2021

Implications of Genetic Variation on Patient Disease and Therapeutic Options

Ashlee Long
University of Alabama at Birmingham

IMPLICATIONS OF GENETIC VARIATION ON PATIENT DISEASE AND THERAPEUTIC OPTIONS

by

ASHLEE LONG

DEEANN WALLIS, CHAIR
MATTHEW ALEXANDER
PETER DETLOFF
SCOTT WILSON
TALENE YACOUBIAN

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2021

IMPLICATIONS OF GENETIC VARIATION ON PATIENT DISEASE AND THERAPEUTIC OPTIONS

ASHLEE LONG

BIOMEDICAL SCIENCES

ABSTRACT

Advancements in sequencing technology have enabled the widespread utilization of genetic testing, increasing the diagnosis and understanding of genetic disease. Clinicians and researchers are now faced with the difficulty of interpreting genetic results and understanding how genetic variation can impact patient disease phenotype. Herein we utilized three different genes (*frataxin*, *neurofibromin 1*, and *NTRK2*), which result in Friedreich's Ataxia (FRDA), Neurofibromatosis Type 1 (NF1), and NTRK2-associated phenotypes respectively, to discuss the impacts of genetic variants on molecular phenotype and explore potential therapeutics. To determine how genetic variation of repeat sequences affects different patient tissues, DNA and protein were extracted from FRDA patient samples and somatic mosaicism of pathogenic variants was analyzed through PCR and Western Blot analyses. We observed a tissue-specific pattern of somatic mosaicism and instability of pathogenically expanded trinucleotide repeat sequences with normal length trinucleotide repeats exhibiting no genetic variation among different tissues. The somatic mosaicism and instability observed represents a potential cause of symptom variability among patients with similar disease-causing variants. Additionally, to determine how different missense mutations impact protein activity we overexpressed NF1 cDNAs that were created with patient-specific missense variants in HEK293 cell-lines to assess protein expression and function. We observed unique effects

on protein stability and function with some missense variants resulting in functional protein at a reduced protein expression. Other variants exhibited high levels of protein but were incapable of proper function. This highlights the importance of understanding how genotype can contribute to phenotype in disorders which are caused by thousands of unique variants in one gene. Additionally, we demonstrated that specific variants within *NTRK2* can result in drastically different functional consequences associated with distinct phenotypes. We utilized an artificial intelligence based software, mediKanren, to predict potential therapeutics to treat the differing functional consequences and were able to partially rescue variant molecular phenotype. This emphasizes the need to determine how different variants within the same gene can result in separate disorders and whether this impacts patient treatment. Overall, these studies highlight the importance of understanding how intragenic mutation heterogeneity can impact patient presentation and that the molecular mechanisms behind these variants can be targeted for personalized treatments through artificial intelligence-based software in precision medicine.

Keywords: genetics, genetic disorders, precision medicine, Friedreich's Ataxia, Neurofibromatosis Type 1, NTRK2, precision medicine

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my mentor Dr. Deeann Wallis for her support, patience, and knowledge. Thank you for allowing me the freedom to pursue different research avenues while providing direction when met with obstacles. Additionally, I would like to thank my committee, Dr. Matthew Alexander, Dr. Peter Detloff, Dr. Scott Wilson, and Dr. Talene Yacoubian for their efforts during my PhD research and their advice offered at each step. The work in this dissertation would not have been possible without the support from the people listed above. Additionally, I'd like to thank the entire UAB community and specifically the CMDB program. The guidance and encouragement I received throughout my PhD career truly helped me navigate and overcome any difficulties. I wish to also thank my family, especially my mom who has been a constant support throughout my life and who has made it possible for me to be where I am today. Finally, I am grateful and appreciative to the patient communities who have graciously donated samples, raised research funding, and are ultimately most impacted by the disorders I have researched.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table *Page*

LIST OF FIGURES

Figures *Page*

SOMATIC INSTABILITY OF THE EXPANDED GAA REPEATS IN FRIEDREICH'S ATAXIA

LIST OF FIGURES (CONTINUED)

Figures *Page*

ANALYSIS OF PATIENT-SPECIFIC *NF1* VARIANTS LEADS TO FUNCTIONAL INSIGHTS FOR RAS SIGNALING THAT CAN IMPACT PERSONALIZED MEDICINE

FUNCTIONAL CHARACTERIZATION AND POTENTIAL THERAPEUTIC AVENUES FOR NOVEL VARIANTS IN THE *NTRK2* GENE

LIST OF FIGURES (CONTINUED)

Figures *Page*

INTRODUCTION

With the sequencing of the human genome, scientists have made great strides in diagnosing and treating rare genetic disorders. As technological advances have increased, our ability to diagnose these disorders is becoming more prevalent in the clinic [1, 2]. Notably, Next Generation Sequencing (NGS) based pre- and postnatal diagnostics have become more widely used. Due to the presence of fetal cell-free DNA (cfDNA) in maternal blood, monogenic disorders can be diagnosed using non-invasive techniques [3]. Additionally, genetic testing has resulted in increased diagnostic yield for adult-onset genetic disorders, specifically those with cardiac disease and cancer predisposition syndromes [4, 5]. This clinical utility for diagnosing both infants and adults has made genetic testing a more frequently utilized tool in medicine.

Physicians also recognize the utility of genetic testing in the clinic, however there are still several problems that need to be addressed. Lerner et al. assessed what clinicians most value before performing genetic testing on a patient and found that the ability to inform clinical management had the highest rating (58.6%) [6]. The ability to provide information regarding disease prevention, reproductive decisions, life planning, and diagnosis were additionally ranked among the highest values of genetic testing [6]. However, the ability to properly interpret complex genetic results in the clinic, especially when confronted with variants of unknown significance, is a factor limiting wide-spread utilization [7, 8]. In order to fully take advantage of the technological advancements in

genomic sequencing, researchers need to work with clinicians to develop better tools and an understanding of how genetic variants affect phenotype.

Genetic testing interpretation in the clinic and variant of significance (VUS) burden

NGS has enabled the identification of more than 4,000 known genes associated with monogenic disorders (Online Mendelian Inheritance in Man (OMIM)) [9-11]. One complicating variable in correctly diagnosing a monogenic disorder in affected patients are variants of unknown significance. On average, roughly 20,000 to 50,000 variants may be identified in coding regions of the genome. While some variants are easily characterized as a Class 1/2 (benign/likely benign) or a Class 4/5 (likely pathogenic/pathogenic) variant, Class 3 variants, the VUSs, represent a challenge in the clinic. These VUSs are typically changes in a gene which have not previously been encountered or do not have a known function. The uncertainty behind these variants make diagnosing and treating patients difficult, as their effects on protein function remain unknown. When confronted by thousands of variants on genomic sequencing it is necessary to narrow down to only clinically relevant variants. In order to quickly exclude variants which are likely benign or benign, several characteristics are analyzed. Frequent detection of a variant in an unaffected population indicates the variant is unlikely to be disease causing. Furthermore, if a variant is found in a symptomatic patient as well as an unaffected family member, then it is unlikely that the variant is clinically significant. Additionally, variants which result in a synonymous amino acid change (ex. nucleotide changes which do not result in amino acid changes due to redundancy of the genetic code) are often benign.

Once the likely benign and benign variants are excluded, doctors are still faced with VUSs. In order to determine whether a VUS contributes to a patient's phenotype, a systematic approach is undertaken. This includes the development of standards and guidelines to determine the likelihood of a variant being pathogenic or benign based on several variant characteristics computational predictions of a variant's effects, segregation analyses, functional studies and other known information [12, 13]. Previously published research can provide information on whether the variant has been observed in other patients who present similarly. Additionally, variants are prioritized if they are a splice site, frameshift, or truncating variant. Nonconservative amino acid substitutions and variants which occur in an evolutionarily conserved amino acid are also likely to be clinically relevant. Additionally, utilizing tools and databases to help interpret genetic variants has become increasingly more advantageous. For example, functional prediction through in-silico modeling is useful at predicting whether a variant is likely to be deleterious or damaging. ClinGen, an NIH resource for clinicians and researchers, is also useful in defining the clinical relevance of genes and variants [14].

Despite current methods for classifying variants, the presence of a VUS on genetic panels still creates many challenges in the clinic. VUSs are commonly encountered; 28-40% of patients who undergo genetic testing for cancer or cardiovascular disorders are found to possess a VUS in a clinically relevant disease gene [15-18]. One aspect that makes VUSs difficult for physicians is the lack of training in interpreting genetic results. Often, patients present to the doctor with a clinical suspicion for a genetic disorder. Therefore, VUS interpretation can be especially challenging when the patient already exhibits the symptoms associated with the genetic variant. Macklin et

al. conducted a survey to assess physician self-reported comfort levels in regard to interpreting results which contain a VUS and found that only approximately 15% of physicians responded correctly to case examples with ~50% incorrectly defining a VUS [19]. Additionally, over 75% of physicians surveyed made an unwarranted recommendation for genetic testing [19].

When recommending genetic testing, it is also important to understand the psychosocial and healthcare impacts that these VUSs can have on the patients themselves and to properly inform the patient of their results. Richter et al. surveyed patients who had received a VUS result and found that one-third of patients failed to recall the clinical significance of their results [20]. Additionally, when genetic counselors were surveyed, only 63.2% felt as though a VUS result was understood by their patients [21]. Overall genetic testing can result in patient distress, with the ambiguity of receiving a VUS diagnosis being particularly stressful [22-25]. The uncertainty of a diagnosis, especially after undergoing genetic testing and being told that there is a genetic change, can be quite distressing for patients. Care needs to be taken in treating the psychosocial impacts of a VUS along with addressing the potential for a genetic disorder diagnosis.

Furthermore, a VUS finding is often used to inform treatment and surveillance in a manner analogous to more established methods that are used with pathogenic variants, particularly in patients who present with the symptoms which can be associated with the genetic variant, sometimes in error. Murray et al. reviewed surgical decisions in patients who had received a VUS result in the *BRCA1* or *BRCA2* gene and found that many pursued cancer risk-reducing measures, including 10.7% who underwent a mastectomy and 20.6% that pursued a bilateral salpingo-oophorectomy [26]. Many of these measures

may have been unnecessary, particularly as the reclassification of previously reported VUSs often occurs. In reanalyzing 75 patients with VUSs in a *BRCA1* and *BRCA2* gene, 30 VUSs (40%) were reclassified as likely benign or benign and 2 VUSs (2.7%) were reclassified as likely pathogenic [27]. Genetic testing is becoming an increasingly common tool in evaluating patients in the clinic and correct interpretation of the results is very important for patient health, treatment, and future medical care. Understanding whether a variant is disease-causing or benign is critical to patient treatment, particularly when preventative measures, such as aggressive prophylactic surgery are the treatments of choice.

Science and translational medicine play an integral role in understanding how VUSs may impact patient phenotype. To aid in classification of unknown variants, functional validation is an integral step in understanding how a variant impacts protein function. Determining the molecular phenotype of a genetic variant provides support for variant pathogenicity [28, 29]. Experiments which are useful in classifying a variant include: rescue/overexpression experiments (comparing WT to variant-sequences) and gene-editing/transgenic expression in model systems (yeast, zebrafish, mouse, etc.) [28, 30]. Additionally, determination of the molecular phenotype gives a better understanding of patient phenotype, allowing for genotype-phenotype correlations to be made. Furthermore, there are several databases, such as GeneMatcher, which aid in determining a variant's contribution on phenotype by providing gene, protein, and phenotypic data correlated to genetic variants [28, 31].

Patient genotype can enhance our understanding of phenotype

A subset of genetic disorders which have a previously established genotype-phenotype correlation, include repeat expansion disorders. Unstable nucleotide repeats represent a unique mutational mechanism that was first described in 1991 as the causative mutations in Fragile X Syndrome and Spinal and Bulbar Muscular Atrophy (SBMA) [32, 33]. Repeat expansion disorders are a unique subtype of genetic disease resulting from the unstable expansion of tri-, tetra-, penta-, hexanucleotide and longer repeats which often result in neurodegenerative and neuromuscular disorders. Interestingly, despite the causative mutation of an expanded nucleotide sequences remaining the same in these disorders, the functional consequences of the expansions can drastically differ. For example, trinucleotide expanded repeat sequences can result in differing molecular features including loss of protein function (ex. Friedreich's Ataxia), altered protein function (ex. Huntington's Disease), or altered RNA function (ex. Myotonic Dystrophy) [34].

Friedreich's ataxia (FRDA) is a trinucleotide repeat disorder affecting 1 in 50,000 individuals that is caused by the homozygous expansion of GAA repeats within the first intron of the *frataxin* (*FXN*) gene [35]. The GAA expansions within the *FXN* gene are associated with transcriptional silencing with consequent frataxin deficiency due to chromatin changes near the repeats [36-38]. FRDA has a wide spectrum of phenotypic symptoms with the first symptoms usually occurring between 5-15 years old and typically including ataxia, sensory impairment, and scoliosis. Symptoms also include cardiac dysfunction and cardiomyopathy with some patients also developing pancreatic issues [39-42]. Symptoms can be highly variable even in patients with the same repeat

disorder and the repeat size determination at time of diagnosis may not always entirely explain the phenotype. Symptom severity correlates with number of GAA repeats in the shorter allele; however, the correlation is far from perfect and unexpected phenotypic variability has been observed [43, 44]. Prior studies have demonstrated that repeat expansion lengths determined from blood-draws or buccal swabs may not accurately reflect repeat length in pathologically relevant tissues, indicating that somatic mosaicism may underlie some phenotypic variability [44-50].

Despite previous studies confirming somatic instability in FRDA, larger studies analyzing the GAA repeat variability in multiple tissues affected by frataxin deficiency in a larger cohort have yet to be undertaken. We hypothesized that the pathological GAA repeats would exhibit somatic mosaicism among clinically and research relevant tissues. Additionally, we hypothesized that the expanded GAA repeats would demonstrate somatic instability over time. Herein, we analyzed somatic instability of the expanded repeats within the FXN gene in heart, cerebral cortex, spinal cord, cerebellar cortex, and pancreatic tissues of 15 FRDA patients. Additionally, we interrogated differences in GAA repeat lengths in cells commonly used in diagnosis and patient research: paired lymphocytes and fibroblasts from 16 FRDA patients. Finally, a longitudinal analysis was performed on the expanded repeats in lymphocytes taken through repeated blood draws of 5 FRDA patients. Changes in tissue-specific repeat lengths of the pathologically expanded GAA repeats may be one of the most important factors underlying clinical variability observed in FRDA patients. The dynamics of repeat length instability among clinically relevant tissues and over time could have critical implications in patient

symptomology and progression as well as in clinical interpretation of research and clinical trials.

While phenotypic severity often correlates with repeat length in repeat expansion disorders, other mutation types can be more difficult to predict. The *neurofibromin* (*NF1*) gene is among the top ten genes with the highest number of variants submitted to ClinVar [16]. Pathogenic variants in the *NF1* gene result in Neurofibromatosis Type 1 (NF1), a neurological tumor predisposition disorder which affects approximately 1 in 3,000 [51]. Patients with this disorder typically exhibit a multifaceted phenotype, including bone dysplasia, learning disabilities, multiple neurofibroma tumors, and malignant tumors. Nearly 2,900 pathogenic variants, with most resulting in lack of NF1 expression, have been reported in the Human Gene Mutation Database (http://www.hgmd.org). There are multiple types of variants observed in *NF1*. Kang et al. identified 250 unique mutations in 339 families and found a mix of frameshift (34%), in-frame insertions or deletions (4%), large deletions (7%), splicing (11%), nonsense (29%), and missense (15%) mutations [52]. Genetic variants occur throughout the *NF1* gene and there does not appear to be a hotspot for mutagenic occurrences. However, while there is a wide spectrum of symptom variability, there are a few genotype-phenotype correlations observed in NF1 patients. Typically, patients with truncating/splicing mutations and large deletions exhibit a higher symptom burden compared to missense mutations and in-frame deletions [52]. Additionally, there are specific *NF1* missense mutations which are associated with unique patient phenotypes. Variants affecting NF1 codons 844-848 as well as R1276 and K1423 result in an aggressive phenotype with severe symptomatic spinal neurofibromas, optic pathway gliomas and a higher predisposition to developing malignancies [53, 54].

However, variants affecting residues M1149, R1809, R1038, and DelM992 result in a mild phenotype with patients exhibiting predominantly café au lait macules (CALMs), skinfold freckling, and no neurofibromas [54-59]. Genetic disorders with missense mutations can have a variable phenotype which may be dependent on the location of the mutation within the gene. Different mutations within the same gene can cause a myriad of complications including but not limited to protein folding, stability, function, or effects on binding partners.

Little is known about the *NF1* gene outside of its function as an inhibitory regulator of Ras signaling. While pathogenic variants occur throughout the gene, the functional consequences of different variants are still to be determined. We hypothesized that determination of variant-specific functional consequences would aid in the ability to predict disorder progression and determine potential therapeutics. Herein we further validated the *mNf1* cDNA expression system developed by Wallis et al. for the study of variant-specific effects on neurofibromin function [60]. Multiple patient-specific variants, many with known genotype-phenotype correlations, spanning the entire *NF1* gene were analyzed. We demonstrated that *Nf1* variants effects on protein stability and Ras signaling are not always linked. Indeed, our study exhibited variant clustering based on multiple factors including severity of presentation, protein function, and protein stability. Such stratification may be useful for understanding patient disease course and targeted care and therapeutics.

Precision medicine

Phenotype-genotype correlations demonstrate the importance of understanding how unique variants can specifically affect the patient. Precision medicine is the customization of healthcare based on a patient's genetics, environment, and lifestyle. The distinct phenotypes in some NF1 patients highlight the benefit of customizing healthcare based on a patient's specific genetic mutation. NF1 patients with a severe phenotype-associated variant would benefit from a more aggressive treatment plan with regular monitoring at shorter intervals compared to patients with a mild phenotype-associated variant where similar treatments/monitoring would put a potentially unneeded financial and psychological burden on the patient while increasing the risk of side effects from unwarranted treatments. Precision medicine is beneficial in other disorders as well. The cancer predisposition genes *BRCA1* and *BRCA2* also exhibit variant-specific phenotypes which would benefit from a more patient-specific precision medicine guide. Al-Mulla et al. demonstrated that founder mutations located in exon 2 of the *BRCA1* gene exhibited a lower penetrance compared to mutations located in exons 11, 13 and 20 [61]. Additionally, the prevalence of breast and ovarian cancers differ significantly between the founder mutations as well and can provide a useful guide for which patients require stricter ovarian and/or breast cancer surveillance [62, 63].

While there are obvious benefits to patient-guided treatment in genetic cancer predisposition disorders, genotype-guided precision medicine is also beneficial in other mendelian disorders. The gene *FBN1* is implicated in cardiovascular disease, is also among the top ten genes with the highest number of unique variants identified, and is associated with several unique phenotypes including Marfan Syndrome (MFS) and

Isolated Ectopia Lentis (IEL) [16, 64] (OMIM: 134797). *FBN1*-associated MFS requires aggressive heart monitoring with frequent echocardiogram studies to prevent sudden-death. However, several *FBN1* variants are implicated in IEL without cardiovascular presentation [65, 66]. In these patients aggressive monitoring would be financially burdensome and place the patient under undue stress and anxiety.

Additionally, precision medicine is utilized to determine patient-specific therapies based on genetic correlations. *NTRK2*-associated phenotypes highlight the importance of understanding the molecular mechanisms behind a phenotype in order to determine therapeutic avenues. *NTRK2* encodes for the tropomyosin related kinase B (TRKB) receptor which binds preferentially to brain derived neurotrophic factor (BDNF). NTRK2 is a member of the neurotrophin receptor kinase family (NTRK) and is important in the regulation of neuronal development and function. Variants located in the *NTRK2* gene are associated with two distinct phenotypes: developmental and epileptic encephalopathy 58 (DEE) (OMIM: 617830) and obesity, hyperphagia, and developmental delay (OHDD) (OMIM: 613886). Interestingly these two phenotypes predominantly occur in a region-specific manner with DEE patients harboring variants within the transmembrane domain of NTRK2 while patients who exhibit OHDD harbor variants in the tyrosine kinase domain [67-69]. This suggests unique mechanisms of pathology behind the phenotypes.

We hypothesized that the two distinct NTRK2-associated phenotypes represent distinct functional consequences. Additionally, we hypothesized that the specific variant effects can be targeted for artificial-intelligence-determined therapeutics. Herein, we described the functional consequences of two distinct NTRK2 variants: Y434C (located in the transmembrane domain) and T720I (located in the tyrosine kinase domain). We

demonstrated that variant Y434C results in upregulation of TRKB and MAPK signaling at baseline levels and is unresponsive to BDNF stimulation. We also demonstrated that variant T720I reacts similarly to other tyrosine kinase domain variants with hypomorphic TRKB-BDNF signaling and additionally resulted in decreased mTor activity [67, 68, 70].

Furthermore, we determined potential therapeutic options for these two variants through the utilization of artificial intelligence (AI) in collaboration with The Precision Medicine Institute at University of Alabama-Birmingham. AI is integral to precision medicine as it allows for the analysis of multiple different health risk factors in the formation of more personalized and patient-specific approaches towards treatment plans and risk prevention [71, 72]. The therapeutics we identified were able to target the distinct functional consequences of the two NTRK2 variants.

The opposing molecular mechanisms behind the two NTRK2 phenotypes highlights the importance of functional studies and understanding how genetic variation affects function. Indeed understanding the functional consequences of different genetic variants can aid in both classifying the pathogenicity of the variant as well as determining potential therapeutic avenues, especially when utilizing tools such as AI [73]. Ultimately, precision medicine guided treatments that can counter the pathological effects of a genetic mutation can be determined in a patient-variant specific manner with greater long-term benefits and reduced side effects.

Conclusions

Advancements in genetic sequencing technology have made genetic testing more common and created more opportunities to better understand human genetics and disease.

In these studies we address the central hypothesis that determining the effects of genetic variation on an individual and tissue-specific level will aid in understanding patient presentation and progression and the development of targeted therapeutics. There are still, however, several obstacles to overcome. The interpretation of genetic results lag behind the advancements in technology. Additionally, caution needs to be taken when considering what psychosocial impacts can result from learning about potential health risk factors and predispositions for the patient. Overall, genetic testing opens the door to greater precision in medicine with the potential for patient-directed healthcare and treatment plans. However, care must be taken when interpreting the results and conveying them to the patient.

SOMATIC INSTABILITY OF THE EXPANDED GAA REPEATS IN FRIEDREICH'S ATAXIA

by

ASHLEE LONG, JILL S. NAPIERALA, URZULA POLAK, LAUREN HAUSER
ARNULF H. KOEPPEN, DAVID R. LYNCH, MAREK NAPIERALA

PLOS ONE

Format adapted for dissertation

Abstract

Friedreich's ataxia (FRDA) is a genetic neurodegenerative disorder caused by transcriptional silencing of the *frataxin* gene (*FXN*) due to expansions of GAA repeats in intron 1. FRDA manifests with multiple symptoms, which may include ataxia, cardiomyopathy and diabetes mellitus. Expanded GAA tracts are genetically unstable, exhibiting both expansions and contractions. GAA length correlates with severity of FRDA symptoms and inversely with age of onset. Thus, tissue-specific somatic instability of long GAA repeats may be implicated in the development of symptoms and disease progression. Herein, we determined the extent of somatic instability of the GAA repeats in heart, cerebral cortex, spinal cord, cerebellar cortex, and pancreatic tissues from 15 FRDA patients. Results demonstrate differences in the lengths of the expanded GAAs among different tissues, with significantly longer GAA tracts detected in heart and pancreas than in other tissues. The expansion bias detected in heart and pancreas may contribute to disease onset and progression, making the mechanism of somatic instability an important target for therapy. Additionally, we detected significant differences in GAA tract lengths between lymphocytes and fibroblast pairs derived from 16 FRDA patients, with longer GAA tracts present in the lymphocytes. This result urges caution in direct comparisons of data obtained in these frequently used FRDA models. Furthermore, we conducted a longitudinal analysis of the GAA repeat length in lymphocytes collected over a span of 7±9 years and demonstrated progressive expansions of the GAAs with maximum gain of approximately 9 repeats per year. Continuous GAA expansions throughout the patient's lifespan, as observed in FRDA lymphocytes, should be considered in clinical trial designs and data interpretation.

Background

Friedreich's ataxia (FRDA), the most common hereditary ataxia, is a multi-organ disorder with an approximate prevalence of 1 in 50,000 [1]. The majority of patients (> 95%) are homozygous for a guanine-adenine-adenine (GAA) repeat expansion in the first intron of the *frataxin* (*FXN*) gene. In FRDA patients, the expanded GAA repeat tracts are associated with locus-specific chromatin changes leading to transcriptional silencing [2-4] of the *FXN* gene and consequent deficiency of the mitochondrial protein, frataxin [1, 5, 6].

FRDA has a variable clinical presentation with onset typically between ages 5 - 15 and the vast majority of patients exhibiting symptoms by age 25 [1, 7, 8]. The first symptoms of FRDA often include ataxia, sensory impairment, and scoliosis. The heart is also affected, resulting in cardiomyopathy and cardiac dysfunction, with 60% of FRDA deaths reflecting cardiac dysfunction [7, 9, 10]. Approximately 10-30% of patients develop dysfunction of the pancreatic beta-cells, including impaired glucose tolerance and diabetes mellitus [8, 9]. The severity of symptoms in FRDA correlates with the size of the expanded GAA repeats, and likely reflects the effects of frataxin deficiency on different tissues.

GAA repeats are polymorphic in unaffected individuals, with lengths ranging from 7-22 copies [1]. The expanded GAA tracts typically contain between 200 and 900 or more repeats, with the majority of alleles containing 700-800 copies of the GAA motif [1]. In unaffected individuals, GAA repeats are stable, with some infrequently occurring contractions or expansions. However, GAA instability can be seen in premutation *FXN* alleles, containing 26-44 uninterrupted repeats [11-14]. Intergenerational instability is observed in the expanded GAA repeats, with a preference for contractions in the paternal line while the maternal line shows both contractions and expansions [15]. Comparative analysis of paired blood and sperm samples from FRDA patients revealed shorter repeat

lengths in sperm, occasionally with GAA tract contractions resulting in unaffected allele sizes [15].

In addition to intergenerational instability, expansions and contractions of GAA repeats in somatic cells have been detected in FRDA, with different tissues containing varying GAA lengths within the same patient [1, 13-22]. Analysis of GAA instability in dorsal root ganglion (DRG) samples from six patients showed a pronounced bias towards repeat expansions [18]. In addition, heterogeneity of the GAA repeat tract was detected in a limited number of samples from other tissues, including blood, cerebrum, cerebellum, spinal cord, brainstem, and heart [18]. *FXN* GAA repeats in the cerebellum showed an expansion bias while the lowest number of large contractions in the repeat tract was observed in DNA isolated from a heart sample [18]. Somatic GAA instability was also detected throughout different regions of the brain, including four cerebral cortical areas, the dentate nucleus and cervical spinal cord [21]. Additionally, two case studies reported differences of GAA tract size in the sural nerve compared to lymphocytes [16, 20]. GAA length heterogeneity was also observed when comparing repeats in DNA samples isolated from paired lymphocytes and cultured fibroblasts of FRDA patients, however, no bias for expansions or contractions was noted [22]. While longitudinal studies of GAA repeat size in an FRDA cohort have not been reported, analyses of expanded GAA repeat tracts in cerebrum, cerebellum, spinal cord, heart, and pancreas of a 24-year-old patient and an 18-week fetus indicated an age-dependent increase of instability, thus suggesting an accumulation of changes in expanded GAA length throughout the lifetime [19].

Along with human studies, somatic instability has been reported in the humanized YAC transgenic YG8 and YG22 as well as the YG8sR FRDA mouse models [17, 23].

These mice harbor relatively short tracts of 190 + 82 (two tandem YAC transgene GAA sequences), 190, and ~200 GAA repeats, respectively [17, 23]. Similar to the results obtained from patient tissues, the mouse tissues showed an expansion bias, particularly in the cerebellum and DRG, as well as an age-dependent increase in mutation load [17, 23]. Lastly, *in vitro* studies on cultured immortalized lymphoblast cell lines and induced pluripotent stem cells (iPSCs) obtained by reprogramming of FRDA fibroblasts demonstrated somatic instability when serially passaged [24-26].

Although somatic instability in FRDA has been reported, systematic analyses of GAA heterogeneity in multiple tissues affected by frataxin deficiency in a larger cohort are lacking. Herein, we determined the somatic instability of the expanded GAA repeats in five tissues from 15 FRDA patients: heart, cerebral cortex, spinal cord, cerebellum, and pancreas. In addition, we compared GAA tract lengths in paired fibroblast and lymphocyte samples isolated from 16 FRDA patients. Finally, we have conducted a longitudinal analysis of GAA repeat tract lengths in lymphocytes based on repeated blood sampling from FRDA patients. Tissue-specific differences and variability in the number of GAA repeats may be one of the most important, but also the most difficult to ascertain, factors underlying clinical variability among FRDA patients. As demonstrated in our study, the dynamics of expanded GAA repeat length over time could have significant implications for correlating disease severity and interpretation of data from clinical trials.

Methods

Patient tissue and DNA samples

All research involving patient/patient tissues has been approved by the Institutional Review Boards at University of Alabama-Birmingham (IRB Protocols: N160923005 and N160922011); Children's Hospital of Philadelphia (IRB Protocol 10-007864). Autopsy specimens of heart, cerebral cortex, spinal cord, cerebellar cortex, and pancreas of 15 FRDA patients were obtained from the FRDA tissue repository maintained at the Veterans Affairs Medical Center in Albany, NY, USA. Tissue samples were divided into aliquots upon arrival and stored at -80°C. Blood and genomic DNA samples from five FRDA patients used in longitudinal analyses were obtained from The Children's Hospital of Pennsylvania (CHOP, Philadelphia, PA). Blood was stored at 4°C until the DNA extraction. Fibroblast lines were derived from skin biopsies performed at CHOP as described in [27], with approvals from CHOP and UAB IRBs. All FRDA patient material used in this study was derived from patients carrying homozygous GAA repeat expansions. Cerebral and cerebellar cortex DNA from unaffected individuals were obtained from The Cooperative Human Tissue Network (Southern Division, University of Alabama at Birmingham) and Dr. Mark Pook (Brunel University, London, UK).

DNA isolation

Genomic DNA from patient tissue samples was isolated using the DNeasy Blood & Tissue kit (cat. 69504; Qiagen). Genomic DNA was isolated from blood samples using the PureLink™ Genomic DNA Mini kit (cat. K1820-00; Invitrogen) or Quick-gDNA™ Blood MidiPrep (cat. D3074; Zymo Research). All isolations were conducted according to manufacturers' recommendations. Genomic DNA concentration, purity, and quality were

determined through measurement on a NanoDrop 2000c Spectrophotometer (ThermoFisher SCIENTIFIC) as well as visualization on a 0.7% agarose gel.

PCR Amplification

Amplification of GAA repeat expansions in the *FXN* gene was performed by PCR using primers GAA_F: 5' – GGCTTGAACTTCCCACACGTGTT and GAA_R: 5' – AGGACCATCATGGCCACACTT as previously described [26, 28, 29]. Amplifications were conducted in 40 or 50µL reactions containing 50-200ng of DNA template. The thermal cycler was programmed for an initial denaturation step of 3 minutes at 94°C, followed by 20 cycles of 20 seconds at 94°C, 30 seconds at 64°C, and 5 minutes at 68°C, followed by 9 cycles of 20 seconds at 94°C and 5 minutes at 68°C, with each subsequent elongation step increased by 15 seconds. A final extension step of 7 minutes at 68°C was then performed.

Amplification of GAA repeats at the 5q23 locus was performed with primers 5q23F: 5' – GTTGCATAGATAAATCAAATTCAT and 5q23R: 5' – ACTCACAGAAAGTATTATTATTCC [30, 31]. Amplifications were conducted in 25µL reactions containing 100ng of DNA template. The thermal cycler was programmed for a denaturation step of 3 minutes at 94°C, followed by 30 cycles which consisted of 30 seconds at 94°C, 30 seconds at 50°C, and 2 minutes at 72°C, with a final extension step of 2 minutes at 72°C.

PCR amplification of the intron1-exon2 region of the *FXN* gene, approximately 9kbp from the GAA repeat tract, was performed using primers In1Ex2F: 5' – AGCACTCGGTTACAGGCACT and In1Ex2R: 5' – GCCCAAAGTTCCAGATTTCC as previously described [32]. Amplifications were performed in 20µL reactions containing

50-100ng of DNA template. The thermal cycler was programmed for an initial denaturation step of 3 minutes at 95°C, followed by 35 cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 30 seconds at 72°C, with a final extension step of 7 minutes at 72°C.

Reactions utilized the FailSafe PCR System with mix D (cat. FS99250; Epicentre) or JumpStart™ REDTaq® ReadyMix™ Reaction Mix (cat. P0982; Millipore Sigma). The amplification products were resolved on 0.9-1% agarose gels stained with ethidium bromide. Lane analyses were performed using Image Lab 5.0 software (BioRad. The length of an expanded GAA tract was determined using the base pair size called by Image Lab 5.0, with the total GAA length calculated by subtracting the subtracting the length of the sequences flanking the GAA repeats, including the length of the PCR primers, from the number of base pairs of the PCR product and dividing the difference by three: [Number of GAA repeats = (length of base pairs of a PCR product - 498)/3].

Western Blot

Lysates were prepared using a RIPA buffer containing 150 mM NaCl, 1% IGEPAL® CA-630, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 50mM Tris-HCl, pH 8, and a protease inhibitor cocktail (cat. P8340; Millipore Sigma). The Bradford Protein Assay Kit (cat. #500-0006; Bio-Rad) was used to determine protein concentration. Fifty micrograms of tissue protein lysate were electrophoresed on NuPAGE™ 4-12% Bis-Tris protein gels (cat. NP0322BOX; ThermoFisher SCIENTIFIC) and transferred onto nitrocellulose membranes (cat. 162-0112, Bio-Rad). Human frataxin was detected with the anti-frataxin polyclonal H-155 antibody (cat. sc-25820; Santa Cruz) and human voltage dependent anion channel (VDAC) was identified with the anti-VDAC polyclonal antibody (cat. #4866; Cell Signaling Technology), both used at 1:1,000 dilutions. Human heat shock

protein (Hsp) 60 was detected with the anti-Hsp60 D307 antibody (cat. #4870; Cell Signaling Technology) at 1:2,000 dilution. GAPDH was identified with the anti-GAPDH antibody (cat. MAB374; Millipore) at 1:25,000 dilution. Primary antibody incubation was performed for at least 12 hours at 4°C. Sheep anti-mouse immunoglobulin (NA931V; GE Healthcare) and donkey anti-rabbit immunoglobulin (NA934V; GE Healthcare), linked to horseradish peroxidase, were used as secondary antibodies at a 1:5,000 dilution for 1 hour at room temperature. Signal was exposed using Amersham ECL Prime Western Blotting Detection Reagent (cat. RPN2232; GE Healthcare) or SuperSignal® West Dura Extended Duration Substrate (cat. 34075; ThermoFisher SCIENTIFIC) and measured using Image Lab 5.0. Quantification was performed using ImageJ software (National Institutes of Health).

