# Translational Genetics: From *Arabidopsis* to Friedreich's ataxia

**A thesis submitted for the degree of Doctor of Philosophy**

**By**

Amanda R Tabib
BBiomedSc(Hons)

March, 2015

School of Biological Sciences
Monash University, Clayton
VIC 3800
Australia

# Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of this thesis.

I declare here the specific contributions made by others, which are presented in this thesis. Dr. Sridevi Sureshkumar generated the natural suppressor Bur-0 lines and Ms Diana Bernal estimated the repeat lengths reported in this thesis. The results presented on the Cal-0 accession are from Dr. Peter McKeown in Charlie Spillane lab. Mr Aleksej Stevanovic designed the artificial microRNA constructs used for cloning in this thesis. Mr Andrei Seleznev carried out all the bioinformatics analysis of the smallRNA data. The idea of looking at smallRNAs came from the work of Mr Hannes Eimer on Arabidopsis and the original large-scale screen on Arabidopsis, from which the compounds were selected, was done by Dr Celine Tasset.

The core theme of this thesis is translational genetics of trinucleotide repeat disorders (TNRD)s. The ideas, development and writing up of all papers in this thesis were the principal responsibility of myself, the candidate, working within the School of Biological Sciences at Monash University, under the supervision of Assoc Professor Sureshkumar Balasubramanian.

Amanda Tabib

# Acknowledgements

The work presented in this thesis was supported by an Australian Postgraduate Award (APA) Scholarship and Monash University. I would like to acknowledge a few individuals who have contributed to the work in this thesis. I would firstly like to thank my supervisor Assoc Professor Sureshkumar Balasubramanian for allowing me to undertake this research in his lab. I would also like to thank the many Post Docs who have helped me in my research through their scientific advice and friendship especially Dr Lynette Fulton, Dr Eduardo Sanchez-Bermejo and Dr Celine Tasset.

I also acknowledge the input of Prof Charles Spillane and Dr Peter McKeown, our collaborators from the National University of Galway (NUiG), Ireland and Dr Joseph Sarsero from the Murdoch Children's Research Institute (MCRI) for his assistance setting up our cell culture facility. I also thank our lab manager Vignesh Sundooramoorthi for his friendship and help in the lab and my previous fellow PhD student, Dr Wangsheng Zhu, who helped me throughout the PhD. He is truly a gifted researcher and a good and encouraging friend to have had in the lab.

I would also like to express my deep appreciation and gratitude to the undergraduate students at Monash who undertook research units in our lab. Especially; Vicky Wong, Yindi Sutton and Luana Colling. The work done by these students was helpful to me in many ways and contributed a lot to my thesis and my PhD experience. I really enjoy teaching and interacting with students, they bring a lot of life and enthusiasm to the lab, which can get pretty repetitive and isolating at times. I hope that they all had a great experience in the lab and that I helped them learn new techniques and skills that will be useful for them in their future careers.

Amanda Tabib

# Abstract

Unstable tri-nucleotide microsatellite repeats are the causal mutation underlying more than 40 severe neuromuscular and neurodegenerative disorders in humans including Huntington's disease, Fragile X syndrome, Myotonic dystrophy and Friedreich's ataxia. These disorders are collectively referred to as tri-nucleotide repeat disorders (TNRD)s. To date there is currently no effective treatment for these disorders, which are mostly progressive, late-onset and ultimately fatal. The mechanisms contributing to these disorders are not fully known and this is largely due to a lack of suitable model systems. Recently, a type of TNRD was discovered in the Bur-0 strain of the model plant *Arabidopsis thaliana*.

The *Arabidopsis* tri-nucleotide repeat (TNR) expansion defect is very similar at the basic molecular level to the human TNRD Friedreich's ataxia (FRDA). Both are caused by dramatic intronic GAA/TTC trinucleotide repeat expansions that result in decreased expression of their affected genes, they both exhibit somatic variation and repeat instability to similar extents. Therefore in this thesis, I take advantage of the Bur-0 TNR expansion model to analyse similarities, define mechanisms and exploit them in a translational approach to identify potential target pathways and small molecules for treatment of FRDA.

In this thesis I analyze the mutational dynamics of the TTC/GAA repeat tract in wild *A. thaliana* populations in Ireland (from where the original Bur-0 accession was collected) and recover wild accessions with the expanded mutant allele. The expanded allelic variant was found to be maintained in wild at a relatively high frequency. I also explore the population structure of the Irish accessions and show that the repeat expansion is present in more than one sub-population of Irish accessions. I then go on to show that certain characteristics known to occur in human disorders, such as 'genetic anticipation' and the involvement of DNA repair pathways, are also occurring in the Bur-0 model. I then take a translational approach from *Arabidopsis* to a human FRDA cell culture model. I exploit the discovery of a novel mechanism discovered in plants in our group, which showed that TTC/GAA repeat expansion can lead to an increase in smallRNAs mapping to genes that harbor them. Taking a translational approach I show that there is an increase in smallRNAs that map to the defective *FXN* locus. I demonstrate that blocking smallRNA pathways actually help to increase *FXN* expression in FRDA cells. Furthermore, I test additional compounds identified through a screen previously done in the lab on the Bur-0 plants and through this identify additional compounds that increase *FXN* expression. Thus these translational approaches have identified potential therapeutic drug candidates for FRDA and potentially other TNRDs.

