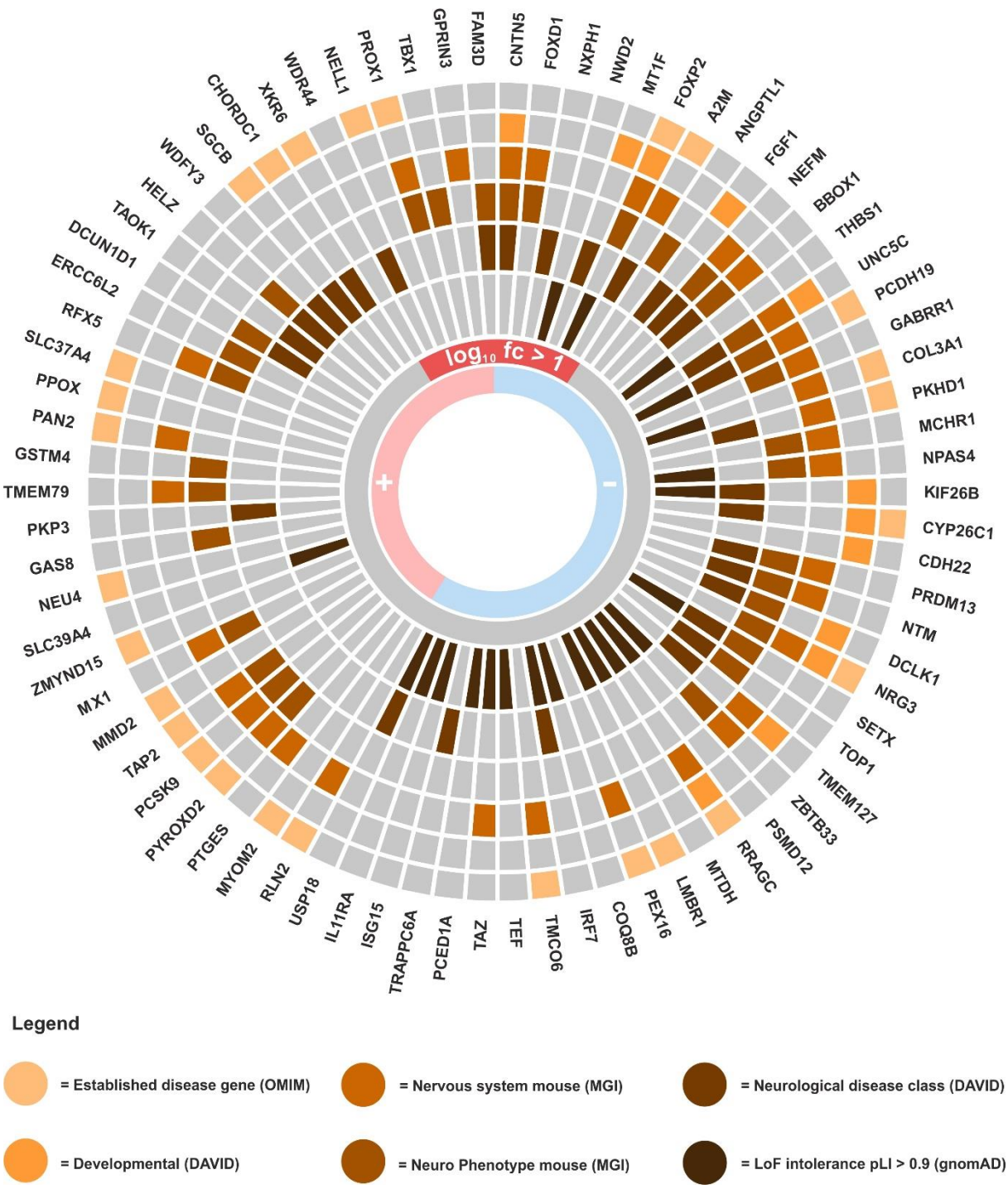

Figure 32 Integrated database search of RNA-Seq THAP1-sensitive candidate genes sorted by fold change. The presence of a GO-term is marked by a coloured square. The innermost ring indicates whether genes are up- (+) or downregulated (-) while the outer ring displays genes with the greatest fold change ($log_2fc > 1$). pLI = probability of loss of function, OMIM = Online Mendelian Inheritance in Man, MGI = Mouse Genome Informatics, DAVID = Database for Annotation, Visualization, and Integrated Discovery.

In conclusion, more than half of all candidate genes were related to neurodevelopmental GO-terms. Four genes were established disease-causes of neurological disorders. An enriched proportion of candidate genes, which was evolutionary constrained against LoF-variants, was found amongst the group of downregulated DEGs. Perhaps the most relevant DEGs based of this search include *FOXP2*, *PCDH19*, *CNTN5*, *NRG3*, *RRAGC*, and *UNC5C*.

5.4.6 Gene expression in the context of reduced penetrance

THAP1 is known for its capacity to bind its own promotor, thereby regulating its own expression (Erogullari et al., 2014). Previous experiments in dystonia patients (e.g., carrying a *KMT2B* mutation in section 5.1) revealed that the underlying cell type is paramount to detect altered disease gene expression. In the following step, the impact of different *THAP1* mutations on the transcriptional regulation in developing neurons (day 44 of cortical differentiation) of *THAP1* itself was assessed (Figure 33). Focusing on the differences between affected and unaffected *THAP1* mutation carriers in iPSCs lines, we can observe a significant *THAP1* expression difference between affected and unaffected family members harbouring the protein-truncating p.Lys158Ansfs*23 mutation. In maturing neurons at day 44, a uniform pattern of transcriptional upregulation of *THAP1* in affected carriers compared to unaffected carriers was detected. Statistical significance was reached in two families (p.Lys158Asnfs*23 and p.Ser21Cys) while p.Arg13His mutation carriers showed similar trends but did not reach statistical significance. Notably, the family harbouring the p.Arg13His is the smallest family with the least samples that were studied.

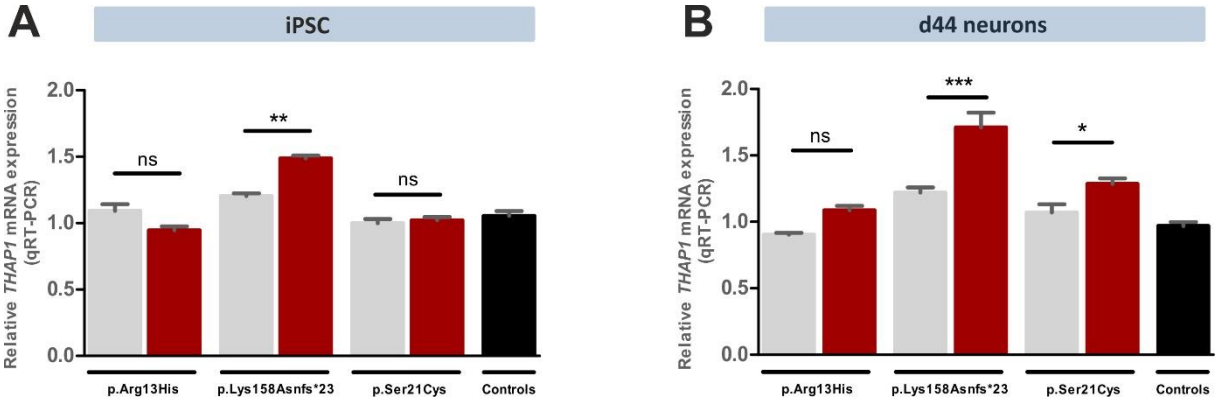

Figure 33 Comparison of *THAP1* gene expression in iPSC lines and developing cortical neurons. *THAP1* expression is grouped by *THAP1* mutations and clinical status and normalized to the house-keeping gene *GAPDH* in (A) iPSC lines and (B) developing neurons day 44 (9 lines of unaffected carriers: grey bars, 9 lines of affected carriers: red bars, and 12 control iPSC lines and 14 control neurons control lines: black bars). Experiments were carried out independently three times with two replicates, values represent mean ± SEM. For statistical analysis, one-way ANOVA followed by Bonferroni's post-hoc test was conducted. *P < 0.05; **P < 0.01; ***P < 0.001

An expressional difference of *THAP1* in 44-day-old *in-vitro* grown cortical neurons revealed a penetrance-linked expression pattern. These results suggest that the established endogenous neuronal model used in this work is a biological correlate to the disease status of *THAP1* mutation carriers. These results are in line with the altered expression observed only in affected *THAP1* mutation carriers (see Figure 29).

6. Discussion

This thesis encompassed four functional projects involving genes causative for dystonia (*KMT2B*, *GNB1*, *VAC14*, and *THAP1*). In the following, these projects will be discussed in the same order as the results were presented.

6.1 *KMT2B* and its role in dystonia

In this project, a novel dystonia patient and daughter of consanguineous parents is reported. Based on the negative family history and her history as a refugee with exhaustive walks, a recessive or environmental cause had primarily been considered. Of note, we found a novel heterozygous *de-novo* 10-bp deletion in *KMT2B*, predicted to result in a premature protein truncation. The childhood-onset in our patient, the initial site of onset in the lower extremities, and a progressive disease course, are fully in keeping with previous reports (Carecchio et al., 2019; Zech et al., 2019, Lange et al., *in preparation*).

The early manifestation of DYT-KMT2B in *KMT2B* mutation carriers suggests a genetic basis of the disorder. However, molecular analyses to elucidate the pathophysiological consequences of *KMT2B* variants are scarce. To examine the effect of the novel frameshift variant (p.Leu1190fs161X), gene expression of *KMT2B* and putative target genes *THAP1* and *TOR1A* were tested in fibroblasts and blood. Additionally, we included DNA from another DYT-KMT2B patient carrying a non-frameshift 1-amino-acid deletion (p.Ser1656del) in the *KMT2B* gene (Lange et al., 2017). Both *KMT2B* variants caused reduced *KMT2B* mRNA levels in fibroblasts compared to controls (Figure 10). This is in line with other reports showing reduced *KMT2B* expression of protein truncating variants (such as p.Glu1009Glyfs*9, p.Gly1048Glufs*132, and p.Ala1563Aspfs*83) in fibroblasts (Meyer et al., 2017; Zech et al., 2016). While nonsense-mediated mRNA decay (NMD) would be a plausible explanation for the reduction of *KMT2B* levels in fibroblasts in the patient harbouring the frameshift mutation, the fact that the other non-frameshift variant caused a similar reduction, suggests other regulatory mechanisms by downstream effectors of KMT2B, or by epigenetic regulation of KMT2B itself. These would also help to explain why mRNA levels after NMD do not cause an expected reduction of 50%, but a *KMT2B* expression that is consistently higher than that. In this study, the gene expression of *THAP1* and *TOR1A* in blood was significantly reduced compared to controls (Figure 10). This is consistent with a report showing significant reductions of THAP1 and TOR1A protein and gene expression levels in fibroblasts of three DYT-KMT2B patients as a result of *KMT2B* frameshift mutations (Meyer et al., 2017). This supports the hypothesis that loss-of-function of KMT2B could lead to disrupted regulation of target genes and proteins, including known dystonia genes. Given the function of KMT2B, this mechanism could be governed by altered histone-methylation and chromatin packaging.

Two observations of this study were unexpected. First, no alteration of *KMT2B* expression in blood was found in the index patient compared to controls (Figure 10). Second, the expression of putative target genes of *KMT2B* (*THAP1* and *TOR1A*) was unaltered in fibroblasts in both patients compared to controls. Both findings could be explained by a tissue-specific action of mutant KMT2B or the effect of the pathogenic variants (truncating vs. non-truncating) on the regulation of target genes. Perhaps, KMT2B protein function must be maintained for the survival in some tissues, and compensatory mechanisms upregulate the expression of the wild-type protein. Additionally, the effect size could be smaller than the detection level. Furthermore, the artificial environment in which fibroblasts are cultured may impose selection pressure that impacts on gene expression. These findings underscore the critical importance of functional studies in relevant cell types such as neurons to dissect the role of pathogenic variants in *KMT2B* and to gain novel insights into the pathophysiology of DYT-KMT2B.

6.2 Impact of *GNB1* variants on dopamine-induced receptor signalling

Pathogenic variants in *GNB1* were previously identified as a novel cause of a neurodevelopmental disorder (Petrovski et al., 2016; Steinrücke et al., 2016). We have identified 16 patients suffering from global developmental delay in conjunction with dystonia, ataxia, and chorea with 14 *GNB1* variants, 10 of which occurred *de novo* (Lohmann et al., 2017). For our functional analysis, we did not include the splice-site and frameshift variants due to the predicted complete loss of function effect on the protein.

Functional analyses of possibly pathogenic variants constitute a crucial basis for the assessment of novel disease-causing variants. The aim of this project was to investigate the functional impact of *GNB1* missense variants using a cellular model of G-protein coupled receptor signalling molecules to differentiate between pathogenic and benign variants. The specific subunits $G\alpha_{olf}$ and $G\gamma_7$ were chosen because of their predominant role in the striatum, a brain region that is of central importance to several movement disorders (Corvol et al., 2001; Gittis and Kreitzer, 2012; Schwindinger et al., 2010). To assess the functional consequence of ten *GNB1* missense variants, dopamine-induced signal propagation was investigated using a cellular BRET assay, which required purified vectors harbouring the variants of interest. Ten *GNB1* missense variants were successfully introduced in vectors. The hydrolysis of the vectors confirmed the absence of gross deletions in the vector backbone, or unwanted side products due to insufficient DNA purification (Figure 11). In the BRET assay, the introduction of dopamine revealed altered functionality in seven missense variants, while three missense variants were not significantly different compared to the wild-type (Figure 12). Based on these results, it can be concluded that seven of ten missense variants are likely pathogenic owing to the altered dopamine-induced signal transduction. Three of ten variants are either benign or might show deficits in the interaction with other effector proteins or cause altered protein functions that were not assayed. Based on the impaired dopamine-induced interaction of $G\beta_1\gamma_7$ with GRK3 as measured by the BRET signal, a loss of function mechanism for the seven missense variants is suggested (Figure 12). Notably, most of the proposed pathogenic variants (R52G, A92T, P94S, P96T, and D118G) were situated at the interface between $G\alpha_{olf}$ and $G\beta_1$ and thus may play a role in the receptor-induced rearrangement. GRKs can phosphorylate active GPCRs to target the receptor for internalization and degradation, thereby regulating the intensity of GPCR signalling to prevent over-activation (Gurevich et al., 2016). As a mechanistic consequence of the reduced binding of $G\alpha_{olf}\beta_1$ to GRK3, enhanced receptor internalization and degradation could be hypothesized. This, in turn, modulates signalling of a number of pathways, including components of the mitogen-activated protein kinase / extracellular signal-regulated kinase (MAPK/ERK) cascade (Luttrell and Miller, 2013). Consistent with this finding, knockout of $G\beta_1$ in mice results in neural tube defects and microcephaly likely via reduced proliferative ERK signalling (Okae and Iwakura, 2010). Furthermore, the defect in neural proliferation in mice can be linked to the global developmental delay observed in many patients with pathogenic variants in *GNB1*. Further studies need to be conducted to prove this mechanistic link between MAPK/ERK signalling and GNB1-related disorders. Notably, no apparent link between the clinical phenotype of the patients and the BRET response amplitude was detected. Although currently there is no causative therapy available, the genetic diagnosis reduces uncertainty in the family by establishing a diagnosis and has major implications for family planning as in autosomal dominantly inherited disorders, there is a 50% chance of unaffected siblings. In conclusion, this study provides functional insights into the perturbed GPCR function that contributes to GNB1-related disorders.

6.3 Clinical spectrum of VAC14-related disorders

Pathogenic variants in *VAC14* have been identified to cause childhood-onset striatonigral degeneration (OMIM#617054) (Lenk et al., 2016). Here, we report a patient with rare compound heterozygous variants in the *VAC14* gene. One variant was transmitted by the mother. Subcloning revealed that the two variants were located on different alleles. It is unknown whether the second variant was either inherited from the father or occurred *de novo*. Mutational analysis of the predicted homodimerization (exons 13 - 16) domain in 160 unresolved dystonia cases did not reveal novel *VAC14* variants. So far, no genetic screen in dystonia patients has been published, however, our results indicate that variants in *VAC14* are causative for the disease only in a small subgroup of dystonia patients. The two missense variants of the index patient were located in exon 16 and therefore situated within the predicted homodimerization and FIG4-binding domain. The majority of published pathogenic missense variants are located within this homodimerization domain while two protein truncating variants are located upstream of the homodimerization domain, suggesting that this region is crucial in the pathogenesis of VAC14-related disorders (Figure 3). Moreover, *in-silico* predictions provide a strong indication that both novel *VAC14* variants have a pathological effect on the protein function.

Our patient presented with a unique phenotype that was characterized by a sudden onset progressive dystonia starting at the age of 13 years. After several years without effective treatment, the patient developed hypersalivation, lost 10 kg of weight, became non-ambulatory, and is currently wheelchair-bound. To date, a total of eleven patients harbouring VAC14 variants have been published (de Gusmao et al., 2019; Kaur et al., 2019; Lenk et al., 2016; Liao et al., 2020; Lyon et al., 2019; Stutterd et al., 2017; Taghavi et al., 2018). Strikingly, several phenotypic features are shared among most *VAC14* mutation carriers. First, after an initially normal psychomotor development, disease onset starts abruptly, usually at a young age (<13 years of age). Second, disease onset often starts in the lower limbs (e.g., clumsy walking or a dystonic gait). And third, the disease is characterized by a progressive course. Patients can become non-ambulatory and non-verbal, suffer from weight loss, or periods of increased irritability that may lead to death. The phenotypic signs of our patient are well in line with published reports of VAC14 mutant carriers (Table 8).

Table 8 Description of VAC14-related genotypes and phenotypes.

	Present study	Lenk et al., 2016		Stutterd et al., 2017		Taghavi et al., 2018		Kaur et al., 2019	Lyon et al., 2019		Liao et al., 2019	de Gusmao et al., 2019
VAC14 (NM_018052)	c.1868G>A c.1942C>T	c.1271G>T c.1528+1G>A	c.1744G>T c.1748C>T	c.1271 G>T c.1096+1 G>C		c.1685C>T		c.1927C>T	c.2005G>T		c.1744G>A c.2042G>A	c.1951A>G
Pregnancy and birth normal	+	+	+	+	n/a	n/a	n/a	+	-	+	+	n/a
Initially normal development	+	+	+	+	+	+	+	+	+	n/a	+	+
AAO [years]	13	3	1.5	3.5	2	13	8	1	2	n/a	2	12
Progressive course	+	+	+	+	+	+	+	+	+	n/a	+	+
Dystonia	+	+	+	+	+	+	+	+	-	n/a	+	+
Walking impaired (e.g. dystonic gait)	+	+	+	+	+	+	+	n/a	+	+	+	+
Speech impaired (dysarthria, slowed speech, non-verbal)	+	+	+	+	+	-	-	n/a	+	+	+	n/a
Hypersalivation	+	+	+	n/a	n/a	n/a	n/a	n/a	-	-	n/a	n/a
Vacuolization	-	+	+	n/a	n/a	n/a	n/a	+	n/a	n/a	n/a	+

AAO = Age at onset, + = present, - = absent, n/a = not applicable

In contrast to the clinical spectrum of the ten described cases, one case was published with two heterozygous *VAC14* variants and Yunis-Varon syndrome (OMIM#216340). Yunis-Varon syndrome, caused by pathogenic mutations in *FIG4*, is characterized by remarkable skeletal defects, juvenile lethality, severe neurological involvement, and enlarged vacuoles in neurons and muscles (Lines et al., 2017). Thus far, skeletal defects have not been part of the clinical presentation in other patients with variants in *VAC14* and could be the outer end of the clinical spectrum of VAC14-related disorders, or could arise from additional molecular effects of dysfunctional FIG4 that may be triggered by specific *VAC14* variants.

The location of the compound heterozygous missense variants within the homodimer domain in our patient raises the hypothesis whether impaired homodimer formation is a functional consequence. To address the functional significance of altered dimerization, a cellular VAC14 overexpression model was generated in SH-SY5Y neuroblastoma cells using lentiviral transduction of mutant and wild-type VAC14. Based on the overlapping clinical phenotype (late age at onset, progressive dystonia, dysarthria, and mutation located in the homodimer domain; Table 8), we included another missense mutation (c.1685C>T, p.Ala562Val; (Taghavi et al., 2018)). The experiment was performed three times in which a trend of enhanced VAC14 dimer formation was observed as a consequence of two missense changes (p.Ala562Val and p.Leu648Phe) compared to VAC14 wild-type after (Figure 15). In contrary, the missense change inherited from the unaffected mother did not reveal altered dimerization in the model, suggesting a milder effect on homodimerization or a different effect of this variant on protein function. Altered homodimerization of VAC14 might affect binding to its binding partners (FIG4 and PIKFYVE) causing dysregulated $PI(3,5)P_2$ synthesis. To prove the pathogenicity of the *VAC14* variants by other methods, the abundance of $PI(3,5)P_2$ could be directly measured by high-pressure liquid chromatography, however, the low abundance of the signalling lipid may complicate the measurement. Another elegant method to evaluate the pathogenicity of *VAC14* variants would involve the quantification of the hyperacidic enlarged lysosomal vesicles by cytometry which Lenk *et al.* has previously established for dysfunctional FIG4 (Lenk et al., 2019).

We next addressed the question whether intracytoplasmic vacuoles are present in the patient's fibroblasts. In this study, no apparent vacuoles were detected (Figure 14). Thus far, enlarged cytoplasmic vacuoles have been found in fibroblasts from four patients with VAC14 mutations (de Gusmao et al., 2019; Kaur et al., 2019; Lenk et al., 2016). However, none of the other investigators report any microscopic analysis on patients' fibroblasts (Liao et al., 2020; Lyon et al., 2019; Stutterd et al., 2017; Taghavi et al., 2018). One possible explanation for the absence of vacuoles in our patient's fibroblasts is that the VAC14 missense variant inherited from the unaffected mother has a less pronounced effect on dimerization and may partly compensate for the altered $PI(3,5)P_2$ turnover that causes the vacuolated phenotype. To determine how VAC14 missense variants influence FIG4 and PIKFYVE binding and intracellular levels of $PI(3,5)P_2$ will be a major goal for future experiments.