Statistical Analysis

Statistical analyses were done using IBM® SPSS® Statistics Version 24 and Excel 2016 software. A p-value less than 0.05 was considered significant for all analyses.

Results

To determine whether somatic instability is present in the expanded GAA tracts of the *FXN* gene in FRDA patient tissues, we analyzed sizes of the repeats in heart, cerebral cortex, spinal cord, cerebellar cortex, and pancreas samples from 15 patients (Table 1). These samples represent a broad spectrum of tissues affected by frataxin deficiency in FRDA patients. Samples were obtained from six males and nine females, with a mean age of disease onset at ~12±7 years and a mean age of death at ~43±18 years (Table 1). The majority of patients died of heart-related causes (60%) and cachexia (27%). Quantitative

western blot analyses demonstrated reduced frataxin levels in patient tissues compared to controls, confirming FRDA status of all samples (S1 Fig).

Table 1. Characterization of the FRDA patient tissues.

Patient	Sex	Age of onset (years)	Age of death (years)	Disease duration (years)	Tissues Available
M1	M	10	24	14	Cc, Sc, Cb
M2	M	8	27	19	H, Cc, Sc, Cb, P
M3	M	9	33	24	H, Cc, Sc, Cb, P
M4	M	9	37	28	H, Cc, Sc, Cb, P
M5	M	7	35	28	H, Cc, Sc, Cb, P
M6	M	16	46	30	H, Cc, Sc, Cb, P
F1	F	12	24	12	H, Cc, Sc, Cb, P
F2	F	10	47	37	H, Cc, Sc, Cb, P
F3	F	17	50	33	H, Cc, Sc, Cb, P
F4	F	5	25	20	H, Cc, Sc, Cb, P
F5	F	7	28	21	H, Cc, Sc, Cb, P
F6	F	7	55	48	H, Cc, Sc, Cb
F7	F	18	67	49	H, Cc, Sc, Cb, P
F8	F	15	69	54	H, Cc, Sc, Cb, P
F9	F	34	77	43	Cc, Sc, Cb, P
Mean ± SD	M, 6; F, 9	12.3 ± 7.2	42.9 ± 17.6	30.7 ± 13.1	
Range		5-34	24-77	12-54	

Abbreviations: F, female; M, male; SD, standard deviation; H, heart; Cc, cerebral cortex; Sc, spinal cord; Cb, cerebellar cortex; P, pancreas.

The sizes of the GAA tracts by PCR differed across five tissues analyzed in samples from all 15 FRDA patients (13 for the heart and pancreas samples; Fig 1A-C and S2 Fig). Typically, several bands or even a broad smear of the PCR products containing expanded GAAs are observed following agarose gel electrophoresis (Fig 1A-C). As a positive control to ensure that the PCR conditions yielded robust amplification of expanded GAAs for each experiment, genomic DNA was isolated from an FRDA fibroblast cell line and used as a template alongside genomic DNA isolated from patient tissues. The fibroblast line used was derived form an individual unrelated to patients from whom the tissues were obtained. Amplification of expanded GAAs using DNA obtained from FRDA fibroblasts revealed

two distinct *FXN* alleles (Fig 1A-C, positive control lane), indicating greater somatic instability in heart, cerebral cortex, spinal cord, cerebellar cortex, and pancreas when compared to homogenous primary cell lines derived from patients. Furthermore, amplification of a polymorphic repeat tract at locus 5q23, previously shown to contain ≥44 GAA triplets [31], revealed a lack of somatic instability between analyzed tissues (Fig 1D-F and S2 Fig). In parallel, PCR amplification of a ~200 bp intron1-exon2 fragment of the *FXN* gene demonstrated that the quality of genomic DNA isolated from all tissues allowed for efficient amplification (Fig 1G-1I and S2 Fig).

Fig 1. Instability analysis of the expanded GAAs in the *FXN* gene across different somatic tissues. Genomic DNA was extracted from heart (H), cerebral cortex (Cc), spinal cord (Sc), cerebellar cortex (Cb) and pancreas (P) tissues and the GAA repeats in the *FXN* locus were amplified by PCR. The results from FRDA patients (**A**) F2 (**B**) F7 and (**C**) M6 shown as examples. (-) represents no-template control and (+) represents positive control for amplification of the expanded GAAs (genomic DNA isolated from fibroblasts obtained from unrelated FRDA fibroblasts). (**D-F**) A GAA repeat tract at the 5q23 locus was amplified by PCR using the same genomic DNA templates used for reactions shown in (**A-C**). (**G-I**) A fragment spanning intron 1 - exon 2 of the *FXN* gene, downstream of the GAA tract, was also amplified using the same templates to serve as a control for genomic DNA quality.

Next, we performed quantitative analyses of PCR products containing expanded GAAs to determine tissue-specific differences in the stability of the repeat tracts. As it was impossible to assign a precise size of each expanded allele (GAA1 and GAA2) in the majority of samples due to the high level of somatic instability, we developed an unbiased GAA length analysis protocol based on defining the minimum, maximum, and mean GAA size of the repeat tracts detected in each sample (Methods). The mean GAA repeat sizes among 15 FRDA patients and 5 different tissues were between 428 and 914 GAAs (Figs 1, 2 and S2, S3 Figs). The median difference between maximum and minimum GAA tract size is 726 GAAs (Figs 1, 2 and S2, S3 Figs).

Fig 2. Quantitative analysis of expanded GAA repeat instability in the *FXN* gene across different tissues. The expanded GAA repeats in the *FXN* gene were amplified from genomic DNA extracted from heart (H), cerebral cortex (Cc), spinal cord (Sc), cerebellar cortex (Cb) and pancreas (P) tissues isolated from FRDA patients. The band intensity of the PCR products and repeat size are shown for FRDA patients (**A**) F2, (**B**) F7, and (**C**) M6. Solid vertical lines represent the mean of all GAA repeat sizes detected across all five tissues for each patient. Gel lanes were manually outlined and gel bands were detected via the Image Lab 5.0's band finder set to high sensitivity. Faint bands of PCR products not detected by the software were manually identified. Band boundaries, accounting for smearing, were automatically outlined by the program with final manual adjustments to include the entire spectrum of PCR products. Multiple PCR analysis with determination of GAA lengths was performed to demonstrate reproducibility of PRC and reliability of measurements (S4 Fig).

Interestingly, comparisons of the GAA length between 5 tissues in a repeated measures analysis revealed that the average GAA size in the heart and pancreas is significantly longer than GAAs present in cerebral cortex, spinal cord, and cerebellar cortex of the same patient (p-value <0.05; Fig 2A-C and S3 Fig). The mean GAA sizes in the heart and pancreas were 752 (standard deviation (SD): 190) and 720 (SD: 167) GAAs compared to 614 (SD: 186), 552 (SD: 168) and 551 (SD: 131) triplet repeats in cerebellar cortex, spinal cord, and cerebral cortex, respectively. The longest mean GAA tracts were detected in either the heart or pancreas samples in 11 out of 14 patients. In contrast, the shortest mean GAA tracts were detected in either the cerebral cortex or the spinal cord tissue samples in 14 out of 15 patients. For all tissues, the number of GAA repeats correlated inversely with age of death, with Pearson's R values ranging from -0.423 (heart) to -0.781 (pancreas). This trend reached statistical significance in spinal cord, cerebellar cortex, and pancreas (p-values <0.05; Fig 3). For comparison, we analyzed lengths of the GAA repeat tracts in cerebral cortex and cerebellar cortex samples from 3 unaffected individuals. The number of GAA repeats in all non-FRDA tissues was less than 30 and no somatic instability was observed (S5 Fig).

Fig 3. Correlative analyses between mean GAA repeat sizes and patient age of death. The mean tissue GAA tract length in FRDA patient (**A**) heart, (**B**) cerebral cortex, (**C**) spinal cord, (**D**) cerebellar cortex, and (**E**) pancreas is plotted against the patient's age at time of death. Statistical significance was reached for spinal cord, cerebellar cortex and pancreas ($p < 0.05$).

Lengths of GAA tracts in paired FRDA fibroblast and lymphocyte samples

Patient fibroblasts and lymphocytes are commonly used in vitro models of human diseases. To determine whether the number of GAAs differ between fibroblasts and lymphocytes obtained from the same FRDA patients, we performed PCR analysis of GAA sizes in 16 paired samples (Table 2). Interestingly, in all 16 patient samples, longer GAA repeats were detected in lymphocytes compared to fibroblasts, in at least one allele (Fig 4A). The *FXN* gene contained larger GAA expansions in 27 out of 32 alleles (84.4%; Fig 4A) when analyzed from lymphocyte DNA. Only 3 *FXN* alleles exhibited GAA contractions in lymphocytes relative to the corresponding GAA alleles in fibroblasts (FA2, FA11 and FA16; Fig 4; Table 2), while repeat lengths of 2 alleles did not change (FA1 and FA13; Fig 4; Table 2). The median increase in repeat number detected in FRDA lymphocytes over the fibroblasts was 75 triplets for GAA1 and 258 repeats for GAA2 corresponding to 21% and 33% expansions, respectively. The maximum detected gain in GAAs reached ~600 triplet repeats (FA5; Fig 4A). The somatic expansions in FRDA lymphocytes versus fibroblasts differed significantly ($p < 0.01$ by paired two sample for means t-test performed for both alleles; Fig 4B). In addition, a correlation exists between the number of GAA repeats gained in lymphocytes relative to size of the repeat tracts in fibroblasts and the number of GAA repeats in lymphocytes, indicating that longer GAA repeat sequences are prone to larger expansions (Fig 4C and 4D). Taken together, these data demonstrate a

significantly greater and length dependent propensity for expansion of GAA repeat tracts in FRDA lymphocytes compared to fibroblasts.

Table 2. GAA repeat length in fibroblasts and lymphocytes of FRDA patients

Patient	Fibroblasts:		Lymphocytes:	
	GAA1	GAA2	GAA1	GAA2
FA1	542	1333	900	1333
FA2	400	967	456	833
FA3	500	750	533	1100
FA4	433	533	542	900
FA5	167	483	750	1083
FA6	458	458	533	533
FA7	233	1100	333	1417
FA8	667	817	833	900
FA9	133	500	153	517
FA10	483	483	533	667
FA11	183	833	133	1100
FA12	133	542	208	967
FA13	500	567	500	833
FA14	625	625	900	900
FA15	600	750	792	1000
FA16	483	933	500	867

Fig 4. Determination of GAA repeat length in paired FRDA patient fibroblasts (F) and lymphocytes (L). (**A**) Agarose gel analysis of GAA repeat sizes in fibroblast/lymphocyte samples isolated from the same individual (FRDA patients FA1 - FA16). (**B**) The mean sizes of the GAA1 and GAA2 alleles between all fibroblast and lymphocyte samples (n = 16) were calculated and compared. A p-value <0.05 denotes a significantly significant difference. (C, D) Correlation between the number of GAA repeats expanded in lymphocytes relative to size of the repeat tracts in fibroblasts and the number of GAA repeats in lymphocytes. The difference between the expanded GAA repeat lengths observed in lymphocyte and fibroblast samples (ΔGAA) was plotted against lymphocyte GAA sizes for each of the 16 FRDA paired samples. The analysis was performed for both alleles (**C**) GAA1 and (**D**) GAA2. The Pearson's correlation coefficient (R) is indicated.

Longitudinal analysis of GAA instability

To determine whether lengths of the expanded GAA tracts change over time in FRDA patient blood samples, we performed a longitudinal analysis of repeat size in peripheral lymphocytes. PCR analyses of GAA length were conducted using samples collected from

five patients over a span of 7-9 years (time point I – initial sample, time point II – second sample; Table 3). Differences in the number of repeats between the two time points were detected in all five patients, in at least one allele (Fig 5A). The GAA repeat sizes determined at the initial time point ranged from ~357 to ~1038 repeats (median of ~800). Interestingly, 8 out of 10 alleles exhibited an increase in size of ~10 repeats or greater and none of the analyzed alleles contracted over time (Fig 5B). The maximum detected expansion of 64 GAAs corresponds to a ~15% increase over 7.3 years (F11; Fig 5). The median expansion size over the 7-9 year range was ~32 GAAs (3.3%), indicating a yearly increase in the number of GAAs of ~4 repeats (maximum detected ~9 GAAs/year; Fig 5C). There is a direct correlation between the number of GAA repeats detected in the *FXN* gene at the time of initial sampling and the change in GAA repeat number over time, indicating a trend for longer repeat tracts to expand at a greater rate than shorter tracts (Fig 5D). Overall, these results demonstrate a bias toward GAA repeat expansion in FRDA patient lymphocytes with a propensity to expand that is dependent on the initial size of the GAA tract.

Table 3. Characterization of FRDA patients

Patient	Age of Onset (years)	Interval between I and II [a]
F10	11	7 y, 9 mo
F11	14	7 y, 4 mo
F12	20	8 y, 3 mo
M7	10	8 y, 8 mo
F13	16	7 y, 4 mo

[a] Repeated blood samplings (I and II) were performed at the intervals indicated in the table.

Fig 5. Assessing time-dependent changes of GAA tract length in FRDA lymphocytes. (**A**) The GAA repeat tract at the *FXN* locus was amplified using genomic DNA extracted from lymphocytes, which were isolated from 5 patients (F10, F11, F12, M7 and F13). Blood samples were taken at an initial timepoint (I) and a second time point 7-9 years after the initial sampling (II) (Table 3). The time-dependent changes in GAA repeat length were

quantitated as (**B**) total GAA repeat gain/loss (ΔGAA) and (**C**) rate of the change (ΔGAA per year). (**D**) The Pearson's correlation coefficient was calculated using the size of the GAA tract at time point I [GAA(I)] and the change of the number of GAAs between time points II and I [ΔGAA (II-I)]. Five pairs of samples (n=10 alleles) were analyzed.

Discussion

Friedreich's ataxia presents with considerable variability in the age of onset, symptoms, and progression. A well-established correlation exists between number of GAA repeats in the smaller allele detected at the diagnosis and the age of onset as well as development of select symptoms, with GAA1 being a better predictor of broadly defined disease severity [33]. It has been estimated that an increase of the GAA1 tract by 100 repeats translates to ~2.3 years earlier disease onset [34]. Typically, the association between GAA number and age of onset reaches R values of 0.6 – 0.75[35]. Although strong, this correlation is far from perfect and numerous "outliers" have been observed with FRDA patients lacking certain disease hallmarks or presenting with symptoms unexpectedly early based on the number of GAAs determined at diagnosis. Prior studies demonstrated that GAA lengths estimated from lymphocytes, skin fibroblasts or buccal cells may not reflect exactly the size of expanded repeat tracts in pathologically relevant tissues, such as DRG neurons or spinal cord [16, 18-22]. Although, some aspects of somatic instability of the expanded GAA repeats in FRDA tissues were reported [13-22], no comprehensive or comparative analyses of GAA tract status in larger cohorts and in various tissue samples have been conducted.

In this work we addressed three separate, unanswered questions: (i) Is there a correlation between the sizes of expanded GAA tracts and tissue type, especially tissues primarily affected in the disease? (ii) Is the number of GAA repeats identical in different

tissues? (iii) Do contractions or expansions of the GAA tracts accumulate over time in tissues? To answer these questions, we thoroughly characterized GAA tracts in a large number of different FRDA and control tissue samples.

Analyses of the GAA repeat region in pancreas, heart, cerebral cortex, cerebellar cortex and spinal cord isolated from 15 different FRDA patients demonstrated significant somatic mosaicism. Unexpectedly, we detected significantly longer GAA tracts in the heart and pancreas when compared to DNA samples isolated from central nervous system tissues. Considering that DNA replication, repair and transcription are three major molecular processes involved in the stimulation or prevention of repeat instability, tissue specific differences in lengths of the GAA tracts can be expected. In fact, activity of the mismatch repair (MMR) system is considered a major factor contributing to GAA repeat instability in human cells. expansions of the GAA repeats in iPSC, reprogrammed from patient fibroblasts, have been correlated to an increase in mismatch repair enzymes when compared to fibroblast cells [36, 37] and overexpression of MMR components stimulated GAA expansions in FRDA fibroblasts [38]. Thus, a combination of the number of replication cycles, tissue-specific transcriptional activity of the *FXN* gene and differences in the activity of various DNA repair systems could be responsible for the significantly longer GAAs detected in pancreas and heart tissues. Interestingly, quantitative Western blot analyses of frataxin expression in different tissues from FRDA patients and unaffected controls indicate a remarkably low level of frataxin in FRDA cardiac tissue relative to nervous system tissue (S1 Fig). Developmentally regulated, tissue specific cell division and replication patterns, aggravated by conditions of decreased frataxin levels, may facilitate expansion of the GAA repeats.

Transcriptome analyses conducted using FRDA mouse models demonstrated a global decrease of gene expression in frataxin deficient tissues [39]. Also, recent studies conducted in our laboratory showed that a subset of genes involved in transcription, translation and DNA repair is downregulated in FRDA cells [40]. Decreased expression of these genes may enable somatic instability of the GAAs in a tissue-specific manner. Somatic expansions that further decrease frataxin levels can in turn reinforce global transcriptome defects and perpetuate expansions of the GAA tract feeding a "somatic instability cycle". Furthermore, difficulties of handling intracellular reactive oxygen species (ROS), characteristic of frataxin deficient cells, may stimulate DNA damage, further augmenting GAA instability, similar to the "toxic oxidation cycle" proposed to explain somatic instability of CAG repeats in Huntington's disease (HD) [41]. Considering an important role of ROS in FRDA pathogenesis, the use of potent mitochondrial ROS scavengers inhibited somatic instability of CAG repeats and delayed phenotype development in the HD mouse model, and may also inhibit progressive somatic expansions of the GAAs in FRDA [41].

Interestingly, results of the longitudinal analyses of the GAA tract in lymphocytes, although conducted on a relatively small pool of expanded alleles, demonstrated a strong, length-dependent bias towards repeat expansions. Eight of 10 alleles analyzed increased in GAA length by up to 64 repeats over a period of 7 – 9 years. This result contrasts with prior in vitro studies on immortalized lymphoblast cells where a predominance for contractions, frequently large, was observed [24]. In fact, in vitro culturing of FRDA cells or model cell lines harboring expanded GAAs results in shortening of the expanded repeats with a notable exception of iPSCs where continuous expansion of the GAA tracts was

observed [25, 26]. This parallel between the FRDA iPSCs and peripheral blood lymphocytes suggest that the specificity of the pluripotent/multipotent state (iPSC and hematopoietic stem cells or lymphoid progenitors) may facilitate GAA expansions. It is difficult to ascertain whether the longer GAA tracts detected in lymphocytes result from progressive expansions of the repeats in this lineage or from contractions in the development and culturing of skin fibroblasts (or both processes at the same time). Results of the longitudinal analyses described in this work strongly favor progressive expansions in lymphocytes or their progenitors. Additionally, to date we have never detected contractions during prolonged culture of FRDA primary skin fibroblasts. The inherent characteristics of the stem cells, including chromatin status, replicative properties, transcriptional and repair activity may contribute to the expansion bias observed in these cells.

Results of our longitudinal studies also indicate that comparative analyses of the GAA length between patients need to be interpreted cautiously as expanded tracts change dynamically over a patient's lifetime. Considering the extreme case of a 15% increase over 7 years (Fig 5), substantial discrepancies in the number of GAAs can exist between a diagnosis made early in life and adulthood. Frequently, symptom correlation studies include the lengths of the GAA1 and GAA2 alleles. Furthermore, enrollment criteria for clinical trials also sometimes consider the number of GAA repeats. Therefore, periodic follow-up determinations of GAA length should be conducted. These data may also be directly beneficial to the FRDA patient community. The GAA repeat number is a frequent topic of discussion during meetings with FRDA patients, both between health professionals and patients as well as among patients themselves. FRDA patients and caregivers

frequently directly associate the repeat tract size with differences in symptoms and disease progression. Thus, additional factors affecting repeat length, such as tissue-specific instability or timing of the molecular diagnosis of the disease as described in this study, may help explain imperfect relationships that sometimes exist between disease status and the number of GAAs detected in blood samples at diagnosis. It is also important to consider secondary genetic modifiers that can contribute to age of onset and disease symptoms, for example, the milder FRDA presentation observed in the Acadian population when compared to patients of other ethnic origins carrying GAA repeat tracts of similar size [22] or earlier age of onset associated with heterozygosity for p.C282Y variant in hemochromatosis (*HFE*) gene [42].

Length polymorphism and somatic instability of the expanded GAA repeat tracts in different tissues represent the first layer of complexity responsible for differences in frataxin expression that can translate to clinical variability between FRDA patients. It has been estimated that the length of the repeat tract accounts for only 36% of the variability in the age of onset with other contributing factors being genetic, epigenetic, or environmental in nature [34]. The *FXN* GAA tract evolved from G/A rich sequences of the ancient Alu elements. It is highly likely that long GAA sequences constantly evolve and change, resulting in introduction of point mutations (interruptions) into the pure GAA motifs. Examples of such interruptions have been recurrently reported in the literature but their actual detailed structure and frequency in *FXN* alleles is unknown. Disturbance of the GAA repeat purity can affect not only expression of the *FXN* gene but also can significantly influence instability of the repeat tract, especially the propensity for continuous expansions.

Although technically challenging, future studies addressing the role of GAA tract interruptions in FRDA heterogeneity are necessary.

Conclusions

Expanded GAA repeats are unstable in FRDA, and expansion bias detected in heart and pancreas is likely to contribute to symptom development and disease progression, making the mechanism of somatic instability an important target for therapy. Differences in the size of the GAA tracts between lymphocytes and fibroblasts urge caution in direct comparisons of data obtained using different model systems. Continuous GAA expansions over time as observed in FRDA lymphocytes should be considered in clinical trials design and data analyses. The kinetics of longitudinal changes in GAA repeat number and potential *cis* and *trans* acting factors affecting this process need to be further investigated.

List of Abbreviations

Cc: Cerebral cortex; Cb: Cerebellar cortex; Ctl: Control F: Female; FRDA: Friedreich's ataxia; FXN: Frataxin; GAA: Guanine-adenine-adenine; GAA1: Shorter guanine-adenine-adenine allele; GAA2: Longer guanine-adenine-adenine allele; GAPDH: Glyceraldehyde 3-phosphate dehydrogenase; H: Heart; HD: Huntington's disease; HSP60: Heat shock protein 60; iPSCs: Induced pluripotent stem cells; M: Male; P: pancreas; PCR: Polymerase chain reaction; ROS: Reactive oxygen species; Sc: Spinal cord; SD: Standard deviation

Acknowledgements

The authors thank all FRDA patients for donating blood samples and skin biopsies.

Authors' contributions

A.L. and U.P. conducted experiments. A.K., L.H. and D.R.L. provided FRDA patient samples. A.L., J.S.N. and M.N. designed the project and wrote the manuscript. All authors read, contributing comments and suggestions, and approved the final manuscript.

References

1. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science. 1996;271(5254):1423-7. PubMed PMID: 8596916.
2. Al-Mahdawi S, Pinto RM, Ismail O, Varshney D, Lymperi S, Sandi C, et al. The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. Hum Mol Genet. 2008;17(5):735-46. doi: 10.1093/hmg/ddm346. PubMed PMID: 18045775.
3. Herman D, Jenssen K, Burnett R, Soragni E, Perlman SL, Gottesfeld JM. Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. Nat Chem Biol. 2006;2(10):551-8. doi: 10.1038/nchembio815. PubMed PMID: 16921367.
4. Saveliev A, Everett C, Sharpe T, Webster Z, Festenstein R. DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. Nature. 2003;422(6934):909-13. doi: 10.1038/nature01596. PubMed PMID: 12712207.
5. Bidichandani SI, Ashizawa T, Patel PI. The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am J Hum Genet. 1998;62(1):111-21. doi: 10.1086/301680. PubMed PMID: 9443873; PubMed Central PMCID: PMCPMC1376805.
6. Campuzano V, Montermini L, Lutz Y, Cova L, Hindelang C, Jiralerspong S, et al. Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. Hum Mol Genet. 1997;6(11):1771-80. PubMed PMID: 9302253.
7. Harding AE. Friedreich's ataxia: a clinical and genetic study of 90 families with an analysis of early diagnostic criteria and intrafamilial clustering of clinical features. Brain. 1981;104(3):589-620. PubMed PMID: 7272714.
8. Pandolfo M. Friedreich ataxia: the clinical picture. J Neurol. 2009;256 Suppl 1:3-8. doi: 10.1007/s00415-009-1002-3. PubMed PMID: 19283344.
9. Delatycki MB, Corben LA. Clinical features of Friedreich ataxia. J Child Neurol. 2012;27(9):1133-7. doi: 10.1177/0883073812448230. PubMed PMID: 22752493; PubMed Central PMCID: PMCPMC3674491.
10. Tsou AY, Paulsen EK, Lagedrost SJ, Perlman SL, Mathews KD, Wilmot GR, et al. Mortality in Friedreich ataxia. J Neurol Sci. 2011;307(1-2):46-9. doi: 10.1016/j.jns.2011.05.023. PubMed PMID: 21652007.
11. Cossee M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel JL, et al. Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proc Natl Acad Sci U S A. 1997;94(14):7452-7. PubMed PMID: 9207112; PubMed Central PMCID: PMCPMC23842.
12. Montermini L, Andermann E, Labuda M, Richter A, Pandolfo M, Cavalcanti F, et al. The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. Hum Mol Genet. 1997;6(8):1261-6. PubMed PMID: 9259271.

13. Pollard LM, Sharma R, Gomez M, Shah S, Delatycki MB, Pianese L, et al. Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic Acids Res. 2004;32(19):5962-71. doi: 10.1093/nar/gkh933. PubMed PMID: 15534367; PubMed Central PMCID: PMCPMC528813.
14. Sharma R, Bhatti S, Gomez M, Clark RM, Murray C, Ashizawa T, et al. The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions. Hum Mol Genet. 2002;11(18):2175-87. PubMed PMID: 12189170.
15. De Michele G, Cavalcanti F, Criscuolo C, Pianese L, Monticelli A, Filla A, et al. Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: pedigree studies and analysis of sperm from patients with Friedreich's ataxia. Hum Mol Genet. 1998;7(12):1901-6. PubMed PMID: 9811933.
16. Bidichandani SI, Garcia CA, Patel PI, Dimachkie MM. Very late-onset Friedreich ataxia despite large GAA triplet repeat expansions. Arch Neurol. 2000;57(2):246-51. PubMed PMID: 10681084.
17. Clark RM, De Biase I, Malykhina AP, Al-Mahdawi S, Pook M, Bidichandani SI. The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. Hum Genet. 2007;120(5):633-40. doi: 10.1007/s00439-006-0249-3. PubMed PMID: 17024371.
18. De Biase I, Rasmussen A, Endres D, Al-Mahdawi S, Monticelli A, Cocozza S, et al. Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. Ann Neurol. 2007;61(1):55-60. doi: 10.1002/ana.21052. PubMed PMID: 17262846.
19. De Biase I, Rasmussen A, Monticelli A, Al-Mahdawi S, Pook M, Cocozza S, et al. Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. Genomics. 2007;90(1):1-5. doi: 10.1016/j.ygeno.2007.04.001. PubMed PMID: 17498922.
20. Machkhas H, Bidichandani SI, Patel PI, Harati Y. A mild case of Friedreich ataxia: lymphocyte and sural nerve analysis for GAA repeat length reveals somatic mosaicism. Muscle Nerve. 1998;21(3):390-3. PubMed PMID: 9486868.
21. Montermini L, Kish SJ, Jiralerspong S, Lamarche JB, Pandolfo M. Somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions in the central nervous system. Neurology. 1997;49(2):606-10. PubMed PMID: 9270608.
22. Montermini L, Richter A, Morgan K, Justice CM, Julien D, Castellotti B, et al. Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. Ann Neurol. 1997;41(5):675-82. doi: 10.1002/ana.410410518. PubMed PMID: 9153531.
23. Anjomani Virmouni S, Ezzatizadeh V, Sandi C, Sandi M, Al-Mahdawi S, Chutake Y, et al. A novel GAA-repeat-expansion-based mouse model of Friedreich's ataxia. Dis Model Mech. 2015;8(3):225-35. doi: 10.1242/dmm.018952. PubMed PMID: 25681319; PubMed Central PMCID: PMCPMC4348561.
24. Bidichandani SI, Purandare SM, Taylor EE, Gumin G, Machkhas H, Harati Y, et al. Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence. Hum Mol Genet. 1999;8(13):2425-36. PubMed PMID: 10556290.

25. Ku S, Soragni E, Campau E, Thomas EA, Altun G, Laurent LC, et al. Friedreich's ataxia induced pluripotent stem cells model intergenerational GAATTC triplet repeat instability. Cell Stem Cell. 2010;7(5):631-7. doi: 10.1016/j.stem.2010.09.014. PubMed PMID: 21040903; PubMed Central PMCID: PMCPMC2987635.
26. Polak U, Li Y, Butler JS, Napierala M. Alleviating GAA Repeat Induced Transcriptional Silencing of the Friedreich's Ataxia Gene During Somatic Cell Reprogramming. Stem Cells Dev. 2016;25(23):1788-800. doi: 10.1089/scd.2016.0147. PubMed PMID: 27615158; PubMed Central PMCID: PMCPMC5155629.
27. Li Y, Polak U, Clark AD, Bhalla AD, Chen YY, Li J, et al. Establishment and Maintenance of Primary Fibroblast Repositories for Rare Diseases-Friedreich's Ataxia Example. Biopreserv Biobank. 2016;14(4):324-9. doi: 10.1089/bio.2015.0117. PubMed PMID: 27002638; PubMed Central PMCID: PMCPMC4991587.
28. Li Y, Lu Y, Polak U, Lin K, Shen J, Farmer J, et al. Expanded GAA repeats impede transcription elongation through the FXN gene and induce transcriptional silencing that is restricted to the FXN locus. Hum Mol Genet. 2015;24(24):6932-43. doi: 10.1093/hmg/ddv397. PubMed PMID: 26401053; PubMed Central PMCID: PMCPMC4654050.
29. Li Y, Polak U, Bhalla AD, Rozwadowska N, Butler JS, Lynch DR, et al. Excision of Expanded GAA Repeats Alleviates the Molecular Phenotype of Friedreich's Ataxia. Mol Ther. 2015;23(6):1055-65. doi: 10.1038/mt.2015.41. PubMed PMID: 25758173; PubMed Central PMCID: PMCPMC4817761.
30. Clark RM, Bhaskar SS, Miyahara M, Dalgliesh GL, Bidichandani SI. Expansion of GAA trinucleotide repeats in mammals. Genomics. 2006;87(1):57-67. doi: 10.1016/j.ygeno.2005.09.006. PubMed PMID: 16316739.
31. P MR, Clark RM, Pollard LM, De Biase I, Bidichandani SI. Replication in mammalian cells recapitulates the locus-specific differences in somatic instability of genomic GAA triplet-repeats. Nucleic Acids Res. 2006;34(21):6352-61. doi: 10.1093/nar/gkl846. PubMed PMID: 17142224; PubMed Central PMCID: PMCPMC1669776.
32. Kim E, Napierala M, Dent SY. Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. Nucleic Acids Res. 2011;39(19):8366-77. doi: 10.1093/nar/gkr542. PubMed PMID: 21745819; PubMed Central PMCID: PMCPMC3201871.
33. Mateo I, Llorca J, Volpini V, Corral J, Berciano J, Combarros O. Expanded GAA repeats and clinical variation in Friedreich's ataxia. Acta Neurol Scand. 2004;109(1):75-8. PubMed PMID: 14653855.
34. Reetz K, Dogan I, Costa AS, Dafotakis M, Fedosov K, Giunti P, et al. Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: a cross-sectional analysis of baseline data. Lancet Neurol. 2015;14(2):174-82. doi: 10.1016/S1474-4422(14)70321-7. PubMed PMID: 25566998.
35. Lynch DR, Perlman SL. Friedreich's ataxia: the European consortium. Lancet Neurol. 2015;14(2):130-1. doi: 10.1016/S1474-4422(14)70327-8. PubMed PMID: 25566997.
36. Du J, Campau E, Soragni E, Ku S, Puckett JW, Dervan PB, et al. Role of mismatch repair enzymes in GAA.TTC triplet-repeat expansion in Friedreich ataxia induced

pluripotent stem cells. J Biol Chem. 2012;287(35):29861-72. doi: 10.1074/jbc.M112.391961. PubMed PMID: 22798143; PubMed Central PMCID: PMCPMC3436184.
37. Hick A, Wattenhofer-Donze M, Chintawar S, Tropel P, Simard JP, Vaucamps N, et al. Neurons and cardiomyocytes derived from induced pluripotent stem cells as a model for mitochondrial defects in Friedreich's ataxia. Dis Model Mech. 2013;6(3):608-21. doi: 10.1242/dmm.010900. PubMed PMID: 23136396; PubMed Central PMCID: PMCPMC3634645.
38. Halabi A, Ditch S, Wang J, Grabczyk E. DNA mismatch repair complex MutSbeta promotes GAA.TTC repeat expansion in human cells. J Biol Chem. 2012;287(35):29958-67. doi: 10.1074/jbc.M112.356758. PubMed PMID: 22787155; PubMed Central PMCID: PMCPMC3436174.
39. Coppola G, Marmolino D, Lu D, Wang Q, Cnop M, Rai M, et al. Functional genomic analysis of frataxin deficiency reveals tissue-specific alterations and identifies the PPARgamma pathway as a therapeutic target in Friedreich's ataxia. Hum Mol Genet. 2009;18(13):2452-61. doi: 10.1093/hmg/ddp183. PubMed PMID: 19376812; PubMed Central PMCID: PMCPMC2694693.
40. Napierala JS, Li Y, Lu Y, Lin K, Hauser LA, Lynch DR, et al. Comprehensive analysis of gene expression patterns in Friedreich's ataxia fibroblasts by RNA sequencing reveals altered levels of protein synthesis factors and solute carriers. Dis Model Mech. 2017;10(11):1353-69. doi: 10.1242/dmm.030536. PubMed PMID: 29125828.
41. Budworth H, Harris FR, Williams P, Lee DY, Holt A, Pahnke J, et al. Suppression of Somatic Expansion Delays the Onset of Pathophysiology in a Mouse Model of Huntington's Disease. PLoS Genet. 2015;11(8):e1005267. doi: 10.1371/journal.pgen.1005267. PubMed PMID: 26247199; PubMed Central PMCID: PMCPMC4527696.
42. Delatycki MB, Tai G, Corben L, Yiu EM, Evans-Galea MV, Stephenson SE, et al. HFE p.C282Y heterozygosity is associated with earlier disease onset in Friedreich ataxia. Mov Disord. 2014;29(7):940-3. doi: 10.1002/mds.25795. PubMed PMID: 24390816.