Overall, the work in this thesis explores the similarities and shows Bur-0 as a potential model for studying various aspects of TNRD research that can potentially be translatable to the human disorders.

# Table of Contents

# CHAPTER 1. Mutational dynamics of TTC/GAA repeat tracts at the *IIL1* locus in *Arabidopsis thaliana*.

## 1.1 INTRODUCTION

### 1.1.1 Microsatellite repeats

Microsatellites are short tandem repeats of 1-6 nucleotides found in the genomes of all organisms. Mutations in simple tandem repeats, small expansions or contractions, occur at rates up to $10^5$ times higher than point mutations as they arise more easily due to slip-strand mispairing rather than incorporation error (Dubrova, Plumb et al. 1998, Kayser, Roewer et al. 2000). Recombination during meiosis also occurs more frequently at these sites which gives rise to extensive variability in microsatellite lengths (Dubrova, Plumb et al. 1998). Expansions and contractions of microsatellites have been suggested to be a main source for rapid evolution of new forms and are believed to be a major source of phenotypic variation in evolution (Fondon and Garner 2004).

Small variations in tandem repeat lengths can inherently give rise to both phenotypic and functional diversification (Levdansky, Romano et al. 2007). Variation of microsatellite lengths within regulatory regions of genes likely plays a critical role in the generation of morphological variation by altering the timing and location of gene expression (Stern and Orgogozo 2008). Repeat length variation in the coding regions of certain developmental genes has been quantitatively associated with significant differences in limb and skull morphology in certain breeds of dog (Fondon and Garner 2004).

Microsatellites within promoters and other cis-regulatory regions of genes have been shown to influence expression of the *vasopressin 1a receptor* (*avpr1a*) gene

in humans and fimbriae formation in the bacteria *Haemophilus influenza* (Moxon, Rainey et al. 1994, Hammock, Lim et al. 2005). Microsatellites within transposable elements are also thought to be involved in regulation of gene expression (Kovtun and McMurray 2008, Ahmed and Liang 2012). Alterations in the number of repeats in certain cell wall proteins can also change physiological properties of the cell surface, providing selective advantages in certain conditions, e.g. in evading host immune systems (Frieman and Cormack 2004, Ma, Jensen et al. 2012).

Aside from these normal roles of microsatellites in influencing gene expression and creating essential diversity, microsatellites of repeating triplet nucleotides, are also the causative basis of a group of heritable genetic diseases in humans, known as tri-nucleotide repeat disorders (TNRD)s. Organisms have developed systems for resisting changes in microsatellite lengths that could be deleterious (Jiricny 2006, Boland and Goel 2010). Despite this, if the repetitive sequences become longer than a crucial threshold length they override genomic safeguards and can expand dramatically through generations and during development of the progeny (Ashley and Warren 1995). These expansions underlie more than 40 severe neuromuscular and neurodegenerative disorders in humans, which differ based on the nucleotides involved, the genes they affect and their location within the gene (Ashley and Warren 1995). Examples of such diseases include Fragile X syndrome (FXS), Huntington's disease (HD), Myotonic dystrophy (DM), Spinobulbar muscular atrophy and Friedreich's ataxia (FRDA).

### 1.1.2 Molecular mechanisms of trinucleotide repeat expansion

Despite two decades of research, the molecular mechanism(s) underlying trinucleotide repeat (TNR) expansion and instability is still largely unknown. So far, flaws in the processes of DNA replication and DNA repair are the primary candidates for TNR instability (Kunkel 1993, Pearson and Sinden 1998, Lahue and Slater 2003, Neumann, Lawson et al. 2010). Kunkel (1993) found that during

DNA replication, strands could misalign in the process of re-annealing in regions rich with repetitive DNA. When misaligned strands continue to replicate, the repeat can expand (Kunkel 1993). Kunkel's concept was the main theory for several years, however subsequent research found that TNR expansions can also occur in terminally differentiated cells, such as neurons, that do not undergo DNA replication (Mangiarini, Sathasivam et al. 1997, Gonitel, Moffitt et al. 2008), suggesting that error prone DNA replication cannot be the only source of TNR instability (Lahue and Slater 2003).