6.4 Clinical and genetic findings in DYT-THAP1 patients

In the main project of this work, four objectives will be discussed including a literature search of phenotypic DYT-THAP1 traits, considerations of the iPSC-derived neuronal model, quality control metrics of the whole transcriptome analysis, and the genetic dysregulation caused by mutations in the *THAP1* gene in an *in-vitro* model of developing cortical neurons.

6.4.1 Clinical spectrum of DYT-THAP1 patients

As part of the first objective, the published literature on phenotypic traits of DYT-THAP1 patients was reviewed. Given the rapidly expanding genetic and phenotypic spectrum of genetic dystonia, MDSGene is addressing the need to manage the complex field of underlying genetic causes by providing a genotype-phenotype database. The database serves as a resource for clinicians correlating the clinical phenotype with an underlying genetic cause to prevent inappropriate care or a diagnostic odyssey for the patient. Furthermore, the database serves as a resource for movement disorder geneticists by predicting the pathogenicity of a variant based on the segregation, *in-silico* prediction, the population frequency, and published functional evidence. In contrast to other phenotypic databases for genetic disorders such as Simulconslt (www.simulconsult.com), Orphanet (www.orpha.net), OMIM (www.omim.org), and ClinVar (www.ncbi.nlm.nih.gov/clinvar), MDSGene specializes on genetically determined movement disorders with more detailed phenotypic information and allows customized phenotypical searches.

We identified 241 THAP1 mutation carries from 163 families in this literature review. However, the actual number of DYT-THAP1 patients worldwide is expected to be significantly higher, which is attributed to selective reporting of novel mutations with unique or unusual phenotypic traits. The results of this systematic literature review are in line with previously published results - red flags for DYT-THAP1 are childhood-onset (2 - 12 years) or adolescence-onset (13 - 20 years) with initial sites of onset in the upper limbs or the neck (Blanchard et al., 2011; Paudel et al., 2012). The Caucasian ethnicity was overrepresented (~57% of all patients), pointing towards a publishing bias of the Western scientific community. Taken together, the systematic review provides detailed knowledge of the genetic and phenotypic spectrum of DYT-THAP1 patients which complements functional studies in this work.

6.4.2 iPSC-derived neurons as a cellular model

The second objective of this project was to generate and characterize DYT-THAP1 patient-derived iPSCs. Patient iPSC-derived neurons carry the same repertoire of genetic variants as the donor and therefore represent the best model to study the pathophysiology and the genetic mechanisms of reduced penetrance. In this study, iPSCs were generated using Sendai virus. In comparison to retroviruses that integrate into the host genome, Sendai viruses are non-integrating, offering the possibility that the iPSC clone is free of virus after a certain number of passages. The characterization of iPSCs is a crucial quality control step prior to differentiation to exclude genetic alterations introduced during reprogramming and to prove the pluripotent state of the iPSC lines. In total, 20 iPSC lines from three families were generated, banked, and successfully characterized with regard to genome integrity, absence of reprogramming factors, expression of pluripotency markers, the ability to form cells of all three germ layers, and the presence of the pathogenic *THAP1* variant (Figure 19 - Figure 21). The analysis of genome integrity is important as previous studies have shown that genetic abnormalities in human iPSC derived from fibroblasts can occur during the process of iPSC generation (Gore et al., 2011; Hussein et al., 2011; Mayshar et al., 2010). Furthermore, the clearance of Sendai virus is particularly relevant for the subsequent differentiation of iPSCs into neurons, as pluripotency factors may impede complete differentiation of iPSCs (Figure 20 and Figure S4).

For this study, control iPSCs were matched for gender and ethnicity and derived from human dermal fibroblasts. This consideration is important as the ancestral background, gender and donor age may alter the global transcriptome and thus create false positive candidate genes by group comparisons (Barral-Arca et al., 2019; Mogil et al., 2018; Storey et al., 2007; Strässler et al., 2018).

In the next step, human iPSC lines were differentiated in glutamatergic cortical neurons which follows the stereotypical temporal order of human corticogenesis *in vivo* (Shi et al., 2012a). In concordance with published literature, we obtained neurons expressing specific cortical marker proteins and morphologically resembling mature neurons at day 90 (Figure 22 and Figure 23) (Shi et al., 2012b). Due to a hypothesized change at early developmental stages, iPSC-derived maturing neural precursor cells were harvested on day 44 to study genetic alterations during early brain development of dystonia patients and controls. The aetiology of dystonia is not yet fully understood – thus, it is conceivable that other neuronal subtypes such as Purkinje cells of the cerebellum, GABAergic neurons of the striatum, dopaminergic neurons of the basal ganglia, or oligodendrocytes contribute to disease manifestation (Balint et al., 2018; Ruiz et al., 2015; Yellajoshyula et al., 2017). Notably, reliable differentiation protocols are not available for every neuronal cell type and many protocols result in mixed cultures. As highly homogeneous cultures are needed for the transcriptome analysis, we decided to use the established cortical differentiation protocol to obtain pure cortical cultures for comparative reasons. Multiple lines of evidence support the notion that the cortico-thalamo-cerebellar network might be impaired in dystonia, suggesting that defects in cortical neurons could contribute to DYT-THAP1 disease pathogenesis (Balint et al., 2018; Ramdhani and Simonyan, 2013).

6.4.3 Genetic dysregulation in DYT-THAP1 cortical neurons

The third objective of this THAP-related part was to investigate iPSC-derived cortical neurons as an endogenous cellular disease model to study the expressional consequences of pathogenic *THAP1* mutations by global transcriptome analysis. In the first paragraph, considerations for the RNA-Seq experimental design are introduced. This will be followed by a short discussion regarding the transcriptome quality metrics, and a hypothesis how the DEGs are linked to dystonia.

Since RNA-Seq was first presented in the year 2008, the tool has shaped our understanding of many aspects of biology (Mortazavi et al., 2008; Stark et al., 2019). For our RNA-Seq experiment, several aspects were considered in our experimental design (Conesa et al., 2016). First, the number of replicates should be appropriate for the experiment. The number of replicates depends on the expected technical and biological variability, but also on resources, and can influence the statistical power to detect DEGs. We chose a total number of 32 samples that included biological and technical replicates to increase the statistical power. Information on biological variability was not available beforehand. Thus, the number of replicates was chosen according to the expected biological variability and limited by the costs for such an analysis. Second, to reduce sequencing noise, the abundant ribosomal RNAs that constitute over 90% of total RNA in mammalian cells and the remnant DNA should be removed. In our experiment, the ribosomal RNA was removed by the Ribo-Zero kit and oligo(dT) beads, while the DNA was efficiently removed by a dedicated DNA removal column (RNeasy Plus Kit) to provide high-quality starting material with minimal DNA impurifications. The RNeasy Kits from QIAGEN deliver high RNA quality comparable to other RNA isolation methods (Scholes and Lewis, 2019). Third, the RNA quality of the starting material should be assessed. In our experiment, the RNA quality was assessed by the gold standard RIN analysis and yielded very good results, which is an important prerequisite for further analysis (Table S5) (Schroeder et al., 2006). Fourth, the library size, sequencing method (single-end or paired-end), and the read length need to be selected (Sims et al., 2014). Deeper sequencing improves the precise

quantification of low-abundant transcripts (Tarazona et al., 2011). Paired-end sequencing is preferable for isoform detection, the discovery of novel transcripts, non-coding RNA, and mapping to repeat regions (Katz et al., 2010). In contrast, single-end sequencing is generally more affordable and often sufficient for well-annotated organisms. We have chosen a deep sequencing 150-bp-paired-end approach (mean of 76.9 Mio raw reads; Table S5) to detect low-abundant transcripts and novel isoforms. The sequencing depth matches standards of the Genotype-Tissue Expression (GTEx) reference databank (78 Mio 76-bp paired-end reads) (Aguet et al., 2017). Fifth, the bioinformatic processing should follow general rules of reproducibility e.g., by creating isolated virtual software environments. The 4SeqGUIv2 pipeline strictly adhered to published recommendations in terms of reproducibility, raw data quality control, and integrated the latest software packages to perform differential gene expression analysis in a protected virtual environment (Beccuti et al., 2018; Conesa et al., 2016).

Quality control is essential to identify contaminants or irregularities of the library to minimize the chances of false-positive findings. Most of the quality metrics were passed by all samples. These analyses included the mean Phred quality scores, duplication levels, and adapter contents, and number of reads aligned to the reference genome (Figure 24 and Figure 26). Noteworthy, two quality metrics deviated from the expected profile (Figure 25). First, the relative GC-content analysis revealed that transcripts with around 60% GC-content were abundantly present in the library, and an increased number of transcripts with 52% GC-content was observed in the R1 sequencing direction. Second, the number of overrepresented sequences was slightly elevated in some samples. The deviation of the GC-content from the expected Gaussian distribution often indicates a technical bias introduced by ribosomal RNA contamination or PCR enrichment bias during library preparation. For instance, after ribosomal depletion using the Ribo-Zero kit, the content of ribosomal RNA can still constitute up to 5% of the total RNA providing a possible explanation for enriched ribosomal sequences (Herbert et al., 2018). The systematic nature of the bias (all samples or all R1 sequencing directions) indicates a technical rather than biological (such as co-culture or contamination) origin of this phenomenon. Therefore, differential gene expression analysis in the present study should not be affected by this bias.

Next, a PCA was performed to gain a first visual impression of the dispersion of the samples. The analysis indicated that mutation carriers of pathogenic variants in *THAP1* could not be discriminated from controls based on the global expression pattern in developing neurons (Figure 28). The high expressional variability among iPSC-derived neuronal samples can result from multiple factors including iPSC generation and differentiation. Several studies report inherent epigenetic differences between pluripotent cell lines that result in variable neuronal populations (Bock et al., 2011; Hu et al., 2010; Osafune et al., 2008). By looking at the phenotype of DYT-THAP1 patients, no gross brain abnormalities in the cortex are observed (e.g., atrophy of the cortex, microcephaly, or global development delay) and except for a more or less severe movement disorder (dystonia), the patients are usually healthy. Dystonia 'only' develops during adolescence and many patients remain completely unaffected in early life, implying that the transcriptional changes are rather subtle. The PCA analysis, however, is meant to show large alterations by reducing the dimensions of a dataset. Therefore, it is feasible, that despite the lack of global gene expression differences in *THAP1* mutation carriers, subtle differences affect cortical neuron function, which can be investigated by differential expression analysis.

To perform differential gene expression analysis, the samples were divided into clinically affected *THAP1* mutation carriers (n = 9), unaffected *THAP1* mutation carriers (n = 9), and controls (n = 11). As a main outcome of this project, the analysis revealed a unique expressional abnormality (104 DEGs; Figure 29) exclusive to manifesting carriers when compared to controls. The identical number of samples from

unaffected carriers did not reveal major expressional differences compared to controls (5 DEGs; Figure 29). Due to the enhanced DEGs in the clinically affected carriers compared to the unaffected carriers, it is concluded that these genes may play a role in the penetrance of DYT-THAP1. To further expand the number of candidate genes, we extended the inclusion criteria by lowering the false discovery rate and included genes from the group comparison of all *THAP1* mutation carriers (n = 18) versus controls. A set of 137 candidate genes was defined and further investigated. An exploratory data analysis approach was conducted to find a unifying biological pattern behind the set of 137 candidate genes. Online-tools for the interpretation of large genetic datasets such as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, EnrichR, and Wikipathway analysis did not reveal a coherent biological pathway. One possible reason for this observation is the low number of input genes that complicate the statistical enrichment of GO clusters. To circumvent this problem, either the false discovery rate of our differential expression analysis would need to be adjusted or more samples would need to be included.

We investigated the genomic distribution of candidate genes, which were partly clustered on chromosome 19, and several pairs of genes in close genetic proximity were both up- or downregulated (Figure 30). A possible explanation could be the binding of the transcription factor THAP1, that led to a shared transcriptional regulation. Although there is the possibility of a chance finding, these genes represent a valuable set of target genes to look for genomic alterations that influence transcription factor binding of THAP1. To link the genomic locations to suggested THAP1 binding sites discovered by the ENCODE project, we compared our candidate genes with >600 promotor regions putatively bound by THAP1 (Feingold et al., 2004). An overlap of five genes was detected among 137 DEGs. A possible explanation for the small overlap is that THAP1 regulates a different set of target genes in developing neurons compared to leukemic cancer cells, which were used in the ENCODE project. Moreover, the pathological *THAP1* mutations in the THAP domain could affect the DNA-binding properties of THAP1 and the regulation of different target genes.

The GO-term classification of the candidate genes established a link of mutant THAP1 to defective nucleus localization, cell adhesion, axon guidance, upregulation of immune response genes as well as oxidoreductases (Figure 31). The defective nucleus localization might be attributed to altered function of the transcription factor THAP1 as the nonsense mutation p.Lys158Asnfs*23 has been shown to reduce nuclear import (Osmanovic et al., 2011). Consistent with results from previous transcriptomic studies in mice, an overlap with the GO terms cell adhesion (Frederick et al., 2019; Zakirova et al., 2018), inflammatory response (Frederick et al., 2019), and axon guidance (Aguilo et al., 2017) were detected. Based on the genes in each GO category and known disease-related genes in mice and humans (Figure 32), several candidate genes involved in nervous system development, axon guidance, and neuronal projection can be prioritized (*FOXP2*, *DCLK1*, *KIF26B, CNTN5*, *CNTNAP3B*, *PCDH19*, and *UNC5C*). A small sub-set of candidate genes was technically validated using qRT-PCR by the Master student Joachim Weber, who was co-supervised by the doctoral applicant, with excellent congruence between RNA-Seq data and qRT-PCR results (Weber, 2019). In the following, the cellular functions of prioritized candidate genes and the corresponding proteins will be discussed in the context of early brain development in DYT-THAP1 patients.

In this study, downregulation of *FOXP2* in iPSC-derived precursor cortical neurons in affected *THAP1* mutation carriers compared to controls was detected (Figure 32; Table S2). *FOXP2* encodes a highly conserved transcription factor in vertebrates that regulates differentiation, patterning, metabolism, and cell-cell regulation (Co et al., 2020). Intriguingly, FOXP2 has been shown to disrupt motor circuits in humans. FOXP2 is known to cause dyspraxia (OMIM#602081), a motor-speech disorder with primary

difficulties in the planning and execution of movements (Lai et al., 2001). The disorder is also characterized by an imprecise control of orofacial movements that impedes language development. Mice with heterozygous *Foxp2* mutations display motor learning deficits (French et al., 2012; Groszer et al., 2008). Furthermore, heterozygous *Foxp2* mutant mice exhibit impaired long-term depression in cortico-striatal brain circuits (Groszer et al., 2008). It has further been demonstrated that the striatal activity is unusually elevated during motor learning on the rotarod (French et al., 2012). Interestingly, overlapping target genes from Foxp2 chromatin immunoprecipitation analysis studies identified the dystonia gene *Gnal* as one of 19 targets (Vernes et al., 2011). The movement disorder phenotype, the role as a transcription factor, the orofacial involvement, and the role in the striatum show remarkable phenotypic overlap with DYT-THAP1 pathology. FOXP2 regulates a plethora of genes implicated in neurite outgrowth and cortico-striatal plasticity in mouse brains (Vernes et al., 2011).

In turn, dynamic processes such as neurite outgrowth are highly dependent on microtubules. In our studies, two microtubule-associated genes, *DCLK1* and *KIF26B*, were downregulated in DYT-THAP1 iPSC-derived precursor neurons compared to controls (Figure 32; Table S2). DCLK1 phosphorylates MAP7 Domain Containing 1 (MAP7D1) to promote axon elongation in cortical neurons (Koizumi et al., 2017). Furthermore, dynamic processes such as cytoskeletal reorganization are essential for brain plasticity (Gordon-Weeks and Fournier, 2014). Alterations in the microtubule network have been linked to dystonia as pathogenic variants in microtubule component ß-tubulin 4A are known to cause DYT-TUBB4A, a complex hereditary dystonia with spasmodic dysphonia as a characteristic feature. The kinesin superfamily proteins (KIF) family, including KIF26B, acts as regulators of microtubule dynamics and cytoskeleton-driven processes (Hirokawa and Tanaka, 2015). It was demonstrated that KIF26B is involved in cell adhesion, cell migration, and polarization, which raises the possibility of an altered cytoskeleton function in DYT-THAP1 mutant cells (Guillabert-Gourgues et al., 2016; Uchiyama et al., 2010). Three members of the KIF family were already implicated in human neuronal diseases: spastic paraplegia 30 (KIF1A; OMIM#610357), cortical dysplasia (KIF2A; OMIM#615411), and fibrosis of extraocular muscles (KIF21A; OMIM#135700) (Erlich et al., 2011; Poirier et al., 2013; Yamada et al., 2003). Microtubule-associated deficits are a plausible link to impaired neurite outgrowth of cortical neurons in DYT-THAP1.

In this study, gene expression of contactin 5 (*CNTN5*) and contactin associated protein family member 3B (*CNTNAP3*) was found to be downregulated in developing iPSC-derived cortical neurons of THAP1 mutation carriers compared to controls (Figure 32; Table S2). Neurite outgrowth and synaptogenesis are also regulated by CNTN5, an adhesion molecule that belongs to the neurexin superfamily of proteins (Zuko et al., 2011). During embryonic development, CNTN5 protein expression was mostly found in the thalamus, the caudate putamen and cortex (Kleijer et al., 2015). The CNTN5 protein is found in nuclei of projection neurons from the thalamus into the cortex (Kleijer et al., 2015). Furthermore, the expression in the thalamus is restricted to glutamatergic neurons. This is in concordance with magnetic resonance diffusion tensor imaging results that identified altered cerebello-thalamo-cortical fibre tracts in DYT-TOR1A dystonia patients to modify the disease penetrance, which are likely developmental in origin (Argyelan et al., 2009). Of interest, iPSC-derived glutamatergic neurons with a heterozygous loss of *CNTN5* have intensified synaptic activity and develop hyperactive neuronal networks (Deneault et al., 2019). These findings are well in line with the hypothesis of dystonia as a network disorder with a perturbed circuit formation of the thalamocortical axis. The role of another member of the contactin family of proteins, CNTNAP3B, is largely elusive - however its paralog CNTNAP3 is abundantly expressed in the basal ganglia during early brain development, especially in the striatum, globus pallidus, and substantia nigra and is involved in motor control and motor learning (Hirata et al., 2016) suggesting a

similar role for CNTNAP3B. Another related gene of the neurexin superfamily, CNTNAP2, causes the cortical dysplasia-focal epilepsy syndrome and Pitt-Hopkins like syndrome 1 (OMIM#610042). The latter is characterized by intellectual delay, motor learning deficits, recurrent seizures, and facial dysmorphisms. The function of the five known contactin-associated proteins is closely connected to the formation of myelinated axons (Zou et al., 2017). Intriguingly, hypomyelination has been proposed as a disease mechanism for DYT-THAP1, which is also linked to abnormal circuit formation (Yellajoshyula et al., 2017). The overall association of the superfamily causing neurologic phenotypes and the involvement in myelination point to a possible role of contactins in the pathogenesis of DYT-THAP1 dystonia.

In our study, *PCDH19* gene expression was downregulated in iPSC-derived precursor cortical neurons in affected *THAP1* mutation carriers compared to controls (Figure 32; Table S2). Protocadherins, such as PCDH19, belong to the cadherin superfamily and play a role in cell-cell adhesion, axon guidance, neurite self-avoidance, and synaptogenesis and are primarily expressed during nervous system development (Garrett and Weiner, 2009; Lefebvre et al., 2012). In mice, conditional knockout of PCDH19 results in a dysregulated spatial distribution of cortical neurons (Hayashi et al., 2017). In humans, PCDH19 is known to cause X-linked early infantile epileptic encephalopathy, which is the second most common monogenic cause of epilepsy (OMIM#300088) (Depienne et al., 2009; Dibbens et al., 2008). In X-linked recessive diseases, typically more affected hemizygous males than affected heterozygous females exist. Conversely, pathogenic variants in *PCDH19* cause epilepsy in females and spare males. On a molecular level, this phenomenon is explained by mixed populations of wild-type and PCDH19 mutant neurons that lead to defects in cortical wiring through random X-inactivation (Pederick et al., 2018). This view is supported by the presence of mosaic *PDH19* deletions in fibroblasts of an affected male (Depienne et al., 2009). Intriguingly, a recent study implicated PCDH19 in regulation of cortical microcircuit assembly (Lv et al., 2019). As we observe only mild expressional differences in our neuronal model, it can be hypothesized that these expressional changes are mosaic and cause a disturbance in the cortical wiring in DYT-THAP1 patients.

UNC5C belongs to the netrin protein family and is known for its function in neural circuit organization by mediating axon repulsion of neuronal growth cones during the development of the nervous system (Poliak et al., 2015). The *UNC5C* gene was found to be downregulated in iPSC-derived precursor cortical neurons in affected *THAP1* mutation carriers compared to controls in this study (Figure 32; Table S2). The growth cone is a dynamic cytoskeletal-rich structure at the tip of an extending axon that senses guidance cues via netrin receptors, such as UNC5C. During brain development growing neurons extend their projections over relatively long distances to reach the correct position, in which UNC5C plays a major role in the dorsal guidance of hindbrain axons (Kim and Ackerman, 2011; Poliak et al., 2015). A study in mice showed that Unc5c is involved in the connection between motor neurons and muscles during early development (Poliak et al., 2015). It was demonstrated that homozygous mutations in *UNC5C* cause an ataxic phenotype and laminar structure defects (Ackerman and Knowles, 1998). Pathogenic variants in another member of the netrin family of genes, netrin G1 *(NTNG1*), is reported to cause an atypical Rett syndrome with or without epilepsy (Borg et al., 2005). Dysfunctional UNC5C causes misrouting of axons in regions without survival signals, which ultimately triggers apoptosis (Llambi et al., 2001). Moreover, pathogenic variants in UN5C5C were suggested to disturb the regulation of neuronal cell survival (Wetzel-Smith et al., 2014). Therefore, UNC5C downregulation could cause altered pathfinding in cortical neurons during neuronal development in DYT-THAP1 patients.