Supporting Information

S1 Fig. Quantitative western blot analyses of frataxin expression. Western blots probed for FXN were performed on (**A,D**) heart, (**B,E**) cerebral cortex, and (**C,F**) cerebellar cortex tissues. Frataxin expression values were normalized using the mitochondrial HSP60 protein. Patient F9 exhibited very low HSP60 expression in cerebellar cortex tissue, therefore GAPDH was used for normalization.

S2 Fig. Instability of the expanded GAAs in the *FXN* gene in different tissues. Genomic DNA was extracted from heart (H), cerebral cortex (Cc), spinal cord (Sc), cerebellar cortex (Cb) and pancreas (P) tissues. Results of PCR analyses of GAA repeat length in FRDA patient tissues (top panel). Analyses of GAA repeat instability in the 5q23 locus in patient tissues (middle panel). Amplification of an intron 1 - exon 2 fragment of the *FXN* gene downstream of the GAA tract as control for genomic DNA quality (bottom panel). (-) represents no-template control and (+) represents positive control (genomic DNA isolated from FRDA fibroblasts). (**A**) F1, (**B**) F4, (**C**) F5, (**D**) F3, (**E**) F6, (**F**) F8, (**G**) F9, (**H**) M1, (**I**) M2, (**J**) M3, (**K**) M5, and (**L**) M4 as described in Table 1.

S3 Fig. Quantitative analysis of expanded GAA instability in the *FXN* gene in different tissues. The expanded GAA repeats in the *FXN* gene were amplified from genomic DNA extracted from heart (H), cerebral cortex (Cc), spinal cord (Sc), cerebellar cortex (Cb) and pancreas (P) tissues isolated from FRDA patients. The band intensity of the PCR products along with the repeat sizes are shown. Solid vertical lines represent the mean of GAA repeat sizes detected across all tissues analyzed. (**A**) F1, (**B**) F4, (**C**) F5, (**D**)

F3, (**E**) F6, (**F**) F8, (**G**) F9, (**H**) M1, (**I**) M2, (**J**) M3, (**K**) M5, and (**L**) M4 as described in Table 1.

S4 Fig. Repeat PCR analysis of the expanded GAAs in the *FXN* gene in different tissues. The expanded GAA repeats in the *FXN* gene were amplified from genomic DNA extracted from heart (H), cerebral cortex (Cc), spinal cord (Sc), cerebellar cortex (Cb) and pancreas (P) tissues isolated from FRDA patients in two independent reactions. Analyses of GAA repeat instability in tissues of FRDA patient F2; (A) experiment 1, (B) experiment 2 and in tissues fo FRDA patient M5 (C) experiment 1, (D) experiment 2. Sizes of the individual PCR products were calculated for each tissue and experiment. No significant differences were observed between sizes of the GAA tracts determined in two independent experiments (t-test, $p > 0.05$).

S5 Fig. GAA repeat tracts in the *FXN* gene of unaffected individuals show no instability. PCR analysis of the GAA repeat region in the *FXN* gene using genomic DNA extracted from the cerebral cortex (Cc) and cerebellar cortex (Cb) tissues of unaffected individuals. (-) represents no-template control and (+) represents positive control (genomic DNA isolated from control fibroblasts).

Figure 1

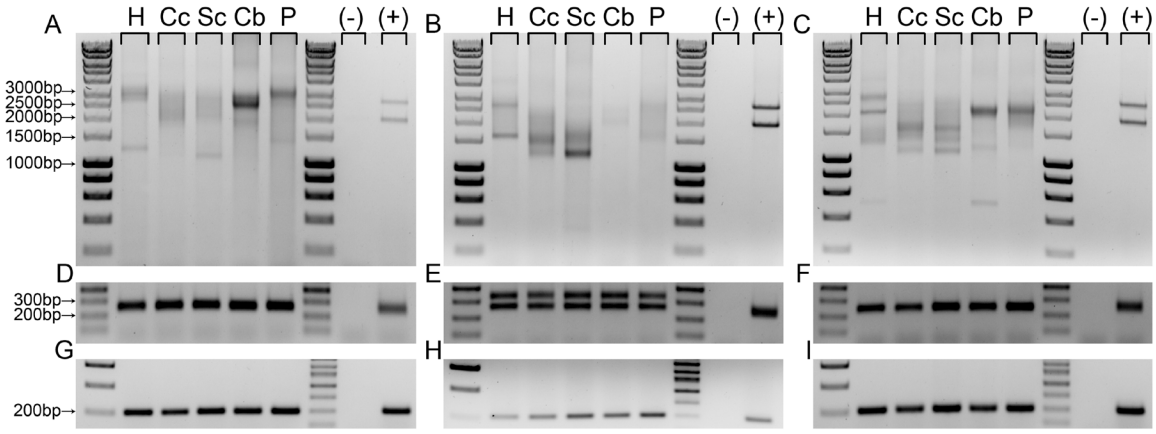

A
H Cc Sc Cb P (-) (+)
3000bp→
2500bp→
2000bp→
1500bp→
1000bp→
B
H Cc Sc Cb P (-) (+)
C
H Cc Sc Cb P (-) (+)
D
300bp→
200bp→
E
F
G
200bp→
H
I

Figure 2

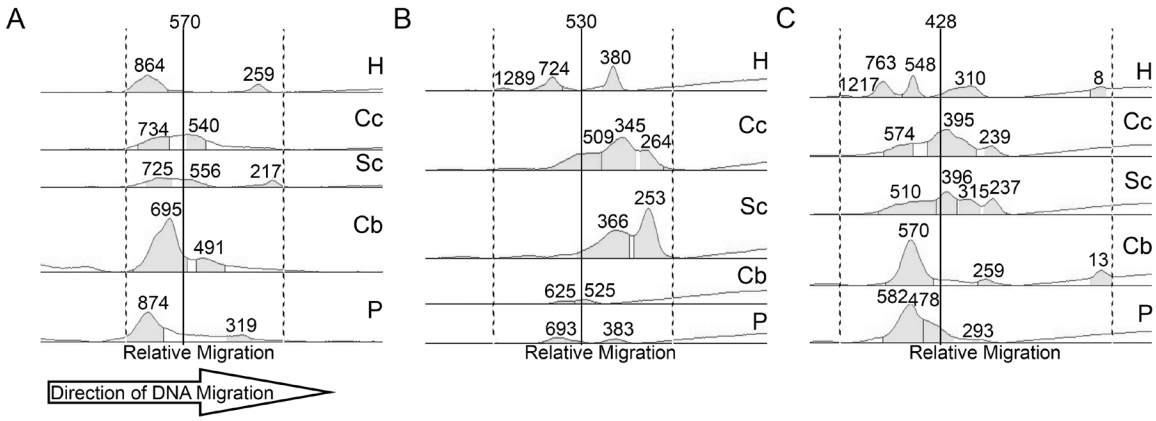

A
570
864
259
H
734
540
Cc
725
556
217
Sc
695
491
Cb
874
319
P
Relative Migration
Direction of DNA Migration
B
530
1289
724
380
H
509
345
264
Cc
253
366
Sc
625
525
Cb
693
383
P
Relative Migration
C
428
1217
763
548
310
8
H
574
395
239
Cc
510
396
315
237
Sc
570
259
13
Cb
582
478
293
P
Relative Migration

Figure 3

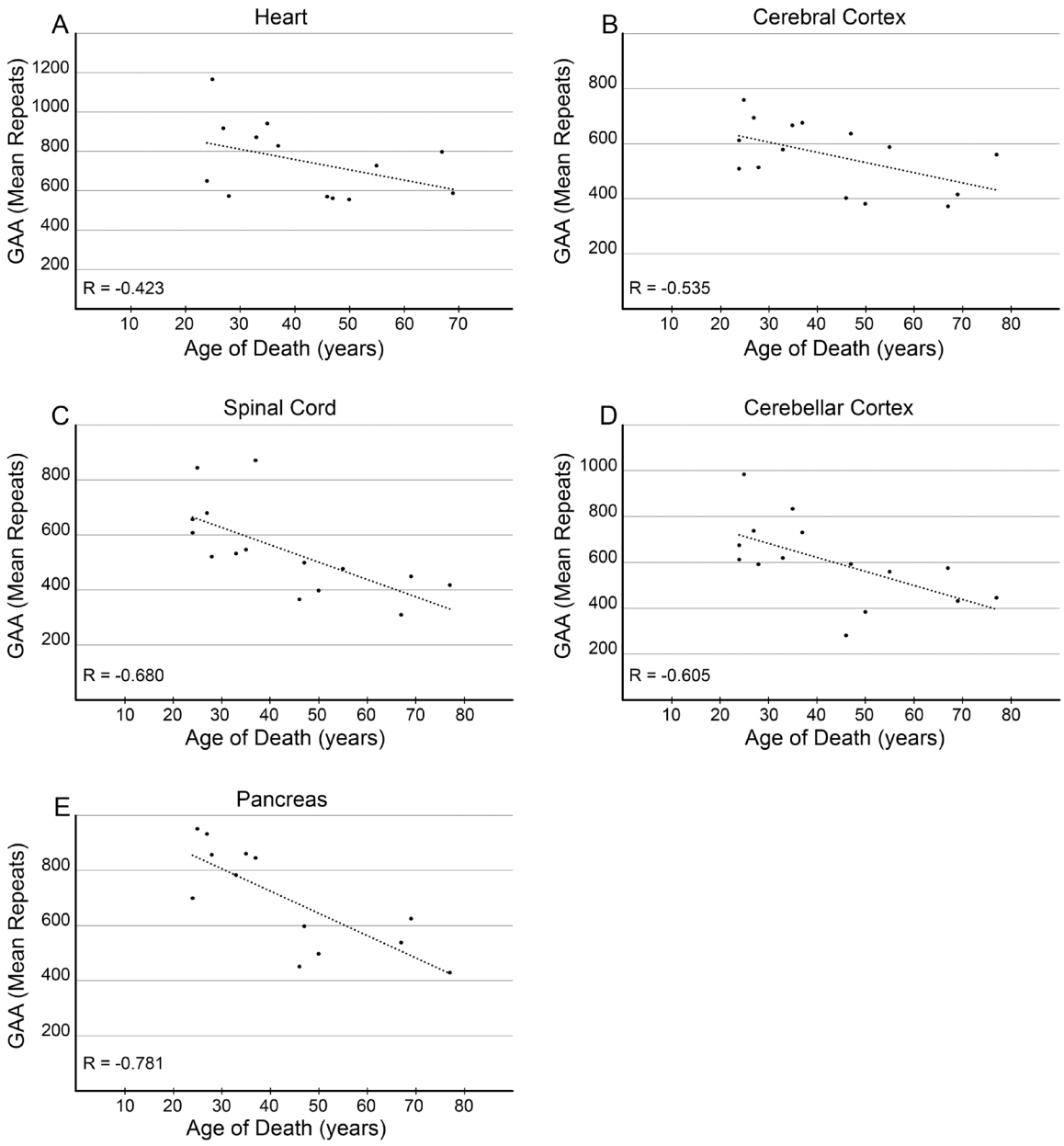

A
Heart
GAA (Mean Repeats)
1200
1000
800
600
400
200
R = -0.423
10 20 30 40 50 60 70
Age of Death (years)
B
Cerebral Cortex
GAA (Mean Repeats)
800
600
400
200
R = -0.535
10 20 30 40 50 60 70 80
Age of Death (years)
C
Spinal Cord
GAA (Mean Repeats)
800
600
400
200
R = -0.680
10 20 30 40 50 60 70 80
Age of Death (years)
D
Cerebellar Cortex
GAA (Mean Repeats)
1000
800
600
400
200
R = -0.605
10 20 30 40 50 60 70 80
Age of Death (years)
E
Pancreas
GAA (Mean Repeats)
800
600
400
200
R = -0.781
10 20 30 40 50 60 70 80
Age of Death (years)

Figure 4

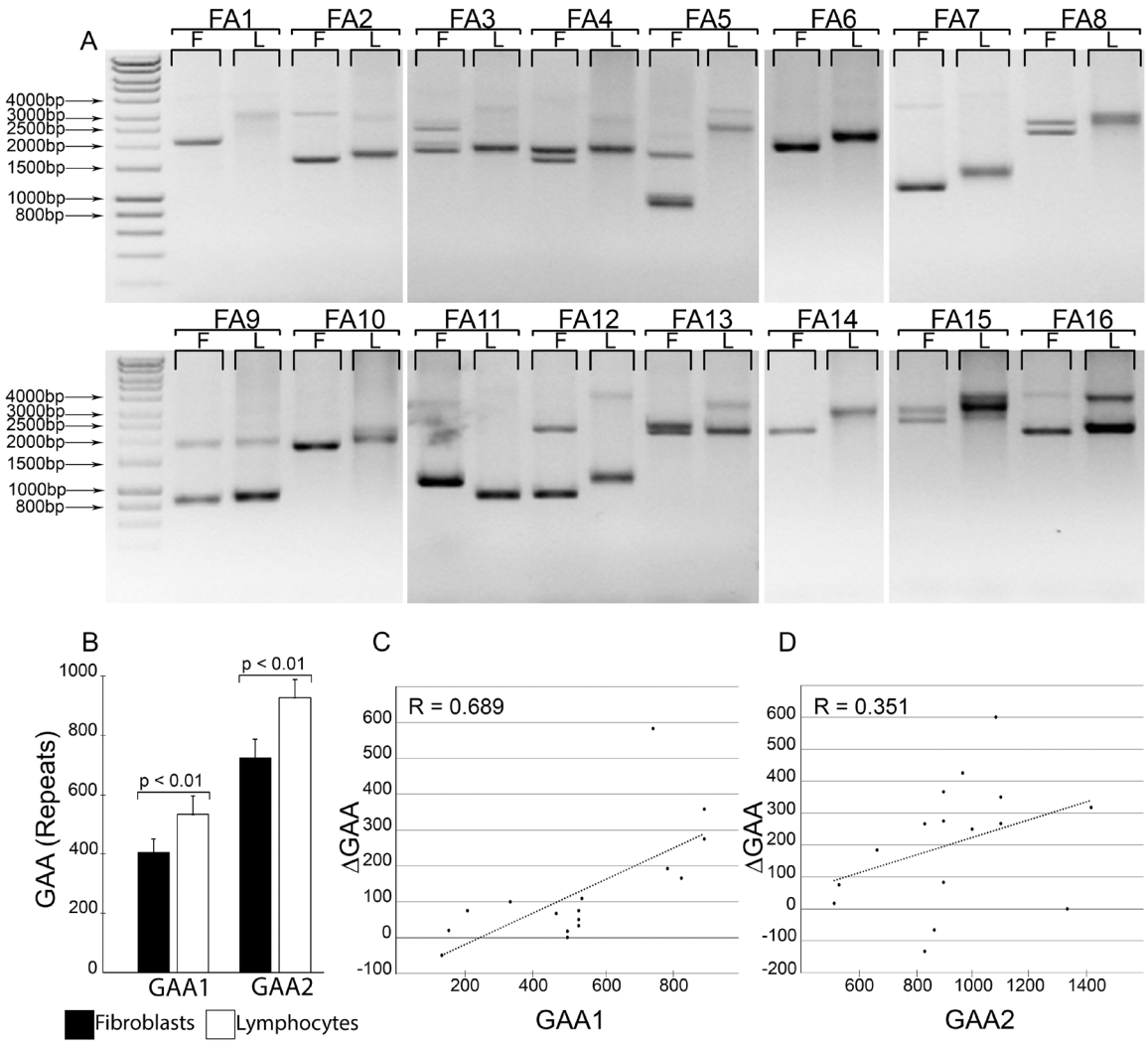
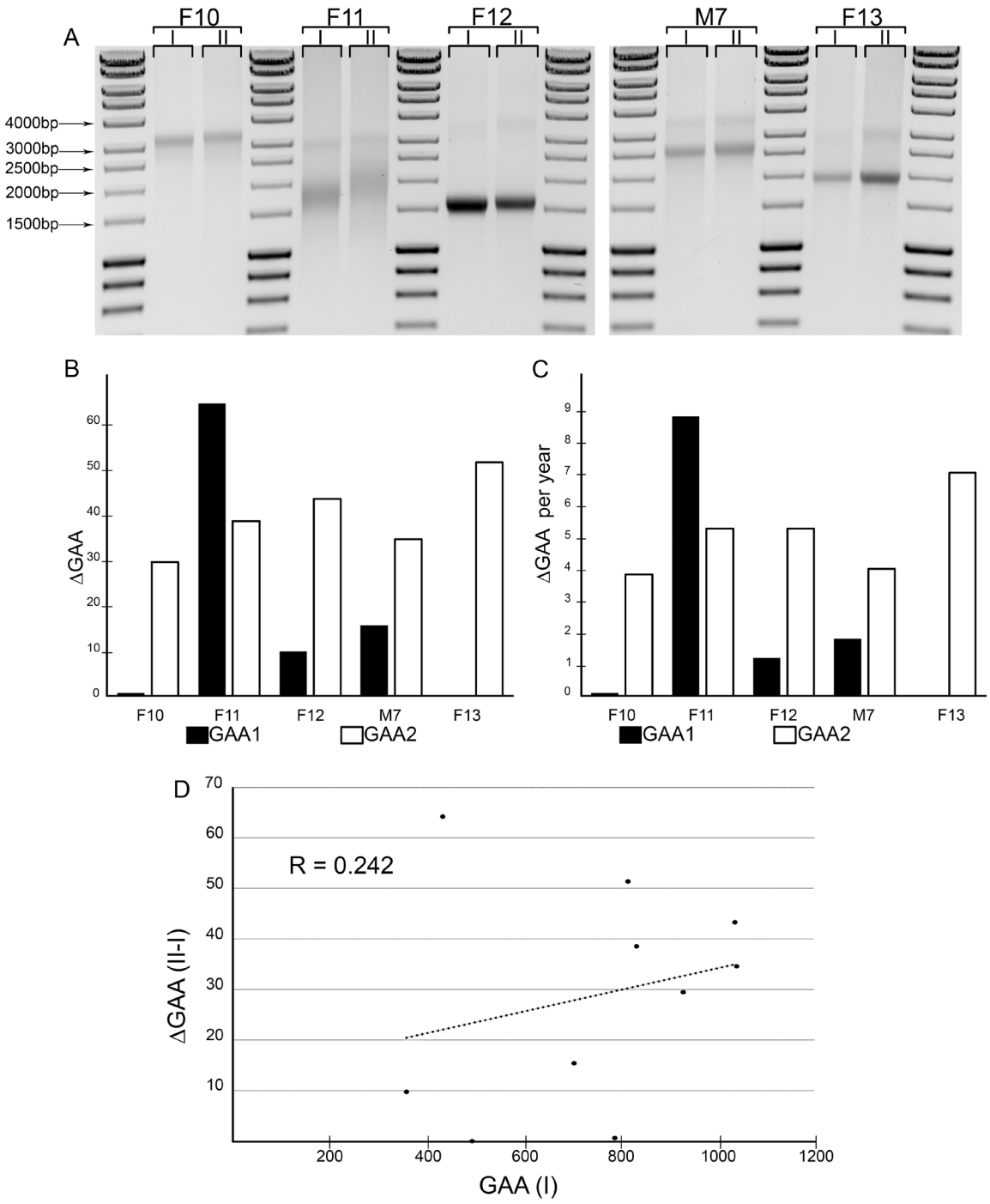
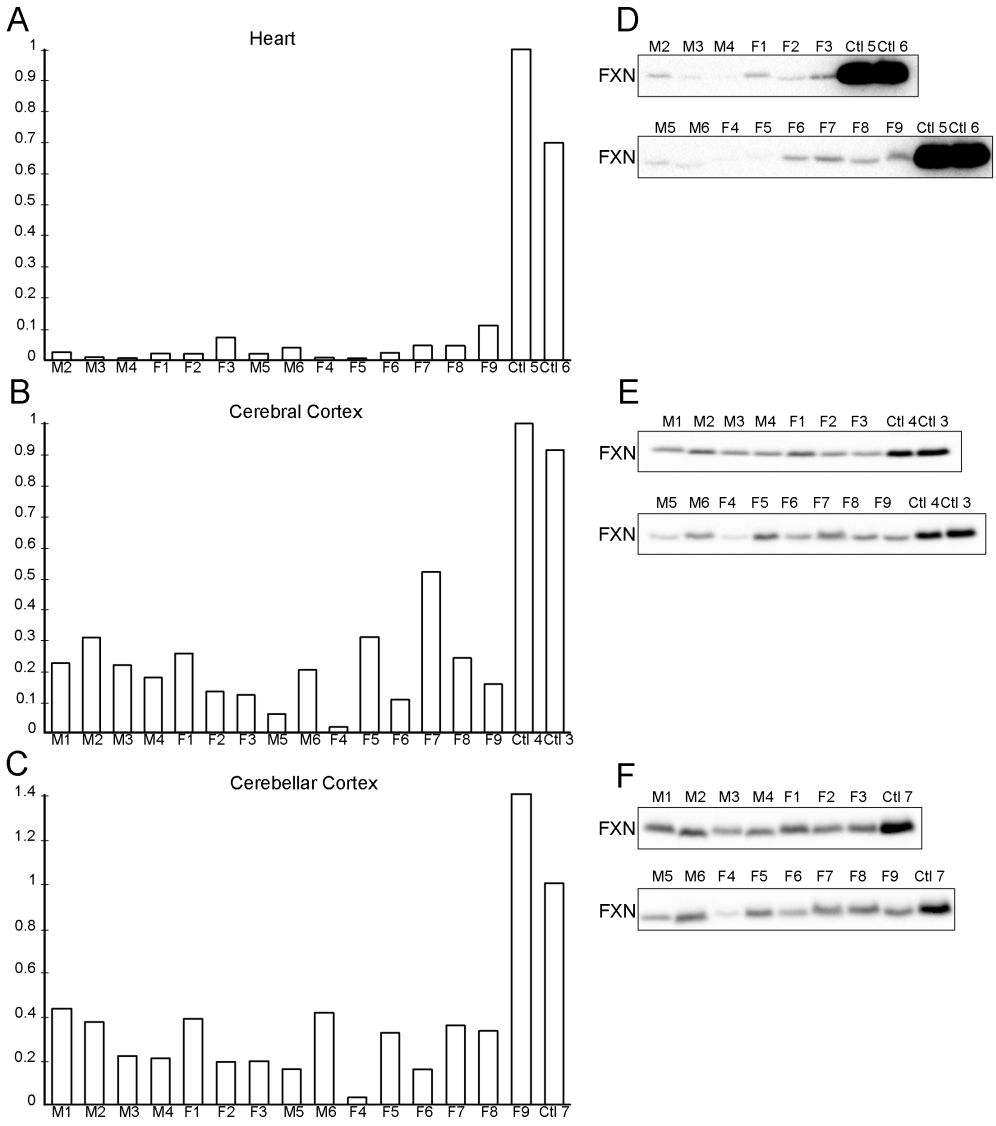

A
FA1
FA2
FA3
FA4
FA5
FA6
FA7
FA8
F
L
4000bp
3000bp
2500bp
2000bp
1500bp
1000bp
800bp
FA9
FA10
FA11
FA12
FA13
FA14
FA15
FA16
B
GAA (Repeats)
p < 0.01
p < 0.01
GAA1
GAA2
Fibroblasts
Lymphocytes
C
R = 0.689
ΔGAA
GAA1
D
R = 0.351
ΔGAA
GAA2

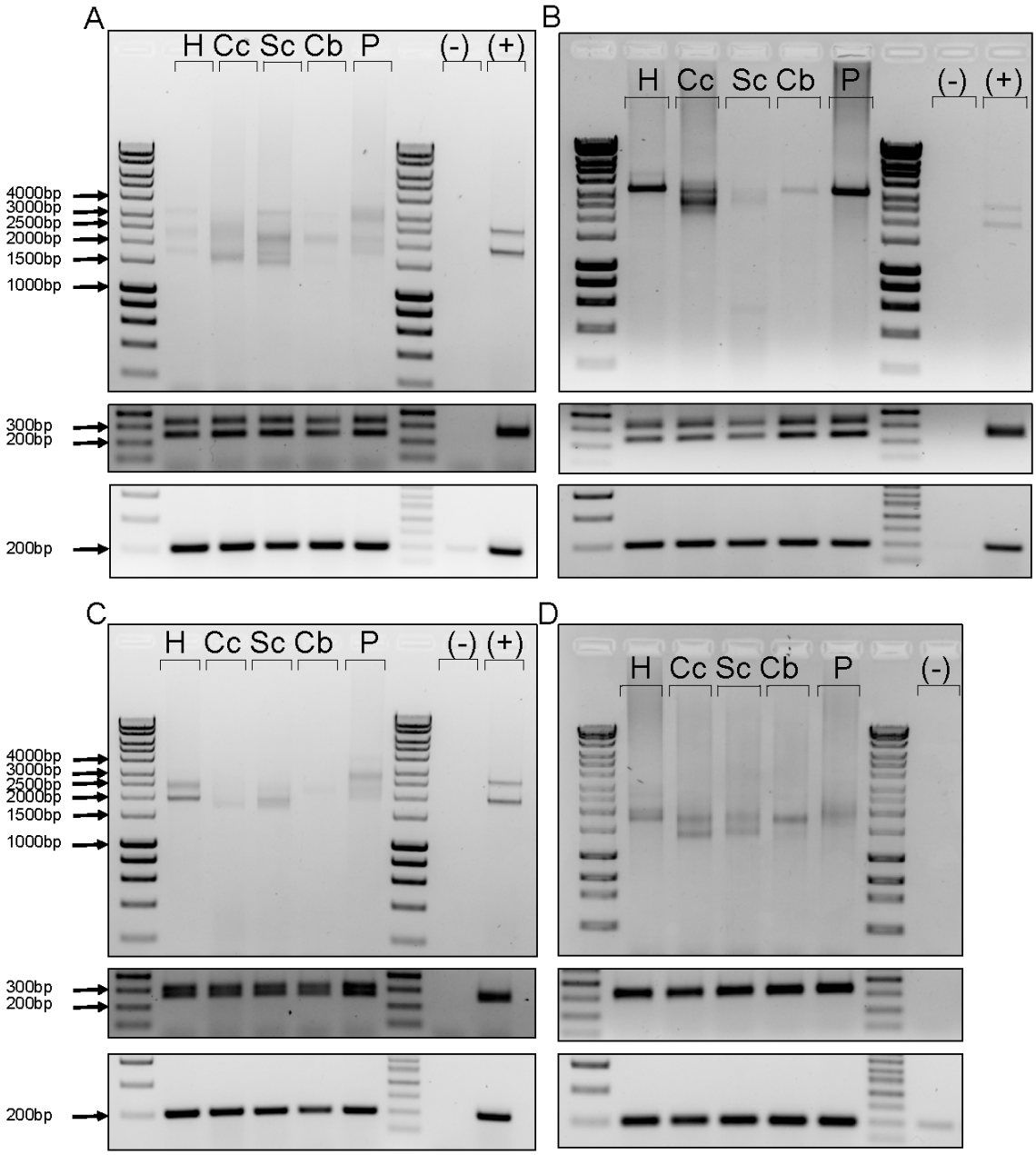
Figure 5


A
F10
F11
F12
M7
F13
I
II
4000bp
3000bp
2500bp
2000bp
1500bp
B
ΔGAA
F10
F11
F12
M7
F13
GAA1
GAA2
C
ΔGAA per year
F10
F11
F12
M7
F13
GAA1
GAA2
D
R = 0.242
ΔGAA (II-I)
GAA (I)

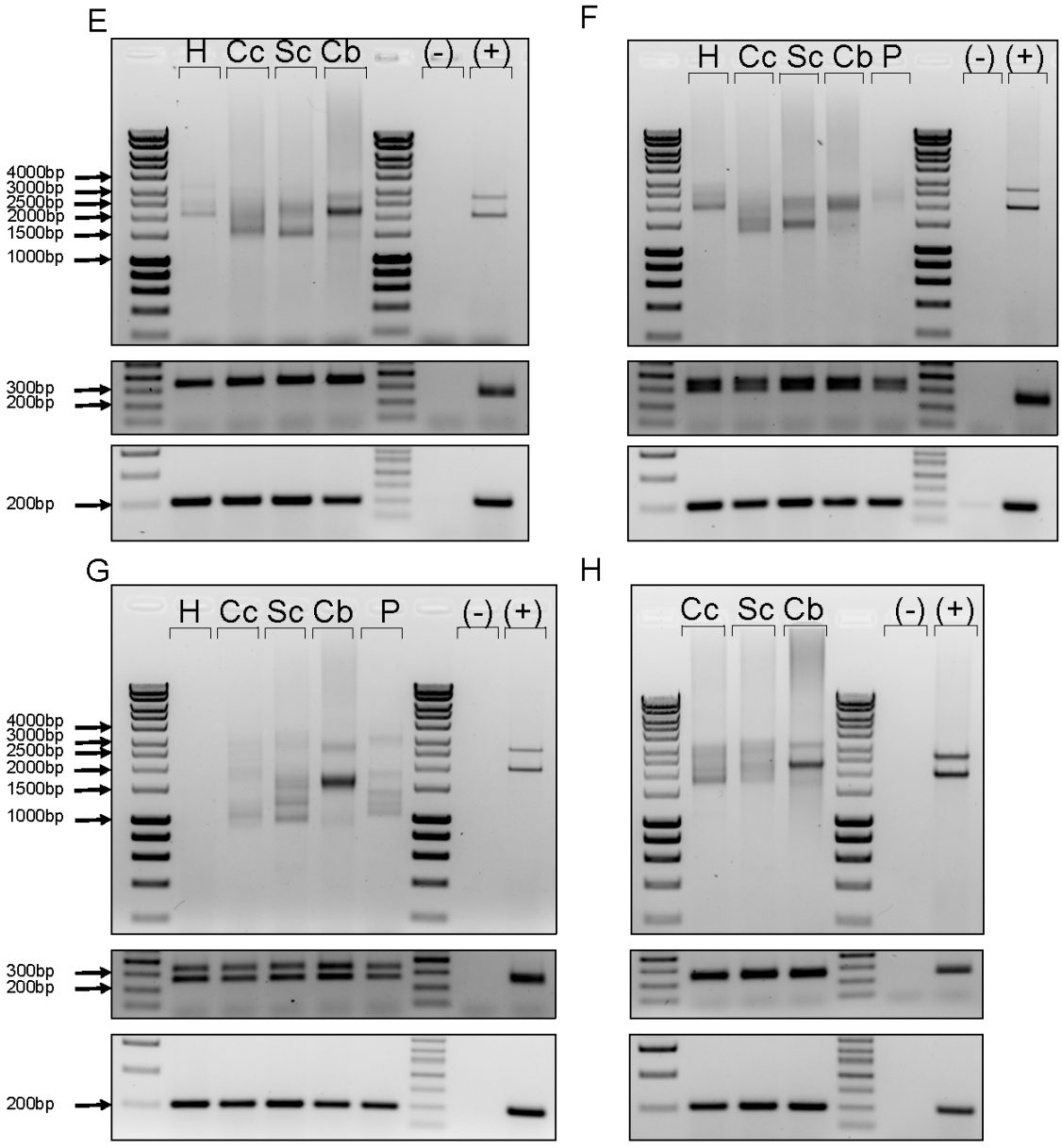
S1 Fig.