Denaturation and renaturation of expanded TNR regions in dsDNA promote DNA 'slip-outs' that are folded into hairpin-like structures. These hairpins kinetically 'trap' repetitive DNA in the otherwise unfavorable slipped-stranded configurations (McMurray 1999, Wells, Dere et al. 2005). Hairpin structures can form *in vivo* and potentially impair replication and repair enzymes (Miret, Pessoa-Brandao et al. 1998, Kerrest, Anand et al. 2009, Lin, Hubert et al. 2009). Thus, formation of these TNR hairpin structures can occur during any process that involves the separation of DNA strands and at present it has been demonstrated that expansions can be mediated through DNA replication, repair and recombination (Liu and Leffak 2012). *In vivo* studies of GAA/TTC repeat tract instability (in the DNA) in *E. coli*, yeast, mice, and human cell lines have implicated processes including; recombination, replication slippage, the direction of replication fork progression, lagging strand errors, Okazaki fragment processing, double strand break repair, mismatch repair and gap filling (Forgacs, Wren et al. 2001, Kovtun and McMurray 2001, Du, Campau et al. 2012). These mechanisms are not mutually exclusive, and any combination of them may result in significant GAA/TTC expansion and a disease symptoms.

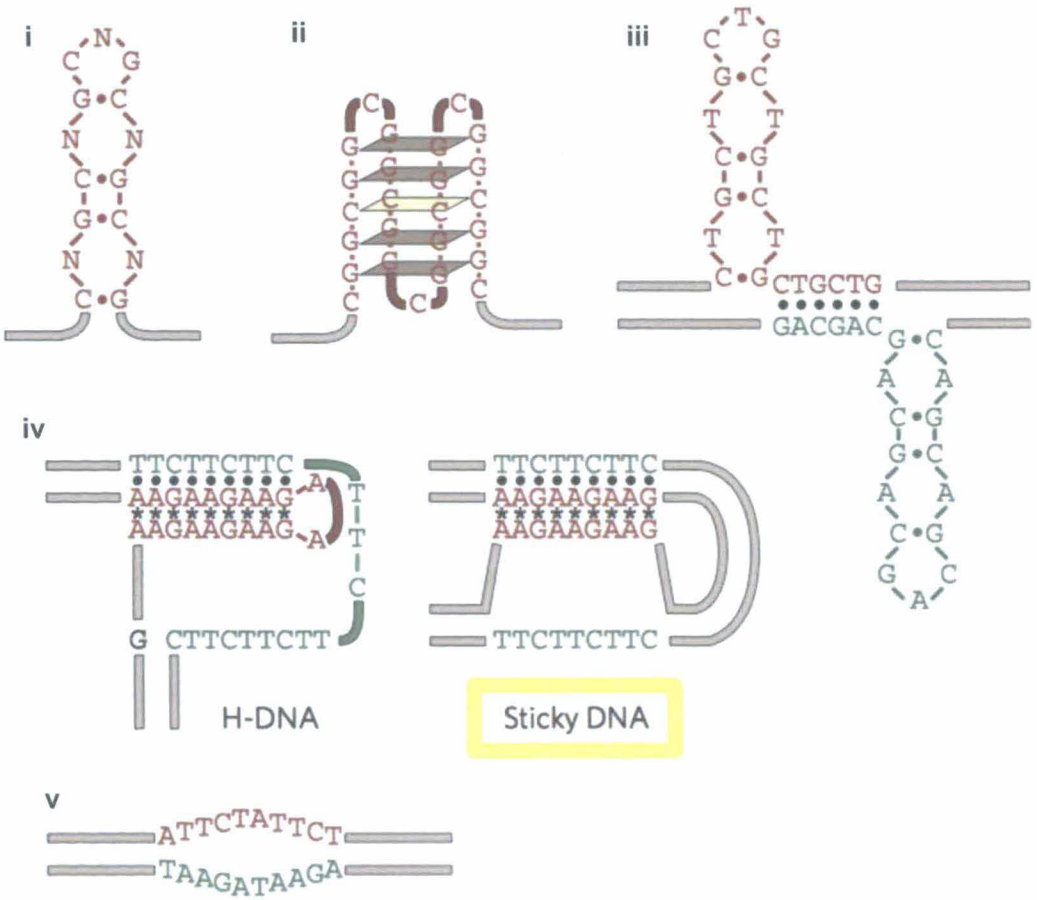



**Figure 1. Unusual DNA structures formed by expandable repeats.** Repetitive DNA can form several unusual structures, examples of which are shown. **(A)** The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. Modified from Mirkin et al, 2007. **(i)** An imperfect hairpin formed by CNG/CNG repeats. **(ii)** A quadruplex-like structure formed by the CGG/CCG repeat. Brown rectangles indicate G quartets, and the yellow rectangle indicates an i motif. **(iii)** A slipped- stranded structure formed by the CTG/CAG repeat. **(iv)** H-DNA and sticky DNA (highlighted in yellow) formed by the GAA/TTC repeat (found in Friedriech's ataxia and 27 °C Bur-0). Only one possible isoform, in which the homopurine strand is donated to the triplex, is shown for both structures. Reverse Hoogsteen pairing is indicated by asterisks. **(v)** A DNA-unwinding element formed by the ATTCT/AGAAT repeat.