Collectively, the THAP1 findings presented in this study reinforce the hypothesis that mild expressional changes present during early development of the brain might be linked to dysregulated wiring during DYT-THAP1 brain development. The normal brain development could be disrupted by dysfunctional neurite outgrowth of cortical projection neurons, which leads to modified wiring of brain circuits. Cortico-striatal connections play a critical role in motor learning as striatal regions integrate input signals from different cortical regions (Plotkin and Goldberg, 2019; Surmeier et al., 2007). While the neurite extension might induce early structural changes in the brain, it is feasible that this disruption affects motor learning later in life. This theory is in line with dystonia mouse models of DYT-TOR1A showing alterations in the cortico-striatal axis leading to impaired motor learning as well as altered motor plasticity (Hewett et al., 2006; Maltese et al., 2018; Martella et al., 2009; Martino et al., 2013). This hypothesis is also consistent with results obtained from neuroimaging of dystonia brains, in which cortico-striato-pallido-thalamic and cerebello-thalamo-cortical projections appear altered compared to controls (Niethammer et al., 2011; Ramdhani and Simonyan, 2013; Vo et al., 2015).

6.4.4 Elucidating genetic modifiers in DYT-THAP1 and future perspectives

The last sub-aim of this project was to find a molecular signature specific to affected *THAP1* mutation carriers that might reveal a genetic contribution leading to disease manifestation. The fact that only five genes were detected as differentially regulated between unaffected *THAP1* carriers and controls in our study demonstrates that there is a genetic signature uniquely present in affected individuals. A total number of 74 candidate genes were identified to be exclusively dysregulated in affected individuals, including genes implicated in brain network organisation (*FOXP2*, *DCLK1*, *KIF26B*, *CNTN5*, *CNTNAP3B*, *PCDH19*, and *UNC5C*). This indicates that these genes, which are dysregulated in early stages of brain development, putatively confer disease risk of manifesting *THAP1* mutation carriers later in life. In concordance with these results, a study investigated cerebello-thalamo-cortical fibre tracts in affected and unaffected DYT-TOR1A patients found that circuit abnormalities are present in affected individuals compared to unaffected individuals, which are believed to underlie disease penetrance (Argyelan et al., 2009). Consistent with the whole transcriptome analysis, we observed an increase of endogenous *THAP1* mRNA expression in maturing neurons of affected THAP1 mutation carriers compared to unaffected THAP1 mutation carriers, suggesting that elevated THAP1 mRNA expression is linked to penetrance and may be reflecting an expressional overcompensation (Figure 33). This allows the conclusion that elevated *THAP1* is linked to the global expressional changes which in turn predisposes for disease manifestation. Molecular mechanisms for two of the three THAP1 variants investigated in this work have been shown to disrupt the function of THAP1. The p.Lys158Asnfs*23 frameshift mutation disrupts the nuclear localization signal, thereby preventing the active nuclear import of the transcription factor THAP1 (Osmanovic et al., 2011). Another study has shown that the p.Arg13His missense mutation was associated with reduced binding of THAP1 to its own promotor leading to an upregulation of *THAP1* in a luciferase reporter gene assay (Erogullari et al., 2014). Thus, both mutations are suggested to result in reduced THAP1 promotor binding and upregulation of *THAP1* expression by different mechanisms. However, this does not explain why *THAP1* levels were only upregulated in manifesting carriers in this study. This indicates that additional regulatory mechanisms are likely to exist, such as environmental stimuli, or genetic variants that affect the expression of genes conferring additional susceptibility to develop the disease. Therefore, future genetic analysis should investigate epigenetic modifications, copy number variants, and single-nucleotide polymorphism in regulatory elements (promotor, enhancer, and silencer) of the DEGs, and THAP1 itself, to find an explanation why these genes are dysregulated in affected THAP1 mutation carriers and to a lesser extent in unaffected individuals. Furthermore, iPSC-derived neuronal models with a CRISPR-Cas mediated knock-out or siRNA-mediated knock-down of the DEGs might help to address the cellular function of the candidate genes. Another approach with the existing RNA-Seq dataset would be to investigate allele-biased expression between affected and unaffected *THAP1* mutation carriers. Differences in allele-biased expression can deliver

plausible explanations for incomplete penetrance. For example, imprinting of one allele has been identified as a disease mechanism in one genetic form of dystonia (DYT-SGCE) and a range of other genetically determined neurodevelopmental conditions such as Prader-Willi syndrome and Angelman syndrome (Cassidy et al., 2012; Kishino et al., 1997; Müller et al., 2002a). Cortical neurons derived from iPSCs have been shown to retain epigenetic information, as shown in a cellular DYT-SGCE model highlighting the suitability of this approach (Grütz et al., 2017). A prominent example of the stochastic monoallelic expression in thalamocortical axon guidance concerns the PCDH family of protocadherins, including the X-linked DEG *PCDH19*, reinforcing the idea that epigenetic regulation could be relevant in DYT-THAP1 (Molnár et al., 2015; Yagi, 2008).

Connecting whole-genome and methylome findings with the transcriptome findings of this work will allow prioritizing genes relevant for experimental follow-up. Cellular neuronal knock-out models of the most promising candidate genes and the identification of regulatory DNA elements of penetrance-associated genes will help to reveal genetic modifiers in DYT-THAP1 dystonia. The elucidation of these mechanisms is important for comprehending the disease pathology and developing novel treatment strategies for DYT-THAP1 dystonia.

7. Conclusions

This thesis addressed several aspects to characterize the cellular impact of genetic variants in dystonia-related genes. Overall, the presented data provide valuable insights into the functional consequences of variants in dystonia-causing genes. The following projects were pursued:

KMT2B

- Gene expression analysis of two novel pathogenic *KMT2B* variants revealed reduced *KMT2B* gene expression in fibroblasts and reduced mRNA levels of putative target genes *THAP1* and *TOR1A* in blood compared to controls suggesting a converging transcriptional regulation of these three dystonia genes. Future studies need to address the transcriptional regulation in neuronal cellular models to prove the cellular mechanism in a cell type relevant for the disease.

GNB1

- By identifying fourteen patients, this study significantly expands the number of *GNB1* mutation carriers. Altered functionality was shown for seven out of ten *GNB1* missense variants by aberrant dopamine-induced signal propagation using a BRET assay, suggesting that deficits in this pathway may underly the pathology in GNB1-related disorders. Further studies are needed in order to explain the phenotypic spectrum of GNB1-related disorders.

VAC14

- Functional assessment of three *VAC14* variants on homodimerization in a cellular overexpression model indicated enhanced VAC14 homodimerization as a consequence of two of these variants, but not for the third variant, compared to the wild-type. The latter one was found together with a VAC14 homodimer-enhancing variant in a newly described dystonia patient from our clinic. The functional results indicate that enhanced VAC14 homodimerization is a possible disease mechanism. In contrast to previous reports, no enlarged vacuoles were detected in our patient's fibroblast culture, possibly due to a compensatory effect of the second milder variant inherited from the mother. Additional work is needed to verify that altered VAC14 homodimerization is a disease mechanism in VAC14-related disorders.

THAP1

- A systematic literature search was conducted to provide a comprehensive phenotypic and demographic overview of DYT-THAP1 patients. Initial dystonic symptoms of DYT-THAP1 patients were most prominently found in the neck and arms, while co-morbidities were generally absent. Additionally, DBS surgery resulted in a marked or moderate improvement in the majority of patients.
- Twenty human dermal fibroblast-derived iPSC lines of ten *THAP1* mutation carriers from three different families harbouring three different pathogenic variants, respectively, were successfully generated and characterized. The analysis included Sendai virus clearance, expression of pluripotency markers, the potential to differentiate into all three germ layers, genome integrity, and retention of the pathogenic variant. These iPSC lines constitute a useful resource for researchers to study the cellular mechanisms underlying reduced penetrance in DYT-THAP1.
- Global transcriptome analysis of iPSC-derived precursors of cortical neurons revealed mild expressional differences exclusively found in affected *THAP1* mutation carriers and not in unaffected *THAP1* mutation carriers compared to controls. In addition, the expression of the disease gene *THAP1* was upregulated in

affected carriers compared to unaffected carriers in maturing cortical neurons. Both findings suggest that a unique expressional signature during early cortical development may affect disease penetrance.

- An exploratory data analysis approach indicated an accumulation of genes related to perturbed cortical networks (*FOXP2*, *DCLK1*, *KIF26B*, *CNTN5*, *CNTNAP3B*, *PCDH19*, and *UNC5C*) that may influence disease penetrance in DYT-THAP1 dystonia. Further studies, including systems biology approaches, are needed to better understand the mechanism of these genes on DYT-THAP1 penetrance and to elucidate the influence of cortical gene expression on DYT-THAP1 pathogenesis.

8. References

Ackerman, S.L., and Knowles, B.B. (1998). Cloning and mapping of the UNC5C gene to human chromosome 4q21-q23. Genomics *52*, 205–208.
Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y.S., Parsana, P., Segrè, A. V., et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.
Aguilo, F., Zakirova, Z., Nolan, K., Wagner, R., Sharma, R., Hogan, M., Wei, C., Sun, Y., Walsh, M.J., Kelley, K., et al. (2017). THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation. Stem Cell Reports *9*, 92–107.
Albanese, A., Bhatia, K., Bressman, S.B., Delong, M.R., Fahn, S., Fung, V.S.C., Hallett, M., Jankovic, J., Jinnah, H.A., and Klein, C. (2013). Phenomenology and Classification of Dystonia : A Consensus Update. Movement Disorders *28*, 863–873.
Almasy, L., Bressman, S.B., Raymond, D., Kramer, P.L., Greene, P.E., Heiman, G.A., Ford, B., Yount, J., De Leon, D., Chouinard, S., et al. (1997). Idiopathic torsion dystonia linked to chromosome 8 in two Mennonite families. Annals of Neurology *42*, 670–673.
Argyelan, M., Carbon, M., Niethammer, M., Ulug, A.M., Voss, H.U., Bressman, S.B., Dhawan, V., Eidelberg, D., Uluğ, A.M., Voss, H.U., et al. (2009). Cerebellothalamocortical connectivity regulates penetrance in dystonia. Journal of Neuroscience *29*, 9740–9747.
Balint, B., Mencacci, N.E., Valente, E.M., Pisani, A., Rothwell, J., Jankovic, J., Vidailhet, M., and Bhatia, K.P. (2018). Dystonia. Nature Reviews Disease Primers *4*, 229–236.
Barbagiovanni, G., Germain, P.L., Zech, M., Atashpaz, S., Lo Riso, P., D'Antonio-Chronowska, A., Tenderini, E., Caiazzo, M., Boesch, S., Jech, R., et al. (2018). KMT2B Is Selectively Required for Neuronal Transdifferentiation, and Its Loss Exposes Dystonia Candidate Genes. Cell Reports *25*, 988–1001.
Barral-Arca, R., Pardo-Seco, J., Bello, X., Martinón-Torres, F., and Salas, A. (2019). Ancestry patterns inferred from massive RNA-seq data. Rna *27*, 857–868.
Baumann, H., Jahn, M., Muenchau, A., Trilck-Winkler, M., Lohmann, K., and Seibler, P. (2018). Generation and characterization of eight human-derived iPSC lines from affected and unaffected THAP1 mutation carriers. Stem Cell Research *33*, 60–64.
Beccuti, M., Cordero, F., Arigoni, M., Panero, R., Amparore, E.G., Donatelli, S., and Calogero, R.A. (2018). SeqBox: RNAseq/ChIPseq reproducible analysis on a consumer game computer. Bioinformatics *34*, 871–872.
Bhatia, K.P., and Marsden, C.D. (1994). The behavioural and motor consequences of focal lesions of the basal ganglia in man. Brain *117*, 859–876.
Blanchard, A., Ea, V., Roubertie, A., Martin, M., Coquart, C., Claustres, M., Béroud, C., and Collod-Béroud, G. (2011). DYT6 dystonia: Review of the literature and creation of the UMD locus-specific database (LSDB) for mutations in the THAP1 gene. Human Mutation *32*, 1213–1224.
Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452.
Bonafé, L., Thöny, B., Penzien, J.M., Czarnecki, B., and Blau, N. (2001). Mutations in the sepiapterin reductase gene cause a novel tetrahydrobiopterin-dependent monoamine-neurotransmitter deficiency without hyperphenylalaninemia. American Journal of Human Genetics *69*, 269–277.
Borg, I., Freude, K., Kübart, S., Hoffmann, K., Menzel, C., Laccone, F., Firth, H., Ferguson-Smith, M.A., Tommerup, N., Ropers, H.H., et al. (2005). Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. European Journal of Human Genetics *13*, 921–927.
Breakefield, X.O., Blood, A.J., Li, Y., Hallett, M., Hanson, P.I., and Standaert, D.G. (2008). The pathophysiological basis of dystonias. Nature Reviews Neuroscience *9*, 222–234.
Bressman, S.B., Raymond, D., Fuchs, T., Heiman, G.A., Ozelius, L.J., and Saunders-Pullman, R. (2009). Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study. The Lancet Neurology *8*, 441–446.
Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L., and Weber, J.L. (1998). Comprehensive human genetic maps: Individual and sex-specific variation in recombination. American Journal of Human

Genetics *63*, 861–869.
Campagne, S., Saurel, O., Gervais, V., and Milon, A. (2010). Structural determinants of specific DNA-recognition by the THAP zinc finger. Nucleic Acids Research *38*, 3466–3476.
Carecchio, M., Invernizzi, F., Gonzàlez-Latapi, P., Panteghini, C., Zorzi, G., Romito, L., Leuzzi, V., Galosi, S., Reale, C., Zibordi, F., et al. (2019). Frequency and phenotypic spectrum of KMT2B dystonia in childhood: A single-center cohort study. Movement Disorders *34*, 1516–1527.
Cassidy, S.B., Schwartz, S., Miller, J.L., and Driscoll, D.J. (2012). Prader-Willi syndrome. Genetics in Medicine *14*, 10–26.
Cayrol, C., Lacroix, C., Mathe, C., Ecochard, V., Ceribelli, M., Loreau, E., Lazar, V., Dessen, P., Mantovani, R., Aguilar, L., et al. (2007). The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. Blood *109*, 584–594.
Charlesworth, G., Plagnol, V., Holmström, K.M., Bras, J., Sheerin, U.M., Preza, E., Rubio-Agusti, I., Ryten, M., Schneider, S.A., Stamelou, M., et al. (2012). Mutations in ANO3 cause dominant craniocervical dystonia: Ion channel implicated in pathogenesis. American Journal of Human Genetics *91*, 1041–1050.
Chen, Y.-Z. (2012). Autosomal Dominant Familial Dyskinesia and Facial Myokymia. Archives of Neurology *69*, 630.
Chen, P., Burdette, A.J., Porter, J.C., Ricketts, J.C., Fox, S.A., Nery, F.C., Hewett, J.W., Berkowitz, L.A., Breakefield, X.O., Caldwell, K.A., et al. (2010). The early-onset torsion dystonia-associated protein, torsinA, is a homeostatic regulator of endoplasmic reticulum stress response. Human Molecular Genetics *19*, 3502–3515.
Chow, C.Y., Landers, J.E., Bergren, S.K., Sapp, P.C., Grant, A.E., Jones, J.M., Everett, L., Lenk, G.M., McKenna-Yasek, D.M., Weisman, L.S., et al. (2009). Deleterious Variants of FIG4, a Phosphoinositide Phosphatase, in Patients with ALS. American Journal of Human Genetics *84*, 85–88.
Clouaire, T., Roussigne, M., Ecochard, V., Mathe, C., Amalric, F., and Girard, J.-P. (2005). The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proceedings of the National Academy of Sciences of the United States of America *102*, 6907–6912.
Co, M., Anderson, A.G., and Konopka, G. (2020). FOXP transcription factors in vertebrate brain development, function, and disorders. Wiley Interdisciplinary Reviews: Developmental Biology 1–25.
Conesa, A., Madrigal, P., Tarazona, S., Gomez-cabrero, D., Cervera, A., Mcpherson, A., Szcze, W., Gaffney, D.J., Elo, L.L., and Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. 1–19.
Corvol, J.C., Studler, J.M., Schonn, J.S., Girault, J.A., and Hervé, D. (2001). Gαolf is necessary for coupling D1 and A2a receptors to adenylyl cyclase in the striatum. Journal of Neurochemistry *76*, 1585–1588.
Corvol, J.C., Muriel, M.P., Valjent, E., Féger, J., Hanoun, N., Girault, J.A., Hirsch, E.C., and Hervé, D. (2004). Persistent increase in olfactory type G-protein α subunit levels may underlie D1 receptor functional hypersensitivity in Parkinson disease. Journal of Neuroscience *24*, 7007–7014.
Das, S.K., Banerjee, T.K., Biswas, A., Roy, T., Raut, D.K., Chaudhuri, A., and Hazra, A. (2007). Community survey of primary dystonia in the city of Kolkata, India. Movement Disorders *22*, 2031–2036.
Deneault, E., Faheem, M., White, S.H., Rodrigues, D.C., Sun, S., Wei, W., Piekna, A., Thompson, T., Howe, J.L., Chalil, L., et al. (2019). CNTN5 -/+ or EHMT2 -/+ human iPSC-derived neurons from individuals with autism develop hyperactive neuronal networks. ELife *8*, 1–26.
Depienne, C., Bouteiller, D., Keren, B., Cheuret, E., Poirier, K., Trouillard, O., Benyahia, B., Quelin, C., Carpentier, W., Julia, S., et al. (2009). Sporadic infantile epileptic encephalopathy caused by mutations in PCDH19 resembles dravet syndrome but mainly affects females. PLoS Genetics *5*.
Dibbens, L.M., Tarpey, P.S., Hynes, K., Bayly, M.A., Scheffer, I.E., Smith, R., Bomar, J., Sutton, E., Vandeleur, L., Shoubridge, C., et al. (2008). X-linked protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment. Nature Genetics *40*, 776–781.
Dimri, S., Basu, S., and De, A. (2016). Use of BRET to study protein-protein interactions in vitro and in vivo. Methods in Molecular Biology *1443*, 57–78.
Djarmati, A., Schneider, S.A., Lohmann, K., Winkler, S., Pawlack, H., Hagenah, J., Brüggemann, N., Zittel, S., Fuchs, T., Raković, A., et al. (2009). Mutations in THAP1 (DYT6) and generalised dystonia with prominent spasmodic dysphonia: a genetic screening study. The Lancet Neurology *8*, 447–452.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
Dulovic-Mahlow, M., Trinh, J., Kandaswamy, K.K., Braathen, G.J., Di Donato, N., Rahikkala, E., Beblo, S., Werber, M., Krajka, V., Busk, Ø.L., et al. (2019). De Novo Variants in TAOK1 Cause Neurodevelopmental Disorders. American Journal of Human Genetics *105*, 213–220.
Endo, W., Ikemoto, S., Togashi, N., Miyabayashi, T., Nakajima, E., Hamano, S. ichiro, Shibuya, M., Sato, R., Takezawa, Y., Okubo, Y., et al. (2019). Phenotype–genotype correlations in patients with GNB1 gene variants, including the first three reported Japanese patients to exhibit spastic diplegia, dyskinetic quadriplegia, and infantile spasms. Brain and Development *1*, 1–6.
Erlich, Y., Edvardson, S., Hodges, E., Zenvirt, S., Thekkat, P., Shaag, A., Dor, T., Hannon, G.J., and Elpeleg, O. (2011). Exome sequencing and disease-network analysis of a single family implicate a mutation in. Genome Research 658–664.
Erogullari, A., Hollstein, R., Seibler, P., Braunholz, D., Koschmidder, E., Depping, R., Eckhold, J., Lohnau, T., Gillessen-Kaesbach, G., Grünewald, A., et al. (2014). THAP1, the gene mutated in DYT6 dystonia, autoregulates its own expression. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms *1839*, 1196–1204.
Feingold, E.A., Good, P.J., Guyer, M.S., Kamholz, S., Liefer, L., Wetterstrand, K., Collins, F.S., Gingeras, T.R., Kampa, D., Sekinger, E.A., et al. (2004). The ENCODE (ENCyclopedia of DNA Elements) Project. Science *306*, 636–640.
Frederick, N.M., Shah, P. V., Didonna, A., Langley, M.R., Kanthasamy, A.G., and Opal, P. (2019). Loss of the dystonia gene Thap1 leads to transcriptional deficits that converge on common pathogenic pathways in dystonic syndromes. Human Molecular Genetics *28*, 1343–1356.
Free, R.B., Hazelwood, L.A., Cabrera, D.M., Spalding, H.N., Namkung, Y., Rankin, M.L., and Sibley, D.R. (2007). D1 and D2 dopamine receptor expression is regulated by direct interaction with the chaperone protein calnexin. Journal of Biological Chemistry *282*, 21285–21300.
French, C.A., Jin, X., Campbell, T.G., Gerfen, E., Groszer, M., Fisher, S.E., and Costa, R.M. (2012). An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning. Molecular Psychiatry *17*, 1077–1085.
Fuchs, T., Gavarini, S., Saunders-Pullman, R., Raymond, D., Ehrlich, M.E., Bressman, S.B., and Ozelius, L.J. (2009). Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. Nature Genetics *41*, 286–288.
Gallea, C., Herath, P., Voon, V., Lerner, A., Ostuni, J., Saad, Z., Thada, S., Solomon, J., Horovitz, S.G., and Hallett, M. (2018). Loss of inhibition in sensorimotor networks in focal hand dystonia. NeuroImage: Clinical *17*, 90–97.
Garrett, A.M., and Weiner, J.A. (2009). Control of CNS synapse development by γ-protocadherin-mediated astrocyte-neuron contact. Journal of Neuroscience *29*, 11723–11731.
Gavarini, S., Cayrol, C., Fuchs, T., Lyons, N., Ehrlich, M.E., Girard, J.P., and Ozelius, L.J. (2010). Direct interaction between causative genes of DYT1 and DYT6 primary dystonia. Annals of Neurology *68*, 549–553.
Gilbertson, T., Humphries, M., and Steele, J.D. (2019). Maladaptive striatal plasticity and abnormal reward-learning in cervical dystonia. European Journal of Neuroscience *50*, 3191–3204.
Girault, J.A. (2012). Signaling in striatal neurons: The phosphoproteins of reward, addiction, and dyskinesia. Progress in Molecular Biology and Translational Science *106*, 33–62.
Gittis, A.H., and Kreitzer, A.C. (2012). Striatal microcircuitry and movement disorders. Trends in Neurosciences *35*, 557–564.
Gordon-Weeks, P.R., and Fournier, A.E. (2014). Neuronal cytoskeleton in synaptic plasticity and regeneration. Journal of Neurochemistry *129*, 206–212.
Gore, A., Li, Z., Fung, H.L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature *471*, 63–67.
Groszer, M., Keays, D.A., Deacon, R.M.J., de Bono, J.P., Prasad-Mulcare, S., Gaub, S., Baum, M.G., French, C.A., Nicod, J., Coventry, J.A., et al. (2008). Impaired Synaptic Plasticity and Motor Learning in Mice with a Point Mutation Implicated in Human Speech Deficits. Current Biology *18*, 354–362.