A
Heart
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
M2 M3 M4 F1 F2 F3 M5 M6 F4 F5 F6 F7 F8 F9 Ctl 5 Ctl 6
B
Cerebral Cortex
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
M1 M2 M3 M4 F1 F2 F3 M5 M6 F4 F5 F6 F7 F8 F9 Ctl 4 Ctl 3
C
Cerebellar Cortex
1.4
1.2
1
0.8
0.6
0.4
0.2
0
M1 M2 M3 M4 F1 F2 F3 M5 M6 F4 F5 F6 F7 F8 F9 Ctl 7
D
M2 M3 M4 F1 F2 F3 Ctl 5 Ctl 6
FXN
M5 M6 F4 F5 F6 F7 F8 F9 Ctl 5 Ctl 6
FXN
E
M1 M2 M3 M4 F1 F2 F3 Ctl 4 Ctl 3
FXN
M5 M6 F4 F5 F6 F7 F8 F9 Ctl 4 Ctl 3
FXN
F
M1 M2 M3 M4 F1 F2 F3 Ctl 7
FXN
M5 M6 F4 F5 F6 F7 F8 F9 Ctl 7
FXN

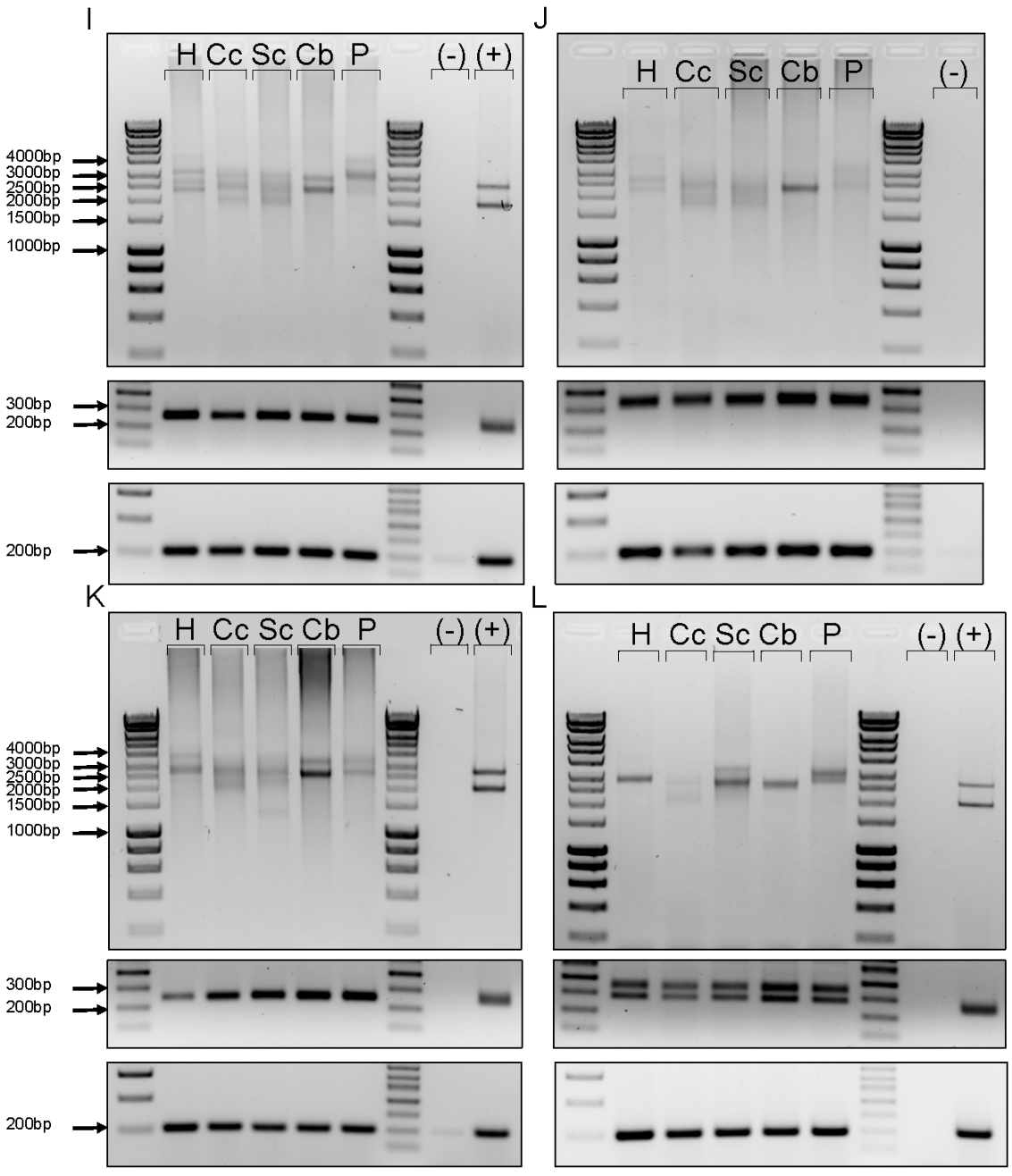
S2 Fig.


A
H Cc Sc Cb P (-) (+)
4000bp
3000bp
2500bp
2000bp
1500bp
1000bp
300bp
200bp
200bp
B
H Cc Sc Cb P (-) (+)
C
H Cc Sc Cb P (-) (+)
4000bp
3000bp
2500bp
2000bp
1500bp
1000bp
300bp
200bp
200bp
D
H Cc Sc Cb P (-)

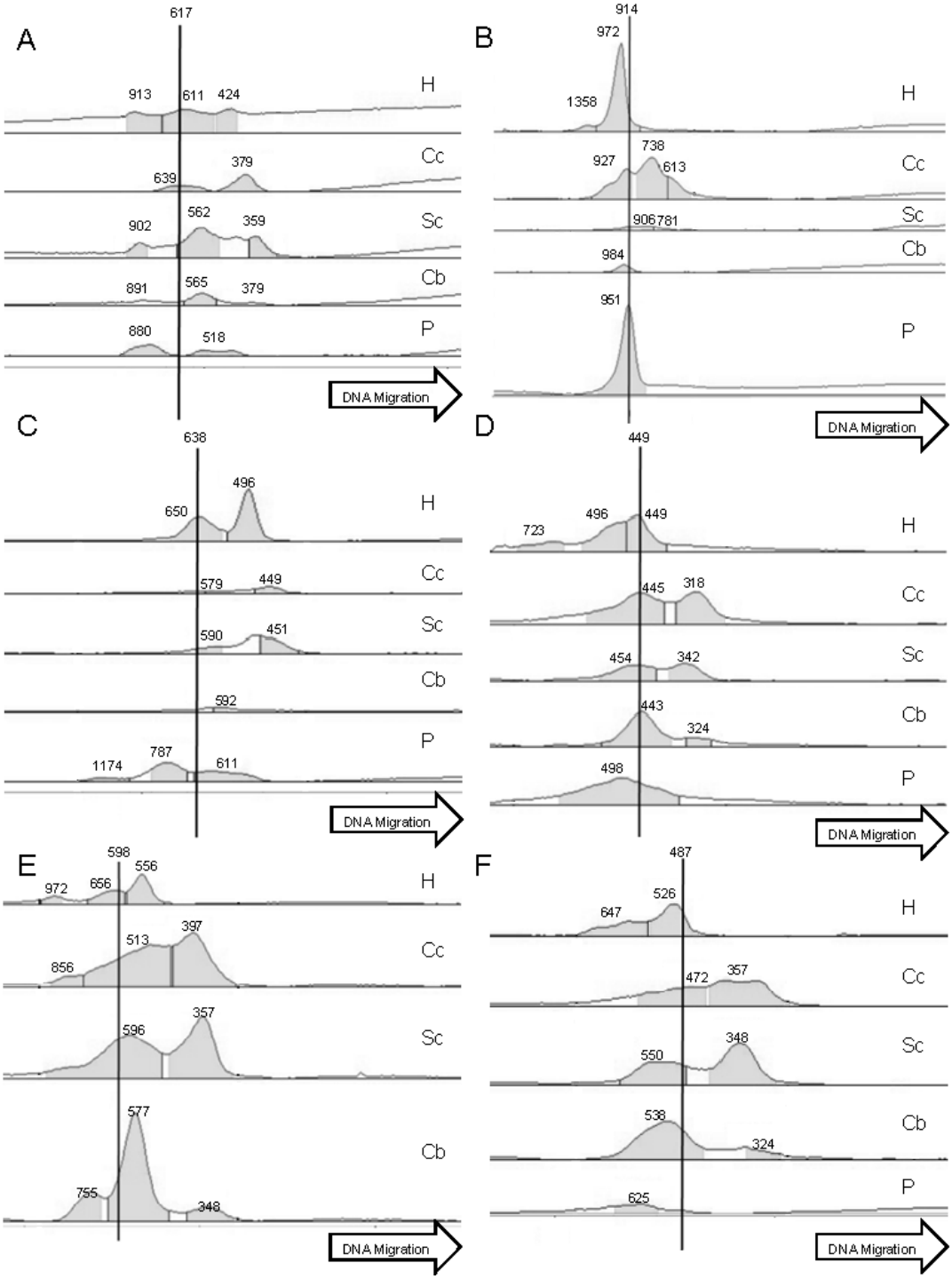
S2 Fig. (Continued)



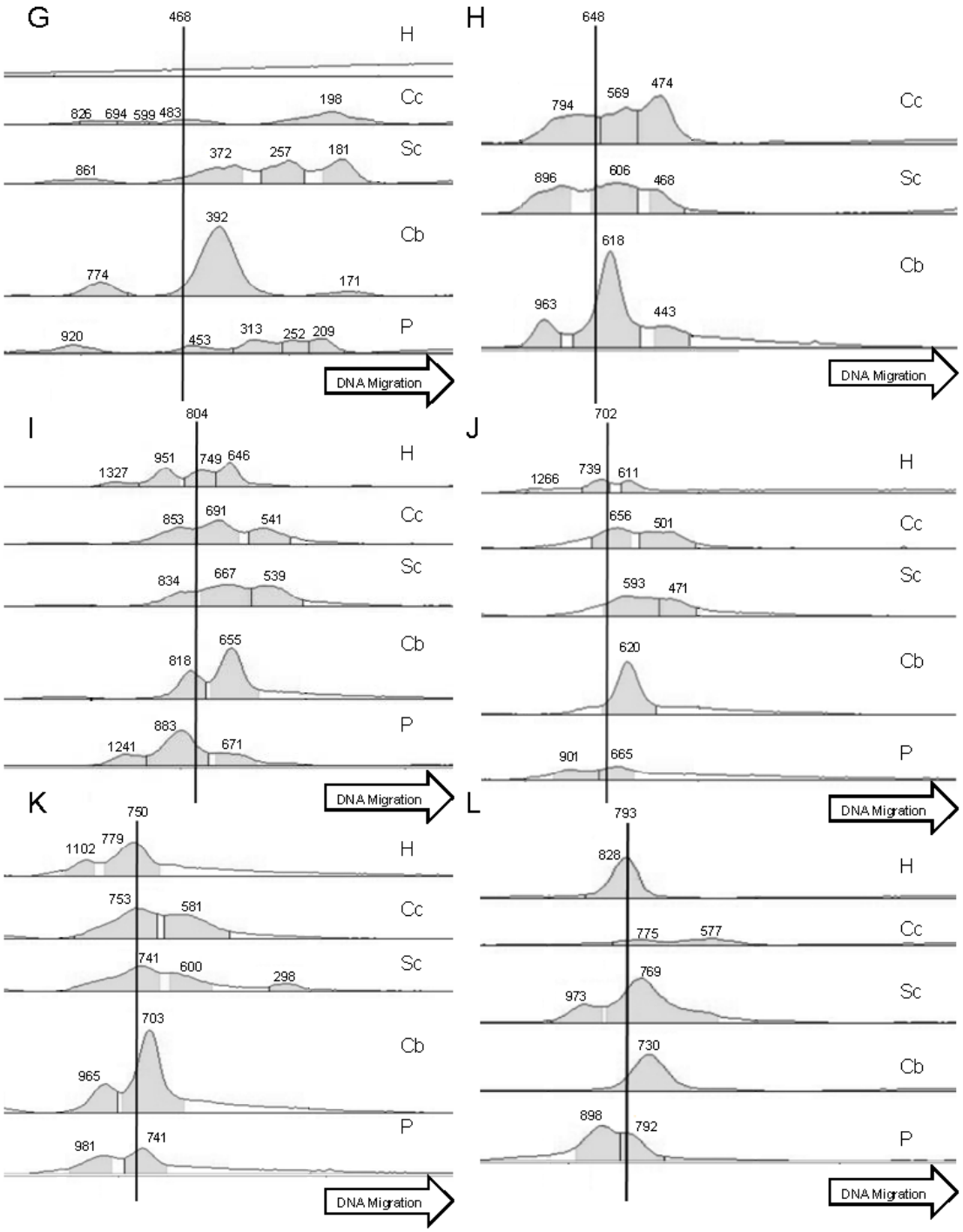
S2 Fig. (Continued)



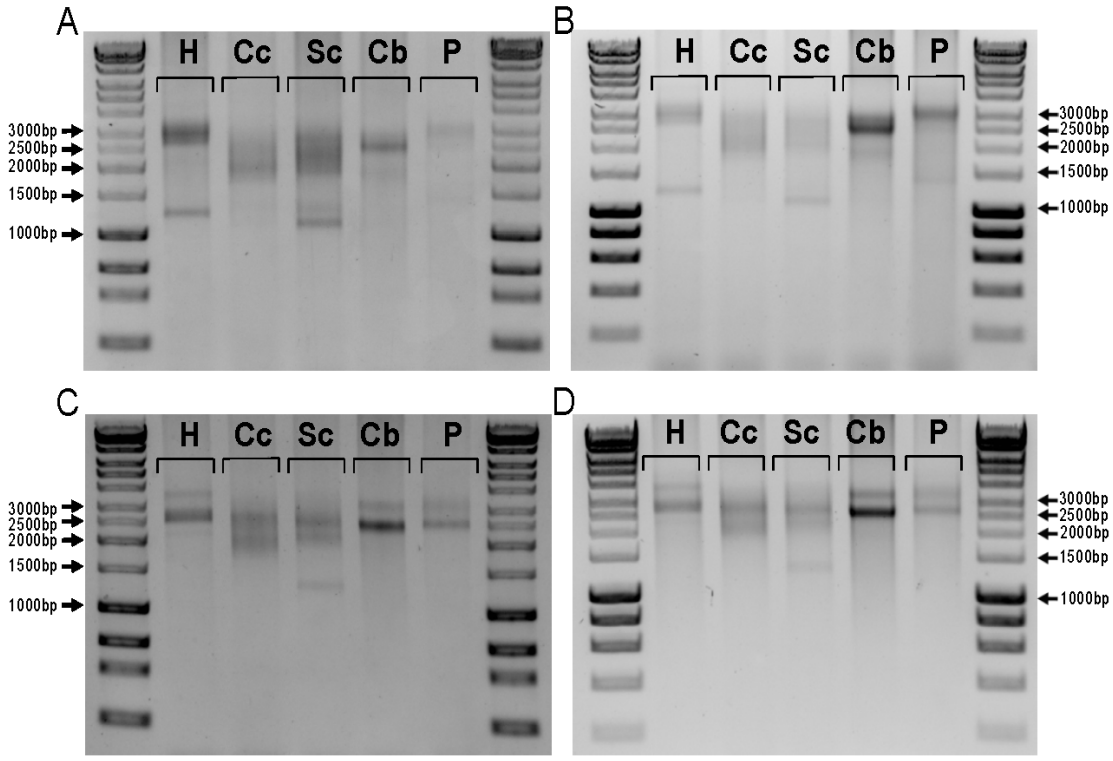
S3 Fig.


A
617
913
611
424
H
639
379
Cc
902
562
359
Sc
891
565
379
Cb
880
518
P
DNA Migration
B
914
972
1358
H
927
738
613
Cc
906
781
Sc
984
Cb
951
P
DNA Migration
C
638
650
496
H
579
449
Cc
590
451
Sc
592
Cb
1174
787
611
P
DNA Migration
D
449
723
496
449
H
445
318
Cc
454
342
Sc
443
324
Cb
498
P
DNA Migration
E
598
556
972
656
H
513
397
856
Cc
596
357
Sc
577
Cb
755
348
DNA Migration
F
487
647
526
H
472
357
Cc
550
348
Sc
538
324
Cb
625
P
DNA Migration

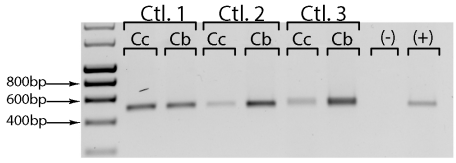
S3 Fig. (Continued)



S4 Fig.


A
H
Cc
Sc
Cb
P
3000bp
2500bp
2000bp
1500bp
1000bp
B
H
Cc
Sc
Cb
P
3000bp
2500bp
2000bp
1500bp
1000bp
C
H
Cc
Sc
Cb
P
3000bp
2500bp
2000bp
1500bp
1000bp
D
H
Cc
Sc
Cb
P
3000bp
2500bp
2000bp
1500bp
1000bp

S5 Fig.



ANALYSIS OF PATIENT-SPECIFIC *NF1* VARIANTS LEADS TO FUNCTINAL INSIGHTS FOR RAS SIGNALING THAT CAN IMPACT PERSONALIZED MEDICINE

by

ASHLEE LONG, HUI LIU, JIAN LIU, MICHAEL DANIEL, DAVID BEDWELL, BRUCE KORF, ROBERT A. KESTERSON, DEEANN WALLIS

In preparation for *Human Mutation*

Format adapted for dissertation

ABSTRACT

We have created a panel of twenty-nine *NF1* variant cDNAs representing benign missense (MS) variants, pathogenic MS variants, many with clinically relevant phenotypes, in-frame deletions, splice variants, and nonsense (NS) variants. We have determined the functional consequences of the variants, assessing their ability to produce mature neurofibromin and restore Ras signaling activity in *NF1* null (-/-) cells. cDNAs demonstrate variant-specific differences in neurofibromin protein levels, suggesting that some variants lead to protein instability or enhanced degradation. When expressed at high levels, some variant proteins are still able to repress Ras activity, indicating that the NF1 phenotype may be due to protein instability. In contrast, other variant proteins are incapable of repressing Ras activity, indicating that some do not functionally engage Ras and stimulate GTP-ase activity. We observed that stability and Ras activity can be mutually exclusive. These assays allow us to categorize variants by functional effects, may help to classify variants of unknown significance, and may have future implications for more directed therapeutics.

INTRODUCTION

Neurofibromatosis type 1 (NF1) is one of the most common autosomal dominant neurological disorders and results in a multifaceted phenotype which includes: bone dysplasia, learning disabilities, benign nerve sheath tumors, and malignant tumors. NF1 is caused by pathogenic variants in the *neurofibromin* (*NF1)* gene, and the clinical phenotypes of individuals with NF1 can vary widely, even among family members with the same genotype. NF1 functions as a GTPase-Activating Protein (GAP), simulating conversion of active Ras-GTP to the inactive form, Ras-GDP (Figure 1). Almost 2,900 pathogenic variants have been reported in the Human Gene Mutation Database (http://www.hgmd.org); most lead to lack of expression of the *NF1* gene product. NF1 does not exhibit a mutagenic hotspot, with variants occurring throughout the gene. A subset (17%) of these variants consists of missense (MS) variants which may result in an unstable or dysfunctional protein (Koczkowska et al., 2019). MS variants in *NF1* are not confined to any specific region of the gene. Some occur in the GAP-related domain (GRD) and might be expected to interfere with GAP function. MS variants occurring outside of this domain may result in other NF1 dysfunction, including protein instability, cellular mis-localization, or the disruption of neurofibromin interaction with other proteins in the cell. Another subset of variants (up to 29% (Kang et al., 2020)) includes nonsense (NS) variants, which result in premature termination of translation, and, in most cases, nonsense-mediated decay of the transcript.

The most critical role of neurofibromin appears to be the regulation of Ras signaling, with loss of neurofibromin function resulting in increased signaling. NF1 tumor cells (*e.g.,* Schwann cells in neurofibromas) exhibit increased Ras signaling as a

consequence of loss of function of both *NF1* alleles (one in the germ line and one somatically-acquired (Figure 1; right) (Cichowski & Jacks, 2001). Therapeutic interventions to date have focused on inhibition of upregulated Ras signaling (*e.g.,* MEK inhibition with selumetinib (Figure 1, right)). While MEK inhibitors have demonstrated effectiveness, not all patients benefit, plexiform neurofibromas do not completely disappear, and there can be significant side effects (Baldo, Magnolato, Barbi, & Bruno, 2021). Therefore, additional treatments that can be used in conjunction with MEK inhibitors are needed.

Only a small portion of the NF1 protein, the GRD, directly interacts with Ras. Some of the key NF1 residues involved in the NF1-Ras interaction include R1276 and K1423 (Yan et al., 2020). R1276, a highly conserved arginine residue, also termed the arginine "finger", aids in the stabilization of the GTP/GDP-NF1 complex (Scheffzek et al., 1997). Mutation of this residue has little effect on Ras binding, but results in complete loss of GAP activity (Ahmadian, Stege, Scheffzek, & Wittinghofer, 1997; Sermon, Lowe, Strom, & Eccleston, 1998). NF1 K1423 forms a salt bridge with Ras D38 to stabilize the protein-protein interaction. In addition to dysregulated Ras signaling due to variants located in the GRD, interference with neurofibromin's localization to the cytoplasmic membrane can result in abnormal interaction with Ras (Stowe et al., 2012). MS variants that occur within the SPRED1-binding domain of NF1 can reduce the affinity for SPRED1 such that neurofibromin fails to traffic to the membrane (Dunzendorfer-Matt, Mercado, Maly, McCormick, & Scheffzek, 2016; Hirata et al., 2016; Yan et al., 2020). However, some pathogenic NF1 variants fail to co-immunoprecipiate with SPRED1, do

not directly interfere with Ras binding, and do retain GAP activity (Dunzendorfer-Matt et al., 2016).

An additional complexity to NF1 interactions is its dimerization activity (Figure 1, left) (Carnes, Kesterson, Korf, Mobley, & Wallis, 2019; Sherekar et al., 2019). While the mechanistic significance has yet to be determined, dimerization provides a potential explanation for the phenotypes observed with many NF1 disease variants. Often heterozygous NS or frameshift (FS) variants are observed in NF1, where there is still a single copy of wild type (WT) *NF1* allele present. The total amount of full-length neurofibromin in some affected individuals with these variants may be considerably less than the predicted 50% of WT levels (Anastasaki, Woo, Messiaen, & Gutmann, 2015). Neurofibromin levels could be drastically lowered if mutant protein dimerizes with WT and then is targeted to the proteasome for degradation (Figure 1). The ubiquitin-proteasome pathway (UPP) controls NF1 levels and both the amplitude and duration of Ras-mediated signaling (Cichowski, Santiago, Jardim, Johnson, & Jacks, 2003). Excessive proteasomal degradation and genetic loss results in NF1 inactivation in sporadic gliomas (McGillicuddy et al., 2009). Proteasomal degradation of NF1 is partially regulated by the binding of both the SAG-SKP1-CUL1-FBXW7 and RBX1/2-CUL3-KBTBD7 complexes with NF1 (Figure 1; left) (Hollstein & Cichowski, 2013; Tan et al., 2011). The CUL3/KBTBD7 complex has been implicated in the pathogenic destabilization of neurofibromin in glioblastomas (Hollstein & Cichowski, 2013).

Here, we further validate our heterologous m*Nf1* cDNA expression system and multiple assays for neurofibromin function and use them to evaluate patient-specific variants (Wallis et al., 2018). We have included many MS variants with known genotype-

phenotype correlations to assess their functional effects on NF1 levels as well as Ras signaling. We demonstrate that effects on stability and Ras signaling can be mutually exclusive functions. We suggest that such stratification of variant effects will have implications for mutation-targeted therapeutics for NF1, neurofibromin-driven breast cancers, and potentially other phenotypes. These assays may also have utility in classifying newly identified variants of unknown significance. As new Ras-independent functions are discovered for neurofibromin, it will be important to assay variants for their effects in new functional assays.

MATERIALS AND METHODS

Cell culture: HEK293 (WT or *NF1* +/+) cells were obtained from ATCC (CRL-1573) and cultured in DMEM + 10% FBS and 1X Pen/Strep using standard culture procedures. *NF1* -/- or null HEK293 cells were previously created through CRISPR Cas9 targeting *NF1* exon 2 (Wallis et al., 2018) . We have chosen to evaluate functional NF1 expression and Ras activity in HEK293 cells because this cell line is well characterized, used historically in NF1 research, easily takes up exogenous DNA, and is easy to culture and scale. HEK293s have all three Ras isoforms and recapitulate Ras signaling; hence, HEK293 cells are an appropriate model system for these assays.

***Nf1* cDNA plasmid development:** The wild type (WT) *Nf1* cDNA plasmid was developed by GeneCopoeia and is commercially available. The full-length mouse cDNA (*mNf1)* produces a >250 kDa neurofibromin protein that is capable of modulating Ras signaling (Wallis et al., 2018). We created an empty vector (EV) control plasmid with the same parental backbone that does not contain the *Nf1* cDNA. We also created a panel of mutant

cDNAs representing variants found in NF1-affected individuals with different clinically relevant phenotypes and assessed their ability to produce mature neurofibromin and restore *Nf1* activity in *NF1*$^{-/-}$ cells. *mNf1* cDNA is appropriate for study because the full-length cDNA sequences of endogenous *hNF1* and *mNf1* have 92% sequence identity; amino acid sequences share 98% identity and human cDNAs have historically been unstable and toxic. Variants were introduced into shuttle vectors by either site-directed mutagenesis or utilization of synthetic DNA fragments and then cloned into the full-length vector using standard enzymes. Each variant plasmid was confirmed by sequencing the entire *Nf1* cDNA insert and all subsequent DNA preparations were validated by spot checking for the variant of interest. Furthermore, multiple plasmid preps were utilized for each variant to mitigate any variability due to quality of DNA.

Transient transections: For Western blots and GTP-Ras assays, cells were transfected using LipoD293 (SignaGen Lab. Cat# SL100668) with up to 1ug of cDNA per 6-well dish seeded with 500,000 cells per well. Assays were performed 48-72 hours later. For titration experiments, WT cDNA was balanced with EV control such that 1000 ng total DNA was transfected.

Western blotting: Cells were lysed with RIPA buffer supplemented with a protease inhibitor cocktail and phosSTOP, and lysates were cleared by centrifugation at 20,000 RPM for 20 minutes at 4^0C. Protein was quantified with a Bradford assay and 50 ug of protein was loaded per well for NF1 blots and 10 ug of protein was loaded for other blots. 8% SDS-polyacrylamide gels were run at 100 V for 2 hours and transferred at 100 V for 2 hours onto PVDF. Blots were probed overnight at 4^0C with primary antibody, washed, and probed for 1 hour at room temperature with secondary. Primary antibodies include N-

Terminal NF1 (Cell Signaling cat# D7R7D 1:1000), tubulin (Abcam cat# ab52866 1:1000), b-actin (Cell Signaling cat# 3700 1:1000), p-ERK (Cell Signaling cat# 9101 1:1000), and total ERK (Cell Signaling cat# 9102 1:1000). Secondary was HRP tagged from Santa Cruz. Chemiluminescent substrate from Bio-Rad was used as per manufacturer's protocols.

RAS-G-LISA Assay: The RAS-G-LISA assay was obtained from Cytoskeleton Inc. and was performed according to the manufacturer's instructions.

Statistical analysis. All assays were repeated a minimum of three times for each variant. For each experiment wild type *Nf1* cDNA (WT) and empty vector (EV) plasmid control with no cDNA insert were included as controls. Depending on whether we intended to show either 1) presence of WT cDNA repressed Ras activity more than absence of *NF1* cDNA (EV) as in our titration experiments or 2) variant cDNA expression or activity was different from WT cDNA, we normalized expression or activity to WT or EV respectively and made comparisons with EV or WT respectively. Normalization to one control allows us to combine data across independent experiments, and combination of multiple repeat experiments controls for the effects of differential transfection efficiencies. Statistical comparisons using student's t-test were made using Excel software to determine which results were statistically significant.

RESULTS

Assay validation with cDNA Titrations: To further validate our cDNA and assay system (Wallis et al., 2018), we performed titration experiments to show a dose-response effect utilizing 0.25 - 1000ng/well of WT mouse *Nf1* cDNA in *NF1* null HEK293 cells

with 500,000 cells balanced with empty vector (EV) control. Neurofibromin/tubulin levels (Figure 2A), GTP-Ras levels (Figure 2B), and pERK/ERK ratios (Figure 2C) all respond in a dose-response manner to the amount of cDNA transfected into the cells. Neurofibromin levels become detectible via Western blot analysis between 1-4 ng/500,000 cells transfected (Figure 2A). GTP-Ras levels were lowered, beyond that of empty vector (0ng WT cDNA), starting at 1-4 ng/500,000 cells transfected, with statistically significant differences observed at 15 ng and above (Figure 2B). Lowered pERK/ERK ratios, a marker for MAPK signaling activity, were also detected at 1-4 ng/500,000 cells transfected, with statistically significant differences from EV at 250 ng (Figure 2C).

Variant Selection and groupings: We utilized a full gene-encompassing panel of *mNf1* cDNAs transfected into an *NF1* null (-/-) cell line to evaluate the functional effects of unique variants (Figure 3 and Table 1). Variants, in addition to WT and EV, were selected based on the following criteria: controls, genotype-phenotype correlations, occurrences in different NF1 domains, and type of mutation. Benign variants, both within and outside of the GRD, were selected due to lack of pathogenicity and as "controls": E1327G, Q1336R, and P2782L. Variants with published genotype phenotype-correlations were prioritized. Variants associated with "mild" phenotypes include delM992, R1038G, M1149V, and R1809C (Koczkowska et al., 2019; Koczkowska, Callens, et al., 2018; Rojnueangnit et al., 2015; Trevisson et al., 2019; Upadhyaya et al., 2007). Variants associated with "severe" phenotypes include L847P, G848R, R1276Q, and K1423E (Koczkowska et al., 2019; Koczkowska, Chen, et al., 2018; Korf, Henson, & Stemmer-Rachamimov, 2005). We also included variants from multiple domains, despite

having “unknown” phenotype associations. C379R falls within the 5’ region with no described domain function, W784R falls within the putative CSRD domain, L1490P falls within the SPRED1 interaction domain, D1623 falls within the Sec14 domain, and L1957P, S1997R, and L2317P all fall within the 3’ region of the protein but not in well-described domains. Additional variants were selected based on the formation of cryptic “splice” sites: Y489C (Messiaen et al., 1999) and G629R. While splicing is not affected in the cDNA system, assessment of functional effects of the subsequent missense variant is a critical first step in the development of antisense therapeutics that might restore normal splicing but leave the variant intact. Additional “nonsense” variants were prioritized based on incidence as well as location throughout the protein.

NF1 levels First we evaluated neurofibromin levels after transfecting cells with equal plasmid concentrations from each representative cDNA. All cDNAs were assayed for NF1 levels by transfecting a consistent 1 ug cDNA into a 6 well plate with 500,000 cells and harvesting cells 48 hours post transfection. Figure 4 demonstrates quantification of NF1/tubulin ratios for all cDNAs and indicates differential variant-specific effects on neurofibromin levels. For example, while some variants remain stable with levels similar to WT (e.g., R1809C), others show much lower levels of neurofibromin; L1490P and D1623G lead to approximately 20% WT levels. We interpret NF1 levels to reflect protein stability, which may be dependent on mutation-targeted proteasomal degradation. All cDNAs should have similar transfection, transcription, and translation efficiencies. Each experiment included both WT and EV control cDNAs. Samples were normalized to tubulin as a load control and WT/tubulin ratios were set at 1.0 in each experiment. All other cDNA/tubulin ratios were reported relative to WT levels. Normalization allowed

comparison across experiments. Western blots of select MS variants show varying full-length NF1 levels (Figure 4B). Western blots of nonsense variants show varying NF1 levels (Figure 4C); note the truncation products run at the anticipated sizes; however, NF1 levels did not correlate with length of prematurely terminated mutant proteins.

Ras Activity: Next we assayed all *Nf1* variant cDNAs for their ability to repress Ras signaling. We evaluated both GTP-Ras levels and pERK/ERK ratios (Figure 5). All cDNAs were assayed by consistently transfecting in 1 ug of cDNA into a 6 well plate with 500,000 cells/well and harvesting protein lysate 48 hours post transfection. Each experiment included both WT and EV control cDNAs. For GTP-Ras levels (blue bars; Figure 5A), each sample was normalized to EV GTP-Ras levels, which were set at 1.0 in each experiment, and all other GTP-Ras levels were reported relative to EV levels. Each cDNA's GTP-Ras level was statistically compared via Student's t-test to the WT cDNA's GTP-Ras level to determine if the variant negatively impacted NF1 ability to repress GTP-Ras levels. Blue asterisks indicate that a variant has statistically significant impaired ability to inhibit GTP-Ras levels; these include: delM992, M1149V, L847P, R1276Q, K1423E, C379R, L1490P, D1623G, S1997R, L2317P, R192X, R461X, R681X, R816X, R1276X, and R1306X.

For pERK/ERK ratios (black bars; Figure 5A), each experiment included both WT and EV control cDNAs, with each sample normalized to the EV cDNA pERK/ERK ratio, which was set at 1.0. Each cDNA's pERK/ERK ratio was statistically compared via Student's t-test to the WT pERK/ERK ratio to determine if the variant negatively impacted its ability to repress pERK activity with black asterisks indicating statistically significant

impaired ability; these include M1149V, L847P, R1276Q, K1423E, L1490P, D1623G, S1997R, L2317P, R192X, R461X, R681X, R816X, R1276X, R1306X, and R1947X.

Activity as a function of stability: Neurofibromin function *in vivo* relies on numerous factors, including stability (abundance of protein available), cellular localization, ability to bind interacting proteins (such as Ras), and ability to stimulate Ras GTPase activity. As we measured two of these factors, neurofibromin levels and GTPase activity, we wanted to determine if the combined factors can lead to variant functional insights. To achieve this, we plotted our *Nf1* WT cDNA titration data (derived from Figure 2) such that neurofibromin/tubulin levels at 1000 ng cDNA was set to a maximum of 1 and plotted on the x-axis and corresponding GTP-Ras activity levels were plotted on the y-axis (Figure 6, gray dots). A trend line was generated (Figure 6, blue dotted line). To evaluate this multidimensional concept, we overlaid variant data onto this plot and categorized variants as we had in Figures 4 and 5 as "Control" (green dots), "Splice" (yellow dots), "Mild" (orange dots), "Severe" (pink dots), and "Unknown" (teal dots). The control variants clustered such that NF1/tubulin ratios were > 0.75 and GTP-Ras levels were < 0.52 (Figure 6 green oval). The cryptic splice variants also clustered at NF1 > 0.89 and GTP-Ras < 0.66. While the mild variants didn't cluster together as tightly, three of the four variants had stable protein levels > 0.8 and two had low GTP-Ras levels < 0.66. These loosely clustered mild variants also cluster with the splice variants (Figure 6 orange ovals). Severe and unknown variants did not form a single cluster. Severe variants R1276Q and K1423E that interact with Ras do maintain stability but cannot suppress Ras and are clustered in the top right (Figure 6 pink oval). Outside of those clusters, we find multiple variants that hug the trend line (blue oval): "Severe" variants L847P and G848R and unknown variants

L2317P, C379R, W784R, and L1957P (these have been individually labeled in the plot). This suggests that given a certain abundance of neurofibromin, the variants can suppress Ras signaling.

DISCUSSION

cDNA model system. Since discovery of the *NF1* gene in 1990 research efforts have been hindered by the lack of a full-length coding cDNA. This is partially due to the size of the gene and toxicity of the human cDNA construct. We are aware of three potentially available *NF1* cDNAs: mouse *Nf1* (Wallis et al., 2018) (isoform 2 with 2839 amino acids), codon optimized human *NF1* (Bonneau, Lenherr, Pena, Hart, & Scheffzek, 2009) (isoform 1 with 2818 aa), and a human *NF1* with mini-intron 35-36 (Cui & Morrison, 2019) (isoforms 1 and 2). As the *mNf1* cDNA is highly homologous to endogenous human *NF1*, we have developed and validated the mouse *Nf1* cDNA expression system that allows us to examine the biochemical effects of any *Nf1* genetic variant. We have been able to perform dose-response studies to titrate in varying amounts of *mNf1* cDNA and are able to detect a clear dose-response in terms of levels of neurofibromin and repression of GTP-Ras levels and pERK/ERK ratios, giving us confidence in both the cDNA and the functional assays.