Until recently triplet repeat expansion causing diseases had only been described in humans. However, in 2009, a TNR associated genetic defect was discovered in the Bur-0 accession of *Arabidopsis thaliana* (hereafter referred to as *A. thaliana)* (Sureshkumar, Todesco et al. 2009). The Bur-0 strain, which originates from Ireland, carries a dramatically expanded intronic TTC/GAA repeat in the *ISOPROPYL MALATE ISOMERASE LARGE SUBUNIT 1* (*IIL1*; At4g13430) gene. The

repeat expansion causes an environment-dependent reduction in *IIL1* expression levels and severely impairs growth of this strain in certain environments such as higher temperatures. The phenotype comprising of these growth defects resulting from the TTC/GAA expansion is referred to as *'irregularly impaired leaves' (iil)*. The Bur-0 TNR expansion shares striking parallels with the human disorder FRDA at the molecular level (Sureshkumar, Todesco et al. 2009). Therefore, by analyzing the Bur-0 model, we can gain an understanding into repeat expansion dynamics and potentially relate this to the human situation.

It is clear that small changes in microsatellite lengths are normal and play important roles in creating diversity and that when these repeat tracts surpass a threshold length, severe pathology can result. The longer the repeat tract the more highly unstable they become with both marked hereditable and somatic variations in repeat lengths occurring (Mirkin 2007). When inherited over generations, triplet repeat expansions exhibit 'genetic anticipation' that is they tend to greatly favour expansion over contraction, leading to larger expansions and hence the associated increased disease severity in the progeny (Carpenter 1994). It is currently unknown why the expanded repeats are so unstable as well as what factors affect and contribute to this instability. This is a significant knowledge gap in TNRD research. Therefore, in this chapter we investigate the effects of environment on the TTC/GAA repeat dynamics in *A. thaliana*.

### 1.1.3 Common mechanisms mediating effects of TNR expansions

At the gene expression level, genes associated with expanded TNRs are either affected by loss of function, leading to protein deficiency, or gain of function, when a mutant protein has a dominant-negative effect on its normal counterpart or acquires a novel deleterious function (Gordenin, Kunkel et al. 1997). Exonic expansions such as in Huntington's disease, Dentatorubropallidoluysian atrophy, Spinobulbar muscular atrophy and several types of inherited cerebellar ataxia, are typically caused by CAG/CUG expansion resulting in long poly gluatamine

(polyQ) tracts (Zoghbi and Orr 2000). This aberrant protein product loses it native function and takes on a toxic gain-of-function role forming insoluble deposits within the cell responsible for the underlying pathology (McMurray 2010).

Conversely, TNR expansions in non-coding regions cause a loss-of-function mutation resulting in decreased gene expression (McMurray 2010). The GAA/TTC repeats in Bur-0 and FRDA are located within introns. Therefore, if the genes were expressed properly the intron bearing expansion should theoretically be spliced out and a normal functional protein result. However, for reasons not fully understood, the mRNA levels and protein levels of frataxin are reduced which ultimately leads to the clinical pathology of FRDA.

FRDA has been studied intensely for the past >15 years and many mechanisms of *Frataxin* (*FXN*) gene silencing have been proposed including; obstructing transcription, abnormal splicing and epigenetic chromatin silencing (Grabczyk, Mancuso et al. 2007, Mirkin 2007). Some data even suggest a potential toxic gain-of-function role, in certain TNRs, at the RNA level involving protein sequestration and RNA interference (RNAi) (Malinina 2005).

It is obvious that there are common underlying mechanisms of TNR expansions including DNA repair mechanisms, effects on protein function & gene expression and TNR instability. Two key questions in FRDA research that remain unanswered are: 'What is the exact mechanism by which the GAA/TCC repeat expansion leads to *FXN* gene silencing and reduced levels of frataxin expression?' And 'What causes the dynamic nature of the GAA/TCC repeat expansions and how does this affect the inheritance and subsequent progression of FRDA disease?' These questions relate to the underlying molecular biology of the expansion mutation itself and may be translatable with the Bur-0 model. The Bur-0 model, which is a natural accession of *A. thaliana*, contains a very similar TNR expansion to that in the *FXN* gene of FRDA patients. That is a large intronic GAA/TTC expansion. In order to use Bur-0 as a model to study the mechanisms of GAA/TTC expansion pathology at the basic molecular level, it is imperative to

first determine whether common mechanisms exist in the Bur-0 model. If so, the results could give translational insights into the human situation.