Grütz, K., Seibler, P., Weissbach, A., Lohmann, K., Carlisle, F.A., Blake, D.J., Westenberger, A., Klein, C., and Grünewald, A. (2017). Faithful SGCE imprinting in iPSC-derived cortical neurons: An endogenous cellular model of myoclonus-dystonia. Scientific Reports *7*, 1–9.
Guillabert-Gourgues, A., Jaspard-Vinassa, B., Bats, M.L., Sewduth, R.N., Franzl, N., Peghaire, C., Jeanningros, S., Moreau, C., Roux, E., Larrieu-Lahargue, F., et al. (2016). Kif26b controls endothelial cell polarity through the Dishevelled/Daam1-dependent planar cell polaritysignaling pathway. Molecular Biology of the Cell *27*, 941–953.
Gurevich, E. V., Gainetdinov, R.R., and Gurevich, V. V. (2016). G protein-coupled receptor kinases as regulators of dopamine receptor functions. Pharmacological Research *111*, 1–16.
de Gusmao, C.M., Stone, S., Waugh, J.L., Yang, E., Lenk, G.M., and Rodan, L.H. (2019). VAC14 Gene-Related Parkinsonism-Dystonia With Response to Deep Brain Stimulation. Movement Disorders Clinical Practice *6*, 494–497.
Hayashi, S., Inoue, Y., Hattori, S., Kaneko, M., Shioi, G., Miyakawa, T., and Takeichi, M. (2017). Loss of X-linked Protocadherin-19 differentially affects the behavior of heterozygous female and hemizygous male mice. Scientific Reports *7*, 1–15.
Helassa, N., Antonyuk, S. V., Lian, L.Y., Haynes, L.P., and Burgoyne, R.D. (2017). Biophysical and functional characterization of hippocalcin mutants responsible for human dystonia. Human Molecular Genetics *26*, 2426–2435.
Hemati, P., Revah-Politi, A., Bassan, H., Petrovski, S., Bilancia, C.G., Ramsey, K., Griffin, N.G., Bier, L., Cho, M.T., Rosello, M., et al. (2018). Refining the phenotype associated with GNB1 mutations: Clinical data on 18 newly identified patients and review of the literature. American Journal of Medical Genetics, Part A *176*, 2259–2275.
Herbert, Z.T., Kershner, J.P., Butty, V.L., Thimmapuram, J., Choudhari, S., Alekseyev, Y.O., Fan, J., Podnar, J.W., Wilcox, E., Gipson, J., et al. (2018). Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. BMC Genomics *19*, 1–10.
Hervé, D. (2011). Identification of a specific assembly of the G protein Golf as a critical and regulated module of dopamine and adenosine-activated cAMP pathways in the striatum. Frontiers in Neuroanatomy *5*, 1–9.
Hervé, D., Le Moine, C., Corvol, J.C., Belluscio, L., Ledent, C., Fienberg, A.A., Jaber, M., Studler, J.M., and Girault, J.A. (2001). Gαolf levels are regulated by receptor usage and control dopamine and adenosine action in the striatum. Journal of Neuroscience *21*, 4390–4399.
Hewett, J.W., Zeng, J., Niland, B.P., Bragg, D.C., and Breakefield, X.O. (2006). Dystonia-causing mutant torsinA inhibits cell adhesion and neurite extension through interference with cytoskeletal dynamics. Neurobiology of Disease *22*, 98–111.
Hirata, H., Umemori, J., Yoshioka, H., Koide, T., Watanabe, K., and Shimoda, Y. (2016). Cell adhesion molecule contactin-associated protein 3 is expressed in the mouse basal ganglia during early postnatal stages. Journal of Neuroscience Research *94*, 74–89.
Hirokawa, N., and Tanaka, Y. (2015). Kinesin superfamily proteins (KIFs): Various functions and their relevance for important phenomena in life and diseases. Experimental Cell Research *334*, 16–25.
Hollins, B., Kuravi, S., Digby, G.J., and Lambert, N.A. (2009). The c-terminus of GRK3 indicates rapid dissociation of G protein heterotrimers. Cellular Signalling *21*, 1015–1021.
Hosack, D.A., Dennis, G., Sherman, B.T., Lane, H.C., and Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE. Genome Biology *4*.
Houlden, H., Schneider, S.A., Paudel, R., Melchers, A., Schwingenschuh, P., Edwards, M., Hardy, J., and Bhatia, K.P. (2010). THAP1 mutations (DYT6) are an additional cause of early-onset dystonia. Neurology *74*, 846–850.
Hsieh, J.J.-D., Ernst, P., Erdjument-Bromage, H., Tempst, P., and Korsmeyer, S.J. (2003). Proteolytic Cleavage of MLL Generates a Complex of N- and C-Terminal Fragments That Confers Protein Stability and Subnuclear Localization. Molecular and Cellular Biology *23*, 186–194.
Hu, B.Y., Weick, J.P., Yu, J., Ma, L.X., Zhang, X.Q., Thomson, J.A., and Zhang, S.C. (2010). Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. Proceedings of the National Academy of Sciences of the United States of America *107*, 4335–4340.

Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Narvä, E., Ng, S., Sourour, M., Hämälä, R., Olsson, C., et al. (2011). Copy number variation and selection during reprogramming to pluripotency. Nature *471*, 58–62.
Ichinose, H., Ohye, T., Takahashi, E. ichi, Seki, N., Hori, T. aki, Segawa, M., Nomura, Y., Endo, K., Tanaka, H., Tsuji, S., et al. (1994). Hereditary progressive dystonia with marked diurnal fluctuation caused by mutations in the GTP cyclohydrolase I gene. Nature Genetics *8*, 236–242.
Jin, N., Chow, C.Y., Liu, L., Zolov, S.N., Bronson, R., Davisson, M., Petersen, J.L., Zhang, Y., Park, S., Duex, J.E., et al. (2008). VAC14 nucleates a protein complex essential for the acute interconversion of PI3P and PI(3,5)P2 in yeast and mouse. EMBO Journal *27*, 3221–3234.
Jinnah, H.A., and Factor, S.A. (2015). Diagnosis and Treatment of Dystonia. Neurologic Clinics *33*, 77–100.
Jinnah, H.A., Neychev, V., and Hess, E.J. (2017). The Anatomical Basis for Dystonia: The Motor Network Model. Tremor and Other Hyperkinetic Movements (New York, N.Y.) *7*, 506.
Jones, W.D., Dafou, D., McEntagart, M., Woollard, W.J., Elmslie, F. V., Holder-Espinasse, M., Irving, M., Saggar, A.K., Smithson, S., Trembath, R.C., et al. (2012). De novo mutations in MLL cause Wiedemann-Steiner syndrome. American Journal of Human Genetics *91*, 358–364.
Kaiser, F.J., Osmanoric, A., Rakovic, A., Erogullari, A., Uflacker, N., Braunholz, D., Lohnau, T., Orolicki, S., Albrecht, M., Gillessen-Kaesbach, G., et al. (2010). The dystonia gene DYT1 is repressed by the transcription factor THAP1 (DYT6). Annals of Neurology *68*, 554–559.
Kamm, C., Fischer, H., Garavaglia, B., Kullmann, S., Sharma, M., Schrader, C., Grundmann, K., Klein, C., Borggräfe, I., Lobsien, E., et al. (2008). Susceptibility to DYT1 dystonia in european patients is modified by the D216h polymorphism. Neurology *70*, 2261–2262.
Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature Methods *7*, 1009–1015.
Kaur, P., Bhavani, G.S.L., Raj, A., Girisha, K.M., and Shukla, A. (2019). Homozygous variant, p.(Arg643Trp) in VAC14 causes striatonigral degeneration: report of a novel variant and review of VAC14-related disorders. Journal of Human Genetics *64*, 1237–1242.
Kim, D., and Ackerman, S.L. (2011). The UNC5C netrin receptor regulates dorsal guidance of mouse hindbrain axons. Journal of Neuroscience *31*, 2167–2179.
Kishino, T., Lalande, M., and Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. Nature Genetics *15*, 70–73.
Kleefstra, T., Kramer, J.M., Neveling, K., Willemsen, M.H., Koemans, T.S., Vissers, L.E.L.M., Wissink-Lindhout, W., Fenckova, M., Van Den Akker, W.M.R., Kasri, N.N., et al. (2012). Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. American Journal of Human Genetics *91*, 73–82.
Kleijer, K.T.E., Zuko, A., Shimoda, Y., Watanabe, K., and Burbach, J.P.H. (2015). Contactin-5 expression during development and wiring of the thalamocortical system. Neuroscience *310*, 106–113.
Klein, C., Baumann, H., Olschewski, L., Hanssen, H., Münchau, A., Ferbert, A., Brüggemann, N., and Lohmann, K. (2019). De-novo KMT2B mutation in a consanguineous family: 15-Year follow-up of an Afghan dystonia patient. Parkinsonism and Related Disorders *64*, 337–339.
Koizumi, H., Fujioka, H., Togashi, K., Thompson, J., Yates, J.R., Gleeson, J.G., and Emoto, K. (2017). DCLK1 phosphorylates the microtubule-associated protein MAP7D1 to promote axon elongation in cortical neurons. Developmental Neurobiology *77*, 493–510.
Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importinα. Journal of Biological Chemistry *284*, 478–485.
Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519–523.
Lange, L.M., Tunc, S., Tennstedt, S., Münchau, A., Klein, C., Assmann, B., and Lohmann, K. (2017). A novel, in-frame KMT2B deletion in a patient with apparently isolated, generalized dystonia. Movement Disorders *32*, 1495–1497.
Lefebvre, J.L., Kostadinov, D., Chen, W. V., Maniatis, T., and Sanes, J.R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. Nature *488*, 517–521.

Lenk, G.M., Ferguson, C.J., Chow, C.Y., Jin, N., Jones, J.M., Grant, A.E., Zolov, S.N., Winters, J.J., Giger, R.J., Dowling, J.J., et al. (2011). Pathogenic mechanism of the FIG4 mutation responsible for charcot-marie-tooth disease CMT4J. PLoS Genetics *7*.
Lenk, G.M., Szymanska, K., Debska-Vielhaber, G., Rydzanicz, M., Walczak, A., Bekiesinska-Figatowska, M., Vielhaber, S., Hallmann, K., Stawinski, P., Buehring, S., et al. (2016). Biallelic Mutations of VAC14 in Pediatric-Onset Neurological Disease. American Journal of Human Genetics *99*, 188–194.
Lenk, G.M., Park, Y.N., Lemons, R., Flynn, E., Plank, M., Frei, C.M., Davis, M.J., Gregorka, B., Swanson, J.A., Meisler, M.H., et al. (2019). CRISPR knockout screen implicates three genes in lysosome function. Scientific Reports *9*, 2–11.
Li, B., and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*.
Li, S., Tiab, L., Jiao, X., Munier, F.L., Zografos, L., Frueh, B.E., Sergeev, Y., Smith, J., Rubin, B., Meallet, M.A., et al. (2005). Mutations in PIP5K3 are associated with François-Neetens mouchetée fleck corneal dystrophy. American Journal of Human Genetics *77*, 54–63.
Liang, C.C., Tanabe, L.M., Jou, S., Chi, F., and Dauer, W.T. (2014). TorsinA hypofunction causes abnormal twisting movements and sensorimotor circuit neurodegeneration. Journal of Clinical Investigation *124*, 3080–3092.
Liao, S., Chen, T., Dai, Y., Wang, Y., Wu, F., and Zhong, M. (2020). Novel VAC14 variants identified in two Chinese siblings with childhood-onset striatonigral degeneration. Molecular Genetics and Genomic Medicine *8*.
Lines, M.A., Ito, Y., Kernohan, K.D., Mears, W., Hurteau-Miller, J., Venkateswaran, S., Ward, L., Khatchadourian, K., McClintock, J., Bhola, P., et al. (2017). Yunis-Varón syndrome caused by biallelic VAC14 mutations. European Journal of Human Genetics *25*, 1049–1054.
Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods *25*, 402–408.
Llambi, F., Causeret, F., Bloch-Gallego, E., and Mehlen, P. (2001). Netrin-1 acts as a survival factor via its receptors UNC5H and DCC. EMBO Journal *20*, 2715–2722.
Lohmann, K., Masuho, I., Patil, D.N., Baumann, H., Hebert, E., Steinrücke, S., Trujillano, D., Skamangas, N.K., Dobricic, V., Hüning, I., et al. (2017). Novel GNB1 mutations disrupt assembly and function of G protein heterotrimers and cause global developmental delay in humans. Human Molecular Genetics *26*, 1078–1086.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 1–21.
Lubitz, S., Glaser, S., Schaft, J., Stewart, A.F., and Anastassiadis, K. (2007). Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. Molecular Biology of the Cell *18*, 2356–2366.
Lüdecke, B., Dworniczak, B., and Bartholomé, K. (1995). A point mutation in the tyrosine hydroxylase gene associated with Segawa's syndrome. Human Genetics *95*, 123–125.
Luttrell, L.M., and Miller, W.E. (2013). Arrestins as regulators of kinases and phosphatases (Elsevier Inc.).
Lv, X., Ren, S.Q., Zhang, X.J., Shen, Z., Ghosh, T., Xianyu, A., Gao, P., Li, Z., Lin, S., Yu, Y., et al. (2019). TBR2 coordinates neurogenesis expansion and precise microcircuit organization via Protocadherin 19 in the mammalian cortex. Nature Communications *10*, 1–15.
Lyon, G.J., Marchi, E., Ekstein, J., Meiner, V., Hirsch, Y., Scher, S., Yang, E., De Vivo, D.C., Madrid, R., Li, Q., et al. (2019). VAC14 syndrome in two siblings with retinitis pigmentosa and neurodegeneration with brain iron accumulation. Cold Spring Harbor Molecular Case Studies *5*, 1–10.
Maltese, M., Martella, G., Imbriani, P., Schuermans, J., Billion, K., Sciamanna, G., Farook, F., Ponterio, G., Tassone, A., Santoro, M., et al. (2017). Abnormal striatal plasticity in a DYT11/SGCE myoclonus dystonia mouse model is reversed by adenosine A2A receptor inhibition. Neurobiology of Disease *108*, 128–139.
Maltese, M., Stanic, J., Tassone, A., Sciamanna, G., Ponterio, G., Vanni, V., Martella, G., Imbriani, P., Bonsi, P., Mercuri, N.B., et al. (2018). Early structural and functional plasticity alterations in a susceptibility period of DYT1 dystonia mouse striatum. ELife *7*, 1–23.

Marras, C., Lang, A., van de Warrenburg, B.P., Sue, C.M., Tabrizi, S.J., Bertram, L., Mercimek-Mahmutoglu, S., Ebrahimi-Fakhari, D., Warner, T.T., Durr, A., et al. (2016). Nomenclature of genetic movement disorders: Recommendations of the international Parkinson and movement disorder society task force. Movement Disorders *31*, 436–457.
Martella, G., Tassone, A., Sciamanna, G., Platania, P., Cuomo, D., Viscomi, M.T., Bonsi, P., Cacci, E., Biagioni, S., Usiello, A., et al. (2009). Impairment of bidirectional synaptic plasticity in the striatum of a mouse model of DYT1 dystonia: Role of endogenous acetylcholine. Brain *132*, 2336–2349.
Martino, D., Gajos, A., Gallo, V., Cif, L., Coubes, P., Tinazzi, M., Schneider, S.A., Fiorio, M., Zorzi, G., Nardocci, N., et al. (2013). Extragenetic factors and clinical penetrance of DYT1 dystonia: An exploratory study. Journal of Neurology *260*, 1081–1086.
Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.C., Yakir, B., Clark, A.T., Plath, K., Lowry, W.E., and Benvenisty, N. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell *7*, 521–531.
Mazars, R., Gonzalez-de-Peredo, A., Cayrol, C., Lavigne, A.C., Vogel, J.L., Ortega, N., Lacroix, C., Gautier, V., Huet, G., Ray, A., et al. (2010). The THAP-Zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase a link between DYT6 and DYT3 dystonias. Journal of Biological Chemistry *285*, 13364–13371.
Meyer, E., Carss, K.J., Rankin, J., Nichols, J.M.E., Grozeva, D., Joseph, A.P., Mencacci, N.E., Papandreou, A., Ng, J., Barral, S., et al. (2017). Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. Nature Genetics *49*, 223–237.
Mogil, L.S., Andaleon, A., Badalamenti, A., Dickinson, S.P., Guo, X., Rotter, J.I., Johnson, W.C., Im, H.K., Liu, Y., and Wheeler, H.E. (2018). Genetic architecture of gene expression traits across diverse populations. PLoS Genetics *14*, 1–21.
Molnár, Z., Garel, S., López-bendito, G., Maness, P., and Price, J. (2015). Through the Embryonic Forebrain. *35*, 1573–1585.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods *5*, 621–628.
Müller, B., Hedrich, K., Kock, N., Dragasevic, N., Svetel, M., Garrels, J., Landt, O., Nitschke, M., Pramstaller, P.P., Reik, W., et al. (2002a). Evidence that paternal expression of the ε-Sarcoglycan gene accounts for reduced penetrance in myoclonus-dystonia. American Journal of Human Genetics *71*, 1303–1311.
Müller, J., Kiechl, S., Wenning, G.K., Seppi, K., Willeit, J., Gasperi, A., Wissel, J., Gasser, T., and Poewe, W. (2002b). The prevalence of primary dystonia in the general community. Neurology *59*, 941–943.
Napolitano, F., Pasqualetti, M., Usiello, A., Santini, E., Pacini, G., Sciamanna, G., Errico, F., Tassone, A., Di Dato, V., Martella, G., et al. (2010). Dopamine D2 receptor dysfunction is rescued by adenosine A2A receptor antagonism in a model of DYT1 dystonia. Neurobiology of Disease *38*, 434–445.
Nery, F.C., Armata, I.A., Farley, J.E., Cho, J.A., Yaqub, U., Chen, P., Da Hora, C.C., Wang, Q., Tagaya, M., Klein, C., et al. (2011). TorsinA participates in endoplasmic reticulum-associated degradation. Nature Communications *2*.
Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature Genetics *42*, 790–793.
Niethammer, M., Carbon, M., Argyelan, M., and Eidelberg, D. (2011). Hereditary dystonia as a neurodevelopmental circuit disorder: Evidence from neuroimaging. Neurobiology of Disease *42*, 202–209.
Okae, H., and Iwakura, Y. (2010). Neural tube defects and impaired neural progenitor cell proliferation in Gβ1-deficient mice. Developmental Dynamics *239*, 1089–1101.
Opal, P., Tintner, R., Jankovic, J., Leung, J., Breakefield, X.O., Friedman, J., and Ozeluis, L. (2002). Intrafamilial phenotypic variability of the DYT1 dystonia: From asymptomatic TOR1A gene carrier status to dystonic storm. Movement Disorders *17*, 339–345.
Oppenheim, H. (1911). Über eine eigenartige Krampfkrankheit des kindlichen und jugendlichen Alters (Dysbasia lordotica progressiva, Dystonia musculorum deformans). Neurologisches Centralblatt *30*, 1090–1107.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nature Biotechnology *26*, 313–315.
Osmanovic, A., Dendorfer, A., Erogullari, A., Uflacker, N., Braunholz, D., Rakovic, A., Vierke, G., Gil-Rodrıguez, C., Muenchau, A., Albrecht, M., et al. (2011). Truncating Mutations in THAP1 Define the Nuclear Localization Signal. Movement Disorders *26*, 1565–1567.
Paudel, R., Hardy, J., Revesz, T., Holton, J.L., and Houlden, H. (2012). Review: Genetics and neuropathology of primary pure dystonia. Neuropathology and Applied Neurobiology *38*, 520–534.
Paudel, R., Li, A., Hardy, J., Bhatia, K.P., Houlden, H., and Holton, J. (2016). DYT6 dystonia: A neuropathological study. Neurodegenerative Diseases *16*, 273–278.
Pederick, D.T., Richards, K.L., Piltz, S.G., Kumar, R., Mincheva-Tasheva, S., Mandelstam, S.A., Dale, R.C., Scheffer, I.E., Gecz, J., Petrou, S., et al. (2018). Abnormal Cell Sorting Underlies the Unique X-Linked Inheritance of PCDH19 Epilepsy. Neuron *97*, 59-66.e5.
Peterson, D.A., Sejnowski, T.J., and Poizner, H. (2010). Convergent evidence for abnormal striatal synaptic plasticity in dystonia. Neurobiology of Disease *37*, 558–573.
Petrovski, S., Küry, S., Myers, C.T., Anyane-Yeboa, K., Cogné, B., Bialer, M., Xia, F., Hemati, P., Riviello, J., Mehaffey, M., et al. (2016). Germline de Novo Mutations in GNB1 Cause Severe Neurodevelopmental Disability, Hypotonia, and Seizures. American Journal of Human Genetics *98*, 1001–1010.
Pisani, A., Martella, G., Tscherter, A., Bonsi, P., Sharma, N., Bernardi, G., and Standaert, D.G. (2006). Altered responses to dopaminergic D2 receptor activation and N-type calcium currents in striatal cholinergic interneurons in a mouse model of DYT1 dystonia. Neurobiology of Disease *24*, 318–325.
Plotkin, J.L., and Goldberg, J.A. (2019). Thinking Outside the Box (and Arrow): Current Themes in Striatal Dysfunction in Movement Disorders. Neuroscientist *25*, 359–379.
Poirier, K., Lebrun, N., Broix, L., Tian, G., Saillour, Y., Boscheron, C., Parrini, E., Valence, S., Pierre, B. Saint, Oger, M., et al. (2013). Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. Nature Genetics *45*, 639–647.
Poliak, S., Morales, D., Croteau, L.P., Krawchuk, D., Palmesino, E., Morton, S., Cloutier, J.F., Charron, F., Dalva, M.B., Ackerman, S.L., et al. (2015). Synergistic integration of netrin and ephrin axon guidance signals by spinal motor neurons. ELife *4*, 1–26.
Powis, Z., Towne, M.C., Hagman, K.D.F., Blanco, K., Palmaer, E., Castro, A., Sajan, S.A., Radtke, K., Feyma, T.J., Juliette, K., et al. (2019). Clinical diagnostic exome sequencing in dystonia: Genetic testing challenges for complex conditions. Clinical Genetics 1–7.
Di Prisco, G.V., Huang, W., Buffington, S.A., Hsu, C.C., Bonnen, P.E., Placzek, A.N., Sidrauski, C., Krnjević, K., Kaufman, R.J., Walter, P., et al. (2014). Translational control of mGluR-dependent long-term depression and object-place learning by eIF2α. Nature Neuroscience *17*, 1073–1082.
Prudente, C.N., Pardo, C.A., Xiao, J., Hanfelt, J., Hess, E.J., LeDoux, M.S., and Jinnah, H.A. (2013). Neuropathology of cervical dystonia. Experimental Neurology *241*, 95–104.
Qian, C., and Zhou, M.M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cellular and Molecular Life Sciences *63*, 2755–2763.
Ramdhani, R.A., and Simonyan, K. (2013). Primary dystonia: conceptualizing the disorder through a structural brain imaging lens. Tremor and Other Hyperkinetic Movements (New York, N.Y.) *3*, 1–11.
Risch, N.J., Bressman, S.B., Senthil, G., and Ozelius, L.J. (2007). Intragenic Cis and Trans modification of genetic susceptibility in DYT1 torsion dystonia. American Journal of Human Genetics *80*, 1188–1193.
Roussigne, M., Cayrol, C., Clouaire, T., Amalric, F., and Girard, J.P. (2003). THAP1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (Par-4) to PML nuclear bodies. Oncogene *22*, 2432–2442.
Ruiz, M., Perez-Garcia, G., Ortiz-Virumbrales, M., Méneret, A., Morant, A., Kottwitz, J., Fuchs, T., Bonet, J., Gonzalez-Alegre, P., Hof, P.R., et al. (2015). Abnormalities of motor function, transcription and cerebellar structure in mouse models of THAP1 dystonia. Human Molecular Genetics *24*, 7159–7170.
Salahpour, A., Espinoza, S., Masri, B., Lam, V., Barak, L.S., and Gainetdinov, R.R. (2012). BRET biosensors to study GPCR biology, pharmacology, and signal transduction. Frontiers in Endocrinology