Assay Performance: We see similar dynamic ranges between the GTP-Ras and pERK/ERK experiments, with WT cDNA able to repress both GTP-Ras and pERK/ERK ratios by about half that seen with EV control. The Morrison lab reported that inactive variants (R1276P) served as better controls than empty vectors (Cui & Morrison, 2019). While our R1276Q and R1276X both displayed similar GTP-Ras levels as our EV, both

showed insignificantly higher pERK/ERK ratios. Indeed, we see significant variability in the pERK/ERK ratios, leading to large error bars. Regardless, each individual cDNA performs similarly between the two assays and across multiple experiments, indicating that the results are reliable. Each variant's ability to affect Ras signaling relative to WT cDNA is consistent between assays; however, we have noted exceptions. delM992 and C379R both have significantly different GTP-Ras levels, whereas pERK/ERK ratios fail to reach significance; and R1748X has a significantly different pERK/ERK ratio, but GTP-Ras doesn't reach significance.

Using our previously published cDNA expression system (Wallis et al., 2018) we can evaluate functional significance of variants in individuals with NF1. These data indicate that each variant has a slightly different functional profile in terms of both protein stability and the ability to inhibit Ras signaling. We incorporated non-pathogenic variants in our assays: E1327G, Q1336R, and P2782L. These three cDNAs performed similarly to the WT cDNA in terms of neurofibromin levels and the ability to inhibit Ras signaling, adding confidence to our functional profile characterizations. Some pathogenic variants result in instability of neurofibromin but still retain GRD function. The best examples of this are with the G848R and L1957P variants, which retain the ability to repress Ras activity (Figure 5) in both of our assays, yet are unstable and produce less than 50% of the neurofibromin that is observed with WT cDNA protein (Figure 4). Some variants result in stable protein but have lost GRD function. For example, K1423E and S1997R both produce neurofibromin levels similar to WT cDNA, yet display significantly elevated Ras activity compared with WT cDNA. We also observed that certain variants demonstrated both unstable protein and loss of function; L1490P and

D1623G exhibited lowered neurofibromin levels, with inability to repress Ras signaling. Nonsense variants resulting in truncated proteins have variable stability. Nonsense variants with truncations after the GRD may maintain GRD function in these overexpression assays. While both R1947X and R2550X show increased Ras signaling, it is not statistically different from WT cDNA.

Genotype-phenotype Correlations and assay results

Genotype-phenotype correlations indicate allele-specific effects for *NF1*. There are multiple MS and small in-frame deletion variants that are associated with unique NF1 phenotypes that may be utilized for further functional analysis. A mild phenotype is associated with the c.2970-2972 delAAT (delM992) single amino acid deletion, consisting of café-au-lait macules (CALMs) and skinfold freckling and lack of neurofibromas (Koczkowska, Callens, et al., 2018; Upadhyaya et al., 2007). Variants involving R1809 are the most frequent recurrent *NF1* variants and present with multiple CALMs, with or without freckling and Lisch nodules, but externally visible plexiform neurofibromas, symptomatic optic pathway glioma or cutaneous or subdermal neurofibromas are not found (Rojnueangnit et al., 2015). Mild phenotypes are also associated with R1038G (Trevisson et al., 2019) and M1149V (Koczkowska et al., 2019). All of these genotypes are associated with Noonan-like facial features. In fact, 31.1%, 29% and 11.5% of individuals with variation at R1809, M1149, and M992, respectively, show Noonan features, in comparison to only 3.4% of "classic" NF affected individuals (Koczkowska et al., 2019). The R1038G cohort, consisting of two families, is too small to meaningfully compare, but Noonan features are also reported in both families (Trevisson et al., 2019). Overall, NF1 affected individuals with delM992, R1038G,

M1149V, and R1809C associated with mild phenotypes lack clinically suspected plexiform, cutaneous, or subcutaneous neurofibromas and are not at risk for malignancy.

In contrast, other MS variants are associated with more severe phenotypes, including increased risk for malignancy and/or spinal neurofibromas. Constitutional MS variants affecting one of five neighboring *NF1* codons—Leu844, Cys845, Ala846, Leu847, and Gly848—located in the cysteine-serine-rich domain (CSRD), typically result in a large number of plexiform and symptomatic spinal neurofibromas, symptomatic optic nerve gliomas, skeletal abnormalities, and malignant neoplasms (Koczkowska, Chen, et al., 2018). R1276Q has been identified in some individuals affected with spinal neurofibromas at all levels and also has been associated with a more severe phenotype (Koczkowska et al., 2019; Korf et al., 2005). K1423E has been associated with a severe phenotype (Koczkowska et al., 2019). Some of these variants (affecting residues G848 and R1276) are observed in individuals with a distinctive phenotype, referred to as "spinal NF." The "spinal NF" phenotype includes few or no cutaneous neurofibromas (cNfs) and a very mild pigmentary phenotype (Burkitt Wright et al., 2013; Ruggieri et al., 2015). These individuals may suffer from a massive internal tumor burden, with neurofibromas at each spinal nerve root and extreme enlargement of most peripheral nerves. These individuals are at great risk of spinal cord compression, pain, and malignant change, and the extreme number of tumors makes surgical treatment difficult or impossible (Koczkowska et al., 2019; Korf et al., 2005). Certain variants correlate with increased incidences of cancer. A greater risk of malignancy for MS variants in codons 844-848 has been reported (Koczkowska, Chen, et al., 2018). Codon 847 is recurrent in NF1 patients with breast cancer (Frayling et al., 2019). A lack of large deletions with an excess of NS and FS

variants has been observed with breast cancer (Frayling et al., 2019; Zheng et al., 2020). Thus, the type of *NF1* variant matters when genotype-phenotype correlations can be made and can influence clinical care.

We had hoped that our functional assays might predict which genotypes could result in specific phenotypes, but a simplistic interpretation is not readily available. Our assays indicate that most of the variants associated with mild phenotypes are hypomorphic alleles. All mild variants are relatively stable and produce neurofibromin at levels above 60% of WT levels. It is not surprising that a certain threshold of neurofibromin must be achieved to have a mild phenotype; however, Ras signaling may be altered depending on the variant. Ras signaling for R1038G and R1809 is elevated, but not statistically different than for WT cDNA. The delM992 variant is unable to completely suppress GTP-Ras or pERK activity; this is statistically significant for GTP-Ras. Finally, M1149V has statistically significant increased Ras signaling. These data suggest that delM992, R1038G, and R1809C act as hypomorphs. In contrast, M1149V appears to have lost the ability to inhibit Ras signaling and an explanation for this unanticipated result for this "mild" variant is not available.

Other genotypes are associated with severe phenotypes: L847P, G848R, R1276Q, and K1423E. cDNAs with these genotypes have highly variable NF1 levels, ranging from 38% - 136% that of WT cDNA. L847P and G848R are located in the CSRD; R1276Q and K1423E are located within the GRD and interact directly with Ras. R1276Q and K1423E are completely unable to suppress Ras signaling and are statistically different from WT cDNA, as would be expected for variants that are critical for Ras binding and GTP-hydrolysis. L847P is also unable to repress Ras signaling;

however, G848R can repress Ras signaling. Its lack of stability likely explains why it is unable to function properly and causes a phenotype.

Given the interdependency of stability and function, we wanted to evaluate NF1 stability as a function of Ras activity and plotted neurofibromin levels with GTP-Ras levels (given that this assay had less variability than the pERK/ERK assay) and drew a trend line. We noted clustering of controls and cryptic splice variants, and that genotypes associated with mild phenotypes also are loosely clustered (orange circle). The presence of the "unknown" variant S1997R within this cluster suggests that individuals with the genotype may have a mild phenotype. In fact, the Leiden Open Variation Database (LOVD) has classified this variant as a variant of uncertain significance (VUS); however, we have identified an individual that meets NF1 diagnostic criteria with this *de novo* variant. Though still adolescent, no cutaneous or plexiform neurofibromas have been identified and the phenotype is thus-far "mild". In addition, we find multiple variants that hug the trend line: variants L847P and G848R (associated with severe phenotypes) and variants L2317P, C379R, W784R, and L1957P. This suggests that given a certain abundance of neurofibromin, the variants are partially able to repress Ras signaling. If the protein could be stabilized *in vivo*, Ras might be repressed and the phenotype rescued. Thus, we demonstrate two distinct clusters of genotypes associated with severe phenotypes, one (pink oval) indicating loss of GTPase function results in pathogenicity and the other indicating loss of stability leads to pathogenicity.

Six NF1 phenotypic subtypes have recently been proposed, and while genotypic data were inadequate to make statistically significant conclusions, particular variants were noted to be consistent with three of the six clusters (Tabata, Li, Knight, Bakker, &

Sarin, 2020). delM992 was consistent with the mild subtype (cluster 1); R1809C was consistent with the freckling-predominant subtype (cluster 2), and L847P and was consistent with the early-onset neural severe (cluster 6) subtype. Ideally, combining genotype and functional data with such phenotypic clustering would be a powerful tool in understanding the phenotypic heterogeneity of NF1. Unfortunately for this cohort (derived from a self-reported registry) only 61 of 2051 participants provided molecular diagnostic data (though ~50% reported that a molecular diagnosis had been made).

Study Limitations: There are several factors that limit our study. First, HEK293 cells are very different from Schwann cells (one of the cell types primarily affected in individuals with NF1). HEK293 cells are derived from human embryonic kidney cells (not neural crest cells; but they maybe neuronal as they share similarities with embryonic adrenal precursor cells (Lin et al., 2014) transformed by incorporation of 4.5 kb of adenovirus 5 genome into human chromosome 19 and carries a modal chromosome number of 64 in 30% of cells. Notably, this increased chromosome number does not affect any of the RAS or RASGAP genes. HEK293s have all three wild type Ras isoforms. Even though the Ras pathway is remarkably conserved, there could be modifiers in this cell line not present in Schwann cells, melanocytes, neurons, and osteoblasts/osteoclasts. Second, while Ras assays are commonly employed to evaluate *NF1* variants, determination and evaluation of alternative functions is critical. Little information is available regarding how variants might affect NF1 dimerization, nuclear localization (or cellular localization in general) or even how they might interact with other binding partners. Recently, a new mechanism whereby NF1 binds the estrogen receptor (ER) and acts as a transcriptional corepressor has emerged; this ER activity is

functionally independent of GAP activity (Zheng et al., 2020). Thus, NF1 is a dual repressor for both Ras and ER signaling. Defining how variants affect these functions will aid our understanding of neurofibromin structure-function. Finally, our assay is an over-expression assay and hence does not reflect endogenous expression levels.

To date, NF1 cannot be cured. While MEK inhibitors can block Ras signaling regardless of mutation type, therapeutics that address the underlying cause of the disease by restoring neurofibromin function to a level that leads to a non-pathogenic phenotype do not yet exist. Various gene and mRNA targeting strategies have been proposed and are being evaluated for their therapeutic potential in NF1 (Leier et al., 2020). Compounds such as proteosome inhibitors could be used to stabilize neurofibromin levels if the protein is being targeted for degradation. Potentiators (analogous to those utilized for CFTR) might be used to directly bind NF1 and stabilize it or prevent it from being degraded. NF1 mimetics might be developed to stimulate Ras GTPase activity. Thus, molecular diagnostics and determination of a variant's stability and function are increasingly relevant to guide clinical care for those with known genotype-phenotype correlations and may also have implications for both classification of variants of uncertain significance (particularly those found in breast cancer) and developing therapeutics. Once variants and their effects are established and categorized, new classes of therapeutics become possible.

Table 1: Panel of *Nf1* cDNAs

DNA	Protein	Type	Domain	Notes
c.574C>T	R192X	NS	5'	
c.1135T>C	C379R	MS	5'	
c.1381C>T	R461X	NS	5'	
c.1466A>G	Y489C	MS	5'	cryptic splice
c.1885G>A	G629R	MS	CSRD	cryptic splice
c.2041C>T	R681X	NS	CSRD	
c.2350T>C	W784R	MS	CSRD	
c.2446C>T	R816X	NS	CSRD	
c.2540T>C	L847P	MS	CSRD	Severe Phenotype
c. 2542G>C	G848R	MS	CSRD	Severe Phenotype
c.2970_2972delAAT	M992del	del	n/a	Mild phenotype
c.3112A>G	R1038G	MS		Mild phenotype
c.3445A>G	M1149V	MS	Tubulin	Mild phenotype
c.3826C>T	R1276X	NS	GRD	
c.3827G>A	R1276Q	MS	GRD	Severe Phenotype; Arginine Finger
c.3916C>T	R1306X	NS	GRD	
c.3980A>G	E1327G	MS	GRD	non pathogenic variant
c.4007A>G	Q1336R	MS	GRD	non pathogenic variant
c.4267A>G	K1423E	MS	GRD	Severe Phenotype; forms a salt bridge w D38 Ras to stabilize interaction; 200-400 fold lower GAP activity (Li 1992); can abolish NF1-Ras interaction (Yan 2020)
c.4469T>C	L1490P	MS	SPRED1/GRD	Likely to destabilize folding of the GAPex region (Yan 2020)
c.4868A>G	D1623G	MS	Sec14	
c.5242C>T	R1748X	NS	PH	
c.5425C>T	R1809C	MS	PH	Mild phenotype
c.5839C>T	R1947X	NS	3'	
c.5870T>C	L1957P	MS	3'	
c.5989A>C	S1997R	MS	3'	
c.6950T>C	L2317P	MS	3'	
c.7648A>T	R2550X	MS	3'	
c.8345C>T	P2782L	MS	3' end	non pathogenic variant

Figure Legends:

Figure 1: NF1 affects multiple signaling pathways to modulate Ras which can be therapeutically targeted through these various pathways. Left side exhibits NF1 dimerization and regulation through the proteasome. Specific NF1 variants might benefit from ubiquitin proteasome pathway (UPP) inhibitors to prevent excess degradation. Right side illustrates the Ras signaling pathway and the use of MEK inhibitors to slow tumor growth resulting from hyperactive Ras signaling.

Figure 2: *Nf1* assay validations with WT mouse cDNA titrations in *NF1* null HEK293 cells. WT NF1 +/+ HEK293 (293 +/+) cells were used as controls. Varying amounts (0-1000ng) of WT plasmid cDNA was transfected into the null cell line. A. Top panel – Quantitation of neurofibromin levels (n = 3). Bottom panel – Representative western blot of titration experiments. B. GTP-Ras activity levels as determined by Ras-GLISA assay (n = 3). Statistically significant differences were seen at 15ng and greater (red asterisk). C. Top panel – Quantitation of pERK/ERK levels (n = 3). Statistically significant differences were observed at 250ng. Bottom panel – Representative western blot. Error bars represent SEM and red asterisks indicate statistical significance.

Figure 3: Schematic of NF1 protein and putative domains with cDNA variants depicted. Red circles – nonsense mutations, blue squares – missense mutations, green triangles – in-frame deletion, and unfilled black circles – non-pathogenic variants. Amino acid positions are notated above protein domains in black text.

Figure 4: Relative neurofibromin protein levels for each cDNA. A. Quantitation of NF1 normalized to tubulin for all the *Nf1* cDNAs and categorized based on variant type. WT/tubulin ratios were set at 1.0 in each experiment with all other variant ratios relative to WT levels. Graph represents an $n > 3$ independent experiments for each cDNA. SEM error bars are shown. B. Representative western blot of select missense mutations showing NF1 and tubulin expression. C. Representative western blot of nonsense mutations showing NF1 and tubulin levels. Note the truncation products run at anticipated sizes.

Figure 5: Ras signaling activity assessed via GTP-Ras and pERK/ERK levels for each cDNA. A. Quantitation of GTP-Ras levels (blue bars) and pERK/ERK ratios (black bars) for all cDNAs. EV levels were set 1.0 and each sample was normalized and reported relative to EV levels. Comparisons to WT and statistical significance were determined via a Student's t-test to determine if the variant negatively impacted NF1's ability to repress Ras signaling. Blue asterisks indicate statistically significant impairment in the ability to inhibit GTP-Ras ($p < 0.05$). Black asterisks indicate statistically significant impaired ability to inhibit pERK/ERK ratios ($p < 0.05$). Graph represents $N>3$ independent experiments for each cDNA. Bars represent SEM. B. Representative image of one western blot of select missense mutations showing pERK and ERK levels. C. Representative image of one western blot of nonsense mutations showing pERK and ERK levels.

Figure 6: Analysis of protein stability and function to group *Nf1* variants. *Nf1* WT protein concentrations and corresponding GTP-Ras levels were plotted (gray dots) to

generate a trend line (blue line). NF1 and GTP-Ras levels for each MS variant were overlaid onto the plot and variants were grouped as controls (green dots), splice (yellow dots), mild (orange dots), severe (pink dots), and unknown (teal dots). Clustering, based on NF1 stability and function is indicated by ovals: control variants – green oval, mild and splice variants – orange oval, select severe variants located in the GRD domain and interacting directly with Ras – pink oval, variants hugging the trend line- blue oval.

References

Ahmadian, M. R., Stege, P., Scheffzek, K., & Wittinghofer, A. (1997). Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nat Struct Biol, 4*(9), 686-689. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9302992

Anastasaki, C., Woo, A. S., Messiaen, L. M., & Gutmann, D. H. (2015). Elucidating the impact of neurofibromatosis-1 germline mutations on neurofibromin function and dopamine-based learning. *Hum Mol Genet, 24*(12), 3518-3528. doi:10.1093/hmg/ddv103

Baldo, F., Magnolato, A., Barbi, E., & Bruno, I. (2021). Selumetinib side effects in children treated for plexiform neurofibromas: first case reports of peripheral edema and hair color change. *BMC Pediatrics, 21*(1), 67. doi:10.1186/s12887-021-02530-5

Bonneau, F., Lenherr, E. D., Pena, V., Hart, D. J., & Scheffzek, K. (2009). Solubility survey of fragments of the neurofibromatosis type 1 protein neurofibromin. *Protein Expr Purif, 65*(1), 30-37. doi:10.1016/j.pep.2008.12.001

Burkitt Wright, E. M., Sach, E., Sharif, S., Quarrell, O., Carroll, T., Whitehouse, R. W., . . . Evans, D. G. (2013). Can the diagnosis of NF1 be excluded clinically? A lack of pigmentary findings in families with spinal neurofibromatosis demonstrates a limitation of clinical diagnosis. *J Med Genet, 50*(9), 606-613. doi:10.1136/jmedgenet-2013-101648

Carnes, R. M., Kesterson, R. A., Korf, B. R., Mobley, J. A., & Wallis, D. (2019). Affinity purification of NF1 protein-protein interactors identifies keratins and neurofibromin itself as binding partners. *Genes, in press*.

Cichowski, K., & Jacks, T. (2001). NF1 tumor suppressor gene function: narrowing the GAP. *Cell, 104*(11239415), 593-604. Retrieved from http://www.hubmed.org/display.cgi?uids=11239415

Cichowski, K., Santiago, S., Jardim, M., Johnson, B. W., & Jacks, T. (2003). Dynamic regulation of the Ras pathway via proteolysis of the NF1 tumor suppressor. *Genes Dev, 17*(4), 449-454. doi:10.1101/gad.1054703

Cui, Y., & Morrison, H. (2019). Construction of cloning-friendly minigenes for mammalian expression of full-length human NF1 isoforms. *Hum Mutat, 40*(2), 187-192. doi:10.1002/humu.23681

Dunzendorfer-Matt, T., Mercado, E. L., Maly, K., McCormick, F., & Scheffzek, K. (2016). The neurofibromin recruitment factor Spred1 binds to the GAP related domain without affecting Ras inactivation. *Proc Natl Acad Sci U S A, 113*(27), 7497-7502. doi:10.1073/pnas.1607298113

Frayling, I. M., Mautner, V. F., van Minkelen, R., Kallionpaa, R. A., Aktas, S., Baralle, D., . . . Upadhyaya, M. (2019). Breast cancer risk in neurofibromatosis type 1 is a

function of the type of NF1 gene mutation: a new genotype-phenotype correlation. *J Med Genet, 56*(4), 209-219. doi:10.1136/jmedgenet-2018-105599

Hirata, Y., Brems, H., Suzuki, M., Kanamori, M., Okada, M., Morita, R., . . . Yoshimura, A. (2016). Interaction between a Domain of the Negative Regulator of the Ras-ERK Pathway, SPRED1 Protein, and the GTPase-activating Protein-related Domain of Neurofibromin Is Implicated in Legius Syndrome and Neurofibromatosis Type 1. *J Biol Chem, 291*(7), 3124-3134. doi:10.1074/jbc.M115.703710

Hollstein, P. E., & Cichowski, K. (2013). Identifying the Ubiquitin Ligase complex that regulates the NF1 tumor suppressor and Ras. *Cancer Discov, 3*(8), 880-893. doi:10.1158/2159-8290.CD-13-0146

Kang, E., Kim, Y. M., Seo, G. H., Oh, A., Yoon, H. M., Ra, Y. S., . . . Lee, B. H. (2020). Phenotype categorization of neurofibromatosis type I and correlation to NF1 mutation types. *J Hum Genet, 65*(2), 79-89. doi:10.1038/s10038-019-0695-0

Koczkowska, M., Callens, T., Chen, Y., Gomes, A., Hicks, A. D., Sharp, A., . . . Messiaen, L. M. (2019). Clinical spectrum of individuals with pathogenic NF1 missense variants affecting p.Met1149, p.Arg1276 and p.Lys1423: genotype-phenotype study in neurofibromatosis type 1. *Hum Mutat*. doi:10.1002/humu.23929

Koczkowska, M., Callens, T., Gomes, A., Sharp, A., Chen, Y., Hicks, A. D., . . . Messiaen, L. M. (2018). Expanding the clinical phenotype of individuals with a 3-bp in-frame deletion of the NF1 gene (c.2970_2972del): an update of genotype-phenotype correlation. *Genet Med*. doi:10.1038/s41436-018-0269-0

Koczkowska, M., Chen, Y., Callens, T., Gomes, A., Sharp, A., Johnson, S., . . . Messiaen, L. M. (2018). Genotype-Phenotype Correlation in NF1: Evidence for a More Severe Phenotype Associated with Missense Mutations Affecting NF1 Codons 844-848. *Am J Hum Genet, 102*(1), 69-87. doi:10.1016/j.ajhg.2017.12.001

Korf, B. R., Henson, J. W., & Stemmer-Rachamimov, A. (2005). Case records of the Massachusetts General Hospital. Case 13-2005. A 48-year-old man with weakness of the limbs and multiple tumors of spinal nerves. *N Engl J Med, 352*(17), 1800-1808. doi:10.1056/NEJMcpc059008

Leier, A., Bedwell, D. M., Chen, A. T., Dickson, G., Keeling, K. M., Kesterson, R. A., . . . Wallis, D. (2020). Mutation-Directed Therapeutics for Neurofibromatosis Type I. *Mol Ther Nucleic Acids, 20*, 739-753. doi:10.1016/j.omtn.2020.04.012

Lin, Y. C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., . . . Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun, 5*, 4767. doi:10.1038/ncomms5767

McGillicuddy, L. T., Fromm, J. A., Hollstein, P. E., Kubek, S., Beroukhim, R., De Raedt, T., . . . Cichowski, K. (2009). Proteasomal and genetic inactivation of the NF1 tumor suppressor in gliomagenesis. *Cancer Cell, 16*(1), 44-54. doi:10.1016/j.ccr.2009.05.009

Messiaen, L. M., Callens, T., Roux, K., Mortier, G. R., Paepe, A. D., Abramowicz, M., . . . Wallace, M. R. (1999). Exon 10b of the NF1 gene represented a mutational hotspot and harbors a recurrent missense mutation y489c associated with aberrant splicing. *Genetics In Medicine, 1*, 248. doi:10.1097/00125817-199909000-00002

Rojnueangnit, K., Xie, J., Gomes, A., Sharp, A., Callens, T., Chen, Y., . . . Messiaen, L. (2015). High Incidence of Noonan Syndrome Features Including Short Stature and Pulmonic Stenosis in Patients carrying NF1 Missense Mutations Affecting p.Arg1809: Genotype-Phenotype Correlation. *Hum Mutat, 36*(11), 1052-1063. doi:10.1002/humu.22832

Ruggieri, M., Polizzi, A., Spalice, A., Salpietro, V., Caltabiano, R., D'Orazi, V., . . . Nicita, F. (2015). The natural history of spinal neurofibromatosis: a critical review of clinical and genetic features. *Clin Genet, 87*(5), 401-410. doi:10.1111/cge.12498

Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F., & Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science, 277*(5324), 333-338. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9219684

Sermon, B. A., Lowe, P. N., Strom, M., & Eccleston, J. F. (1998). The importance of two conserved arginine residues for catalysis by the ras GTPase-activating protein, neurofibromin. *J Biol Chem, 273*(16), 9480-9485. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9545275

Sherekar, M., Han, S. W., Ghirlando, R., Messing, S., Drew, M., Rabara, D., . . . Esposito, D. (2019). Biochemical and structural analyses reveal that the tumor suppressor neurofibromin (NF1) forms a high-affinity dimer. *J Biol Chem*. doi:10.1074/jbc.RA119.010934

Stowe, I. B., Mercado, E. L., Stowe, T. R., Bell, E. L., Oses-Prieto, J. A., Hernandez, H., . . . McCormick, F. (2012). A shared molecular mechanism underlies the human rasopathies Legius syndrome and Neurofibromatosis-1. *Genes Dev, 26*(13), 1421-1426. doi:10.1101/gad.190876.112

Tabata, M. M., Li, S., Knight, P., Bakker, A., & Sarin, K. Y. (2020). Phenotypic heterogeneity of neurofibromatosis type 1 in a large international registry. *JCI Insight, 5*(16). doi:10.1172/jci.insight.136262

Tan, M., Zhao, Y., Kim, S. J., Liu, M., Jia, L., Saunders, T. L., . . . Sun, Y. (2011). SAG/RBX2/ROC2 E3 ubiquitin ligase is essential for vascular and neural development by targeting NF1 for degradation. *Dev Cell, 21*(6), 1062-1076. doi:10.1016/j.devcel.2011.09.014

Trevisson, E., Morbidoni, V., Forzan, M., Daolio, C., Fumini, V., Parrozzani, R., . . . Clementi, M. (2019). The Arg1038Gly missense variant in the NF1 gene causes a mild phenotype without neurofibromas. *Mol Genet Genomic Med, 7*(5), e616. doi:10.1002/mgg3.616

Upadhyaya, M., Huson, S. M., Davies, M., Thomas, N., Chuzhanova, N., Giovannini, S., . . . Messiaen, L. (2007). An absence of cutaneous neurofibromas associated with a 3-bp inframe deletion in exon 17 of the NF1 gene (c.2970-2972 delAAT): evidence of a clinically significant NF1 genotype-phenotype correlation. *Am J Hum Genet, 80*(1), 140-151. doi:10.1086/510781

Wallis, D., Li, K., Lui, H., Hu, K., Chen, M. J., Li, J., . . . Kesterson, R. A. (2018). Neurofibromin (NF1) genetic variant structure-function analyses using a full-length mouse cDNA. *Hum Mutat, 39*, 816-821. doi:10.1002/humu.23421

Yan, W., Markegard, E., Dharmaiah, S., Urisman, A., Drew, M., Esposito, D., . . . Simanshu, D. K. (2020). Structural Insights into the SPRED1-Neurofibromin-

KRAS Complex and Disruption of SPRED1-Neurofibromin Interaction by Oncogenic EGFR. *Cell Rep, 32*(3), 107909. doi:10.1016/j.celrep.2020.107909

Zheng, Z. Y., Anurag, M., Lei, J. T., Cao, J., Singh, P., Peng, J., . . . Chang, E. C. (2020). Neurofibromin Is an Estrogen Receptor-alpha Transcriptional Co-repressor in Breast Cancer. *Cancer Cell, 37*(3), 387-402 e387. doi:10.1016/j.ccell.2020.02.003

Acknowledgements: This work was partially supported through the Gilbert Family Foundation's Gene Therapy Initiative grant numbers: 563676 to DMB and 563624 to DW.

Conflict of Interest:

Conflicts of Interest are as follows; the remaining co-authors do not have any conflicts of interest.

Bruce Korf:

- Chair, Children's Tumor Foundation Medical Advisory Committee
- Chair, External Advisory Committee for NTAP and also for NF Research Initiative
- Member, advisory committees for AstraZeneca and Springworks

Robert Kesterson is a lead advisor for

- Infixion

David Bedwell is consultant for

- PTC Therapeutics, Inc.

Deeann Wallis, Robert Kesterson, and Bruce Korf are inventors on US Provisional Patent Application No. 62/903,521 - Exon skipping to treat Neurofibromatosis Type 1.

Data Sharing: Data Availability Statement: Data available on request from the authors**.**

Figure 1

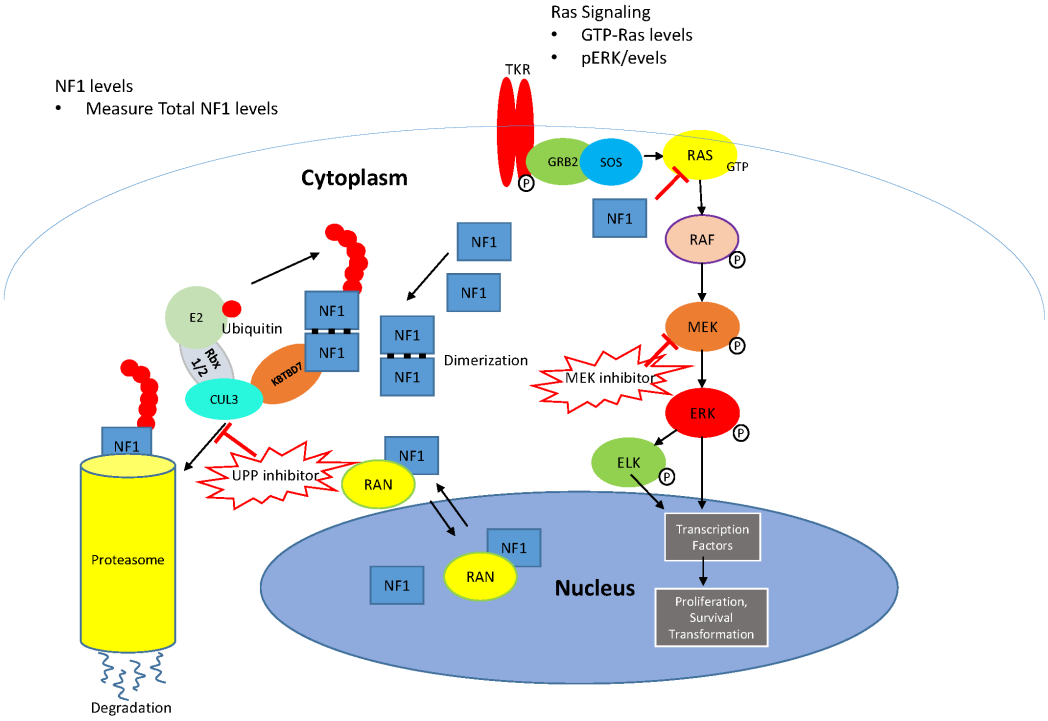

Ras Signaling
• GTP-Ras levels
• pERK/evels
NF1 levels
• Measure Total NF1 levels
TKR
Cytoplasm
GRB2
SOS
RAS
GTP
NF1
RAF
MEK
ERK
ELK
Transcription Factors
Proliferation, Survival Transformation
Nucleus
RAN
Dimerization
MEK inhibitor
UPP inhibitor
E2
Ubiquitin
Rbx 1/2
CUL3
KBTBD7
Proteasome
Degradation

Figure 2

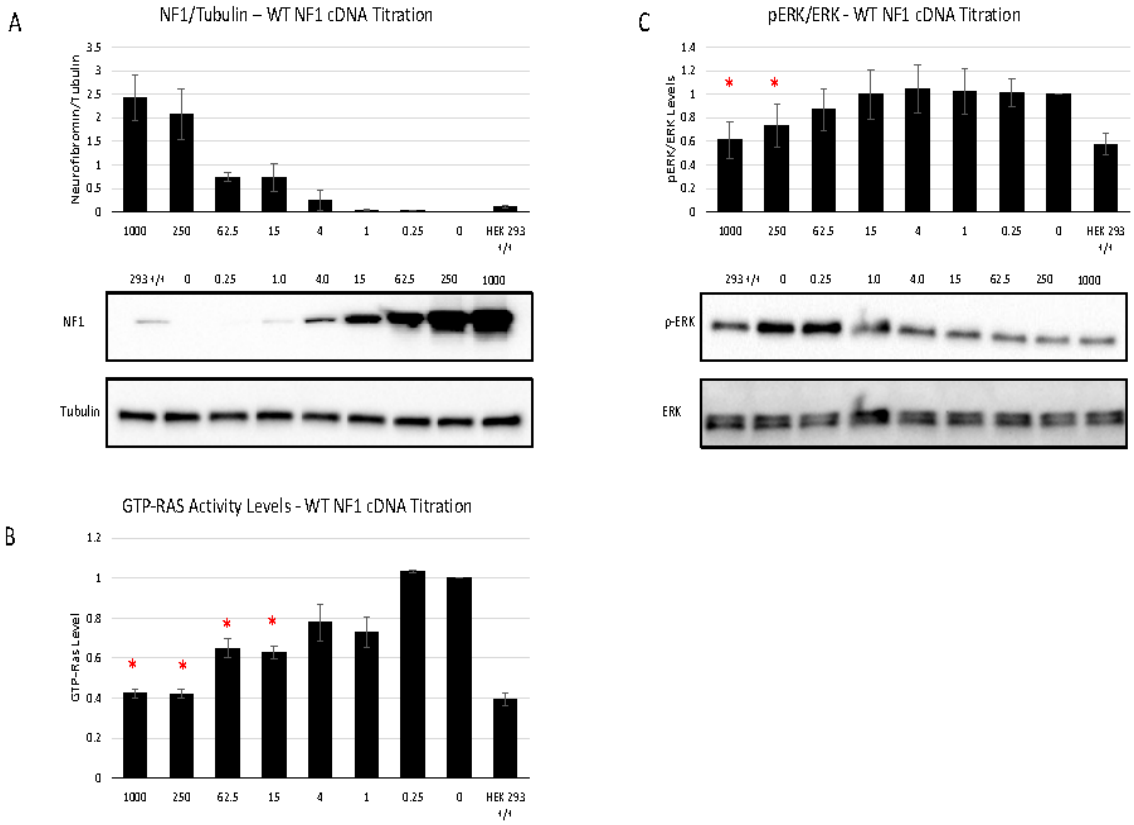

A
NF1/Tubulin – WT NF1 cDNA Titration
Neurofibromin/Tubulin
1000 250 62.5 15 4 1 0.25 0 HEK 293 +/+
293 +/+ 0 0.25 1.0 4.0 15 62.5 250 1000
NF1
Tubulin
C
pERK/ERK - WT NF1 cDNA Titration
pERK/ERK Levels
1000 250 62.5 15 4 1 0.25 0 HEK 293 +/+
293 +/+ 0 0.25 1.0 4.0 15 62.5 250 1000
p-ERK
ERK
B
GTP-RAS Activity Levels - WT NF1 cDNA Titration
GTP-Ras Level
1000 250 62.5 15 4 1 0.25 0 HEK 293 +/+