### 1.1.4 Bur-0, an example of a triplet expansion associated genetic defect

Having a natural model of a triplet repeat expansion associated genetic defect in *A. thaliana* holds enormous potential in tri-nucleotide repeat disorder research. Studying a particular phenotypic trait of an organism in that organism's natural environment enables us to understand the trait in the context of evolution. Evolutionary forces do not allow disadvantageous alleles to last within populations in large numbers. Commonly, if a particular seemingly deleterious phenotypic variant is observed in a population more frequently than expected, it either has some advantage associated with it or is neutral in that given environment.

For example, sickle cell anemia is a crippling inheritable haemoglobin disorder that occurs at much greater frequencies within certain ethnic groups (Weatherall and Clegg 2001, Grosse, Odame et al. 2011). The defective allele is maintained at a much higher frequency in African populations where the common parasitic infection malaria is endemic. In such populations the sickle cell allelic variant Hb S, offers an adaptive advantage by conferring a resistance to malarial infection (Gong, Parikh et al. 2013).

Analysis of the length of the TTC/GAA triplet repeat in the *IIL1* gene across 95 globally distributed *A. thaliana* stains obtained from the stock center revealed, a normal distribution in repeat length ranging from 0-36 repeats with the exception of the Bur-0 strain, which possessed 400+ repeats (Sureshkumar, Todesco et al. 2009). This suggested that this triplet repeat expansion is considerably rare and thus far, unique to the Bur-0 strain, which is the only strain from Ireland available in the stock center. The Bur-0 accession was originally collected from Burren national park, County Clare, Ireland in 1958 by

the collector D. Ratcliffe. Given that the Bur-0 accession contains the only known triplet repeat expansion described outside humans, the main goal in this chapter was to investigate whether this expansion is still recoverable in Ireland and if so analyse the mutational dynamics of the *IIL1* TTC/GAA repeats in Irish strains of *A. thaliana*.

Plants are sessile organisms and therefore need to display plasticity and be able to adapt to the environment they are in. Over time, through genotype by environment interaction (GXE), positive selection (the acquisition and propagation of advantageous mutations) or negative selection (removal of detrimental alleles) will take place, enabling species to adapt to their local environment. The Burren region, from which Bur-0 originates, has a unique ecology owing to the effects of the last glacial period (Hennessy 2010). As the ice melted, kartification (solutional erosion) occurred and the underlying limestone was left with deep cracks known as 'grikes' and isolated rocks called 'Clints' this gives the characteristic 'limestone pavements' that cover the Burren. Aside from this unique glacio-karsk landscape, the limestone makes the soil unusually alkaline and together these unique characteristics support the unusual array of Mediterranean, arctic and alpine plants growing side-by-side, some of which are exceedingly rare (Webb 1961, Scannell 1983). Given this unique ecology, we sort to undertake a population study in Ireland, where by wild *A. thaliana* populations would be collected from the Burren and surrounding regions, and the mutational dynamics/repeat variability within their *IIL1* TTC/GAA repeat tracts analysed, thus investigating the effects of the local environment on the triplet repeat. To achieve this we employed 2 aims:

**AIM 1:** To investigate whether the triplet expansion is still present in Ireland and if so, analyse the tri-nucleotide repeat instability at the *IIL1* locus in the Irish accessions of *A. thaliana*.

**AIM 2:** To analyse the impacts of various abiotic environmental stresses on the *A. thaliana* '*iil*' phenotype.

### 1.1.5 Investigating TTC/GAA mutational dynamics in wild Irish accessions of *A. thaliana*

In the 1st aim, we investigate *IIL1* TTC/GAA microsatellite repeat variability in a population study of wild *A. thaliana* collected from across Ireland. As mentioned above, a global analysis on the TTC/GAA repeat lengths has revealed that the *IIL1* triplet expansion is considerably rare and thus far found to be unique to the Bur-0 accession. Though, aside from the Bur-0 accession, there are no other *A. thaliana* accessions from Ireland commercially available at the stock center. This together with the unusual ecology of the Burren region prompted us to investigate the TTC/GAA repeat dynamics in Irish *A. thaliana*.

In this study we had the primary objective of determining whether the repeat expansion allele for *IIL1* TTC/GAA was still present in the wild, and a secondary objective to investigate the relationship between repeat length variability and the geographical location from which the Ireland-wide *A. thaliana* samples originated. That is, taking a genotype by environment (GXE) type approach to determine if any local environment had an influence on repeat length.