3, 1–9.
Saunders-Pullman, R., Raymond, D., Senthil, G., Kramer, P., Ohmann, E., Deligtisch, A., Shanker, V., Greene, P., Tabamo, R., Huang, N., et al. (2007). Narrowing the DYT6 dystonia region and evidence for locus heterogeneity in the amish-mennonites. American Journal of Medical Genetics, Part A *143*, 2098–2105.
Schirinzi, T., Sciamanna, G., Mercuri, N.B., and Pisani, A. (2018). Dystonia as a network disorder: A concept in evolution. Current Opinion in Neurology *31*, 498–503.
Scholes, A.N., and Lewis, J.A. (2019). Comparison of RNA Isolation Methods on RNA-Seq: Implications for Differential Expression and Meta-Analyses. BioRxiv 728014.
Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: An RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology *7*, 1–14.
Schwindinger, W.F., Murphree Mihalcik, L.J., Giger, K.E., Betz, K.S., Stauffer, A.M., Linden, J., Herve, D., and Robishaw, J.D. (2010). Adenosine A2A receptor signaling and Golf assembly show a specific requirement for the γ7 subtype in the striatum. Journal of Biological Chemistry *285*, 29787–29796.
Sharma, N. (2019). Neuropathology of Dystonia. Tremor and Other Hyperkinetic Movements (New York, N.Y.) *9*, 569.
Shen, E., Shulha, H., Weng, Z., and Akbarian, S. (2014). Regulation of histone H3K4 methylation in brain development and disease. Philosophical Transactions of the Royal Society B: Biological Sciences *369*.
Shi, Y., Kirwan, P., and Livesey, F.J. (2012a). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nature Protocols *7*, 1836–1846.
Shi, Y., Kirwan, P., Smith, J., Robinson, H.P.C.C., and Livesey, F.J. (2012b). Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nature Neuroscience *15*, 477–486.
Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. Nature Reviews Genetics *15*, 121–132.
Siokas, V., Aloizou, A.M., Tsouris, Z., Michalopoulou, A., Mentis, A.F.A., and Dardiotis, E. (2019). Risk factor genes in patients with dystonia: A comprehensive review. Tremor and Other Hyperkinetic Movements *9*, 1–10.
Song, W., Chen, Y., Huang, R., Chen, K., Pan, P., Yang, Y., and Shang, H.F. (2011). Novel THAP1 gene mutations in patients with primary dystonia from Southwest China. Journal of the Neurological Sciences *309*, 63–67.
Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. Nature Reviews Genetics *20*, 631–656.
Steeves, T.D., Day, L., Dykeman, J., Jette, N., and Pringsheim, T. (2012). The prevalence of primary dystonia: A systematic review and meta-analysis. Movement Disorders *27*, 1789–1796.
Steinrücke, S., Lohmann, K., Domingo, A., Rolfs, A., Bäumer, T., Spiegler, J., Hartmann, C., and Münchau, A. (2016). Novel GNB1 missense mutation in a patient with generalized dystonia, hypotonia, and intellectual disability. Neurology: Genetics *2*, 7–9.
Storey, J.D., Madeoy, J., Strout, J.L., Wurfel, M., Ronald, J., and Akey, J.M. (2007). Gene-expression variation within and among human populations. American Journal of Human Genetics *80*, 502–509.
Strässler, E.T., Aalto-Setälä, K., Kiamehr, M., Landmesser, U., and Kränkel, N. (2018). Age Is Relative—Impact of Donor Age on Induced Pluripotent Stem Cell-Derived Cell Functionality. Frontiers in Cardiovascular Medicine *5*, 1–8.
Stutterd, C., Diakumis, P., Bahlo, M., Fanjul Fernandez, M., Leventer, R.J., Delatycki, M., Amor, D., Chow, C.W., Stephenson, S., Meisler, M.H., et al. (2017). Neuropathology of childhood-onset basal ganglia degeneration caused by mutation of VAC14. Annals of Clinical and Translational Neurology *4*, 859–864.
Surmeier, D.J., Ding, J., Day, M., Wang, Z., and Shen, W. (2007). D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends in Neurosciences *30*, 228–235.
Taghavi, S., Chaouni, R., Tafakhori, A., Azcona, L.J., Firouzabadi, S.G., Omrani, M.D., Jamshidi, J., Emamalizadeh, B., Shahidi, G.A., Ahmadi, M., et al. (2018). A Clinical and Molecular Genetic Study of

50 Families with Autosomal Recessive Parkinsonism Revealed Known and Novel Gene Mutations. Molecular Neurobiology *55*, 3477–3489.
Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: A matter of depth. Genome Research *21*, 2213–2223.
Tellmann, G. (2006). The E-Method: A highly accurate technique for gene-expression analysis. Nature Methods *3*.
Uchiyama, Y., Sakaguchi, M., Terabayashi, T., Inenaga, T., Inoue, S., Kobayashi, C., Oshima, N., Kiyonari, H., Nakagata, N., Sato, Y., et al. (2010). Kif26b, a kinesin family gene, regulates adhesion of the embryonic kidney mesenchyme. Proceedings of the National Academy of Sciences of the United States of America *107*, 9240–9245.
Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. Science *347*.
Vernes, S.C., Oliver, P.L., Spiteri, E., Lockstone, H.E., Puliyadi, R., Taylor, J.M., Ho, J., Mombereau, C., Brewer, A., Lowy, E., et al. (2011). FOXP2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genetics *7*.
Vo, A., Sako, W., Niethammer, M., Carbon, M., Bressman, S.B., Ulug, A.M., and Eidelberg, D. (2015). Thalamocortical connectivity correlates with phenotypic variability in dystonia. Cerebral Cortex *25*, 3086–3094.
Weber, J. (2019). Transkriptomanalyse bei betroffenen und nicht betroffenen THAP1-Mutationsträgern: Auf der Suche nach Faktoren der reduzierten Penetranz. University of Lübeck, Lübeck, Germany.
Wetzel-Smith, M.K., Hunkapiller, J., Bhangale, T.R., Srinivasan, K., Maloney, J.A., Atwal, J.K., Sa, S.M., Yaylaoglu, M.B., Foreman, O., Ortmann, W., et al. (2014). A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. Nature Medicine *20*, 1452–1457.
Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M.M., and Sexton, P.M. (2018). Mechanisms of signalling and biased agonism in G protein-coupled receptors. Nature Reviews Molecular Cell Biology *19*, 638–653.
Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. Protein and Cell *2*, 202–214.
Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T.R., and Min, J. (2018). DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. Structure *26*, 85-95.e3.
Yagi, T. (2008). Clustered protocadherin family. Development Growth and Differentiation *50*, 358–364.
Yamada, K., Andrews, C., Chan, W.M., McKeown, C.A., Magli, A., De Berardinis, T., Loewenstein, A., Lazar, M., O'Keefe, M., Letson, R., et al. (2003). Heterozygous mutations of the kinesin KIF21A in congenital fibrosis of the extraocular muscles type 1 (CFEOM1). Nature Genetics *35*, 318–321.
Yellajoshyula, D., Liang, C.C., Pappas, S.S., Penati, S., Yang, A., Mecano, R., Kumaran, R., Jou, S., Cookson, M.R., and Dauer, W.T. (2017). The DYT6 Dystonia Protein THAP1 Regulates Myelination within the Oligodendrocyte Lineage. Developmental Cell *42*, 52-67.e4.
Yoda, A., Adelmant, G., Tamburini, J., Chapuy, B., Shindoh, N., Yoda, Y., Weigert, O., Kopp, N., Wu, S.C., Kim, S.S., et al. (2015). Mutations in G protein β subunits promote transformation and kinase inhibitor resistance. Nature Medicine *21*, 71–75.
Yokoi, F., Dang, M.T., Li, J., Standaert, D.G., and Li, Y. (2011). Motor deficits and decreased striatal dopamine receptor 2 binding activity in the striatum-specific Dyt1 conditional knockout mice. PLoS ONE *6*.
Yunis, E., and Varón, H. (1980). Cleidocranial Dysostosis, Severe Micrognathism, Bilateral Absence of Thumbs and First Metatarsal Bone, and Distal Aphalangia: A New Genetic Syndrome. American Journal of Diseases of Children *134*, 649–653.
Zakirova, Z., Fanutza, T., Bonet, J., Readhead, B., Zhang, W., Yi, Z., Beauvais, G., Zwaka, T.P., Ozelius, L.J., Blitzer, R.D., et al. (2018). Mutations in THAP1/DYT6 reveal that diverse dystonia genes disrupt similar neuronal pathways and functions. PLoS Genetics *14*, 1–25.
Zech, M., Boesch, S., Maier, E.M., Borggraefe, I., Vill, K., Laccone, F., Pilshofer, V., Ceballos-Baumann, A., Alhaddad, B., Berutti, R., et al. (2016). Haploinsufficiency of KMT2B, Encoding the Lysine-Specific

Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. American Journal of Human Genetics *99*, 1377–1387.
Zhang, Y., Zolov, S.N., Chow, C.Y., Slutsky, S.G., Richardson, S.C., Piper, R.C., Yang, B., Nau, J.J., Westrick, R.J., Morrison, S.J., et al. (2007). Loss of Vac14, a regulator of the signaling lipid phosphatidylinositol 3,5-bisphosphate, results in neurodegeneration in mice. Proceedings of the National Academy of Sciences of the United States of America *104*, 17518–17523.
Zhao, Y., Xiao, J., Gong, S., Clara, J.A., and LeDoux, M.S. (2013). Neural expression of the transcription factor THAP1 during development in rat. Neuroscience *231*, 282–295.
Zittel, S., Moll, C.K.E., Brüggemann, N., Tadic, V., Hamel, W., Kasten, M., Lohmann, K., Lohnau, T., Winkler, S., Gerloff, C., et al. (2010). Clinical neuroimaging and electrophysiological assessment of three DYT6 dystonia families. Movement Disorders *25*, 2405–2412.
Zou, Y., Zhang, W.F., Liu, H.Y., Li, X., Zhang, X., Ma, X.F., Sun, Y., Jiang, S.Y., Ma, Q.H., and Xu, D.E. (2017). Structure and function of the contactin-associated protein family in myelinated axons and their relationship with nerve diseases. Neural Regeneration Research *12*, 1551–1558.
Zuko, A., Bouyain, S., Van Der Zwaag, B., and Burbach, J.P.H. (2011). Contactins: Structural aspects in relation to developmental functions in brain disease (Elsevier Inc.).

9. Appendix

9.1 List of abbreviations

AAO	Age at onset
ACTB	Actin beta
ADYY5	Adenylate cyclase 5
ANO3	Anoctamin 3
BDNF	Brain derived neurotrophic factor
BRAT1	BRCA1 associated ATM activator 1
BRET	Bioluminescence resonance energy transfer
BRN2 (POU3F2)	POU class 3 homeobox 2
CADD	Combined Annotation Dependent Depletion
cDNA	Complementary DNA
CNTN5	Contactin 5
CNTNAP3B	Contactin associated protein like 3B
CTIP2	COUP-TF-interacting protein 2
DBS	Deep brain stimulation
DCLK1	Doublecortin like kinase 1
ddH_2O	Double-distilled water
DEG	Differentially expressed gene
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribose nucleoside triphosphate
dT	Desoxythymidin
DYT	Locus designation for dystonia
E.coli	*Escherichia coli*
E8	Essential 8™ medium
EB	Embryoid body
EDTA	Ethylenediamine tetraacetic acid
ENCODE	Encyclopaedia of DNA elements
eQTL	Expression quantitative trait loci
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
FGF2	Fibroblast growth factor 2
FIG4	FIG4 phosphoinositide 5-phosphatase
FOXD1	Forkhead box D1
FOXP2	Forkhead box P2
GABA	Gamma-aminobutyric acid
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GC-content	Guanine-cytosine content
GNAL	G protein subunit alpha L
GNB1 / $G\beta_1$	Guanine nucleotide-binding protein beta 1 gene / protein
gnomAD	Genome Aggregation Database
GO	Gene ontology
GPCR	G-protein-coupled receptor
GRK3	G-protein-coupled receptor kinase 3
GTP	Guanosine-5'-triphosphate
HEK293T	Human embryonic kidney cell 293T
HPCA	Hippocalcin
HPLC-water	High pressure liquid chromatography grade water
iPSC	Induced pluripotent stem cell
IQR	Interquartile range

KIF26B	Kinesin family member 26B
KLF4	Kruppel like factor 4
KMT	Lysine methyltransferase
KMT2B	Lysine methyltransferase 2B
KOS	Acronym for Klf4, Oct3/4, and Sox2
KSR	Neural differentiation medium
LoF	Loss of function
MAP2	Microtubule associate protein 2
MDSGene	Movement Disorder Society Genetic mutation database
MEM-NEAA	Minimum Essential Medium Non-Essential Amino Acids
$MgCl_2$	Magnesium chloride
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid
mTeSR1	mTeSR™1 medium
MYC	MYC Proto-Oncogene
NANOG	nanog homeobox
NIuc	NanoLuc® luciferase
NMM	Neural maintenance medium
NRG3	Neuregulin 3
OCT4 (POU5F1)	POU class 5 homeobox 1
OMIM	Online Mendelian Inheritance in Man
Opti-MEM	Opti-Minimum Essential Medium
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCDH19	Protocadherin 19
PCR	Polymerase chain reaction
PDL/LA	Poly-D-lysine/laminin
PenStrep	Penicillin/streptomycin
Phred Score	Quality measure of nucleobases
$PI(3,5)P_2$	Phosphatidylinositol 3,5-bisphosphate
PIKFYVE	Phosphoinositide kinase, FYVE-type zinc finger containing
pLI	Probability of loss of function
PLO/LA	Poly-L-ornithine/laminin
PRKRA	Protein activator of interferon induced protein kinase EIF2AK2
qRT-PCR	Quantitative real-time PCR
Read	Sequence of base pairs used in automated sequencing
RIN	RNA integrity number
RNA	Ribonucleic acid
RNA-Seq	RNA-Sequencing
RRAGC	Ras related GTP binding C
SETX	Senataxin
SGCE	Sarcoglycan epsilon
SH-SY5Y	Neuroblastoma cells
SLC	Solute carrier
SNP	Single-nucleotide polymorphism
SOX17	SRY-box transcription factor 17
SOX2	SRY-box transcription factor 2
SRY	Sex determining region Y
SSEA4	Stage-specific embryonic antigen 4
STAR	Spliced Transcripts Alignment to a Reference
STR	Short tandem repeat
TBE	Tris-Borate-EDTA
TBS-T	Tris-buffered saline with Tween 20

THAP1	THAP domain containing 1
TOPO-TA	Topoisomerase-linked vector
TOR1A	Torsin family 1 member A
TPM	Transcripts per kilobase million
TUJ1	Class III β-tubulin
UNC5C	Unc-5 netrin receptor C
VAC14	VAC14 component of PIKFYVE complex
vGLUT1	Vesicular glutamate transporter 1
WD40	WD40 domain-containing 40
WDR44	WD repeat domain 44
WT	Wild-type
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta

Units

µ	Micro
Å	Angstrom
bp	Base pair
g	Gram
h	Hour
kb	Kilo base pairs
kDa	Kilo Dalton
l	Liter
M	Molar mass
min	Minute
ml	Millilitre
s	Second
U	Unit
x g	Centrifugal force

Amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S

Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

9.2 Supplementary material

Table S1 Standardized search terms used for the systematic literature search on PubMed (www.ncbi.nlm.nih.gov/pubmed).

Gene	Search Term
DYT-TOR1A	dystonia* AND (TOR1A OR DQ2 OR DYT1 OR 9q34.11) AND (gene* OR genetic* OR mutation* OR mutated OR varia*) AND "english"[Language] AND ("2018/10/01"[Date - Publication] : "3000"[Date - Publication])
DYT-THAP1	dystonia* AND (THAP1 OR DYT6 OR 8p11.21) AND (gene* OR genetic* OR mutation* OR mutated OR varia*) AND "english"[Language] AND ("2018/10/01"[Date - Publication] : "3000"[Date - Publication])
DYT-ANO3	dystonia* AND (ANO3 OR TMEM16C OR C11ORF25 OR anoctamin OR 11p14.3-p14.2) AND (gene* OR genetic* OR mutation* OR mutated OR varia*) AND "english"[Language] AND ("2018/10/01"[Date - Publication] : "3000"[Date - Publication])
DYT-GNAL	dystoni* AND (GNAL OR Galphaolf OR DYT25 OR 18p11.21) AND (gene* OR genetic* OR mutation* OR mutated OR varia*) AND "english"[Language] AND ("2018/10/01"[Date - Publication] : "3000"[Date - Publication])
DYT/PARK-PRKRA	(dystoni*) AND (PRKRA OR PACT OR RAX OR interferon-inducible OR interferon-induced OR “PKR protein activator” OR 2q31 OR EIF2AK2 OR DYT16 OR HSD14) AND "english"[Language] AND ("2018/10/01"[Date - Publication] : "3000"[Date - Publication])
DYT-HPCA	(dystoni*) AND (HPCA OR DYT2 OR hippocalcin OR 1p35.1) AND (gene* OR genetic* OR mutation* OR mutated OR varia*) AND "english"[Language] AND ("2018/10/01"[Date - Publication] : "3000"[Date - Publication])

Table S2 Combined list of candidate DEGs by group comparison of affected (AI), unaffected and all THAP1 mutation carrier (AllTHAP1) vs controls (Ctr) in alphabetical order. Following parameters were applied using DESeq2: $\log_2$fc ≥ 0.8 and adj. p-value ≤ 0.1 as well as $\log_2$fc ≤ 0.8 and adj. p-value ≤ 0.05.