Figure 3

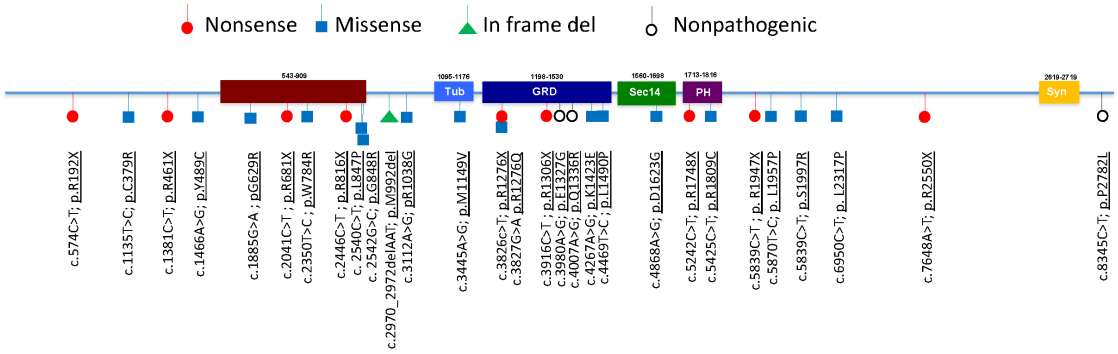

Nonsense
Missense
In frame del
Nonpathogenic
543-909
1095-1176
Tub
1198-1530
GRD
1560-1698
Sec14
1713-1816
PH
2619-2719
Syn
c.574C>T; p.R192X
c.1135T>C; p.C379R
c.1381C>T; p.R461X
c.1466A>G; p.Y489C
c.1885G>A ; pG629R
c.2041C>T ; p.R681X
c.2350T>C ; p.W784R
c.2446C>T ; p.R816X
c. 2540C>T; p.L847P
c. 2542G>C; p.G848R
c.2970_2972delAAT; p.M992del
c.3112A>G; pR1038G
c.3445A>G; p.M1149V
c.3826c>T; p.R1276X
c.3827G>A p.R1276Q
c.3916C>T ; p.R1306X
c.3980A>G; p.E1327G
c.4007A>G; p.Q1336R
c.4267A>G; p.K1423E
c.4469T>C ; p.L1490P
c.4868A>G; p.D1623G
c.5242C>T; p.R1748X
c.5425C>T; p.R1809C
c.5839C>T ; p. R1947X
c.5870T>C; p. L1957P
c.5839C>T; p.S1997R
c.6950C>T; p. L2317P
c.7648A>T; p.R2550X
c.8345C>T; p.P2782L

Figure 4

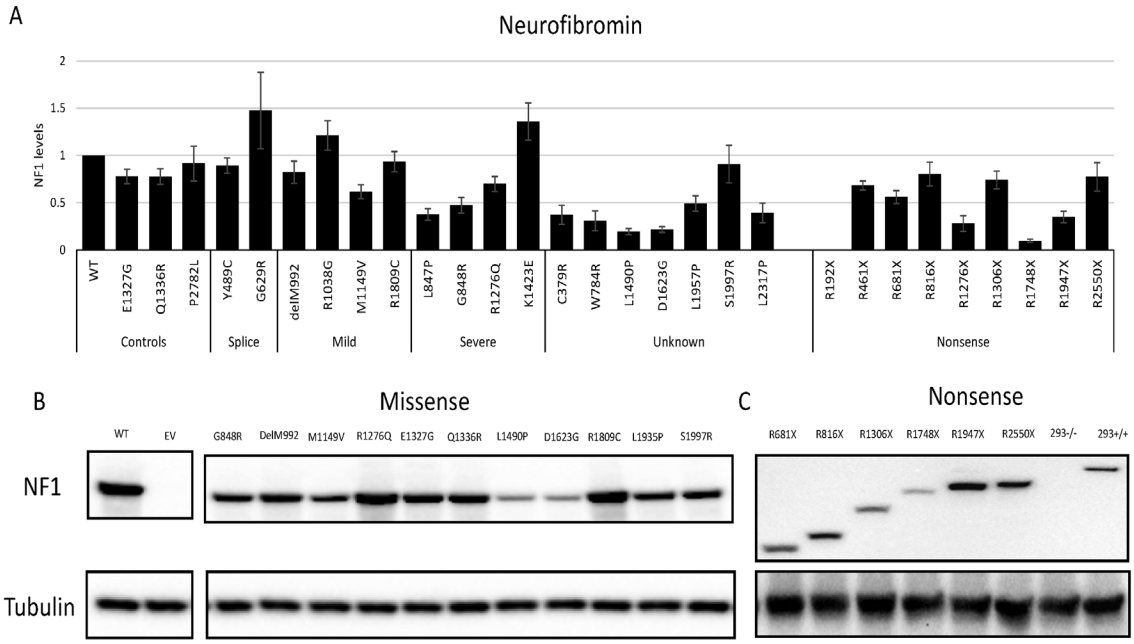

A
Neurofibromin
NF1 levels
2
1.5
1
0.5
0
WT
E1327G
Q1336R
P2782L
Y489C
G629R
delM992
R1038G
M1149V
R1809C
L847P
G848R
R1276Q
K1423E
C379R
W784R
L1490P
D1623G
L1957P
S1997R
L2317P
R192X
R461X
R681X
R816X
R1276X
R1306X
R1748X
R1947X
R2550X
Controls
Splice
Mild
Severe
Unknown
Nonsense
B
Missense
WT
EV
G848R
DelM992
M1149V
R1276Q
E1327G
Q1336R
L1490P
D1623G
R1809C
L1935P
S1997R
NF1
Tubulin
C
Nonsense
R681X
R816X
R1306X
R1748X
R1947X
R2550X
293-/-
293+/+

Figure 5

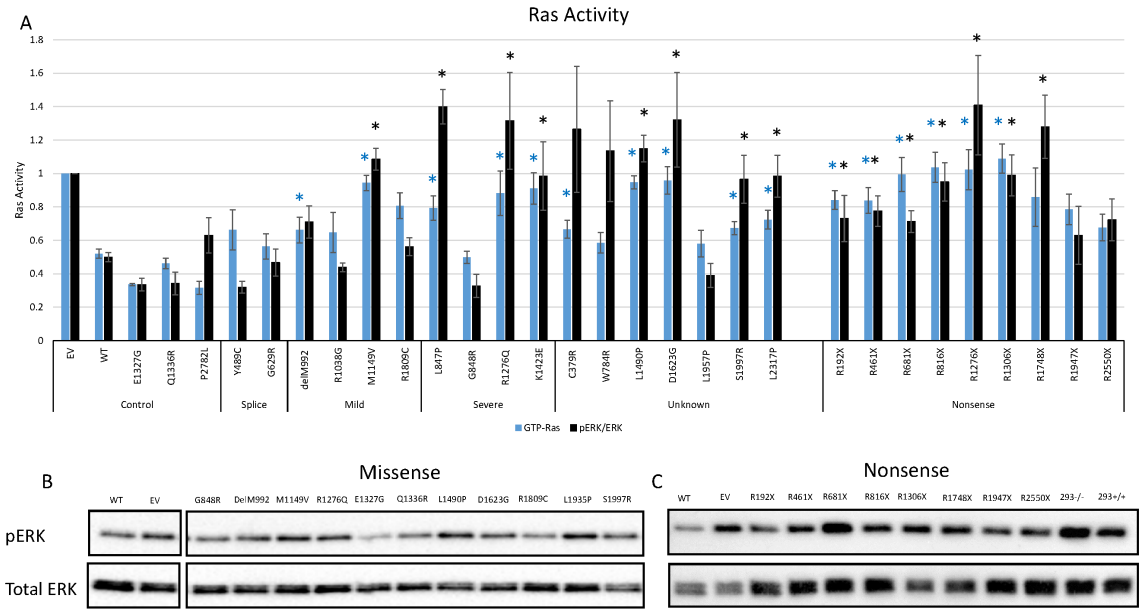
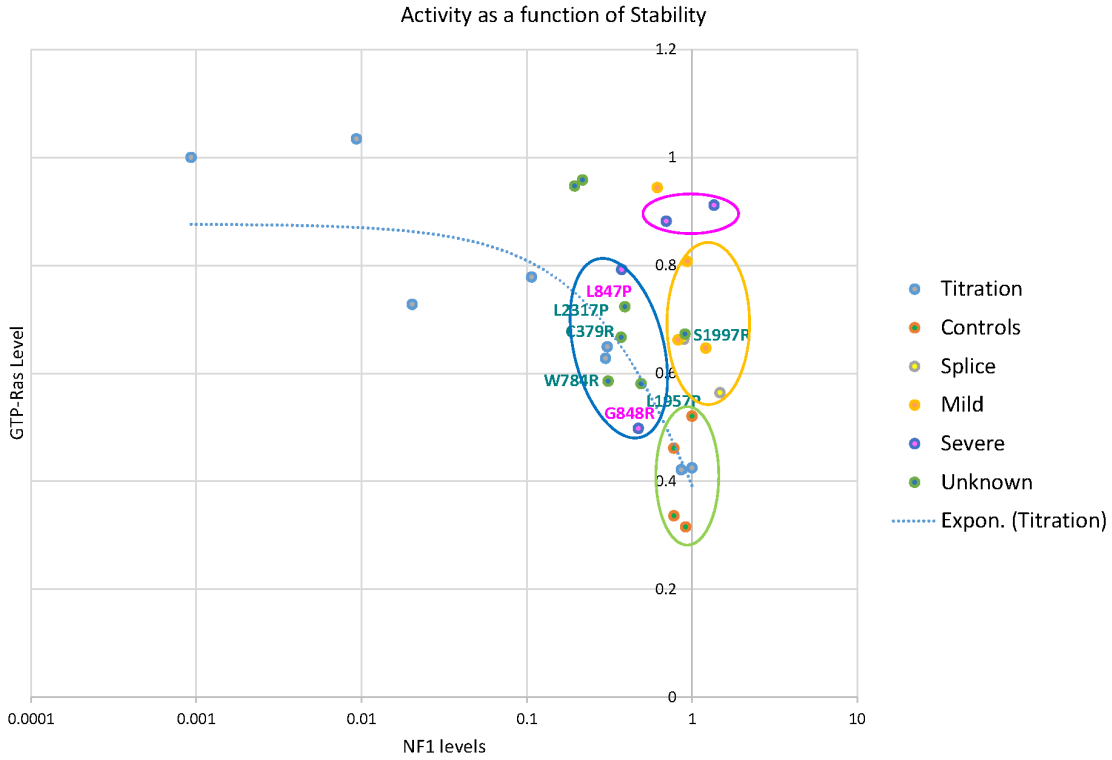

A
Ras Activity
Ras Activity
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
EV
WT
E1327G
Q1336R
P2782L
Y489C
G629R
delM992
R1038G
M1149V
R1809C
L847P
G848R
R1276Q
K1423E
C379R
W784R
L1490P
D1623G
L1957P
S1997R
L2317P
R192X
R461X
R681X
R816X
R1276X
R1306X
R1748X
R1947X
R2550X
Control
Splice
Mild
Severe
Unknown
Nonsense
GTP-Ras
pERK/ERK
B
Missense
WT
EV
G848R
DelM992
M1149V
R1276Q
E1327G
Q1336R
L1490P
D1623G
R1809C
L1935P
S1997R
pERK
Total ERK
C
Nonsense
WT
EV
R192X
R461X
R681X
R816X
R1306X
R1748X
R1947X
R2550X
293-/-
293+/+

Figure 6


Activity as a function of Stability
GTP-Ras Level
NF1 levels
0.0001
0.001
0.01
0.1
1
10
0
0.2
0.4
0.6
0.8
1
1.2
L847P
L2317P
C379R
W784R
G848R
S1997R
Titration
Controls
Splice
Mild
Severe
Unknown
Expon. (Titration)

FUNCTIONAL CHARACTERIZATION AND POTENTIAL THERAPEUTIC AVENUES FOR VARIANTS IN THE NTRK2 GENE CAUSING DEVELOPMENTAL AND EPILEPTIC ENCEPHALOPATHIES

by

ASHLEE LONG, ANDREW CROUSE, ROBERT A. KESTERSON, MATTHEW MIGHT, DEEANN WALLIS

In preparation for *American Journal of Medical Genetics Part B*

Format adapted for dissertation

Abstract

Variants within the *Neurotrophic Tyrosine Kinase Receptor Type 2* (*NTRK2)* gene have been discovered to play a role in developmental and epileptic encephalopathies, a group of debilitating conditions for which little is known about cause or treatment. Here we determine the functional consequences of two variants: p.Tyr434Cys (Y434C) (located in the transmembrane domain) and p.Thr720Ile (T720I) (located in the catalytic domain). Wild-type (WT) and variant cDNAs were constructed and transfected into HEK293 cells. In cell culture, variant Y434C exhibited ligand-independent activation of TRKB signaling with an associated abnormal response to brain derived neurotrophic factor stimulation and increased levels of phosphorylated ERK and ELK1 activity. Expression of variant T720I, resulted in decreased TRKB signaling with reduced mTor activity as determined by decreased levels of phosphorylated S6. With the deleterious mechanisms characterized, we utilized mediKanren (a novel artificial intelligence tool) to identify therapeutics to compensate for the pathological effects. Downregulation of TRKB, through inhibition with mediKanren-predicted compound 1NM-PP1 led to decreased MEK activity. Upregulation of TRKB signaling by mediKanren-predicted valproic acid led to subsequent increase of mTor activity. Overall, our results provide further characterization of the pathogenicity of these two variants in the *NTRK2* gene. Indeed, Y434C is the first patient-specific *NTRK2* variant with demonstrated hypermorphic activity. Furthermore, we observed that variants Y434C and T720I result in distinct functional consequences that require distinct therapeutic strategies. These data suggest the possibility that unique mutations within different regions of the *NTRK2* gene results in separate clinical presentations, representing distinct genetic disorders requiring unique therapeutics.

Introduction

The advent of next generation sequencing technologies and the reduction of costs has enabled the wide-spread use of patient DNA sequencing in the clinic. Increasingly, clinicians are faced with genetic changes which have either unknown clinical significance (which may prohibit providing a molecular diagnosis) or unknown impact on gene function (which undermines the ability to develop therapeutic strategies). *In vitro* cellular assays can be used to determine 1.) if a variant impacts gene function and thus provide support for its clinical significance and 2.) how a variant impacts gene function and thus how to compensate for it with potential therapeutics. Here we investigate the impact of two variants within the *NTRK2* gene whose clinical significance has been determined, but whose functional impact has not [1].

NTRK2, located on chromosome 9q21 encodes for the tropomyosin related kinase B (TRKB) receptor protein that localizes to the central and peripheral nervous systems [2-4]. TRKB belongs to the tyrosine protein kinase superfamily, sharing homology with TRKA and TRKC, which are receptors for a group of neurotrophic factors, including brain derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3) and neurotrophin-4 (NT-4). TRKB binds mainly to BDNF and NT-4 and with a lower affinity to NT-3 [5-7]. TRK receptor-neurotrophin signaling is important in the regulation of neurogenesis, proliferation and differentiation of neural precursors, as well as axon and dendrite growth [8-11].

Disruption of the fine balance of BDNF/TRKB signaling results in a wide range of pathologies. Down regulation of BDNF/TRKB signaling has been observed in certain psychiatric states, including schizophrenia as well as increased stress-phenotypes, such as anxiety [12-16] as well as neurodegenerative disorders such as Alzheimer's and Huntington disease [12; 17-19]. Increased expression of TRKB is associated with multiple types of malignancies including brain, lung, and blood cancers [12; 20; 21]. Increased activation of TRKB via BDNF results in the activation of the MAPK, PI3K, and JAK-STAT3 signaling pathways, promoting tumor growth and metastasis [8; 12; 20; 22; 23]. Of note, dysfunctional BDNF/TRKB signaling has also been implicated in the pathophysiology of epilepsy. Increased TRKB signaling, through increased BDNF-mediated activation, results in epileptogenesis [24-28]. This is supported by the observation that mice with a heterozygous

Bdnf gene deletion exhibit suppressed epileptogenesis [28]. Several studies have also analyzed the influence that *NTRK2* single nucleotide polymorphisms (SNPs) have on the development of epilepsy. Patients with specific *NTRK2* SNPs exhibited a younger age of seizure onset with a greater incidence of drug-resistant epilepsy [29; 30].

While general changes in BDNF/TRKB expression or signaling have been associated with complex common diseases, specific pathological variants in NTRK2 have been identified in patients with rare neurological genetic disorders [1]. However, the specific underlying molecular mechanisms of pathology of the Mendelian disorders remains unknown. Whole genome sequencing performed on patients diagnosed with severe, early-onset epilepsy and intellectual disability found two *de novo* NTRK2 variants, Y434C located in the transmembrane domain and T720I located in the catalytic domain [1]. Y434C was observed in four individuals and is associated with developmental and epileptic encephalopathy (DEE) (OMIM: 617830), a severe neurodevelopmental disorder characterized by the onset of infantile spasms and refractory seizures in the first days or months of life [1]. T720I was observed in only one patient with DEE and hyperphagia. Interestingly, other NTRK2 variants including nearby p.Tyr722Cys have been associated with obesity, hyperphagia, and developmental delay (OMIM: 613886) a neurodevelopmental disorder characterized by global developmental delay and hyperphagia resulting in severe early-onset obesity and sometimes associated with other neurological function defects including absence seizures [1; 31; 32]. Hence, the two variants (Y434C and T720I) display overlapping, yet potentially distinct phenotypes.

Given the multifaceted role of TRKB, these phenotypes may result from different underlying molecular mechanisms that would require consideration for effective therapy. Herein, we describe in vitro functional data indicating that these variants differed in their impact on TRKB signaling. *NTRK2* Y434C exhibited ligand-independent activation of TRKB signaling with increased MAPK signaling activity. In contrast, *NTRK2* T720I exhibits hypomorphic activity with decreased MAPK and mTOR signaling activity. Additionally, we explored potential therapeutic avenues through utilization of the artificial intelligence-based software program mediKanren. Treatment with compound 1NM-PP1, a selective tyrosine kinase inhibitor, abrogated the increased activity observed with the *NTRK2* Y434C variant. In contrast, treatment with valproic acid, a commonly utilized

anti-epileptic drug, increased mTor activity in variant *NTRK2* T70I. Our data indicates that these two *NTRK2* variants represent distinct genetic disorders and propose potential therapeutic avenues which address their specific mechanisms of action.

Methods

NTRK2 plasmids

We obtained NTRK2 ORF cDNA in the mammalian expression vector pcDNA3.1+ (GeneScript ref. OHu24187D) with custom mutagenesis by GeneScript to create the Y434C and T720I mutations. Both WT and variant constructs were confirmed through sequencing once plasmids were prepared in our lab. DNA for transfections was prepared using Qiagen's Endotoxin free plasmid kits.

Cell culture and transfections

HEK293 cells were obtained from ATCC (CRL-1573) and cultured in DMEM (ref. 11995-065; Gibco) + 10% FBS and 1X penicillin-streptomycin using standard culture procedures. The HEK293 cell line was selected for analysis of TRKB expression and activity because it is well characterized, easily takes up exogenous DNA, and is easy to culture and scale. HEK293s do not express TrkB endogenously, but have all three Ras isoforms and recapitulate Ras-MAPK signaling [33]; hence, HEK293 cells are an appropriate model system for these assays. Cells were plated in a 6-well plate, and at approximately 80% confluency cells were transfected using LipoD293 Reagent (ref. #SL00668; SignaGen Laboratories) with 1ug of *NTRK2* wild-type and variant cDNAs, purchased from GeneScript. Drug treatments were performed at 24 hrs post-transfection. At 48 hrs post-transfection, cells were serum starved for 24 hrs and protein lysates were harvested. BDNF (ref. 450-02-10UG; PeproTech) stimulation was performed 5 minutes prior to protein harvest.

MediKanren and drug treatments

Potential treatments were chosen using mediKanren, an artificial intelligence agent developed by The Hugh Kaul Precision Medicine Institute at The University of Alabama at Birmingham as part of the NIH NCATS Biomedical Data Translator consortium [34]. Operationally, mediKanren is a biomedical reasoning system composed of a custom knowledge-graph database and a constraint-based logic reasoning engine, developed from the miniKanren family of relational-reasoning logic languages. The reasoning engine finds (bio)logical relationships between biomedical concepts that include molecules, drugs, proteins, pathways, genes, diseases, symptoms, phenotypes. Valproic acid (ref. 2815; Tocris) was flagged as a potential downregulator of TRKB, and PP1 Analog II, 1NM-PP1 (ref. 529581-1MG; EMD Millipore) was flagged as a potential upregulator of TRKB activity. Additional drug treatments included Selumetinib (ref. AZD6244; Selleckchem), a MEK inhibitor.

Western blotting and antibodies

Protein lysates were prepared using Pierce™ RIPA Buffer (ref. 89900; Thermo Scientific), a protease inhibitor cocktail (ref. 04 693 116 001; Roche), and a phosphatase inhibitor (ref. 04 906 837 001; Roche). Protein concentration was determined using Pierce™ BCA Protein Assay Kit (ref. 23225; Thermo Fisher Scientific). 50 micrograms of protein lysate were electrophoresed on Mini-PROTEAN TGX 4-20% gels (ref. #4561093; BIO-RAD) and transferred onto polyvinylidene fluoride membranes (ref. #1620177; BIO-RAD) and blocked with powdered milk. Human phospho- and total TRKB was detected with the monoclonal anti-P-TrkA (Y490) TrkB (Y516) antibody and anti-TrkB antibody respectively (ref. 4619S; Cell Signaling Technology: ref. 4606S; Cell Signaling Technology). Human phospho- and total ERK was detected with the anti-P-p44/42 MAPK (T202/Y204) antibody and the anti-p44/42 MAPK antibody respectively (ref. 9101L; Cell Signaling Technology: ref. 9102S; Cell Signaling Technology). Human phospho- and total S6 was detected with the anti-P-S6 ribosomal Protein (S240/244) antibody and the anti-S6 Ribosomal Protein (5610) antibody (ref. 2215S; Cell Signaling Technology: ref. 2217S;

Cell Signaling Technology). Tubulin was detected with the monoclonal anti-alpha-Tubulin antibody (ref. 3873S; Cell Signaling Technology). All antibodies were used at a 1:1000 dilution, except for tubulin which was used at a 1:5,000 dilution. Primary antibody incubation was performed for at least 12 hours at 4C. Goat polyclonal anti-rabbit immunoglobulin (ref. ab97051; Abcam), linked to horseradish peroxidase, was used as a secondary antibody at a 1:5,000 dilution for at least 1 hour at room temperature. Signal was exposed using Clarity™ Western ECL Substrate (ref. #170-5061; BIO-RAD) and measured using a ChemiDoc™ MP Imaging System (BIO-RAD) and Image Lab 6.0.1. Quantification was performed using ImageJ software (National Institutes of Health).

Luciferase reporter assay

ELK1 activity was analyzed using the ELK-TAD Luciferase Reporter HEK293 stable cell line (ref. SL-0040-FP; Signosis). Cell culture and transfection of full length NTRK2 cDNAs were performed as stated above. Reporter activity was measured using a BioTek Synergy 2 plate reader and Gen5 Microplate Reader and Imager Software 3.02.

Statistical analyses

Statistical analyses were done using Excel Office 365. Three or more separate experiments were analyzed for each condition and a Student's t-test was performed to determine significance. For all analyses, a p-value less than 0.05 was considered significant.

Results

Cells overexpressing Y434C and T720I exhibit an impaired response to BDNF stimulation at the TRKB receptor

To examine the effects of *NTRK2* variants Y434C and T720I compared to wildtype (WT) on TRKB signaling, we transfected full-length WT or mutant cDNA constructs into HEK293 cells. TRKB and phospho-TRKB (pTRKB) expression was confirmed through

western blot analysis (Fig. 1a) with an empty vector (EV) as a control. For the T720I variant, total TrkB is modestly reduced by ~20% ($p<0.05$) suggesting possible reduction in protein stability or half-life. In addition, variant T720I exhibited more than two-fold reduced pTRKB expression (normalized to total-TrkB) compared to WT (p-value < 0.05, Fig. 1a, b) while Y434C expressed pTRKB at similar levels to WT. We analyzed how ligand-mediated activation of the TRKB receptor is affected using BDNF stimulation in cells expressing variant NTRK2 protein. Transfected cells overexpressing either Y434C, T720I or WT were stimulated with BDNF at a concentration of 10ng/mL, 5 minutes prior to protein harvest. Cells transfected with WT *NTRK2* and stimulated with BDNF exhibited an approximate two-fold increase in pERK expression, an indication of activated TRKB mediated Ras-MAP kinase (MAPK) activity, when compared to unstimulated cells (Fig. 1c, d). Interestingly, the two variants had opposite impacts on pERK signaling. The cells expressing the Y434C variant with or without BDNF-stimulation had expression levels of pERK similar to the BDNF-stimulated WT expressing cells indicating possible ligand independent constitutive activation of TRKB. Cells expressing the T720I variant with or without BDNF-stimulation expressed pERK at a similar level as unstimulated WT expressing cells suggesting loss or suppression of activation response (Fig. 1c, d).

Previous studies have analyzed TRKB activation of MAPK activity in variant Y722C located just two amino acids from T720I in the NTRK2 catalytic domain and demonstrated that the Y722C variant resulted in impaired pERK/ERK ratios over a range of BDNF treatments [31]. To explore possible similarities between the variants' impact, we treated cells overexpressing WT or T720I NTRK2 with increasing concentrations of BDNF at 0, 0.5, 10, 50, and 200ng/mL. While both WT and T720I cells show increasing pERK/ERK ratios with increasing BDNF concentrations (Fig. 2a, b), the T720I expressing cells demonstrate a loss in magnitude of response which seems to plateau at around 10 ng/ml BDNF, suggesting a blunted maximal response. These results are in agreement with what has previously been published on the *NTRK2* variant Y722C which exhibits a hypomorhpic activity and similar patient phenotypes [31]. While the only statistically significant difference was seen at the highest dose, the average signaling from variant T720I was lower at all concentrations.

NTRK2 ***variants Y434C and T720I exhibit differing functional consequences on the basal activity of TRKB downstream signaling molecules***

We analyzed the consequences of each variant on downstream intracellular signaling, focusing on Ras-MAP kinase (MAPK) and PI3K-mTOR signaling. We used pERK/ERK ratios and ELK-1 transcriptional activity to evaluate the Ras pathway and pS6 to evaluate mTOR activity. Upon overexpression of the NTRK2 Y434C variant, baseline activity of pERK signaling was significantly increased, greater than two-fold, compared to WT expressing cells (p-value < 0.05, Fig 1 c, d and Fig. 3 a, b). In contrast, pERK levels of T720I were significantly lower than WT (Fig1 c, d and Fig 3 a, b). These data suggest that RAS-signaling activity is increased by the Y434C variant and decreased in the T720I variant. Additionally, ELK1 expression, another downstream molecule of the TRKB-MAPK signaling pathway, was significantly increased by greater than two-fold, upon Y434C overexpression compared to WT (p-value < 0.01, Fig. 3c). In contrast, ELK1 signaling in cells overexpressing the NTRK2 T720I variant were significantly lower than WT.

Next, we determined the consequences of the two NTRK2 variants on mTOR activity by analyzing their effects on pS6. In cells overexpressing variant T720I, pS6 was significantly decreased by 60% compared to WT (p-value, 0.05, Fig. 3d, e) with no change observed in Y434C overexpressing cells. These data indicate that *NTRK2* variants Y434C and T720I result in distinct regulatory consequences. The Y434C variant exhibits BDNF independent signaling through the RAS pathway, but does not impact PI3K pathway, while the T720I variant decreases signaling through both the RAS and mTOR pathways.

These results provide evidence that cells expressing variant T720I respond to BDNF stimulation of the TRKB receptor at approximately 50% reduced levels compared to WT. The decreased response to BDNF stimulation and the observed lower levels of MAPK activity support a downregulation mechanism of action for this NTRK2 variant. These data also support that cells overexpressing variant Y434C exhibit an increase in basal activity of TRKB signaling which is independent from BDNF stimulation and provides evidence that this variant results in constitutive TRKB activation compared to WT NTRK2.

MediKanren predicted treatments 1NM-PP1 and Valproic Acid attenuate the effects of NTRK2 Y434C and T720I variants respectively

Given the differing molecular impact of the two variants, different therapeutic strategies may be needed for patients. mediKanren, an artificial intelligence agent developed internally as part of the NIH NCATS Biomedical Data Translator consortium (https://ncats.nih.gov/translator), was used to identify potential novel compounds that compensate for the changes in signaling. mediKanren is a biomedical reasoning system composed of a custom knowledge-graph database and a constraint-based logic reasoning engine, developed from the miniKanren family of relational-reasoning logic languages [35]. Its database incorporates over 180 distinct sources of biomedical knowledge, including the scientific literature (Pubmed), data on approved and unapproved drugs (via FDA and drugbank), and genes (NCBI, UniProt) with the goal of identifying therapeutic candidates for both complex and rare disorders for which no effective standard of care exists. (Source code for mediKanren software is available on github (https://github.com/webyrd/mediKanren).

mediKanren was utilized to determine possible therapeutic avenues for NTRK2 Y434C with a focus on chemical substances that negatively regulate NTRK2 in order to attenuate the increased MAPK signaling observed in cells expressing Y434C. One potential compound predicted by mediKanren is 1NM-PP1; a small-molecule inhibitor that is known to downregulate TRKB through inhibition of BDNF/TRKB autophosphorylation [36]. We preferentially decided to test 1NM-PP1, over the commonly used TRK inhibitor K252a (which was also predicted by MediKanren), due to previous literature demonstrating compound 1NM-PP1's selectiveness and reduced toxicity [37]. Cells overexpressing either WT or Y434C variant cDNAs were treated with three dosages of 1NM-PP1 ranging from 3-10uM (Fig. 4a-d). At doses of 3uM and higher, 1NM-PP1 significantly decreased pTRKB levels by 50% or greater compared to untreated cells. At 5uM and greater, 1NM-PP1 significantly decreased pERK levels by at least 20% or greater in both WT and Y434C expressing cells (p-value < 0.05, Fig. 4a-d). These data provide evidence that the constitutive TRKB activation observed in variant Y434C overexpressing cells can be reduced when treated with the selective TRKB inhibitor 1NM-PP1 resulting in decreased TRKB-MAPK signaling. This provides further support that the increased

MAPK signaling observed with the Y434C variant is due to a ligand-independent mechanism which cannot be entirely rescued through simple pTRKB inhibition. While 1NM-PP1 is not an FDA approved drug the results suggest a therapeutic strategy for patients with this type of variant to decrease pTRKB levels.

To determine whether the excessive pERK activity can be inhibited in cells overexpressing NTRK2 variants, particularly Y434C, we treated them with Selumetinib, a known MEK inhibitor. As anticipated, cells overexpressing WT, Y434C, or T720I NTRK2 exhibited significantly decreased pERK activity with Selumetinib at both 1uM and 10 uM dosages (Sup. Fig. 1a, b). This indicates that these cellular pathways are responsive to modulation downstream as indicated by the MEK inhibitor Selumetinib.

mediKanren was similarly used to determine potential treatments for *NTRK2* variant T720I. For this variant, therapeutics which were predicted to upregulate TRKB signaling were targeted. mediKanren identified valproic acid (VPA), a compound with known anticonvulsant activity and a common treatment used in epilepsy. A patient harboring the T720I variant has previously been shown to have a positive response to VPA, but the mechanism of action was unknown [1]. Variant T720I overexpressing cells were treated with three concentrations of VPA ranging from 500-1000uM (Fig. 5a, b). Increasing doses of VPA increased pS6 in T720I overexpressing cells up to greater than two-fold when compared to untreated cells (p-value < 0.05, Fig. 5a, b). A trend toward increased pERK activity with treatment of VPA was also observed in T720I expressing cells, but this failed to reach significance. Thus, the observed decreased pS6 in cells overexpressing T720I can be compensated for with VPA.

Discussion

We characterize the functional consequences of two *NTRK2* variants found in patients with rare genetic disorders: Y434C and T720I (Fig. 6). We provide evidence that variant T720I, located in the tyrosine kinase domain of *NTRK2*, functions similarly to other previously described *NTRK2* variants located in the catalytic domain and results in a hypomorphic allele with decreased activation of MAPK and mTOR signaling. Variant Y434C, in contrast, exhibits increased and ligand-independent MAPK signaling. Our data

provides a possible explanation for the mechanisms underlying the different patient phenotypes. Additionally, these data provide proof of principle in the utilization of mediKanren's ability to use underlying molecular mechanisms to predict possible therapeutic strategies that would compensate for changes in signaling or gene expression.