It has been suggested that certain stresses in the local environment may contribute to the repeat instability in other TNRDs (Mangiarini, Sathasivam et al. 1997, Kennedy, Evans et al. 2003). For example, a resent study in murine model of Huntington's disease (HD) revealed an accumulation of larger expansions in tissues that are major sites of HD pathology such as the brain striatum (Kennedy, Evans et al. 2003). Larger expansions also accumulate in the muscle of Myotonic dystrophy (DM1) patients (Ashizawa, Dubel et al. 1993, Morales, Couto et al. 2012). Thus the local environment seems to be impacting the mutational dynamics.

In order to analyze the *IIL1* TTC/GAA repeat variability within natural populations of Irish *A. thaliana*, a population genetic study was undertaken over

a two-year period from May-June 2011 and May-June 2012. Seed and DNA samples were collected from wild *A. thaliana* plants from over 200 different locations across Ireland, with samples also collected from Scotland.

### 1.1.6 Investigating affects of abiotic stress on *A. thaliana iil* phenotype and TTC/GAA instability

In the 2nd aim we employ Bur-0 plants, with 400+ TTC/GAA repeats and previously described spontaneous natural suppressors (NS) in the Bur-0 background with the *IIL1* TTC/GAA repeat tract harboring ~61 and 213 repeats (Bernal and Balasubramanian, personal communication)
(Sureshkumar, Todesco et al. 2009). These plants were grown on soil and MS plates under more than 10 different stress conditions, in order to determine which conditions expose the cryptic *IIL1* TNR expansion defects.

IIL1 is an enzyme with broad specificity that catalyzes reversible hydroxyacid isomerizations via dehydration/hydration reactions. IIL1-mediated isomerization events are involved in the biosynthesis of leucine and isoleucine as well as glucosinolates. However, aside from this, little is known concerning IIL1 function. At 23 °C the *IIL1* gene expression level in Bur-0 is similar to that in the reference strain Col-0, but at elevated temperature conditions, such as 27 °C, *IIL1* expression in Bur-0 is greatly reduced to ≤25% when compared to the levels in Col-0 at 27 °C (Fig. 8) as described previously (Sureshkumar, Todesco et al. 2009).

Since the TTC/GAA causing growth defects in *A. thaliana* Bur-0 are cryptic and not seen under standard growth conditions, it is interesting to establish whether disease phenotypes due to the *IIL1* repeat expansion occur under any other environmental conditions and abiotic stresses. This would tell us whether the TTC/GAA repeat expansion associated *iil* defects are a general stress response or specific to higher temperatures and thus provide insights into the intrinsic

nature of the triplet expansion associated genetic defect in *A. thaliana*.

Recovery of the *iil* or any other cryptic phenotype associated with the TNR expansion in other stress conditions will also help characterize the nature of the TNR defect. At the DNA level, repeat length dynamics under different stress conditions can give insights into the nature of instability. Impacts of various abiotic stress conditions on *IIL1* gene expression can also give insights into the mechanisms undelrying TNR expansion and its associated effects.

## 1.2 RESULTS

### 1.2.1 Population genetic study of wild Irish *A. thaliana*.

***AIM 1:*** *To investigate whether the triplet expansion is still present in Ireland and if so, analyse the tri-nucleotide repeat instability at the IIL1 locus in the Irish accessions of A. thaliana*

### 1.2.2 A collection of Irish populations of *A. thaliana*

We made two sets of targeted collections of *A. thaliana* accessions from Irish populations. The first collection was from the Burren region from where the Bur-0 strain was originally collected (www.arabidopsis.org). The second set of collections comprised additional accessions from other parts of the island of Ireland. More than 530 samples of leaf tissue and seeds of *A. thaliana* were collected from 150 distinct locations (Fig 2) across Ireland, spanning an area of ~260 x 265 Km with a minimum-maximum distance of 0.1- 47 km respectively.