No.	Gene ID	$\log_2$fc	Adj. p-value	Ensemble gene ID	Group comparison	Full gene name
1	A2M	-0.97	0.054	ENSG00000175899	AI vs Ctr	Alpha-2-macroglobulin
2	AC133919.1	1.12	0.031	ENSG00000283182	AI vs Ctr	Novel trancript
3	ADCK1	0.53	0.005	ENSG00000063761	AI vs Ctr	AarF domain containing kinase 1
4	ANGPTL1	-0.96	0.075	ENSG00000116194	AI vs Ctr	Angiopoietin like 1
5	ATP4A	0.93	0.049	ENSG00000105675	AI vs Ctr	ATPase H+/K+ transporting subunit alpha
6	BAALC	-0.89	0.025	ENSG00000164929	AI vs Ctr	BAALC binder of MAP3K1 and KLF4
7	BBOX1	-0.92	0.086	ENSG00000129151	AI vs Ctr	Gamma-butyrobetaine hydroxylase 1
8	BRAT1	0.33	0.034	ENSG00000106009	AllTHAP1 vs Ctr	BRCA1 associated ATM activator 1
9	CAB39	-0.24	0.016	ENSG00000135932	AllTHAP1 vs Ctr	Calcium binding protein 39
10	CDH22	-0.87	0.065	ENSG00000149654	AI vs Ctr	Cadherin 22
11	CGGBP1	-0.27	0.037	ENSG00000163320	AI vs Ctr	CGG triplet repeat binding protein 1
12	CHORDC1	-0.34	0.044	ENSG00000110172	AllTHAP1 vs Ctr	Cysteine and histidine rich domain containing 1
13	CNTN5	-1.21	0.016	ENSG00000149972	AI vs Ctr	Contactin 5
14	CNTNAP3B	0.99	0.025	ENSG00000154529	AI vs Ctr	Contactin associated protein like 3B
15	COL3A1	-0.89	0.076	ENSG00000168542	AI vs Ctr	Collagen type III alpha 1 chain
16	COQ8B	0.35	0.038	ENSG00000123815	AI vs Ctr	Coenzyme Q8B
17	CTC-435M10.3	0.84	0.050	ENSG00000255730	AI vs Ctr	Novel protein
18	CYP26C1	-0.87	0.075	ENSG00000187553	AI vs Ctr	Cytochrome P450 family 26 subfamily C member 1
19	DCLK1	-0.57	0.040	ENSG00000133083	AI vs Ctr	Doublecortin like kinase 1

20	DCUN1D1	-0.24	0.044	ENSG00000043093	AllTHAP1 vs Ctr	Defective in cullin neddylation 1 domain containing 1
21	DHDH	0.85	0.077	ENSG00000104808	AI vs Ctr	Cihydrodiol dehydrogenase
22	DHRS12	0.42	0.049	ENSG00000102796	AI vs Ctr	Dehydrogenase/reductase 12
23	DNPEP	0.40	0.008	ENSG00000123992	AI vs Ctr	Aspartyl aminopeptidase
24	DUS2	0.32	0.045	ENSG00000167264	AI vs Ctr	Cihydrouridine synthase 2
25	ECM2	-0.90	0.093	ENSG00000106823	AI vs Ctr	Extracellular matrix protein 2
26	ERCC6L2	-0.23	0.044	ENSG00000182150	AllTHAP1 vs Ctr	ERCC excision repair 6 like 2
27	F8A3	-0.89	0.094	ENSG00000277150	AI vs Ctr	Coagulation factor VIII associated 3
28	FAM118B	-0.37	0.005	ENSG00000197798	AI vs Ctr	Family with sequence similarity 118 member B
29	FAM3D	0.87	0.022	ENSG00000198643	UI vs Ctr	Family with sequence similarity 3 member D
30	FAM50B	0.88	0.084	ENSG00000145945	AI vs Ctr	Family with sequence similarity 50 member B
31	FGF1	-0.95	0.080	ENSG00000113578	AI vs Ctr	Fibroblast growth factor 1
32	FIS1	-0.36	0.031	ENSG00000214253	AI vs Ctr	Fission, mitochondrial 1
33	FOXD1	-1.14	0.025	ENSG00000251493	AI vs Ctr	Forkhead box D1
34	FOXP2	-1.03	0.058	ENSG00000128573	AI vs Ctr	Forkhead box P2
35	FYTTD1	-0.29	0.012	ENSG00000122068	AllTHAP1 vs Ctr	Forty-two-three domain containing 1
36	GABRR1	-0.90	0.078	ENSG00000146276	AI vs Ctr	Gamma-aminobutyric acid type A receptor rho1 subunit
37	GAS8	1.38	0.000	ENSG00000141013	AI vs Ctr	Growth arrest specific 8
38	GPRIN3	-0.80	0.044	ENSG00000185477	AllTHAP1 vs Ctr	GPRIN family member 3
39	GSTM4	0.48	0.044	ENSG00000168765	AllTHAP1 vs Ctr	Glutathione S-transferase mu 4
40	GTPBP3	0.34	0.034	ENSG00000130299	AllTHAP1 vs Ctr	GTP binding protein 3, mitochondrial
41	HAUS4	0.46	0.038	ENSG00000092036	AllTHAP1 vs Ctr	HAUS augmin like complex subunit 4
42	HELZ	-0.27	0.038	ENSG00000198265	AllTHAP1 vs Ctr	Helicase with zinc finger
43	IL11RA	0.77	0.049	ENSG00000137070	AI vs Ctr	Interleukin 11 receptor subunit alpha
44	INHBC	0.83	0.085	ENSG00000175189	AI vs Ctr	Inhibin subunit beta C
45	IRF7	0.43	0.037	ENSG00000185507	AI vs Ctr	Interferon regulatory factor 7
46	ISG15	0.71	0.001	ENSG00000187608	AI vs Ctr	ISG15 ubiquitin like modifier
47	KIF12	0.92	0.085	ENSG00000136883	AI vs Ctr	Kinesin family member 12
48	KIF26B	-0.88	0.068	ENSG00000162849	AI vs Ctr	Kinesin family member 26B
49	LMBR1	-0.23	0.046	ENSG00000105983	AI vs Ctr	Limb development membrane protein 1
50	LYPD5	1.05	0.040	ENSG00000159871	AI vs Ctr	LY6/PLAUR domain containing 5
51	MCHR1	-0.89	0.092	ENSG00000128285	AI vs Ctr	Melanin concentrating hormone receptor 1
52	ME3	0.66	0.044	ENSG00000151376	AllTHAP1 vs Ctr	Malic enzyme 3
53	MMD2	1.03	0.025	ENSG00000136297	AI vs Ctr	Monocyte to macrophage differentiation associated 2
54	MT1F	-1.06	0.045	ENSG00000198417	AI vs Ctr	Metallothionein 1F
55	MTDH	-0.27	0.017	ENSG00000147649	AI vs Ctr	Metadherin
56	MX1	1.05	0.001	ENSG00000157601	AI vs Ctr	MX dynamin like GTPase 1
57	MX2	1.20	0.016	ENSG00000183486	AI vs Ctr	MX dynamin like GTPase 2
58	MYO15B	0.89	0.092	ENSG00000266714	AI vs Ctr	Myosin XVB
59	MYOM2	0.89	0.081	ENSG00000036448	AI vs Ctr	Myomesin 2
60	NEFM	-0.93	0.075	ENSG00000104722	AI vs Ctr	Neurofilament medium
61	NELL1	-0.72	0.038	ENSG00000165973	AllTHAP1 vs Ctr	Neural EGFL like 1
62	NEU4	1.21	0.001	ENSG00000204099	AI vs Ctr	Neuraminidase 4
63	NPAS4	-0.89	0.079	ENSG00000174576	AI vs Ctr	Neuronal PAS domain protein 4

64	NPIPB9	0.88	0.077	ENSG00000196993	AI vs Ctr	Nuclear pore complex interacting protein family member B9
65	NRG3	-0.50	0.046	ENSG00000185737	AI vs Ctr	Neuregulin 3
66	NTM	-0.79	0.046	ENSG00000182667	AI vs Ctr	Neurotrimin
67	NWD2	-1.07	0.041	ENSG00000174145	AI vs Ctr	NACHT and WD repeat domain containing 2
68	NXPH1	-1.07	0.041	ENSG00000122584	AI vs Ctr	Neurexophilin 1
69	OLFML1	-0.97	0.075	ENSG00000183801	AI vs Ctr	Olfactomedin like 1
70	PAN2	0.36	0.031	ENSG00000135473	AllTHAP1 vs Ctr	Poly(A) specific ribonuclease subunit PAN2
71	PARP3	0.55	0.041	ENSG00000041880	AI vs Ctr	Poly(ADP-ribose) polymerase family member 3
72	PCDH19	-0.90	0.054	ENSG00000165194	AI vs Ctr	Protocadherin 19
73	PCED1A	0.59	0.000	ENSG00000132635	AI vs Ctr	PC-esterase domain containing 1A
74	PCSK9	0.93	0.083	ENSG00000169174	AI vs Ctr	Proprotein convertase subtilisin/kexin type 9
75	PCYT1B	-0.83	0.097	ENSG00000102230	AI vs Ctr	Phosphate cytidylyltransferase 1, choline, beta
76	PDE1C	-0.81	0.084	ENSG00000154678	AI vs Ctr	Phosphodiesterase 1C
77	PEX16	0.33	0.038	ENSG00000121680	AI vs Ctr	Peroxisomal biogenesis factor 16
78	PFDN4	-0.37	0.025	ENSG00000101132	AI vs Ctr	Prefoldin subunit 4
79	PI15	-0.91	0.083	ENSG00000137558	AI vs Ctr	Peptidase inhibitor 15
80	PKHD1	-0.89	0.100	ENSG00000170927	AI vs Ctr	PKHD1 ciliary IPT domain containing fibrocystin/polyductin
81	PKP3	1.98	0.000	ENSG00000184363	AI vs Ctr	Plakophilin 3
82	PLCXD3	-0.86	0.058	ENSG00000182836	AI vs Ctr	Phosphatidylinositol specific phospholipase C X domain containing 3
83	PPIE	0.26	0.005	ENSG00000084072	AllTHAP1 vs Ctr	Peptidylprolyl isomerase E
84	PPOX	0.31	0.034	ENSG00000143224	AllTHAP1 vs Ctr	Protoporphyrinogen oxidase
85	PRDM13	-0.87	0.085	ENSG00000112238		PR/SET domain 13
86	PROX1	-0.77	0.044	ENSG00000117707	AllTHAP1 vs Ctr	Prospero homeobox 1
87	PSMD12	-0.31	0.046	ENSG00000197170	AI vs Ctr	Proteasome 26S subunit, non-ATPase 12
88	PTGES	0.90	0.000	ENSG00000148344	AI vs Ctr	Prostaglandin E synthase
89	PYROXD2	0.92	0.005	ENSG00000119943	AI vs Ctr	Pyridine nucleotide-disulphide oxidoreductase domain 2
90	RBAK	-0.22	0.001	ENSG00000146587	AllTHAP1 vs Ctr	RB associated KRAB zinc finger
91	RFX5	0.25	0.034	ENSG00000143390	AllTHAP1 vs Ctr	Regulatory factor X5
92	RLN2	0.85	0.054	ENSG00000107014	AI vs Ctr	Relaxin 2
93	RP11-1212A22.4	0.95	0.041	ENSG00000233024	AI vs Ctr	Nuclear pore complex interacting protein family, member A9
94	RP11-872D17.8	0.93	0.085	ENSG00000254979	AI vs Ctr	Novel protein
95	RP11-903H12.5	0.92	0.078	ENSG00000259171	AI vs Ctr	Novel protein, ANG-RNASE4 readthrough
96	RRAGC	-0.28	0.041	ENSG00000116954	AI vs Ctr	Ras related GTP binding C
97	SCNN1D	0.85	0.077	ENSG00000162572	AI vs Ctr	Rodium channel epithelial 1 delta subunit
98	SDR39U1	0.32	0.031	ENSG00000100445	AI vs Ctr	Short chain dehydrogenase/reductase family 39U member 1
99	SETX	-0.41	0.045	ENSG00000107290	AI vs Ctr	Senataxin
100	SGCB	-0.33	0.041	ENSG00000163069	AllTHAP1 vs Ctr	Sarcoglycan beta
101	SLC25A18	-0.82	0.086	ENSG00000182902	AI vs Ctr	Solute carrier family 25 member 18
102	SLC30A9	-0.23	0.038	ENSG00000014824	AI vs Ctr	Solute carrier family 30 member 9
103	SLC37A4	0.30	0.041	ENSG00000137700	AllTHAP1 vs Ctr	Solute carrier family 37 member 4
104	SLC38A5	0.84	0.092	ENSG00000017483	AI vs Ctr	Solute carrier family 38 member 5

105	SLC39A4	1.09	0.031	ENSG00000147804	AI vs Ctr	Solute carrier family 39 member 4
106	SLC43A1	0.90	0.088	ENSG00000149150	AI vs Ctr	Solute carrier family 43 member 1
107	SMIM15	-0.25	0.005	ENSG00000188725	AI vs Ctr	Small integral membrane protein 15
108	SYTL5	-1.02	0.058	ENSG00000147041	AI vs Ctr	Synaptotagmin like 5
109	TAB1	0.31	0.044	ENSG00000100324	AllTHAP1 vs Ctr	TGF-beta activated kinase 1 (MAP3K7) binding protein 1
110	TAOK1	-0.26	0.031	ENSG00000160551	AllTHAP1 vs Ctr	TAO kinase 1
111	TAP2	0.95	0.004	ENSG00000204267	AI vs Ctr	Transporter 2, ATP binding cassette subfamily B member
112	TAZ	0.52	0.016	ENSG00000102125	AI vs Ctr	Tafazzin
113	TBC1D26	0.95	0.071	ENSG00000255104	AI vs Ctr	Novel transcript
114	TBX1	-0.77	0.041	ENSG00000184058	AllTHAP1 vs Ctr	T-box transcription factor 1
115	TEF	0.48	0.005	ENSG00000167074	AI vs Ctr	TEF transcription factor, PAR bZIP family member
116	THBS1	-0.92	0.083	ENSG00000137801	AI vs Ctr	Thrombospondin 1
117	TMCO6	0.45	0.031	ENSG00000113119	AI vs Ctr	Transmembrane and coiled-coil domains 6
118	TMEM127	-0.35	0.033	ENSG00000135956	AI vs Ctr	Transmembrane protein 127
119	TMEM184C	-0.21	0.044	ENSG00000164168	AllTHAP1 vs Ctr	Transmembrane protein 184C
120	TMEM79	0.49	0.044	ENSG00000163472	AllTHAP1 vs Ctr	Transmembrane protein 79
121	TOP1	-0.37	0.012	ENSG00000198900	AI vs Ctr	DNA topoisomerase I
122	TRAPPC6A	0.60	0.025	ENSG00000007255	AI vs Ctr	Trafficking protein particle complex 6A
123	UBE2F-SCLY	0.92	0.091	ENSG00000258984	AI vs Ctr	UBE2F-SCLY readthrough (NMD candidate)
124	UBE2V1	-0.36	0.021	ENSG00000244687	AI vs Ctr	Ubiquitin conjugating enzyme E2 V1
125	UNC5C	-0.91	0.092	ENSG00000182168	AI vs Ctr	Unc-5 netrin receptor C
126	USP18	0.82	0.075	ENSG00000184979	AI vs Ctr	Ubiquitin specific peptidase 18
127	USP21	0.29	0.005	ENSG00000143258	AllTHAP1 vs Ctr	Ubiquitin specific peptidase 21
128	WDFY3	-0.29	0.044	ENSG00000163625	AllTHAP1 vs Ctr	WD repeat and FYVE domain containing 3
129	WDR44	-0.49	0.044	ENSG00000131725	AllTHAP1 vs Ctr	WD repeat domain 44
130	XKR5	0.68	0.044	ENSG00000275591	AllTHAP1 vs Ctr	XK related 5
131	XKR6	-0.39	0.044	ENSG00000171044	AllTHAP1 vs Ctr	XK related 6
132	ZBTB33	-0.34	0.004	ENSG00000177485	AI vs Ctr	Zinc finger and BTB domain containing 33
133	ZMYND15	1.08	0.031	ENSG00000141497	AI vs Ctr	Zinc finger MYND-type containing 15
134	ZNF208	0.83	0.071	ENSG00000160321	AI vs Ctr	Zinc finger protein 208
135	ZNF260	-0.26	0.030	ENSG00000254004	AllTHAP1 vs Ctr	Zinc finger protein 260
136	ZNF662	1.36	0.003	ENSG00000182983	AI vs Ctr	Zinc finger protein 662
137	ZP3	1.02	0.058	ENSG00000188372	AI vs Ctr	Zona pellucida glycoprotein 3

Table S3 Biomaterial resource of all harvested cell pellets and microscope slides at different time points during cortical differentiations.

Identifier	iPSC passage at neural induction	No. pellets remaining after extraction (day harvested)	No. pellets remaining (day harvested)	No. pellets remaining (day harvested)	Microscope slides, 4% (w/v) formaldehyde-fixed on day
C16-3969-2	P16	3 (46)	3 (76)	-	90
C13-4011-1	P25	3 (44)	3 (78)	3 (90)	88
C17-4011-2	P19	3 (44)	3 (77)	-	86
C15-3735-5	P10	3 (43)	3 (76)	3 (89)	89
C12-3735-21	P14	3 (44)	3 (77)	3 (90)	90
C17-3736-13	P9	3 (44)	3 (77)	3 (79)	86
C15-3737-1	P16	3 (43)	3 (76)	3 (89)	89
C17-3737-3	P16	3 (44)	3 (77)	-	86
C7-10249-4	P26	3 (46)	-	2 (92)	92
C8-10249-5	P13	3 (45)	1 (76)	2 (92)	91
C7-10250-2	P15	3 (46)	1 (79)	2 (92)	92
C11-10250-10	P13	3 (44)	3 (76)	3 (90)	90
C11-10252-6	P14	3 (44)	3 (76)	3 (90)	90
C12-10252-8	P15	3 (44)	3 (77)	3 (90)	90
C17-10253-1	P12	3 (44)	3 (77)	-	86
C12-10253-11	P17	3 (44)	3 (77)	3 (90)	90
C10-10260-1	P17	3 (44)	3 (75)	3 (91)	91
C16-10260-12	P14	3 (46)	3 (76)	-	90
C14-062030203A	P42	3 (43)	3 (77)	-	90
C10-065030301A	P31	3 (44)	3 (75)	3 (91)	91
C10-065030508A	Px+2	3 (44)	3 (75)	3 (91)	91
C18-065030608A	Px+7	5 (45)	-	-	-
C8-0670301	P24	3 (45)	1 (76)	2 (92)	91
C7-0680301	Px+8	3 (46)	-	2 (92)	92
C9-084030201A	P49	3 (45)	-	-	93
C8-086030101	P46	3 (45)	1 (76)	2 (92)	91
C9-086030302A	P39	3 (45)	2 (75)	3 (93)	93
C18-08903030704A	Px+6	3 (45)	-	-	-
C8-156030101A	Px+10	3 (45)	1 (76)	-	91
C14-163030108A	Px+8	3 (43)	3 (77)	3 (90)	90
C13-841030107A	Px+5	3 (44)	-	3 (90)	88
C13-856030704	P25	3 (44)	3 (78)	3 (90)	88

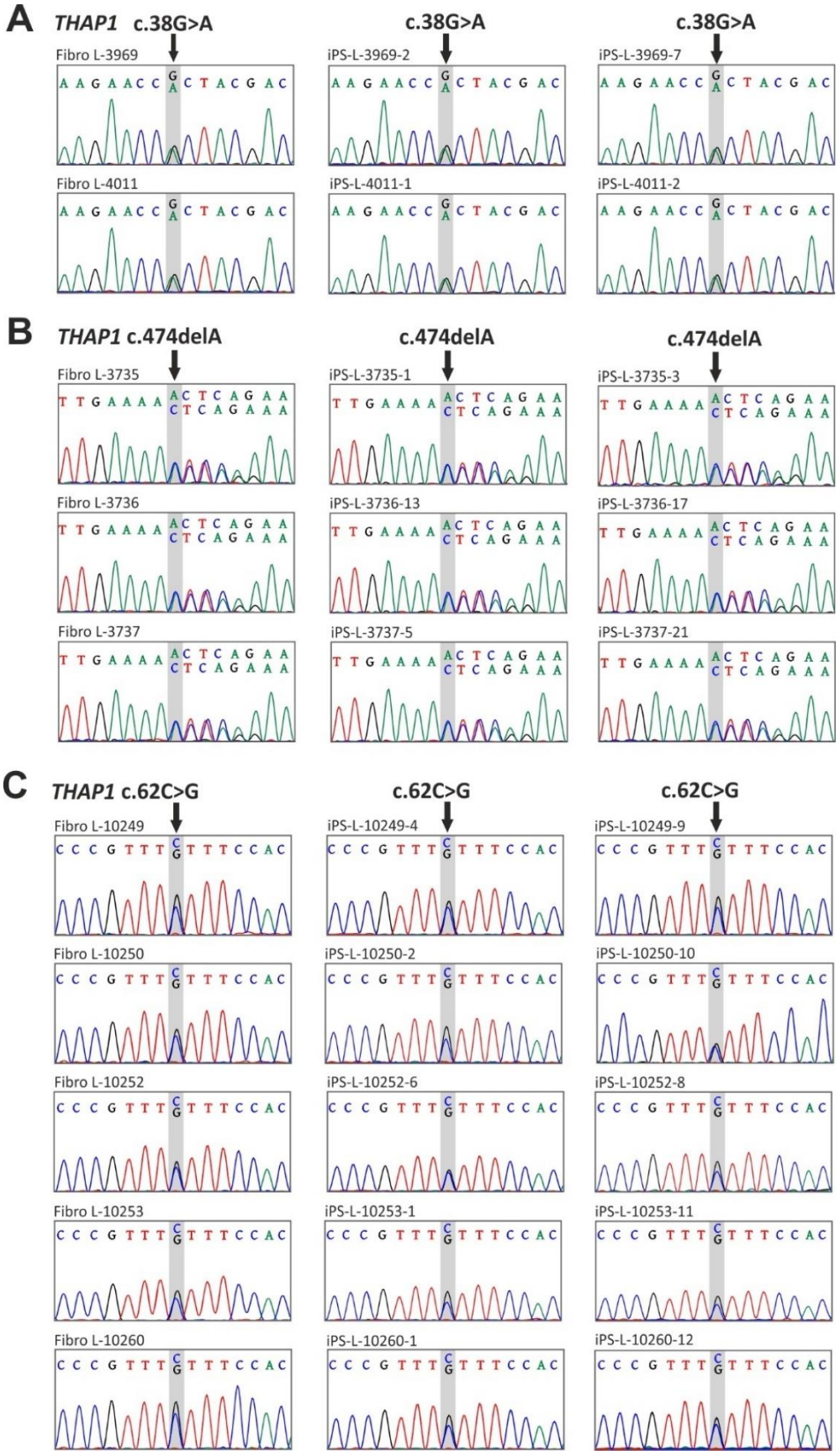

Figure S1 Mutation analysis of two iPSC clones and corresponding fibroblasts of *THAP1* mutation carriers by Sanger sequencing. The grey background indicates the pathogenic *THAP1* (NM_018105.3) variants (A) c.38G>A (p.Arg13His), (B) c.474delA (p.Lys158Asnfs*23), and (C) c.62C>G (p.Ser21Cys). Samples are labelled with an L-identifier.