Variant T720I is a previously unstudied *NTRK2* variant located in the tyrosine kinase domain. The majority of *NTRK2* variants previously described also occur in the tyrosine kinase catalytic domain and result in impaired TRKB signaling: P660L, R691H, R696K, S714F, R715Q, R715W, Y722C, and T821A [31; 32; 38]. These variants predominately exhibited decreased BDNF-NTRK2 signaling consistent with hypomorphic activity [31; 32; 38]. We observed decreased MAPK and mTOR signaling which is in agreement with these other variants. Interestingly, hypomorphic *NTRK2* alleles have predominantly been linked to obesity. Patients that harbor the *NTRK2* Y722C and T720I variants exhibit hyperphagia with severe and early onset obesity [1; 31]. It was reported that variant Y722C also resulted in impaired neurite growth and decreased MAPK-AKT activation [31; 32; 38]. Additionally, a genetic analysis performed on extreme obesity patients identified a previously unreported *NTRK2* variant also located in the tyrosine kinase domain: S667W [39]. While clustering and the recurrence of rare variants in the catalytic domain of TrkB (and not elsewhere) raise the possibility that they confer specific properties to the protein, it is difficult to rule out the possibility that the variants induce a dominant negative effect like that reported for the truncated TrkB that results in inhibition of the full length allele and sequestration of BDNF [40]. However, conditional deletion of *TrkB* using *Rgs-Cre* in mouse models and a TrkB hypomorphic mutant in which TrkB is expressed at ~25% of the normal amount throughout the body both resulted in the development of hyperphagic obesity [41; 42], suggesting that the alleles are hypomorphic and not dominant negative. Notably, mouse models with patient-specific *NTRK2* variants have not yet been reported. Interestingly only the Y722C and T720I variants show seizure activity in addition to the obesity phenotype [1; 32]. Perhaps this particular region of the TK domain is essential for this function. Indeed, this appears to be the first report where decreased BDNF/TRKB signaling lead to an epileptic phenotype.

To determine potential therapeutic options, we utilized mediKanren to predict treatments that potentially result in increased NTRK2 signaling. Valproic acid (VPA), a well-known anti-epileptic drug, was predicted by mediKanren as a potential therapeutic

and several studies indicate that VPA stimulates TRKB and MAPK activity using similar doses to this study [43-46]. Other studies suggest that VPA may result in decreased BDNF/TRKB signaling [47]. Additionally, previous studies have demonstrated that VPA is effective in promoting neurite growth [43]. Overexpression of T720I resulted in attenuated TRKB signaling with a subsequent decrease in TRKB mediated mTOR activity, but mediKanren-predicted VPA treatment increased pS6 with a trend towards increasing MAPK activity as well. Notably, while an *NTRK2* T720I patient responded favorably to VPA treatment, patients harboring Y434C, the second variant characterized in this study, did not exhibit the same favorable response [1]. Interestingly, *NTRK2* has been linked to drug-resistant epilepsy [29]. However, the drug-resistance may be allele specific, and these differences could potentially be explained by the differing functional consequences of each variant.

NTRK2 variant Y434C exhibited an upregulation of BDNF/TRKB signaling. Increased activation of TRKB signaling has already been suggested to promote epilepsy and increased NTRK2 is associated with drug-resistant epilepsy [25; 29; 48]. In cells overexpressing Y434C, constitutive ligand-independent TRKB signaling with associated increased MAPK activity was observed. The variant's location in the transmembrane domain could explain its unique effects. Additionally, patients with the Y434C variant exhibit a different phenotype with the development of severe and early onset seizures but lacking the hyperphagic obesity observed in hypomorphic *NTRK2* variants. mediKanren-predicted 1NM-PP1 significantly reduced TRKB receptor activation and decreased downstream signaling molecules which represents a potential anti-seizure therapeutic avenue as previously suggested [36; 48]. Interestingly, while compound 1NM-PP1 has previously been shown to selectively target and inhibit mutant NTRK2 with TRKB kinase inhibition, we observed a response in WT expressing cells with reductions in pTRKB activity and pERK. While the literature observes that 1NM-PP1 does exhibit greater selectivity towards specific mutants over WT, inhibition of WT protein has been observed in previous studies [49; 50].

Notably, this is the first characterization of an *NTRK2* patient variant located outside of the tyrosine kinase domain that exhibits increased MAPK signaling when compared to WT. While entirely speculative, there are several possible mechanisms that

might account for the gain of function effect. It is classically understood that ligand binding of the TRKB receptor induces signal transduction through dimerization of the receptor, tyrosine kinase activation, and internalization of the ligand-receptor complex. Internalization seems critical in signal transduction and the internalized receptor remains phosphorylated and activated with its extracellular domain bound to the ligand inside the signaling endosomes [51; 52]. It is possible that the variant does not require ligand binding to induce tyrosine kinase activity or internalization. While ligand binding induces structural changes in the receptor dimers which propagate to their transmembrane domains and push the transmembrane domains further apart [53], Y434C could create such an "activated" conformation and allow cross-phosphorylation to occur without ligand binding. Alternatively, Trks can interact laterally and form dimers in the absence of ligand. Indeed, TrkB exhibits a propensity for homodimerization [53]. Further, TrkB may also signal without dimerization at the plasma membrane [33]. Indeed monomer rates of ~90% were seen and BDNF binds monomeric TrkB on the plasma membrane to activate ERK1/2 signaling [33]. Thus, the variant could alter the spatial organization of the receptor in terms of its localization to the plasm membrane or its internalization within the endosome resulting in increased signaling from the plasma membrane.

In summary, our data emphasizes the importance of understanding how multiple variants, even within the same gene, affect downstream functions differently. This study highlights the need for this type of precision medicine approach in determining patient-specific treatments.

Funding

Funding provided by the UAB-HudsonAlpha Center for Genomic Medicine: Pilot project in Genomic Medicine

Acknowledgements None

Competing interests None

Ethics approval statement Not Applicable

Contributorship Statement (Specific contribution of authors in the paper if applicable): Author contributorship is as follows: Conception or design of the work: MM, AC, and DW. Data collection: AL. Data analysis and interpretation: AL, AC, RAK, and DW. Drafting the article: AL, AC, MM, and DW. Critical revision of the article: AC, RAK, and DW. Final approval of the version to be published: DW.

References

1. Hamdan FF, Myers CT, Cossette P, Lemay P, Spiegelman D, Laporte AD, Nassif C, Diallo O, Monlong J, Cadieux-Dion M, Dobrzeniecka S, Meloche C, Retterer K, Cho MT, Rosenfeld JA, Bi W, Massicotte C, Miguet M, Brunga L, Regan BM, Mo K, Tam C, Schneider A, Hollingsworth G; Deciphering Developmental Disorders Study, FitzPatrick DR, Donaldson A, Canham N, Blair E, Kerr B, Fry AE, Thomas RH, Shelagh J, Hurst JA, Brittain H, Blyth M, Lebel RR, Gerkes EH, Davis-Keppen L, Stein Q, Chung WK, Dorison SJ, Benke PJ, Fassi E, Corsten-Janssen N, Kamsteeg EJ, Mau-Them FT, Bruel AL, Verloes A, Õunap K, Wojcik MH, Albert DVF, Venkateswaran S, Ware T, Jones D, Liu YC, Mohammad SS, Bizargity P, Bacino CA, Leuzzi V, Martinelli S, Dallapiccola B, Tartaglia M, Blumkin L, Wierenga KJ, Purcarin G, O'Byrne JJ, Stockler S, Lehman A, Keren B, Nougues MC, Mignot C, Auvin S, Nava C, Hiatt SM, Bebin M, Shao Y, Scaglia F, Lalani SR, Frye RE, Jarjour IT, Jacques S, Boucher RM, Riou E, Srour M, Carmant L, Lortie A, Major P, Diadori P, Dubeau F, D'Anjou G, Bourque G, Berkovic SF, Sadleir LG, Campeau PM, Kibar Z, Lafrenière RG, Girard SL, Mercimek-Mahmutoglu S, Boelman C, Rouleau GA, Scheffer IE, Mefford HC, Andrade DM, Rossignol E, Minassian BA, Michaud JL. (2017). High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. Am J Hum Genet 101, 664-685.
2. Martin-Zanca, D., Oskam, R., Mitra, G., Copeland, T., and Barbacid, M. (1989). Molecular and biochemical characterization of the human trk proto-oncogene. Mol Cell Biol 9, 24-33.
3. Klein, R., Parada, L.F., Coulier, F., and Barbacid, M. (1989). trkB, a novel tyrosine protein kinase receptor expressed during mouse neural development. EMBO J 8, 3701-3709.
4. Nakagawara, A., Liu, X.G., Ikegaki, N., White, P.S., Yamashiro, D.J., Nycum, L.M., Biegel, J.A., and Brodeur, G.M. (1995). Cloning and chromosomal localization of the human TRK-B tyrosine kinase receptor gene (NTRK2). Genomics 25, 538-546.
5. Klein, R., Lamballe, F., Bryant, S., and Barbacid, M. (1992). The trkB tyrosine protein kinase is a receptor for neurotrophin-4. Neuron 8, 947-956.
6. Squinto, S.P., Stitt, T.N., Aldrich, T.H., Davis, S., Bianco, S.M., Radziejewski, C., Glass, D.J., Masiakowski, P., Furth, M.E., Valenzuela, D.M., Distefano, P.S., Yancopoulous, G.D. (1991). trkB encodes a functional receptor for brain-derived neurotrophic factor and neurotrophin-3 but not nerve growth factor. Cell 65, 885-893.
7. Klein, R., Nanduri, V., Jing, S.A., Lamballe, F., Tapley, P., Bryant, S., Cordon-Cardo, C., Jones, K.R., Reichardt, L.F., and Barbacid, M. (1991). The trkB tyrosine protein kinase is a receptor for brain-derived neurotrophic factor and neurotrophin-3. Cell 66, 395-403.

8. Huang, E.J., and Reichardt, L.F. (2001). Neurotrophins: roles in neuronal development and function. Annu Rev Neurosci 24, 677-736.
9. Huang, E.J., and Reichardt, L.F. (2003). Trk receptors: roles in neuronal signal transduction. Annu Rev Biochem 72, 609-642.
10. Bartkowska, K., Paquin, A., Gauthier, A.S., Kaplan, D.R., and Miller, F.D. (2007). Trk signaling regulates neural precursor cell proliferation and differentiation during cortical development. Development 134, 4369-4380.
11. Wei, Z., Liao, J., Qi, F., Meng, Z., and Pan, S. (2015). Evidence for the contribution of BDNF-TrkB signal strength in neurogenesis: An organotypic study. Neurosci Lett 606, 48-52.
12. Gupta, V.K., You, Y., Gupta, V.B., Klistorner, A., and Graham, S.L. (2013). TrkB receptor signalling: implications in neurodegenerative, psychiatric and proliferative disorders. Int J Mol Sci 14, 10122-10142.
13. Lin, Z., Su, Y., Zhang, C., Xing, M., Ding, W., Liao, L., Guan, Y., Li, Z., and Cui, D. (2013). The interaction of BDNF and NTRK2 gene increases the susceptibility of paranoid schizophrenia. PLoS One 8, e74264.
14. Prowse, N., Dwyer, Z., Thompson, A., Fortin, T., Elson, K., Robeson, H., Fenner, B., and Hayley, S. (2020). Early life selective knockdown of the TrkB receptor and maternal separation modulates adult stress phenotype. Behav Brain Res 378, 112260.
15. Thompson Ray, M., Weickert, C.S., Wyatt, E., and Webster, M.J. (2011). Decreased BDNF, trkB-TK+ and GAD67 mRNA expression in the hippocampus of individuals with schizophrenia and mood disorders. J Psychiatry Neurosci 36, 195-203.
16. Zorner, B., Wolfer, D.P., Brandis, D., Kretz, O., Zacher, C., Madani, R., Grunwald, I., Lipp, H.P., Klein, R., Henn, F.A., Gass P. (2003). Forebrain-specific trkB-receptor knockout mice: behaviorally more hyperactive than "depressive". Biol Psychiatry 54, 972-982.
17. Chen, Z., Simmons, M.S., Perry, R.T., Wiener, H.W., Harrell, L.E., and Go, R.C. (2008). Genetic association of neurotrophic tyrosine kinase receptor type 2 (NTRK2) With Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 147, 363-369.
18. Gines, S., Paoletti, P., and Alberch, J. (2010). Impaired TrkB-mediated ERK1/2 activation in huntington disease knock-in striatal cells involves reduced p52/p46 Shc expression. J Biol Chem 285, 21537-21548.
19. Ginsberg, S.D., Che, S., Wuu, J., Counts, S.E., and Mufson, E.J. (2006). Down regulation of trk but not p75NTR gene expression in single cholinergic basal forebrain neurons mark the progression of Alzheimer's disease. J Neurochem 97, 475-487.
20. Meng, L., Liu, B., Ji, R., Jiang, X., Yan, X., and Xin, Y. (2019). Targeting the BDNF/TrkB pathway for the treatment of tumors. Oncol Lett 17, 2031-2039.
21. Miloszewska, J., Przybyszewska, M., Gos, M., Swoboda, P., and Trembacz, H. (2013). TrkB expression level correlates with metastatic properties of L1 mouse sarcoma cells cultured in non-adhesive conditions. Cell Prolif 46, 146-152.
22. Kimura S, Harada T, Ijichi K, Tanaka K, Liu R, Shibahara D, Kawano Y, Otsubo K, Yoneshima Y, Iwama E, Nakanishi Y, Okamoto I. (2018). Expression of brain-

derived neurotrophic factor and its receptor TrkB is associated with poor prognosis and a malignant phenotype in small cell lung cancer. Lung Cancer 120, 98-107.
23. Yuzugullu, H., Von, T., Thorpe, L.M., Walker, S.R., Roberts, T.M., Frank, D.A., and Zhao, J.J. (2016). NTRK2 activation cooperates with PTEN deficiency in T-ALL through activation of both the PI3K-AKT and JAK-STAT3 pathways. Cell Discov 2, 16030.
24. Binder, D.K., Croll, S.D., Gall, C.M., and Scharfman, H.E. (2001). BDNF and epilepsy: too much of a good thing? Trends Neurosci 24, 47-53.
25. Heinrich, C., Lahteinen, S., Suzuki, F., Anne-Marie, L., Huber, S., Haussler, U., Haas, C., Larmet, Y., Castren, E., and Depaulis, A. (2011). Increase in BDNF-mediated TrkB signaling promotes epileptogenesis in a mouse model of mesial temporal lobe epilepsy. Neurobiol Dis 42, 35-47.
26. Isackson, P.J., Huntsman, M.M., Murray, K.D., and Gall, C.M. (1991). BDNF mRNA expression is increased in adult rat forebrain after limbic seizures: temporal patterns of induction distinct from NGF. Neuron 6, 937-948.
27. Iughetti, L., Lucaccioni, L., Fugetto, F., Predieri, B., Berardi, A., and Ferrari, F. (2018). Brain-derived neurotrophic factor and epilepsy: a systematic review. Neuropeptides 72, 23-29.
28. Kokaia, M., Ernfors, P., Kokaia, Z., Elmer, E., Jaenisch, R., and Lindvall, O. (1995). Suppressed epileptogenesis in BDNF mutant mice. Exp Neurol 133, 215-224.
29. Almoguera, B., McGinnis, E., Abrams, D., Vazquez, L., Cederquist, A., Sleiman, P.M., Dlugos, D., Hakonarson, H., and e, M.E.R.G. (2019). Drug-resistant epilepsy classified by a phenotyping algorithm associates with NTRK2. Acta Neurol Scand 140, 169-176.
30. Torres CM, Siebert M, Bock H, Mota SM, Krammer BR, Duarte JÁ, Bragatti JA, Castan JU, de Castro LA, Saraiva-Pereira ML, Bianchin MM. (2017). NTRK2 (TrkB gene) variants and temporal lobe epilepsy: A genetic association study. Epilepsy Res 137, 1-8.
31. Yeo, G.S., Connie Hung, C.C., Rochford, J., Keogh, J., Gray, J., Sivaramakrishnan, S., O'Rahilly, S., and Farooqi, I.S. (2004). A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. Nat Neurosci 7, 1187-1189.
32. Sonoyama T, Stadler LKJ, Zhu M, Keogh JM, Henning E, Hisama F, Kirwan P, Jura M, Blaszczyk BK, DeWitt DC, Brouwers B, Hyvönen M, Barroso I, Merkle FT, Appleyard SM, Wayman GA, Farooqi IS. (2020). Human BDNF/TrkB variants impair hippocampal synaptogenesis and associate with neurobehavioural abnormalities. Sci Rep 10, 9028.
33. Zahavi, E.E., Steinberg, N., Altman, T., Chein, M., Joshi, Y., Gradus-Pery, T., and Perlson, E. (2018). The receptor tyrosine kinase TrkB signals without dimerization at the plasma membrane. Sci Signal 11.
34. Michael Patton, G.R., William E. Byrd, Matthew Might. (2020). [MKMEDI] mediKanren: A System for Bio-medical Reasoning. In. (miniKanren Workshop.
35. Friedman, D.P., Byrd, W.E., Kiselyov, O., and Hemann, J. (2018). The Reasoned Schemer, Second Edition (MIT Press).

36. Oh, H., Lewis, D.A., and Sibille, E. (2016). The Role of BDNF in Age-Dependent Changes of Excitatory and Inhibitory Synaptic Markers in the Human Prefrontal Cortex. Neuropsychopharmacology 41, 3080-3091.
37. Chen, X., Ye, H., Kuruvilla, R., Ramanan, N., Scangos, K.W., Zhang, C., Johnson, N.M., England, P.M., Shokat, K.M., and Ginty, D.D. (2005). A chemical-genetic approach to studying neurotrophin signaling. Neuron 46, 13-21.
38. Gray, J., Yeo, G., Hung, C., Keogh, J., Clayton, P., Banerjee, K., McAulay, A., O'Rahilly, S., and Farooqi, I.S. (2007). Functional characterization of human NTRK2 mutations identified in patients with severe early-onset obesity. Int J Obes (Lond) 31, 359-364.
39. Stahel, P., Sud, S.K., Lee, S.J., Jackson, T., Urbach, D.R., Okrainec, A., Allard, J.P., Bassett, A.S., Paterson, A.D., Sockalingam, S., Dash, S. (2019). Phenotypic and genetic analysis of an adult cohort with extreme obesity. Int J Obes (Lond) 43, 2057-2065.
40. Fenner, B.M. (2012). Truncated TrkB: beyond a dominant negative receptor. Cytokine Growth Factor Rev 23, 15-24.
41. Liao, G.Y., Li, Y., and Xu, B. (2013). Ablation of TrkB expression in RGS9-2 cells leads to hyperphagic obesity. Mol Metab 2, 491-497.
42. Xu, B., Goulding, E.H., Zang, K., Cepoi, D., Cone, R.D., Jones, K.R., Tecott, L.H., and Reichardt, L.F. (2003). Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor. Nat Neurosci 6, 736-742.
43. Yuan, P.X., Huang, L.D., Jiang, Y.M., Gutkind, J.S., Manji, H.K., and Chen, G. (2001). The mood stabilizer valproic acid activates mitogen-activated protein kinases and promotes neurite growth. J Biol Chem 276, 31674-31683.
44. Michaelis M, Suhan T, Michaelis UR, Beek K, Rothweiler F, Tausch L, Werz O, Eikel D, Zörnig M, Nau H, Fleming I, Doerr HW, Cinatl J Jr. (2006). Valproic acid induces extracellular signal-regulated kinase 1/2 activation and inhibits apoptosis in endothelial cells. Cell Death Differ 13, 446-453.
45. Ghiglieri V, Sgobio C, Patassini S, Bagetta V, Fejtova A, Giampà C, Marinucci S, Heyden A, Gundelfinger ED, Fusco FR, Calabresi P, Picconi B. (2010). TrkB/BDNF-dependent striatal plasticity and behavior in a genetic model of epilepsy: modulation by valproic acid. Neuropsychopharmacology 35, 1531-1540.
46. Theunissen, P.T., Robinson, J.F., Pennings, J.L., de Jong, E., Claessen, S.M., Kleinjans, J.C., and Piersma, A.H. (2012). Transcriptomic concentration-response evaluation of valproic acid, cyproconazole, and hexaconazole in the neural embryonic stem cell test (ESTn). Toxicol Sci 125, 430-438.
47. Dedoni, S., Marras, L., Olianas, M.C., Ingianni, A., and Onali, P. (2019). Downregulation of TrkB Expression and Signaling by Valproic Acid and Other Histone Deacetylase Inhibitors. J Pharmacol Exp Ther 370, 490-503.
48. Liu, G., Kotloski, R.J., and McNamara, J.O. (2014). Antiseizure effects of TrkB kinase inhibition. Epilepsia 55, 1264-1273.
49. Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C.J., McLauchlan, H., Klevernic, I., Arthur, J.S., Alessi, D.R., and Cohen, P. (2007). The selectivity of protein kinase inhibitors: a further update. Biochem J 408, 297-315.
50. Bartkowiak, B., Yan, C.M., Soderblom, E.J., and Greenleaf, A.L. (2019). CDK12 Activity-Dependent Phosphorylation Events in Human Cells. Biomolecules 9.

51. Howe, C.L., Valletta, J.S., Rusnak, A.S., and Mobley, W.C. (2001). NGF signaling from clathrin-coated vesicles: evidence that signaling endosomes serve as a platform for the Ras-MAPK pathway. Neuron 32, 801-814.
52. Grimes, M.L., Zhou, J., Beattie, E.C., Yuen, E.C., Hall, D.E., Valletta, J.S., Topp, K.S., LaVail, J.H., Bunnett, N.W., and Mobley, W.C. (1996). Endocytosis of activated TrkA: evidence that nerve growth factor induces formation of signaling endosomes. J Neurosci 16, 7950-7964.
53. Ahmed, F., and Hristova, K. (2018). Dimerization of the Trk receptors in the plasma membrane: effects of their cognate ligands. Biochem J 475, 3669-3685.

Figure Legends

Figure 1. Overexpression of *NTRK2* variants exhibits variable consequences on TRKB activation. A) Representative Western of cells overexpressing of *NTRK2* variants (WT (wildtype), EV (empty vector), Y434C or T720I), probed for phospho-TRKB and total TRKB. B.) Quantification of Western blots (N=3) normalized to WT. C-D.) Activation of the TRKB receptor was analyzed through stimulation with and without brain derived neurotrophic factor (BDNF, 10ng/mL). Downstream signaling consequences were analyzed. C.) Representative Western probed with phospho-ERK and total-ERK. D.) quantification performed (N=3) (*p-value ≤ 0.05; **p-value ≤ 0.01).

Figure 2. *NTRK2* variant T720I (T720I) exhibits hypomorphic activity in response to TRKB activation. Upon overexpression of WT or T720I NTRK2, activation of the TRKB receptor was analyzed through stimulation with BDNF at different doses (0, 0.5, 10, 50, and 200ng/mL). Downstream signaling consequences were analyzed through Western blots probed with phospho-ERK and total-ERK A.) Representative Western B.) quantification performed and normalized to untreated WT (N=3) (*p-value ≤ 0.05; **p-value ≤ 0.01).

Figure 3. Effects of *NTRK2* variants on downstream signaling molecules. Western blots, performed on protein lysates overexpressing WT, EV, Y434C, or T720I NTRK2, were probed for phospho-ERK, total ERK, A-B.) Representative Western and quantification of western blots with pERK/total ERK (N=3) C.) A luciferase reporter analysis of ELK1 transcriptional activity was performed on cells over-expressing WT, EV, Y434C or T720I NTRK2 and normalized to WT D-E.) Representative Western probed and quantification (N=3) for pS6/total S6 normalized to WT NTRK2. (*p-value ≤ 0.05; **p-value ≤ 0.01).

Figure 4. MediKanren predicated treatment rescues NTRK2 variant cellular phenotypes. A-B.) Variant Y434C overexpressing cells were treated with compound 1NM-PP1 (0, 3, 5, and 10 uM) and protein lysate harvested. A.) Representative Western

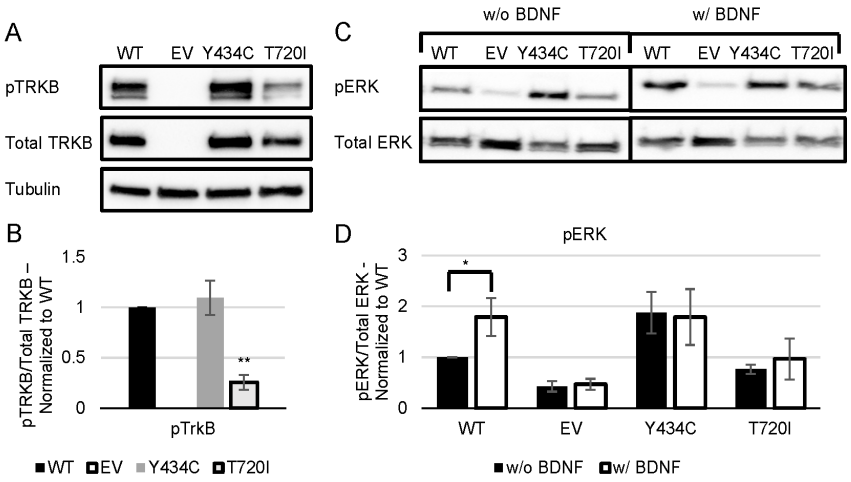
blots were probed for phospho-TRKB, total TRKB, phospho-ERK and total ERK B.) with quantification normalized to untreated (N=3). C-D.) NTRK2 WT overexpressing cells were treated with compound 1NM-PP1, protein lysate harvested and representative Western blot probed similar to variant Y434C with quantification normalized to untreated (N=3) (*p-value ≤ 0.05; **p-value ≤ 0.01).

Figure 5. MediKanren predicted treatment rescues NTRK2 Variant T720I cellular phenotypes. A-B.) Variant T720I overexpressing cells were treated with Valproic Acid (0, 500, 750, and 1000uM) and protein lysate harvested. C.) Representative Western blot probed for phospho-S6, total S6, phospho-ERK and total ERK D.) with quantification normalized to untreated (N=3) (*p-value ≤ 0.05; **p-value ≤ 0.01).

Figure 6. Diagram of NTRK2 signaling, effect of each VUS, and select inhibitor response. BDNF binds the TRKB Receptor (NTRK2) and signals through both the mTOR and MAPK pathways. While the T720I variant represses signaling activity as seen through lowered pERK, and pS6 levels, VPA treatment normalizes their levels. Conversely the Y434C variant results in ligand–independent activation and increased MAPK signaling and NM-PP1 treatment inhibits receptor activation and MAPK signaling.

Supplemental Figure 1. MEK inhibition with Selumetinib abrogates increased MAPK activity observed with *NTRK2* variant pTry434Cys (Y434C). A-C) Upon overexpression of *NTRK2* variants, cells were treated with the MEK inhibitor Selumetinib (0, 1 or 10 uM) and protein lysates harvested. A-B.) Representative Western blots probed for phospho-ERK and total ERK with C.) quantification of Western blots normalized to untreated WT (N=3) (*p-value ≤ 0.05; **p-value ≤ 0.01).

Figure 1


A
WT EV Y434C T720I
pTRKB
Total TRKB
Tubulin
B
pTRKB/Total TRKB – Normalized to WT
1.5
1
0.5
0
**
pTrkB
WT EV Y434C T720I
C
w/o BDNF
w/ BDNF
WT EV Y434C T720I WT EV Y434C T720I
pERK
Total ERK
D
pERK
pERK/Total ERK - Normalized to WT
3
2
1
0
*
WT
EV
Y434C
T720I
w/o BDNF w/ BDNF

n ≥ 3

Figure 2

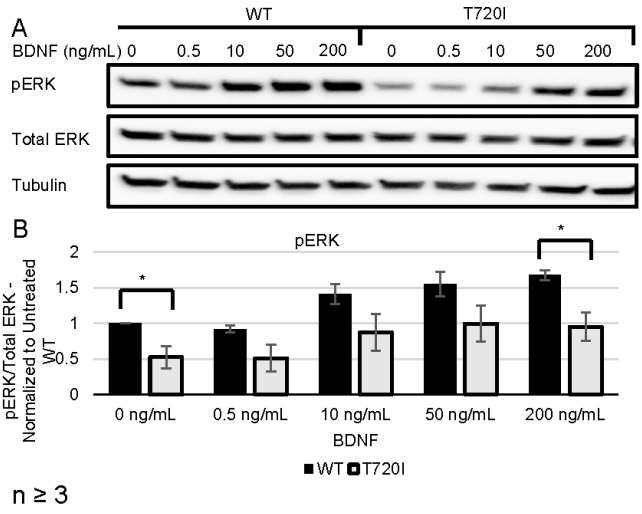
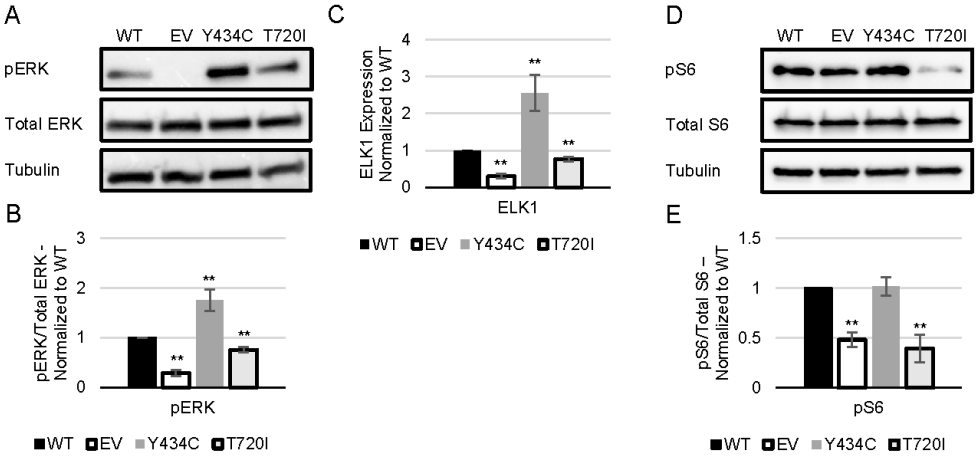
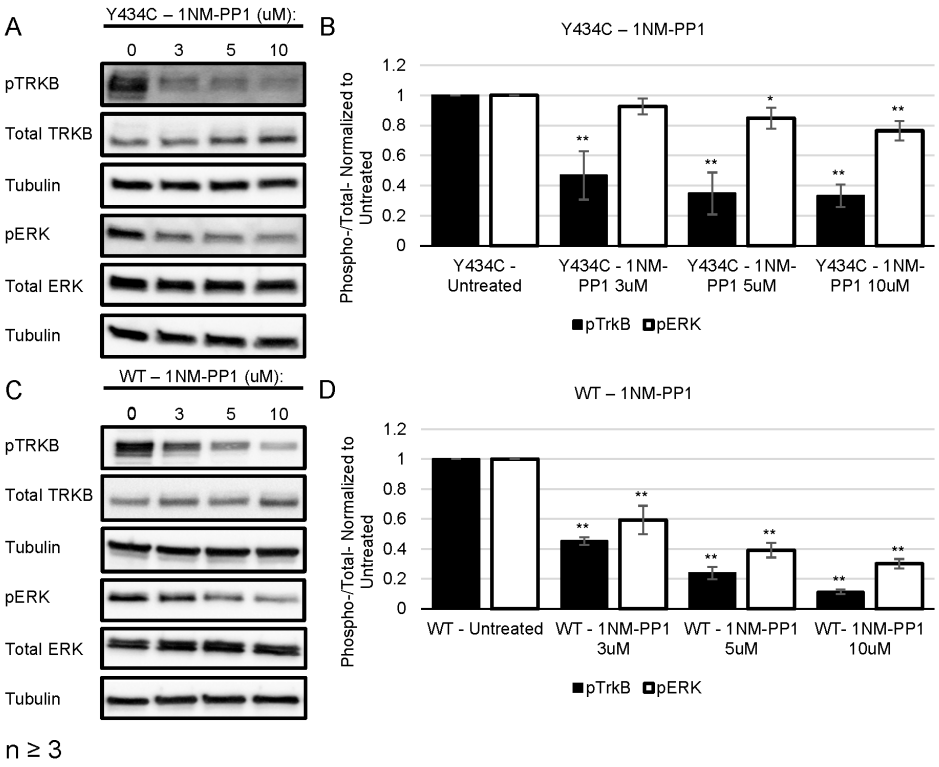

A
WT
T720I
BDNF (ng/mL) 0 0.5 10 50 200 0 0.5 10 50 200
pERK
Total ERK
Tubulin
B
pERK
pERK/Total ERK - Normalized to Untreated WT
2
1.5
1
0.5
0
*
*
0 ng/mL
0.5 ng/mL
10 ng/mL
50 ng/mL
200 ng/mL
BDNF
WT T720I
n ≥ 3

Figure 3


A
WT
EV
Y434C
T720I
pERK
Total ERK
Tubulin
B
pERK/Total ERK - Normalized to WT
pERK
WT
EV
Y434C
T720I
C
ELK1 Expression Normalized to WT
ELK1
D
pS6
Total S6
Tubulin
E
pS6/Total S6 – Normalized to WT
pS6

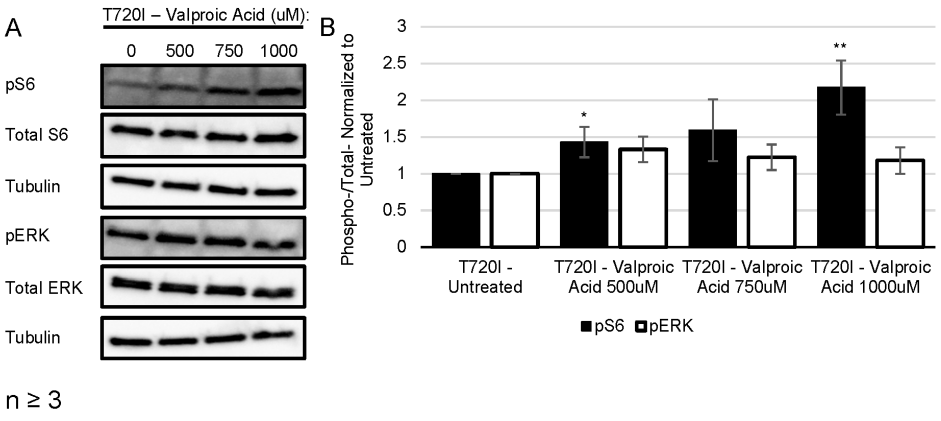
n ≥ 3

Figure 4


A
Y434C – 1NM-PP1 (uM):
0 3 5 10
pTRKB
Total TRKB
Tubulin
pERK
Total ERK
Tubulin
B
Y434C – 1NM-PP1
Phospho-/Total- Normalized to Untreated
1.2
1
0.8
0.6
0.4
0.2
0
Y434C - Untreated
Y434C - 1NM-PP1 3uM
Y434C - 1NM-PP1 5uM
Y434C - 1NM-PP1 10uM
pTrkB pERK
C
WT – 1NM-PP1 (uM):
0 3 5 10
pTRKB
Total TRKB
Tubulin
pERK
Total ERK
Tubulin
D
WT – 1NM-PP1
Phospho-/Total- Normalized to Untreated
1.2
1
0.8
0.6
0.4
0.2
0
WT - Untreated
WT - 1NM-PP1 3uM
WT - 1NM-PP1 5uM
WT- 1NM-PP1 10uM
pTrkB pERK

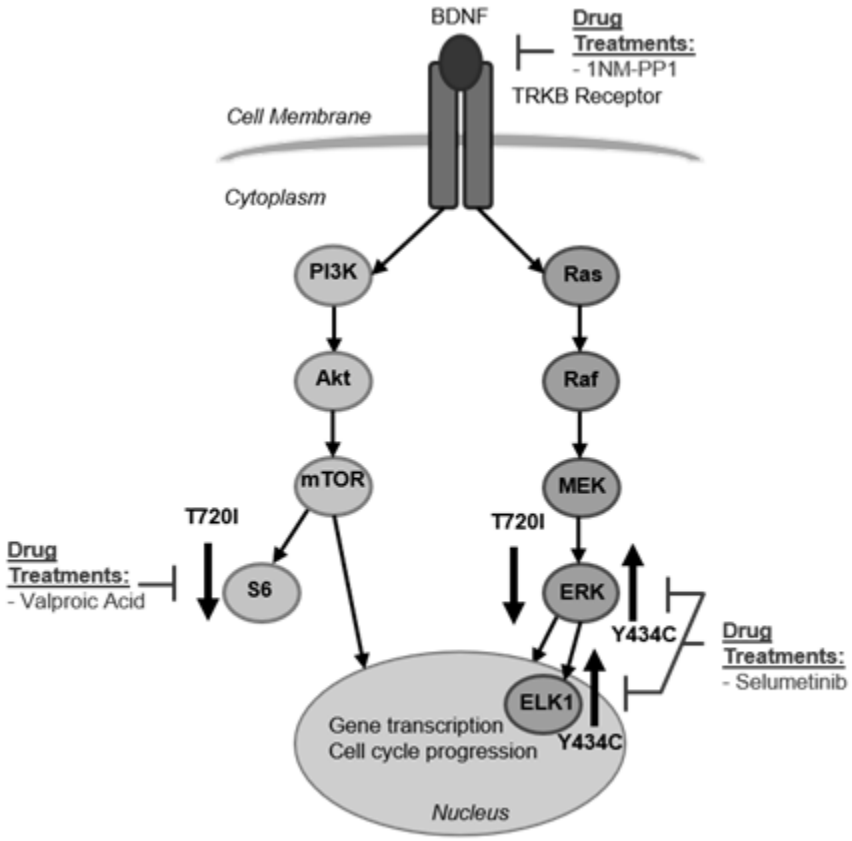
n ≥ 3

Figure 5

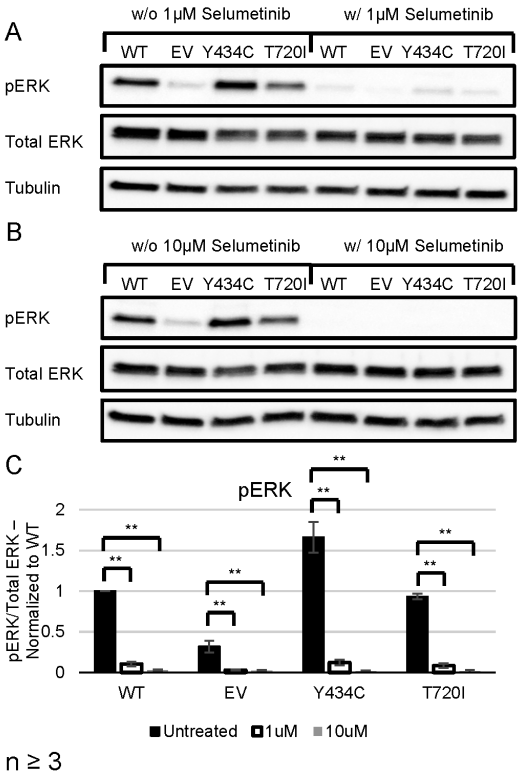

A
T720I – Valproic Acid (uM):
0
500
750
1000
pS6
Total S6
Tubulin
pERK
Total ERK
Tubulin
B
Phospho-/Total- Normalized to Untreated
3
2.5
2
1.5
1
0.5
0
*
**
T720I - Untreated
T720I - Valproic Acid 500uM
T720I - Valproic Acid 750uM
T720I - Valproic Acid 1000uM
pS6
pERK

n ≥ 3

Figure 6


BDNF
Drug Treatments:
- 1NM-PP1
TRKB Receptor
Cell Membrane
Cytoplasm
PI3K
Ras
Akt
Raf
mTOR
MEK
T720I
T720I
Drug Treatments:
- Valproic Acid
S6
ERK
Y434C
Drug Treatments:
- Selumetinib
ELK1
Gene transcription
Cell cycle progression
Y434C
Nucleus

S1 Fig.