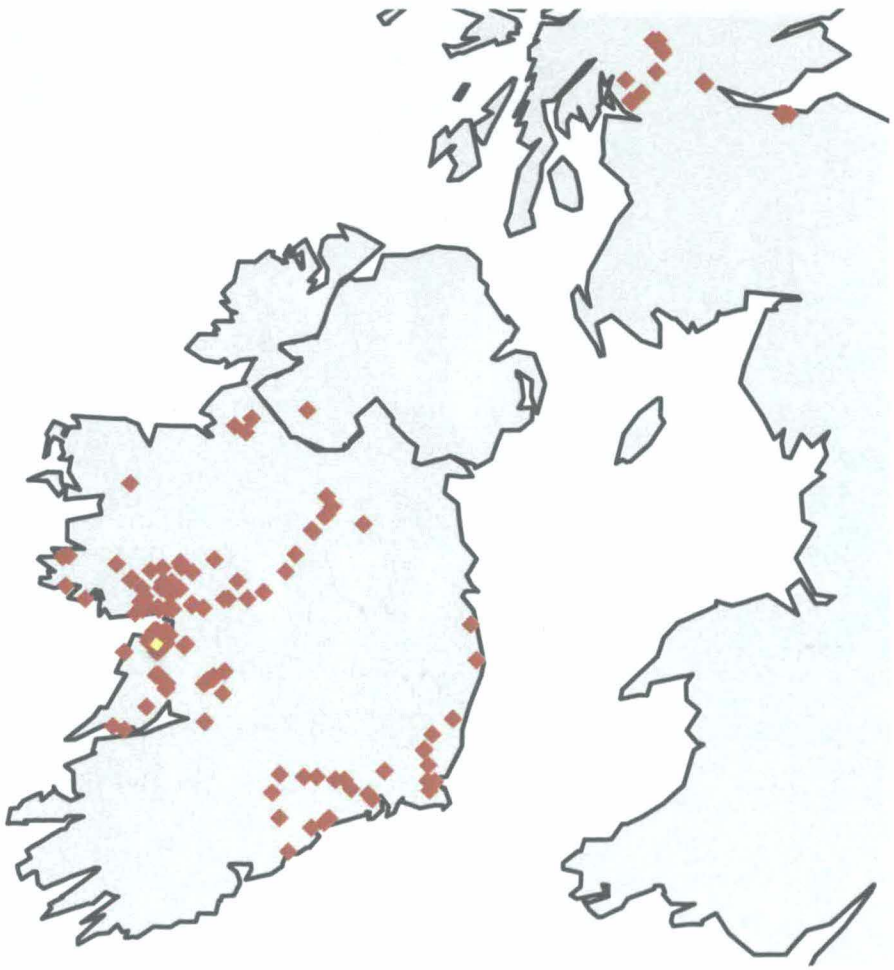


**Figure 2. Geographical location of *A. thaliana* populations of Ireland and Scotland surveyed in this work.** Yellow dot indicates approximate location of lab strain Bur-0 collected in 1958. Red dots indicate wild populations of *A. thaliana* collected in 2011/2012 and used in this study

Several individual plants were sampled from each established population wherever these could be found. Demographic factors of: geographic location (latitude and longitude), altitude (in m above sea level), and any observations relating to the substrate and growth habit were recorded from each collection site (Appendix - Table S1). Phenotypic variation was evident in the natural settings including differences in leaf morphology and plant size (Fig 3). Plants growing in more challenging microenvironments such as the bases of walls appeared to show evidence of stress (Fig 3), suggesting genotypic interactions with the local microenvironment.


**Figure 3. Variation in *A. thaliana* phenotypes at various collection sites in Ireland. (A & B)** Phenotypic characteristics of populations established in rocky limestone rich environments within the Burren. **(C, F & G)** Populations established in open soil; **(D, E, & H)** populations growing in more restricted niches displaying typical stressed characteristics. **(J)** Leaf samples from a collection of accessions from south east Ireland displaying strongly serrated leaves.