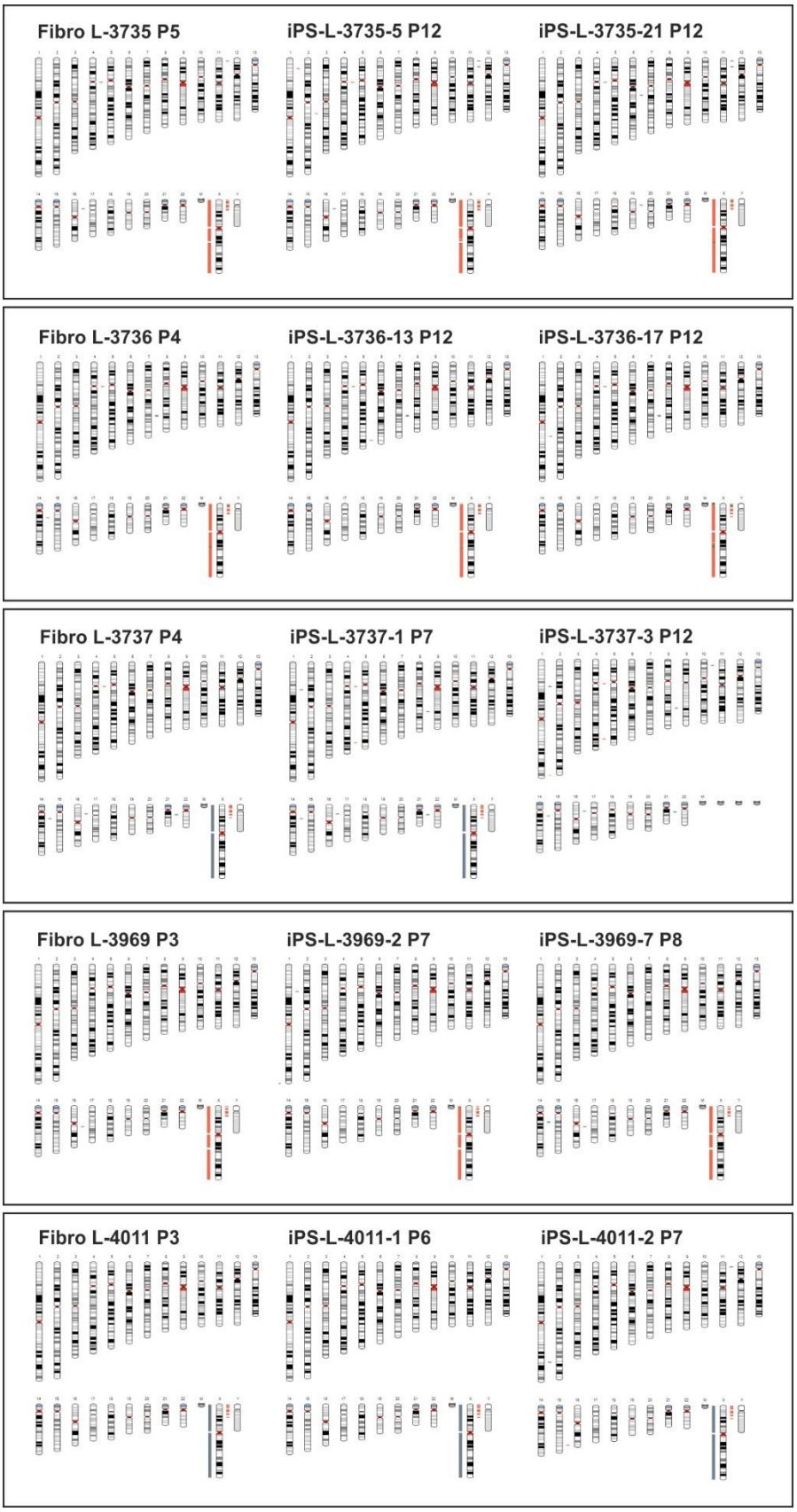

Figure S2 continued

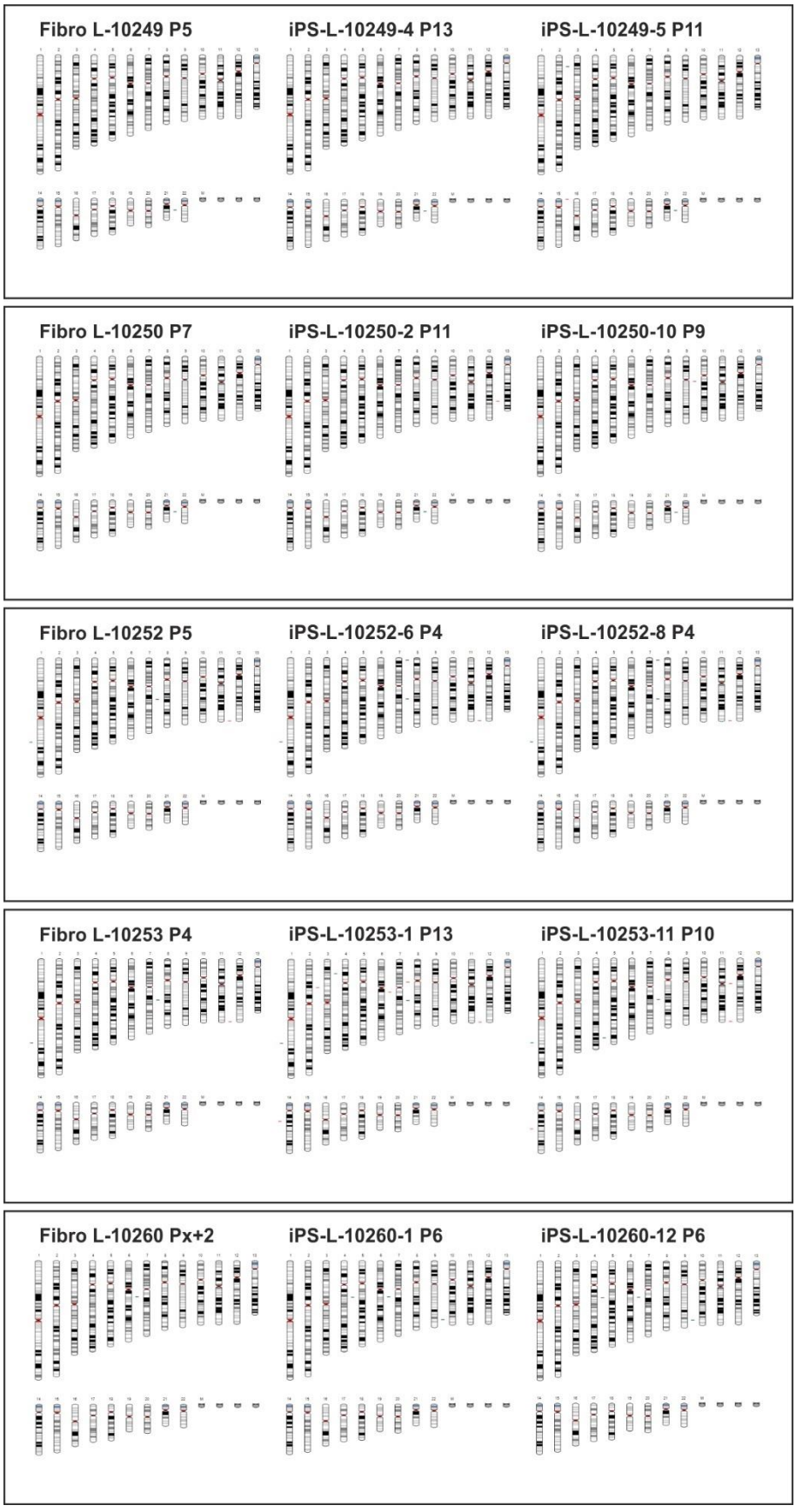

Figure S2 Karyotype analysis by SNP genotyping of parental fibroblasts and two iPSC lines using the OmniExpress-24 BeadChip (~710,000 marker) of the indicated cell culture passage. No large chromosomal aberrations or copy number variants were detected. KaryoStudio software incompatibilities with the raw data led to a lack of the allosomes in some cases. Manual screening confirmed no abnormalities within these regions.

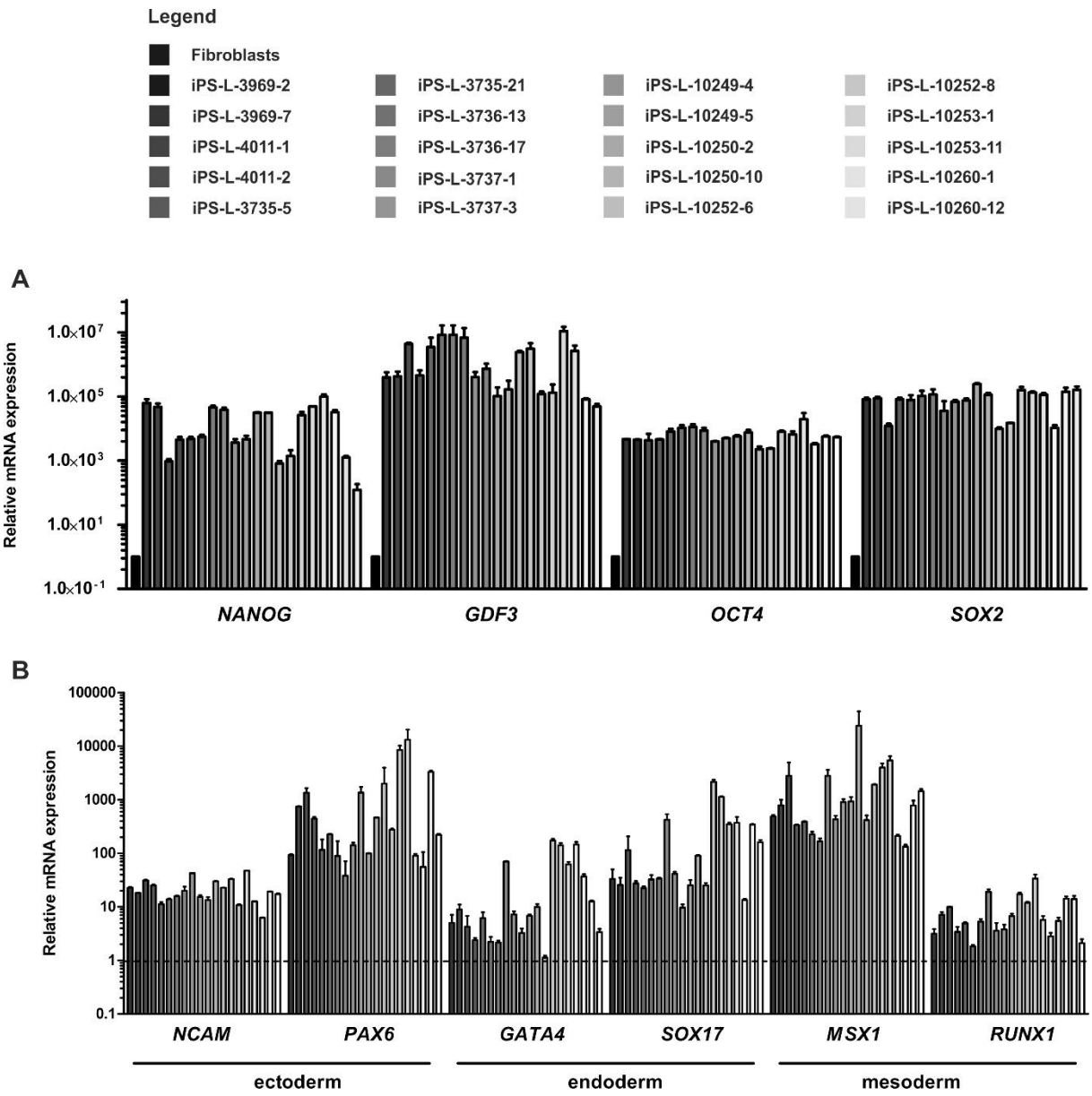

Figure S3 Confirmation of pluripotency of twenty generated iPSC lines by qRT-PCR. (A) Relative mRNA levels of endogenous pluripotency markers *NANOG*, *GDF3*, *OCT4,* and *SOX2* compared to control fibroblasts normalized to *ACTB*. (B) Relative mRNA expression of specific germ layer markers (*NCAM*, *PAX6*, *GATA4*, *SOX17*, *MSX1*, and *RUNX1*) upon spontaneous differentiation into embryoid bodies compared to the respective iPSC line and normalized to *ACTB*.

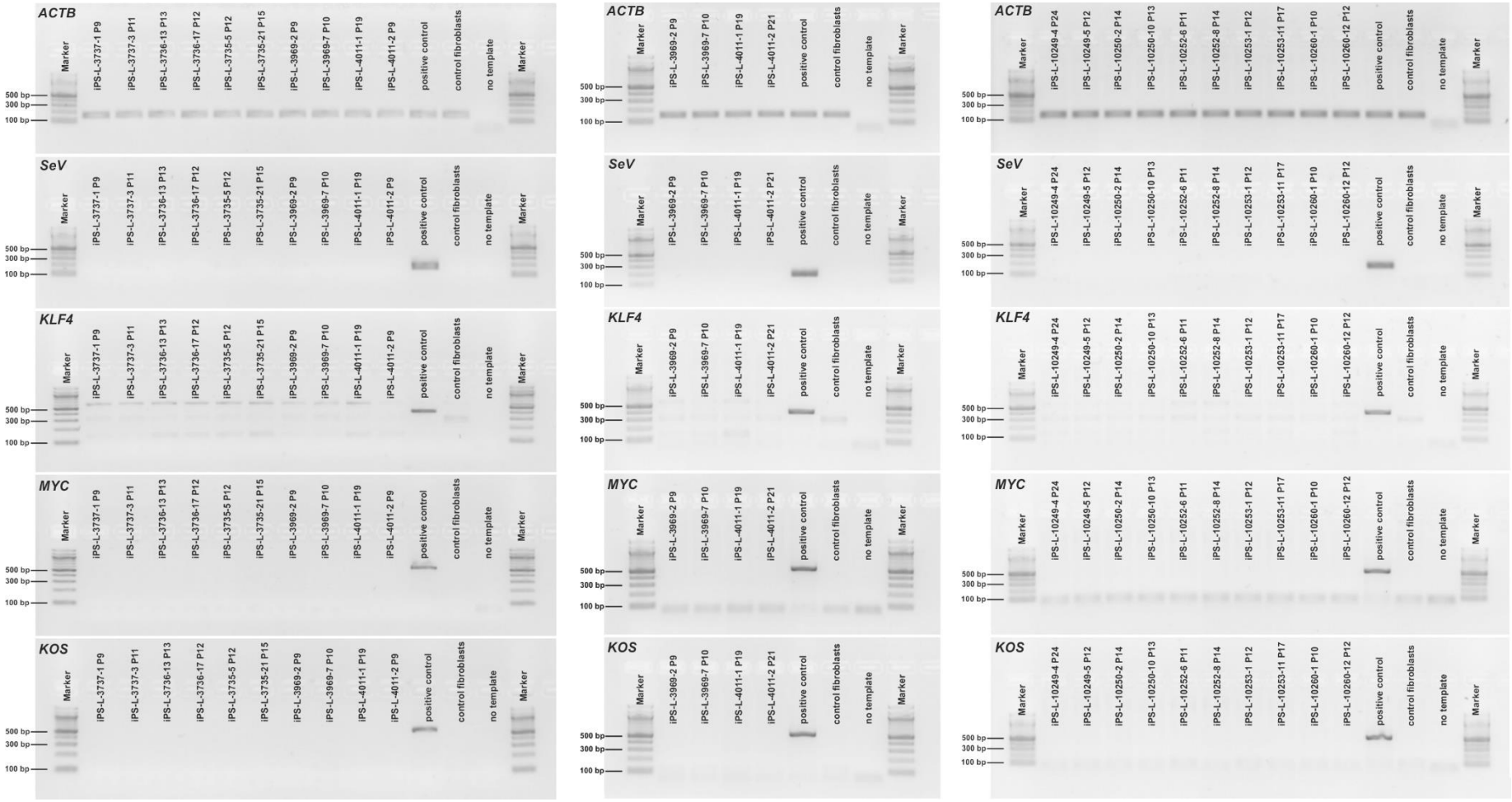

Figure S4 Clearance of Sendai virus components and in iPSC cultures at the indicated passage number. Sendai transgene-specific PCR products were separated by size on an agarose gel. Positive control: fibroblasts after seven days post Sendai virus infection. Primer sequences are listed in Section 4.2.7. *ACTB* = ß-ACTIN (1551 bp), SeV = Sendai virus (181 bp), *KLF4* = Kruppel-like factor 4 (410 bp), *MYC* = MYC Proto-Oncogene (532 bp), *KOS* = Acronym for the genes *hKLF4*, *hOCT*, and *hSOX2* (528 bp).