A
w/o 1μM Selumetinib
w/ 1μM Selumetinib
WT EV Y434C T720I WT EV Y434C T720I
pERK
Total ERK
Tubulin
B
w/o 10μM Selumetinib
w/ 10μM Selumetinib
WT EV Y434C T720I WT EV Y434C T720I
pERK
Total ERK
Tubulin
C
pERK
pERK/Total ERK – Normalized to WT
2
1.5
1
0.5
0
**
WT
EV
Y434C
T720I
Untreated
1uM
10uM

n ≥ 3

CONCLUSIONS AND FUTURE DIRECTIONS

Increasing utilization of genomic sequencing has raised as many questions as it has answered. Scientists are discovering that there is still much to learn about the human genome and its role in disease. One of the main difficulties in interpreting genetic results is the fact that genetic variation is extremely common and functional assessment of variants can be time consuming and costly. Whole-genome sequencing (WGS) of >2,500 individuals found a total of over 88 million variants [1]; though only a few have known functions. Single nucleotide polymorphisms made up the majority of variants found (84.7 million), short insertions/deletions (indels) were the second most frequent variant observed (3.6 million), and 60,000 structural variants were also seen [1]. In fact, some studies have found the number of individual single-nucleotide variants to be even greater [2]. Additionally, the 1000 Genome Project Consortium has demonstrated that among the 179 individuals sequenced, each person carries approximately 250-300 loss-of-function variants [3]. It was estimated that, on average, each individual genome was shown to possess 50-100 genic variants classified as disease-causing [3]. Thus, determining which variants are functional and/or clinically significant and understanding how genetic variation can impact disease have become important topics in genetics research.

Pathogenicity of certain variants is more easily established. For example, tandem repeats have been well established as a causative mutation in repeat expansion disorders. Further, the length of pathogenic expansions is known to positively correlate with the

onset and severity of disease. In Friedreich's Ataxia, GAA1, which is the shorter allele, is a good predictor of disease severity, with every increase of 100 GAA repeats estimated to result in ~2.3 years earlier disease onset [4]. Despite this strong correlation between pathogenically expanded GAA number and age of onset, some FRDA patients present with an unexpected age of onset and a more variable phenotypic spectrum than would be predicted based on repeat lengths at time of diagnosis [5, 6].

Our study offers tissue-specific mosaicism and somatic instability as potential contributing factors for the disparity observed between the GAA repeat length at time of diagnosis and the symptom variability observed in Friedreich's Ataxia. Previous studies have demonstrated that GAA repeat lengths determined from lymphocytes or buccal cells at time of diagnosis may not be representative of the expanded repeat length in disease-relevant tissues [5-11]. While several studies have already analyzed somatic instability of the trinucleotide repeats in FRDA patient tissues, they are limited by small sample sizes and patient tissue access. Our study analyzed repeat lengths in a large cohort of 15 patients utilizing 5 phenotypically relevant tissues: heart, cerebral cortex, spinal cord, cerebellar cortex, and pancreas [9]. This enabled us to analyze tissue-specific mosaicism in FRDA repeat lengths with greater confidence and answer if there is a correlation between sizes of the pathological GAAs and different clinically relevant tissue types. Additionally, we addressed whether the number of GAA repeats is identical in different tissues. We observed somatic mosaicism of the expanded repeats among five different tissues in all 15 FRDA patients. Interestingly, a tissue-specific pattern was observed with heart and pancreatic tissue exhibiting longer repeat lengths compared to cerebral cortex and spinal cord tissues. While the mechanism for longer repeat lengths in heart and

pancreas tissues remains unknown, western blot analysis of frataxin in FRDA patient heart tissues exhibited lower expression when compared to cerebral and cerebellar cortex tissues. This agrees with the expected association of longer repeats negatively correlating with frataxin expression. Interestingly in frataxin deficient tissues, transcriptome analyses conducted on FRDA mouse models and patient cells exhibited a global decrease in gene expression [12]. Additionally, in FRDA cells, genes involved in transcription, translation and DNA repair were downregulated [13]. Furthermore, increased reactive oxygen species (ROS), observed in frataxin deficient cells, may stimulate DNA damage and further GAA repeat instability similar to what has been proposed to explain the somatic instability in Huntington's disease [9, 14]. Further research should be undertaken to understand how global transcriptional differences in FRDA cells along with increased ROS and DNA damage result in somatic instability of the pathological repeats in FRDA. Additionally, interrogating whether these differences could aid in determining potential therapeutics would be a logical next step.

Our study also sought to answer the question of whether contractions or expansions in the expanded GAA repeats accumulate as a patient ages. To determine how GAAs change over time, repeat testing on lymphocytes drawn from 5 FRDA patients over 7-9 years was performed. Notably, the repeated blood draws demonstrated that the majority of expanded GAA repeats exhibit a length-dependent bias towards expansion, with longer repeat lengths exhibiting greater somatic instability. Of note, one patient in our study exhibited a 15% increase in repeat length over 7 years. No contractions were observed in any of the patient alleles. Interestingly, these results contrast with previous research studies of GAA repeats in immortalized lymphoblast cells where contractions

were predominantly observed [15]. Culturing of FRDA cells and model cell lines typically results in shortening of the expanded repeats as well, however, iPSC culturing results in continuous expansion of the GAA repeats [16, 17]. It is possible that pluripotent/multipotent state observed in iPSCs as well as hematopoietic stem cells or lymphoid progenitors may contribute to repeat expansion [9].

Additionally, our study highlights the importance of understanding how somatic mosaicism and instability of the GAA repeats can confound research results and interpretation. Significant discrepancies between GAA repeat lengths at time of diagnosis and adulthood can exist, suggesting that reassessment of GAA repeat lengths may be needed to fully predict patient symptom presentation and progression throughout disease course. Further research needs to be undertaken to determine how repeat expansions over time can affect clinical trial and research interpretation. Also, understanding why certain patients exhibit larger expansions over their life time compared to other patients can help aid our understanding of FRDA disease progression. Overall, our study emphasizes the importance that somatic mosaicism and instability can have on patient symptom severity and progression. Additionally, these results suggest potential therapeutic avenues in targeting somatic instability to potentially slow and/or stall disease progression in patients.

Germline variation and how it impacts protein stability and function may also dictate phenotype. We demonstrated through our study on NF1 missense variants that unique germline *NF1* variants can affect function. In order to determine how different *NF1* variants impact protein, we sought to further validate a mouse *Nf1* cDNA expression system recently developed [18]. Previous research utilizing human *NF1*

cDNA has proven challenging, partially due to size and toxicity of the full-length *NF1* cDNA construct. For our study, we utilized *mNf1* cDNA which is highly homologous to hNF1 and allows us to examine the molecular consequences of different *NF1* variants. Our study is the first of its kind to create a large panel of patient specific *NF1* variants and analyze their effects on NF1 expression and its ability to inhibit Ras activity in order to create a functional profile.

Additionally, our study was also able to analyze variant effects in association with patient phenotypes. As there are several phenotype-genotype correlations already described in NF1 patients [19-26] we attempted to correlate function and phenotype as we evaluated whether phenotypes cluster with specific molecular consequences. We assessed various *NF1* mutations based on NF1 expression (protein abundance) and Ras inhibition (protein function). There are several variants associated with a mild phenotype: DelM992, R1809, R1038G, and M1149V [19, 21, 23-26]. Patients harboring one of these variants typically exhibit café-au-lait macules (CALMS), with or without skinfold freckling and Lisch nodules. Most notable, these variants do not result in tumor development with an associated increased risk of malignancy. Additionally, R1809, M1149, and M992 exhibit Noonan-like features more frequently than “classic” NF affected individuals [21]. When variant NF1 function was assayed, we observed that most mild-associated variants are hypomorphic with relatively stable NF1 levels with at least approximately 60% of wild-type levels. Interestingly, Ras signaling differed among our mild phenotype variants. Ras signaling was not significantly elevated in R1038G and R1809, however M1149V and delM992 resulted in differing degrees of significant Ras activity elevation. Variant M1149V did stand out with its unexpected complete loss of

Ras inhibition. Of note, many NF1 variants associated with a mild phenotype (R1809C, M1149V, and DelM992) clustered together when neurofibromin levels with GTP-Ras levels were plotted and a trend line drawn. Interestingly, the "unknown" variant S1997R also clustered with the mild variants which is consistent with the limited knowledge available on this variant's patient phenotype.

Our study also included several severe-phenotype associated variants: L847P, G848R, R1276Q, and K1423E. Variants affecting NF1 codons 844-848 are located in the cysteine-serine-rich domain (CSRD) and are associated with multiple plexiform and spinal neurofibromas with a greater risk of malignancy [22]. R1276Q and K1423E are located in the gap-related domain (GRD) and are also associated with multiple symptomatic spinal neurofibromas. Interestingly, the severe variant G848R was able to inhibit RAS activity, demonstrating a functional protein. However, G848R NF1 protein expression was decreased compared to WT. We observed a second, distinct grouping of two severe variants, R1276Q and K1423E, which did not cluster with the other severe phenotype variants. However, the consequences of variant R1276 and K1423 are well-established, with inactivation of the GRD domain of the NF1 protein resulting in inability to inhibit Ras activity [27-29].

Our study highlights the importance of understanding the functional consequences of variants and how correlation to phenotype can aid in the proposal of a categorization system for unique missense variants in NF1. Future research, analyzing other signaling consequences in NF1 variants can aid in further categorizing variants based on functional effects and enable more predictive reliability in determination of phenotype. This can

then be used to inform likely symptom presentation and progression as well as to develop more-directed therapeutic avenues.

Understanding the molecular consequences or function of a variant is especially important when different variants within the same gene have opposing effects. In our study on the *NTRK2* gene we were able to functionally characterize two rare and novel variants and utilize these molecular phenotypes to determine potential therapeutic avenues. In this study we analyzed two unique variants in the *NTRK2* gene in order to determine the molecular mechanisms underlying their distinct phenotypes [30]. Variant T720I is located in the tyrosine kinase domain of the NTRK2 protein. We demonstrated that this variant results in a hypomorphic downregulation of TRKB-BDNF signaling with consequent decreased MAPK and mTOR signaling as observed through decreased pERK and pS6 expression. Interestingly, patients with this variant exhibit unique symptoms, notably hyperphagia with early-onset obesity. Patients with *NTRK2* variant Y722C, very close to our variant of interest and also located in the tyrosine kinase domain, likewise exhibit hyperphagia with severe associated obesity as well [30, 31]. In addition to T720I, other hypomorphic *NTRK2* allele variants, also located in the tyrosine kinase domain, have been previously described [30-33]. Variant Y434C is located in the transmembrane domain of the NTRK2 protein. We demonstrate that this variant may result in a ligand-independent upregulation of TRKB signaling. While stimulation with BDNF does not demonstrate a noticeable response, cells overexpressing the Y434C variant exhibit increased MAPK signaling as demonstrated through increased pERK levels and ELK1 transcriptional activity. Notably, the Y434C variant is the first described in NTRK2 to exhibit hypomorphic activity.

Our study was able to target these distinct functional consequences of the two NTRK2 variants to determine potential therapy avenues through the utilization of artificial intelligence (AI). In collaboration with the UAB's Precision Medicine Institute, we utilized an AI-based software, mediKanren, to determine potential therapeutic avenues for our two rare patient variants [34]. Since the founding of artificial intelligence (AI) in 1950 by Alan Turning, AI has advanced over the past few decades to become a useful tool in medicine [35, 36]. Initially, AI started with fuzzy logic utilizing "if, then rules" while accounting for real-world situations where things are not "black or white", but rather varying shades of gray and accommodating for the imprecision and ambiguity that is inherent in medicine [35-37]. AI has since evolved into more advanced technologies including several subfields: machine learning (ML) and deep learning (DL) [35]. ML is able to identify patterns to analyze specific situations and then utilizes the information to "learn" [35, 38]. ML has led to the development of DL with an artificial neural network (ANN) which is analogous to the biological nervous system and allows AI to perform parallel computations [35, 36]. This neural network allows large amounts of data to be analyzed and incorporated in the determination of potential therapeutics and more precision-mediated patient care. Since its inception, AI has had increasing utility in medicine and the clinic. Precision medicine has grown through ANN, allowing AI to make decisions based on individual patient factors including genes, function, and environment [39, 40].

In our study, we queried mediKanren to predict potential chemical substances which would result in either the upregulation (for the hypomorphic T720I variant) or downregulation (for variant Y434C) of NTRK2 TRKB signaling. MediKanren predicted

valproic acid (VPA) as a therapeutic to increase NTRK2 activity and upregulate MAPK activity. Interestingly, VPA was previously shown to result in an efficacious response in a patient harboring the T720I variant [30]. Indeed, VPA treated T720I overexpressing cells exhibited a partial rescue of molecular phenotype with a significant increase in mTOR activity and a trend towards increased MAPK activity. Interestingly, this compound may also represent a potential therapy in the obese patient cohort through targeting of TRKB-BDNF associated appetite control. Further studies may be warranted to determine whether this commonly used medication can have off-target effects which may aid in the treatment of obesity and obesity-related metabolic syndromes.

MediKanren predicted compound 1NM-PP1, a selective kinase inhibitor, to treat variant Y434C. Upon treatment with 1NM-PP1, Y434C overexpressing cells exhibited decreased phosphorylation of TRKB with a subsequent decrease in pERK. Of note, patients harboring the Y434C variant had previously been treated with anti-epileptic drugs (AEDs), including VPA which interestingly resulted in an unfavorable response [30]. Treatment with a kinase inhibitor has been previously untried. This represents a novel therapeutic option in epilepsy and may be useful in epileptic seizures which had previously been shown to be drug-refractory. These studies emphasize the value that variant functional characterization to determine molecular phenotype can have in the utilization of precision medicine and patient treatments.

There are several limitations to these studies. For example, we were unable to correlate GAA repeat instability to phenotype due to lack of clinical information. Ideally associating repeat length with relative symptomology would aid in determining how tissue mosaicism affects phenotype and FRDA severity and would lend strength to our

study's conclusions. Additionally, utilization of HEK293 with overexpression of variant-specific cDNAs contributes to certain limitations in these studies. While these cells are useful as a convenient model system in research and easily manipulated. Ideally, conducting follow-up analyses in clinically relevant cell line would be the next appropriate step in future studies.

In summary, our combined studies highlight the importance of understanding genetic variation and how individual differences contribute to phenotype. Somatic variation is often an understudied contribution to symptom presentation, either through repeat length polymorphism, the addition of a second hit mutation, genetic modifiers, and more. Understanding how these tissue-specific variations occur can aid in understanding phenotype as well as understanding the development and progression of distinct symptoms. Furthermore, determining how genetic variants affect protein function can also aid in understanding patient phenotype as well as guiding for potential therapies. Overall, understanding genetic variation and the functional impacts on the protein of unique genetic variants is becoming increasingly recognized as the next steps in precision medicine and better patient care.

REFERENCES

1. Evans, J.P., B.C. Powell, and J.S. Berg, *Finding the Rare Pathogenic Variants in a Human Genome.* JAMA, 2017. **317**(18): p. 1904-1905.
2. Gilissen, C., et al., *Disease gene identification strategies for exome sequencing.* Eur J Hum Genet, 2012. **20**(5): p. 490-7.
3. Hayward, J. and L.S. Chitty, *Beyond screening for chromosomal abnormalities: Advances in non-invasive diagnosis of single gene disorders and fetal exome sequencing.* Semin Fetal Neonatal Med, 2018. **23**(2): p. 94-101.
4. Green, R.C., et al., *ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing.* Genet Med, 2013. **15**(7): p. 565-74.
5. Taylor, M.R., J.G. Edwards, and L. Ku, *Lost in transition: challenges in the expanding field of adult genetics.* Am J Med Genet C Semin Med Genet, 2006. **142C**(4): p. 294-303.
6. Lerner, B., et al., *The value of genetic testing: beyond clinical utility.* Genet Med, 2017. **19**(7): p. 763-771.
7. Hamilton, A.B., et al., *Factors influencing organizational adoption and implementation of clinical genetic services.* Genet Med, 2014. **16**(3): p. 238-45.
8. Manolio, T.A., et al., *Implementing genomic medicine in the clinic: the future is here.* Genet Med, 2013. **15**(4): p. 258-67.
9. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
10. Roest Crollius, H., et al., *Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence.* Nat Genet, 2000. **25**(2): p. 235-8.
11. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
12. Duzkale, H., et al., *A systematic approach to assessing the clinical significance of genetic variants.* Clin Genet, 2013. **84**(5): p. 453-63.
13. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genet Med, 2015. **17**(5): p. 405-24.
14. Rehm, H.L., et al., *ACMG clinical laboratory standards for next-generation sequencing.* Genet Med, 2013. **15**(9): p. 733-47.
15. Fatkin, D. and R. Johnson, *Variants of Uncertain Significance and "Missing Pathogenicity".* J Am Heart Assoc, 2020. **9**(3): p. e015588.
16. Federici, G. and S. Soddu, *Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers.* J Exp Clin Cancer Res, 2020. **39**(1): p. 46.

17. Pottinger, T.D., et al., *Pathogenic and Uncertain Genetic Variants Have Clinical Cardiac Correlates in Diverse Biobank Participants.* J Am Heart Assoc, 2020. **9**(3): p. e013808.
18. Tung, N., et al., *Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer.* J Clin Oncol, 2016. **34**(13): p. 1460-8.
19. Macklin, S.K., et al., *Physician interpretation of variants of uncertain significance.* Fam Cancer, 2019. **18**(1): p. 121-126.
20. Richter, S., et al., *Variants of unknown significance in BRCA testing: impact on risk perception, worry, prevention and counseling.* Ann Oncol, 2013. **24 Suppl 8**: p. viii69-viii74.
21. Petrucelli, N., et al., *Clinical interpretation and recommendations for patients with a variant of uncertain significance in BRCA1 or BRCA2: a survey of genetic counseling practice.* Genet Test, 2002. **6**(2): p. 107-13.
22. Clift, K., et al., *Patients' views on variants of uncertain significance across indications.* J Community Genet, 2020. **11**(2): p. 139-145.
23. Macklin, S., et al., *The Psychosocial Impact of Carrying a Debated Variant in the GLA Gene.* J Genet Couns, 2018. **27**(1): p. 217-224.
24. Solomon, I., et al., *Lynch Syndrome Limbo: Patient Understanding of Variants of Uncertain Significance.* J Genet Couns, 2017. **26**(4): p. 866-877.
25. Wynn, J., et al., *Examining the Psychosocial Impact of Genetic Testing for Cardiomyopathies.* J Genet Couns, 2018. **27**(4): p. 927-934.
26. Murray, M.L., et al., *Follow-up of carriers of BRCA1 and BRCA2 variants of unknown significance: variant reclassification and surgical decisions.* Genet Med, 2011. **13**(12): p. 998-1005.
27. So, M.K., et al., *Reinterpretation of BRCA1 and BRCA2 variants of uncertain significance in patients with hereditary breast/ovarian cancer using the ACMG/AMP 2015 guidelines.* Breast Cancer, 2019. **26**(4): p. 510-519.
28. Rodenburg, R.J., *The functional genomics laboratory: functional validation of genetic variants.* J Inherit Metab Dis, 2018. **41**(3): p. 297-307.
29. Strande, N.T. and J.S. Berg, *Defining the Clinical Value of a Genomic Diagnosis in the Era of Next-Generation Sequencing.* Annu Rev Genomics Hum Genet, 2016. **17**: p. 303-32.
30. Bonjoch, L., et al., *Approaches to functionally validate candidate genetic variants involved in colorectal cancer predisposition.* Mol Aspects Med, 2019. **69**: p. 27-40.
31. Sobreira, N., et al., *GeneMatcher: a matching tool for connecting investigators with an interest in the same gene.* Hum Mutat, 2015. **36**(10): p. 928-30.
32. La Spada, A.R., et al., *Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy.* Nature, 1991. **352**(6330): p. 77-9.
33. Verkerk, A.J., et al., *Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome.* Cell, 1991. **65**(5): p. 905-14.
34. Gatchel, J.R. and H.Y. Zoghbi, *Diseases of unstable repeat expansion: mechanisms and common principles.* Nat Rev Genet, 2005. **6**(10): p. 743-55.

35. Campuzano, V., et al., *Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion.* Science, 1996. **271**(5254): p. 1423-7.
36. Al-Mahdawi, S., et al., *The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues.* Hum Mol Genet, 2008. **17**(5): p. 735-46.
37. Herman, D., et al., *Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia.* Nat Chem Biol, 2006. **2**(10): p. 551-8.
38. Saveliev, A., et al., *DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing.* Nature, 2003. **422**(6934): p. 909-13.
39. Tsou, A.Y., et al., *Mortality in Friedreich ataxia.* J Neurol Sci, 2011. **307**(1-2): p. 46-9.
40. Delatycki, M.B. and L.A. Corben, *Clinical features of Friedreich ataxia.* J Child Neurol, 2012. **27**(9): p. 1133-7.
41. Pandolfo, M., *Friedreich ataxia: the clinical picture.* J Neurol, 2009. **256 Suppl 1**: p. 3-8.
42. Harding, A.E., *Friedreich's ataxia: a clinical and genetic study of 90 families with an analysis of early diagnostic criteria and intrafamilial clustering of clinical features.* Brain, 1981. **104**(3): p. 589-620.
43. Bidichandani, S.I., et al., *Very late-onset Friedreich ataxia despite large GAA triplet repeat expansions.* Arch Neurol, 2000. **57**(2): p. 246-51.
44. Montermini, L., et al., *Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion.* Ann Neurol, 1997. **41**(5): p. 675-82.
45. Clark, R.M., et al., *The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model.* Hum Genet, 2007. **120**(5): p. 633-40.
46. De Biase, I., et al., *Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients.* Ann Neurol, 2007. **61**(1): p. 55-60.
47. De Biase, I., et al., *Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life.* Genomics, 2007. **90**(1): p. 1-5.
48. Machkhas, H., et al., *A mild case of Friedreich ataxia: lymphocyte and sural nerve analysis for GAA repeat length reveals somatic mosaicism.* Muscle Nerve, 1998. **21**(3): p. 390-3.
49. Montermini, L., et al., *Somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions in the central nervous system.* Neurology, 1997. **49**(2): p. 606-10.
50. Long, A., et al., *Somatic instability of the expanded GAA repeats in Friedreich's ataxia.* PLoS One, 2017. **12**(12): p. e0189990.
51. Friedman, J.M., *Epidemiology of neurofibromatosis type 1.* Am J Med Genet, 1999. **89**(1): p. 1-6.
52. Kang, E., et al., *Phenotype categorization of neurofibromatosis type I and correlation to NF1 mutation types.* J Hum Genet, 2020. **65**(2): p. 79-89.
53. Koczkowska, M., et al., *Genotype-Phenotype Correlation in NF1: Evidence for a More Severe Phenotype Associated with Missense Mutations Affecting NF1 Codons 844-848.* Am J Hum Genet, 2018. **102**(1): p. 69-87.

54. Koczkowska, M., et al., *Clinical spectrum of individuals with pathogenic NF1 missense variants affecting p.Met1149, p.Arg1276, and p.Lys1423: genotype-phenotype study in neurofibromatosis type 1.* Hum Mutat, 2020. **41**(1): p. 299-315.
55. Batalla, A., et al., *Genotype-phenotype correlation in type 1 neurofibromatosis: pMet992del mutation and milder disease.* Pediatr Dermatol, 2018. **35**(5): p. e268-e271.
56. Upadhyaya, M., et al., *An absence of cutaneous neurofibromas associated with a 3-bp inframe deletion in exon 17 of the NF1 gene (c.2970-2972 delAAT): evidence of a clinically significant NF1 genotype-phenotype correlation.* Am J Hum Genet, 2007. **80**(1): p. 140-51.
57. Pinna, V., et al., *p.Arg1809Cys substitution in neurofibromin is associated with a distinctive NF1 phenotype without neurofibromas.* Eur J Hum Genet, 2015. **23**(8): p. 1068-71.
58. Rojnueangnit, K., et al., *High Incidence of Noonan Syndrome Features Including Short Stature and Pulmonic Stenosis in Patients carrying NF1 Missense Mutations Affecting p.Arg1809: Genotype-Phenotype Correlation.* Hum Mutat, 2015. **36**(11): p. 1052-63.
59. Trevisson, E., et al., *The Arg1038Gly missense variant in the NF1 gene causes a mild phenotype without neurofibromas.* Mol Genet Genomic Med, 2019. **7**(5): p. e616.
60. Wallis, D., et al., *Neurofibromin (NF1) genetic variant structure-function analyses using a full-length mouse cDNA.* Hum Mutat, 2018. **39**(6): p. 816-821.
61. Al-Mulla, F., et al., *Age-dependent penetrance of different germline mutations in the BRCA1 gene.* J Clin Pathol, 2009. **62**(4): p. 350-6.
62. Bogdanova, N.V., et al., *High frequency and allele-specific differences of BRCA1 founder mutations in breast cancer and ovarian cancer patients from Belarus.* Clin Genet, 2010. **78**(4): p. 364-72.
63. Plakhins, G., et al., *Genotype-phenotype correlations among BRCA1 4153delA and 5382insC mutation carriers from Latvia.* BMC Med Genet, 2011. **12**: p. 147.
64. Loeys, B.L., D.M. Matthys, and A.M. de Paepe, *Genetic fibrillinopathies: new insights in molecular diagnosis and clinical management.* Acta Clin Belg, 2003. **58**(1): p. 3-11.
65. Zhou, Y., et al., *Genotype variant screening and phenotypic analysis of FBN1 in Chinese patients with isolated ectopia lentis.* Mol Med Rep, 2021. **23**(4).
66. Yang, Y., et al., *Novel p.G1344E mutation in FBN1 is associated with ectopia lentis.* Br J Ophthalmol, 2021. **105**(3): p. 341-347.
67. Gray, J., et al., *Functional characterization of human NTRK2 mutations identified in patients with severe early-onset obesity.* Int J Obes (Lond), 2007. **31**(2): p. 359-64.
68. Yeo, G.S., et al., *A de novo mutation affecting human TrkB associated with severe obesity and developmental delay.* Nat Neurosci, 2004. **7**(11): p. 1187-9.
69. Hamdan, F.F., et al., *High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies.* Am J Hum Genet, 2017. **101**(5): p. 664-685.
70. Sonoyama, T., et al., *Human BDNF/TrkB variants impair hippocampal synaptogenesis and associate with neurobehavioural abnormalities.* Sci Rep, 2020. **10**(1): p. 9028.

71. Uddin, M., Y. Wang, and M. Woodbury-Smith, *Artificial intelligence for precision medicine in neurodevelopmental disorders.* NPJ Digit Med, 2019. **2**: p. 112.
72. Johnson, K.B., et al., *Precision Medicine, AI, and the Future of Personalized Health Care.* Clin Transl Sci, 2021. **14**(1): p. 86-93.
73. Michael Patton, G.R., William E. Byrd, Matthew Might, *[MKMEDI] mediKanren: A System for Bio-medical Reasoning*. 2020: miniKanren Workshop.
74. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
75. Telenti, A., et al., *Deep sequencing of 10,000 human genomes.* Proc Natl Acad Sci U S A, 2016. **113**(42): p. 11901-11906.
76. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
77. Reetz, K., et al., *Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: a cross-sectional analysis of baseline data.* Lancet Neurol, 2015. **14**(2): p. 174-82.
78. Coppola, G., et al., *Functional genomic analysis of frataxin deficiency reveals tissue-specific alterations and identifies the PPARgamma pathway as a therapeutic target in Friedreich's ataxia.* Hum Mol Genet, 2009. **18**(13): p. 2452-61.
79. Napierala, J.S., et al., *Comprehensive analysis of gene expression patterns in Friedreich's ataxia fibroblasts by RNA sequencing reveals altered levels of protein synthesis factors and solute carriers.* Dis Model Mech, 2017. **10**(11): p. 1353-1369.
80. Budworth, H., et al., *Suppression of Somatic Expansion Delays the Onset of Pathophysiology in a Mouse Model of Huntington's Disease.* PLoS Genet, 2015. **11**(8): p. e1005267.
81. Bidichandani, S.I., et al., *Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence.* Hum Mol Genet, 1999. **8**(13): p. 2425-36.
82. Ku, S., et al., *Friedreich's ataxia induced pluripotent stem cells model intergenerational GAATTC triplet repeat instability.* Cell Stem Cell, 2010. **7**(5): p. 631-7.
83. Polak, U., et al., *Alleviating GAA Repeat Induced Transcriptional Silencing of the Friedreich's Ataxia Gene During Somatic Cell Reprogramming.* Stem Cells Dev, 2016. **25**(23): p. 1788-1800.
84. Klose, A., et al., *Selective disactivation of neurofibromin GAP activity in neurofibromatosis type 1.* Hum Mol Genet, 1998. **7**(8): p. 1261-8.
85. Yan, W., et al., *Structural Insights into the SPRED1-Neurofibromin-KRAS Complex and Disruption of SPRED1-Neurofibromin Interaction by Oncogenic EGFR.* Cell Rep, 2020. **32**(3): p. 107909.
86. Li, Y., et al., *Somatic mutations in the neurofibromatosis 1 gene in human tumors.* Cell, 1992. **69**(2): p. 275-81.
87. Kaul, V., S. Enslin, and S.A. Gross, *History of artificial intelligence in medicine.* Gastrointest Endosc, 2020. **92**(4): p. 807-812.
88. Ramesh, A.N., et al., *Artificial intelligence in medicine.* Ann R Coll Surg Engl, 2004. **86**(5): p. 334-8.

89. Steimann, F., *On the use and usefulness of fuzzy sets in medical AI.* Artif Intell Med, 2001. **21**(1-3): p. 131-7.
90. Ghahramani, Z., *Probabilistic machine learning and artificial intelligence.* Nature, 2015. **521**(7553): p. 452-9.