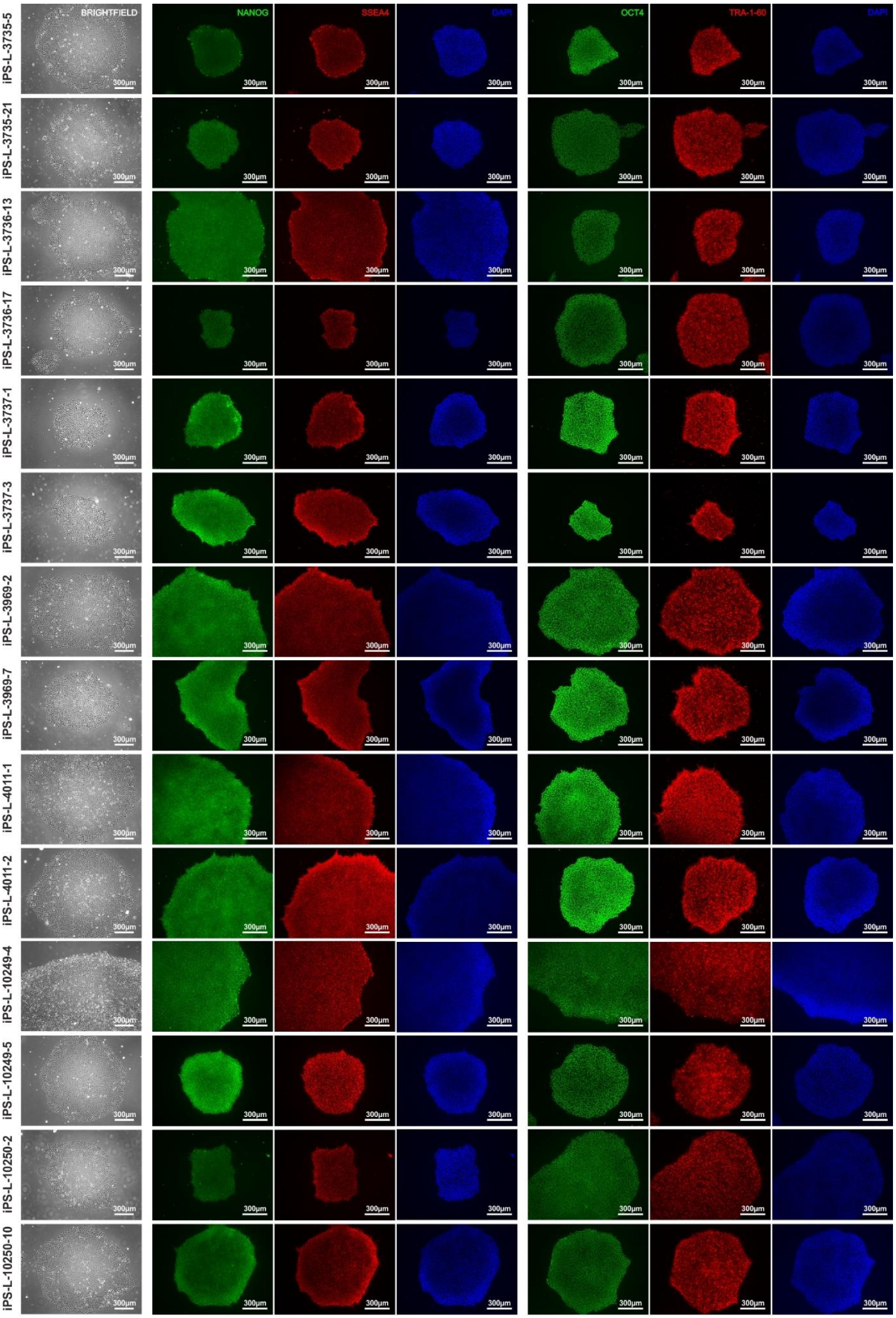

Figure S5 continued

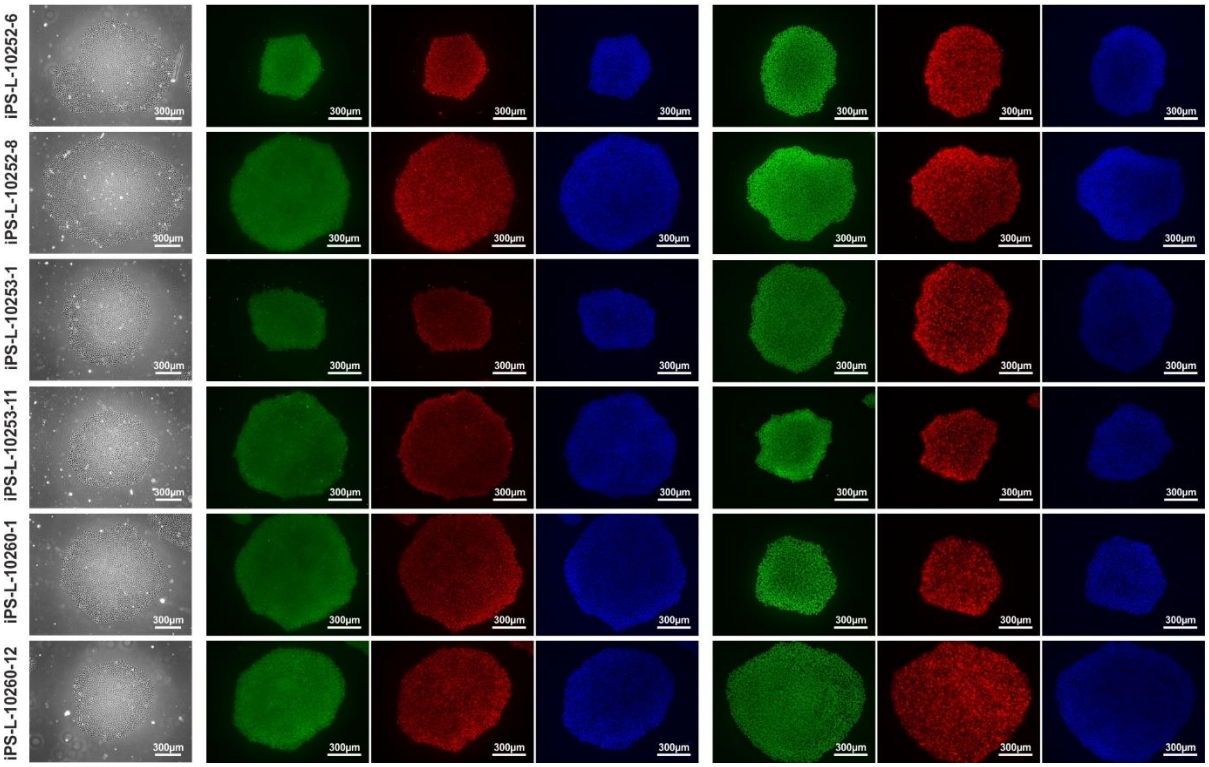

Figure S5 Immunofluorescence staining and brightfield micrographs of iPSC colonies of *THAP1* mutation carriers. Immunostaining shows the presence of pluripotency markers NANOG, SSEA4, OCT4, and TRA-1-60 (for antibody concentrations see section 4.2.5).

Table S4 STR analysis of four polymorphic markers in genomic DNA extracted from iPSC-derived neurons (day 44) and blood. In cases where blood was not available, iPSC derived neurons were compared to the respective iPSC line instead.

Locus / Identifier	D1S2663	D6S305	D14S306	D20S888	DYS390
CEPH1*	201-193	230-222	206-198	159-155	221-221
CEPH2*	203-183	228-228	194-194	163-147	0-0
Blood L-3735	197-193	228-204	206-202	155-153	213-213
C15-3735-5	197-193	228-204	206-202	155-153	213-213
C12-3735-21	197-193	228-204	206-202	155-153	213-213
Blood L-3736	205-193	228-204	206-198	155-153	213-213
C17-3736-13	205-193	228-228	206-198	155-153	213-213
Blood L-3737	207-193	228-204	206-202	159-155	0-0
C15-3737-1	207-193	228-204	206-202	159-155	0-0
C17-3737-3	207-193	228-204	206-202	159-155	0-0
Blood L-3969	201-199	226-226	198-194	163-163	221-221
C16-3969-2	201-199	226-226	198-194	163-163	221-221
Blood L-4011	199-195	230-226	198-194	163-159	0-0
C13-4011-1	199-195	230-226	198-194	163-159	0-0
C17-4011-2	199-195	230-226	198-194	163-159	0-0
Blood L-4993	191-191	218-204	206-206	161-147	0-0
C9-084030201A	191-191	218-204	206-206	161-147	0-0
Blood L-5872	201-201	228-218	210-198	161-155	213-213
C8-0670301	201-201	228-218	210-198	161-155	213-213
Blood L-5991	197-193	208-204	206-206	161-149	0-0
C14-062030203A	197-193	208-204	206-206	161-149	0-0
Blood L-6004	209-199	228-226	210-198	159-149	0-0
C18-08903030704A	209-199	228-226	210-198	159-149	0-0
Blood L-6069	191-183	222-218	198-194	159-147	221-221
C8-156030101A	191-183	222-218	198-194	159-147	221-221
Blood L-6353	197-193	230-204	210-198	155-141	0-0
C8-086030101	197-193	230-204	210-198	155-141	0-0
Blood L-6396	191-191	222-204	206-202	147-145	0-0
C7-0680301	191-191	222-204	206-202	147-145	0-0
C9-086030302A	191-191	222-204	206-202	147-145	0-0
Blood L-6534	197-191	218-218	210-202	147-145	217-217
C10-065030508A	197-191	218-218	210-202	147-145	217-217
C10-065030301A	197-191	218-218	210-202	147-145	217-217
C18-065030608A	197-191	218-218	210-202	147-145	217-217
Blood L-4646	203-199	224-204	202-202	163-159	213-213
C14-163030108A	203-199	224-204	202-202	163-159	213-213
iPS-841	193-193	228-218	206-178	159-155	213-213
C13-841030107A	193-193	228-218	206-178	159-155	213-213
iPS-856	199-195	232-216	202-194	167-159	0-0
C13-856030704	199-195	232-216	202-194	167-159	0-0

*Characterized control individuals recruited by the Centre d'Etude du Polymorphisme Humain (CEPH)
The underlined marker did not match between blood and iPSC-derived neurons (D6S305). A sample mix-up in L-3736 was excluded by further testing of six polymorphic markers (D1S2845, D1S2870, D21S1909, D3S3647, D5S2090, and D19S878) that showed matching results between DNA from blood and neuronal cultures.

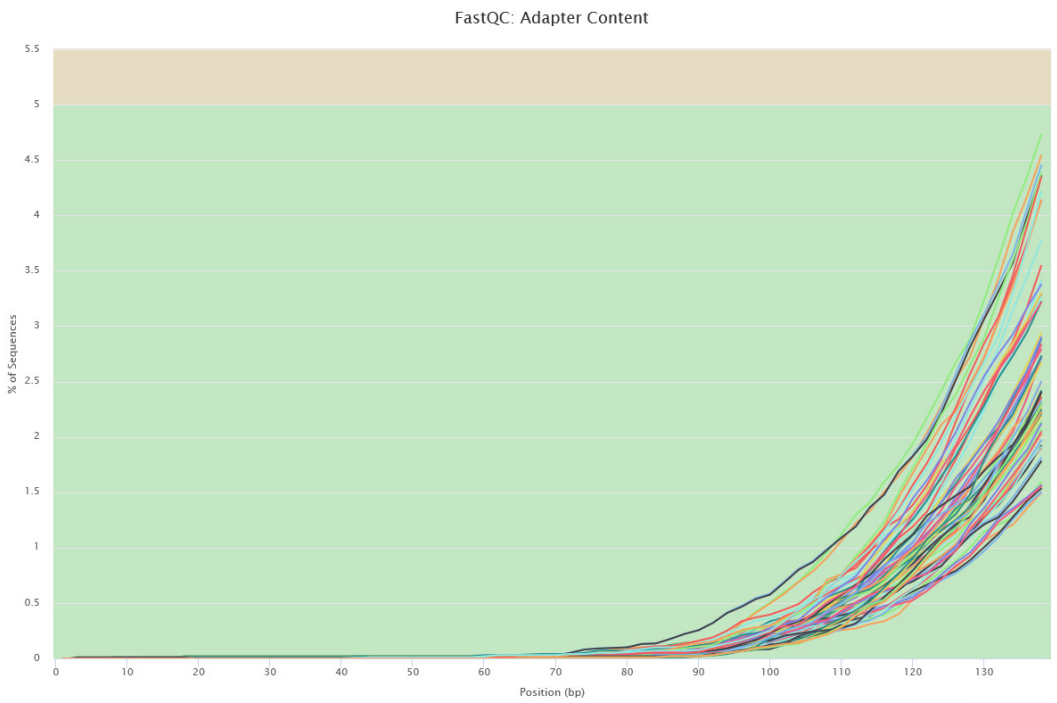

Figure S6 Adapter content after trimming with the package Skewer. The coloured lines show the percentage of adapter sequences in 32 iPSC-derived cortical neurons. The green background marks the accepted proportion of adapter sequences (<5%). The plot was generated by the tool Multi-QC.

Table S5 Summary of RNA quality metrics including RNA integrity number and absorbance ratios, quality control values by Multi-QC (raw reads and GC-content), and mapping statistics by the STAR aligner of all samples used for RNA-Seq.

Mutational status	Identifier	RNA integrity number (RIN)	RNA conc. [ng/µl]	Absorbance 260/280nm	Absorbance 260/230nm	Total number of raw reads	GC-content [%]	Number of uniquely mapped reads (STAR)
THAP1 p.Arg13His	C16-3969-2	10.0	1223	2.04	2.20	92359838	49.19	84987039
	C13-4011-1	9.8	1375	2.03	2.22	78111220	51.10	69259851
	C17-4011-2	10.0	1297	2.05	2.14	82686956	50.64	75000002
THAP1 p.Lys158Asnfs*23	C15-3735-5	9.9	1176	2.04	2.19	85031076	49.80	77004717
	C12-3735-21	9.7	1243	2.03	2.20	73067783	51.41	65046518
	C17-3736-13	10.0	412	1.98	2.77	73020180	52.10	63963077
	C15-3737-1	9.9	877	2.04	2.05	71547420	49.09	65927632
	C17-3737-3	10.0	1241	2.04	2.34	71852421	49.69	64679217
THAP1 p.Ser21Cys	C7-10249-4	10.0	640	1.89	1.66	68787525	53.59	56941916
	C8-10249-5	9.9	1285	2.03	2.23	81516097	52.12	70441082
	C7-10250-2	10.0	433	1.97	2.15	79346956	51.34	69238537
	C11-10250-10	9.9	580	2.05	2.16	58627385	52.87	49652417
	C11-10252-6	10.0	674	2.05	2.17	82301796	54.30	68855248
	C12-10252-8	9.6	1366	2.03	2.12	86128027	52.14	76288913
	C17-10253-1	9.5	2228	1.97	2.29	75523685	50.80	67034462
	C12-10253-11	9.9	1519	2.03	2.20	81370331	54.61	66592040
	C10-10260-1	10.0	1430	2.10	2.08	94207571	52.38	80868582
	C16-10260-12	10.0	1282	2.03	2.21	79526827	48.92	73453720
Wild-type	C14-062030203A	9.9	408	1.98	2.04	75045202	50.93	65716320
	C10-065030301A	10.0	1201	2.04	2.20	71626016	51.91	62786381
	C10-065030508A	10.0	170	2.06	1.72	79693850	52.00	69438088
	C18-065030608A	10.0	1679	2.01	2.31	73940045	50.48	66038221
	C8-0670301	10.0	1146	2.05	2.13	67517398	50.96	58217533
	C7-0680301	10.0	424	1.99	1.47	70102790	52.52	59054336
	C9-084030201A	10.0	335	2.02	2.11	90218561	52.52	78334557
	C8-086030101	10.0	118	2.06	1.22	71579269	51.05	61264535
	C9-086030302A	10.0	296	2.04	1.66	77943641	52.60	66810100
	C18-08903030704A	10.0	410	1.99	2.77	71557360	50.13	64045930
	C8-156030101A	10.0	198	2.05	1.90	77212566	53.12	62990633
	C14-163030108A	9.9	1145	2.04	2.21	78822713	49.05	71406882
	C13-841030107A	9.8	1462	2.04	2.21	72884311	51.94	62238253
	C13-856030704	9.7	652	2.06	2.16	68693680	51.36	59268248
Mean [range]		9.9 [9.5 - 10]	935 [118 - 2228]	2.02 [1.89 - 2.10]	2.10 [1.22 - 2.77]	76.9 Mio [58.6-94.2]	51.5 [48.9 - 54.6]	67.2 Mio [49.7 - 85.0]

9.3 Publications

Original articles

1. **Baumann H**, Tunc S, Münchau A, Brüggemann N, Lohmann K Compound-heterozygous VAC14 variants in a patient with progressive dystonia. (*in preparation for Parkinsonism Relat Disord*)
2. **Baumann H**, Weber, J, Busch, H, Trilck-Winkler M, Münchau A, Kostic A, Klein C, Kaiser FJ, Seibler P, Lohmann K. Whole transcriptome analysis in human iPSC-derived neurons of affected and unaffected DYT-THAP1 patients. (*in preparation for Mov Disord)*
3. Lange L*, Junker J*, Löns S*, **Baumann H***, Olschewski L, Schaake S, Petkovic S, Kasten M, Westenberger A, Domingo A, Madoev H, König I, Marras C, Ozelius LJ, Camergos S, Klein C, Lohmann K. Genotype-Phenotype Relations for the isolated Dystonia Genes TOR1A, THAP1, GNAL, ANO3, KMT2B, PRKRA, and HPCA: MDSGene Systematic Review. (*in preparation for Mov Disord*)
4. Dulovic-Mahlow M, Trinh J, Kandaswamy KK, Braathen GJ, Di Donato N, Rahikkala E, Beblo S, Werber M, Krajka V, Busk ØL, **Baumann H**, Al-Sannaa NA, Hinrichs F, Affan R, Navot N, Al Balwi MA, Oprea G, Holla ØL, Weiss MER, Jamra RA, Kahlert AK, Kishore S, Tveten K, Vos M, Rolfs A, Lohmann K. De Novo Variants in TAOK1 Cause Neurodevelopmental Disorders. Am J Hum Genet. 2019 Jul 3;105(1):213-220. (IF: 9.924)
5. Dulovic-Mahlow M, Gajos A, **Baumann H**, Pozojevic J, Kaiser FJ, Bogucki A, Lohmann K. Highly reduced penetrance in a family with a THAP1 nonsense mutation: Role of THAP1 expression? *Parkinsonism Relat Disord.* 2019 May 25. pii: S1353-8020(19)30257-3. (IF: 4.360)
6. Tunc S, Dulovic-Mahlow M, **Baumann H**, Baaske MK, Jahn M, Junker J, Münchau A, Brüggemann N, Lohmann K. Spinocerebellar Ataxia Type 28-Phenotypic and Molecular Characterization of a Family with Heterozygous and Compound-Heterozygous Mutations in AFG3L2. *Cerebellum.* 2019 Aug;18(4):817-822. (IF: 3.413)
7. Klein C, **Baumann H**, Olschewski L, Hanssen H, Münchau A, Ferbert A, Brüggemann N, Lohmann K. De-novo KMT2B mutation in a consanguineous family: 15-Year follow-up of an Afghan dystonia patient. *Parkinsonism Relat Disord.* 2019 Mar 25. pii:S1353-8020(19)30111-7 (IF: 4.360)
8. Trinh J, Lohmann K, **Baumann H**, Balck A, Borsche M, Brüggemann N, Dure L, Dean M, Volkmann J, Tunc S, Prasuhn J, Pawlack H, Imhoff S, Lill CM, Kasten M, Bauer P, Rolfs A, International Parkinson's Disease Genomics Consortium (IPDGC), Klein C. Utility and implications of exome sequencing in early-onset Parkinson's disease. *Mov Disord.* 2019 Jan;34(1):133-137 (IF: 8.061)
9. **Baumann H**, Jahn M, Münchau A, Trilck-Winkler M, Lohmann K, Seibler P. Generation and characterization of eight human-derived iPSC lines from affected and unaffected THAP1 mutation carriers. *Stem Cell Res.* 2018 Dec;33:60-64 (IF: 3.929)
10. Lohmann K, Masuho I, Patil DN, **Baumann H**, Hebert E, Steinrücke S, Trujillano D, Skamangas NK, Dobricic V, Hüning I, Gillessen-Kaesbach G, Westenberger A, Savic-Pavicevic D, Münchau A, Oprea G, Klein C, Rolfs A, Martemyanov KA. Novel GNB1 mutations disrupt assembly and function of G protein heterotrimers and cause global developmental delay in humans. *Hum Mol Genet.* 2017 Mar 15;26(6):1078-1086. (IF: 4.902)
11. **Baumann H**, Wolff S, Münchau A, Hagenah JM, Lohmann K, Klein C. Evaluating the role of TMEM230 variants in Parkinson's disease. *Parkinsonism Relat Disord.* 2017 Feb;35:100-101 (IF: 4.721)

The published articles have a cumulative impact factor (IF) of 39.741 and an average of 4.968.

*These authors have contributed equally to this work.

Abstracts

1. **Baumann H**, Trilck-Winkler M, Grosse M, Münchau A, Kostic V, Klein C, Kaiser FJ, Seibler P, Lohmann K. Expression analysis of candidate genes in a cell model of affected and unaffected *THAP1* mutation carriers and controls. Movement Disorder Society Conference 2019 in Nice, France
2. **Baumann H**, Trilck M, Grosse M, Münchau A, Kostic A, Klein C, Kaiser FJ, Seibler P, Lohmann K. Gene expression analysis in cortical neurons differentiated from 32 induced pluripotent stem cell (iPSC) lines of *THAP1* mutation carriers and controls. European Society of Human Genetics Conference 2019 in Gothenburg, Sweden
3. **Baumann H**, Trilck M, Jahn M, Münchau A, Kostic V, Klein C, Seibler P, Lohmann K. Generation and in-depth characterization of 20 induced pluripotent stem cell (iPSC) lines from 10 dystonia patients and healthy carriers of *THAP1* mutations. European Society of Human Genetics Conference 2018 in Milan, Italy
4. **Baumann H**, Trilck M, Seibler P, Kaiser F, Lohmann K. Generation of patient iPSC-derived cortical neurons to identify genetic modifiers of reduced penetrance in DYT6 dystonia. Research Workshop Technologies and Resources 2017 in Lübeck, Germany
5. **Baumann H**, Masuho I, Dipak P, Steinrücke S, Hebert E, Dobricic V, Hüning I, Gillessen-Kaesbach G, Westenberger A, Dusanka S, Münchau A, Klein C, Rolfs, A, Martemyanov K, Lohmann K. Characterization of GNB1 mutations as a cause of global developmental delay in combination with dystonia, ataxia, or chorea in children. Movement Disorder Congress 2017 in Vancouver, Canada

Oral presentations

1. **Baumann H**, Trilck-Winkler M, Grosse M, Münchau A, Kostic V, Klein C, Kaiser FJ, Seibler P, Lohmann K. In-vitro-Genexpressionsanalyse in 32 humanen, aus iPS-Zellen differenzierten, kortikalen Neuronen von *THAP1*-Mutationsträgern und Kontrollen. German Society of Neurology Congress 2019 in Stuttgart, Germany
2. **Baumann H**, Rare genetic mutations – How to prove their functional relevance. Center for Rare Diseases Workshop 2017 in Plön, Germany
3. **Baumann H**, Lill C, Rolfs A, Brüggemann N, Klein C, Lohmann K. Exome sequencing in 55 patients with Parkinson's disease reveals mutations in *TMEM230*, *PLA2G6* and other genes. Human Genetics Conference 2017 in Lübeck, Germany

9.4 Acknowledgements

I am truly grateful to my supervisor Professor Katja Lohmann for her kind support, inspiring leadership, and thoughtful guidance throughout my thesis. Thank you for being ever available and supportive during this time.

I am grateful to Professor Klein for giving me the possibility to work in the field of neurogenetics and for this intriguing topic. It is a great pleasure working in such an interdisciplinary team that is full-heartedly dedicated to translational science.

I would like to express my sincere thanks to my second referee Professor Henrik Oster for kindly agreeing to act as a second reviewer of this work. Moreover, I would like to thank Professor Charli Kruse as the chairman of the examination committee.

For his continuous support during my two years in the iPS-lab I would like to extend my thanks to Professor Philip Seibler. For the supervision and technical support, I would like to thank Michaela, Insa, Franziska, and Beryll.

Thank you Professor Hauke Busch and Professor Frank Kaiser for the fruitful collaboration and valuable interdisciplinary input.

I thank all the colleagues from the Institute of Neurogenetics for the excellent and warm working atmosphere – my special thanks go to Frauke, Julia, Evelyn, Sarah, Victor, Charles, Marija, Kerstin, Lara, Leonoara, Joachim, Sofia, and Aloysius for keeping up this spirit every day. I would also like to extend my thanks to all present and former colleagues and collaborators I have not named explicitly.

I owe a debt of gratitude to my family for a life-long support and encouragement. Thank you Meike and Thalke for your unshakable belief in me and for maintaining your positive way of thinking at all times.

Frederike there aren't enough words to say thank you for everything. I am glad to share my life with such a wonderful, passionate and kind person. I am looking forward to every adventure that may await us.