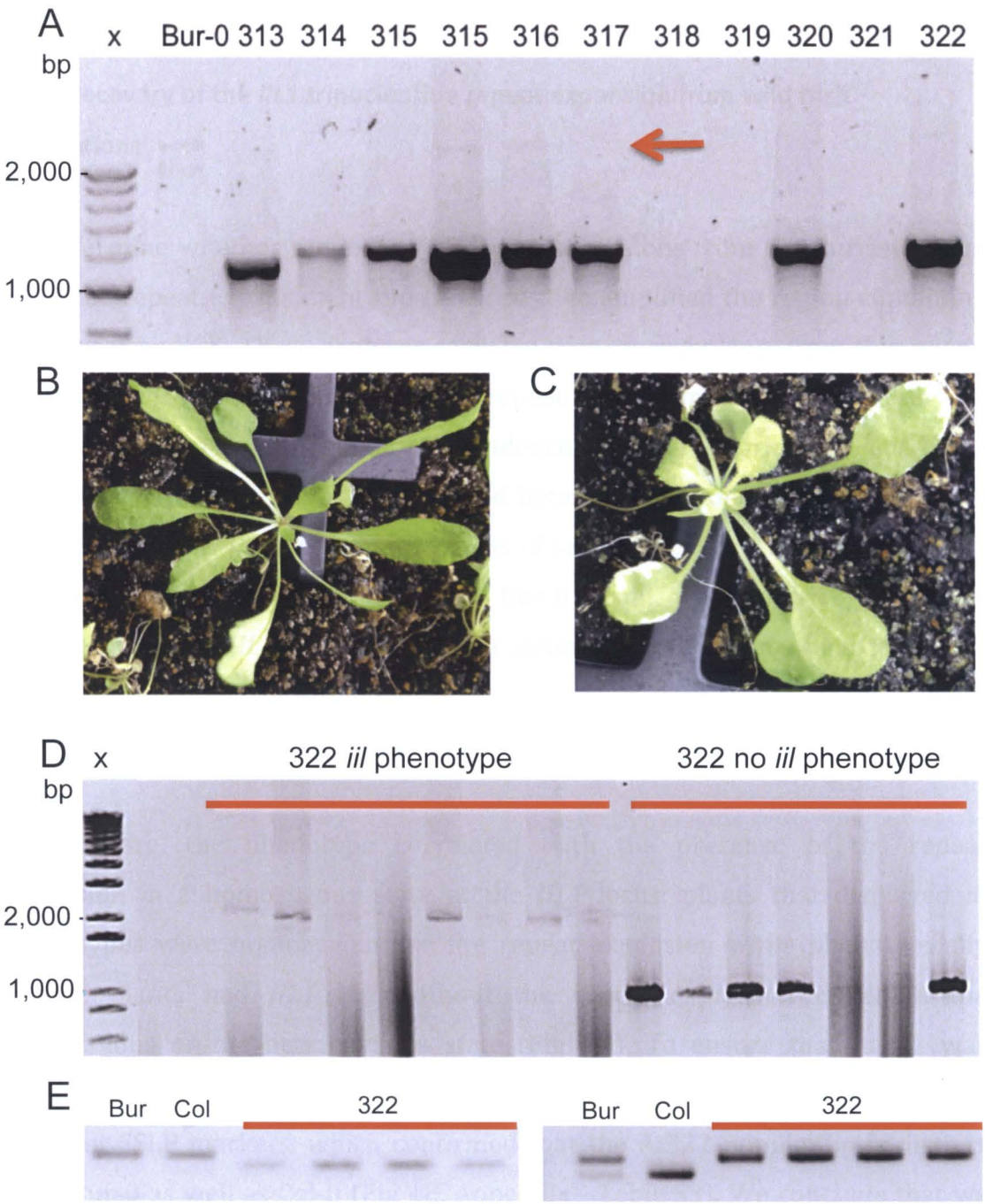



**Figure 4. Recovery of the *IIL1* triplet repeat expansion from Ireland. (A)** PCR analysis of Bur-0 and wild accessions At313-322. Red arrow indicates potentially heterozygous plants harbouring dramatically expanded *IIL1* TTC/GAA allele. Seeds collected from the potential heterozygotes were grown at 27 °C SD and clear segregation for the *iil* phenotype was observed in the progeny. **(B, C)** Representative pictures for the Irish strains displaying either the *iil* (B) or normal (without *iil*) (C) phenotypes. **(D)** PCR analysis of the *IIL1* TTC/GAA repeat length of individuals with and without the *iil* phenotype. **(E)** Genotypic analysis using two SSLP markers on Bur-0, Col-0 and At322 plants that segregated for the *iil* phenotype. Results from two markers are shown, which show that At322 plants are different in their genetic background to both Bur-0 and Col-0, but similar to each other despite the presence or absence of the *IIL1* TTC/GAA expansion.

### 1.2.3 Recovery of the *IIL1* trinucleotide repeat expansion from wild Irish populations

To determine whether the extant wild Irish accessions from the Burren region carry the repeat expansion at the *IIL1* locus, we amplified the region containing the repeat by PCR. These analyses revealed that several plants from this region contained the repeat expansion. The repeat expansion was often present in either the heterozygous state or with evidence for somatic variation (Fig 4A). To compare the phenotypes of an identified heterozygous population to the Bur-0 accession, we grew plants from the seeds of sample At322, at 27 °C under short day (SD) conditions. At 27 °C SD, the Bur-0 strain, which carries the repeat expansion at the *IIL1* locus, displays a strong growth-arrest phenotype with narrow tapered leaves. We found that plants grown from the At322 seeds were segregating for the *iil* phenotype (Fig 4B, C).

Furthermore, the phenotype correlated with the presence of the repeat expansion in a homozygous state at the *IIL1* locus: plants that displayed *iil* phenotypes were homozygous for the repeat expansion while phenotypically normal plants had *IIL1* loci without the repeat expansion either in the homozygous or the heterozygous state (Fig 4D). To ensure that At322 was independent from the Bur-0 accession, we genotyped these plants using 23 different SSLP markers, which confirmed that the At322 population is distinct from Bur-0 as well as Col-0 (Fig 4E, Appendix - Table S3). We conclude that we have recovered the expanded TTC/GAA triplet repeat in the *IIL1* gene, which is persisting in nature within wild Irish populations of *A. thaliana.*

To determine the frequency of the *IIL1* repeat expansion in Irish populations, we undertook a screen that combined a phenotypic analysis at higher temperatures with a PCR analysis. After screening several individuals representing each of the 131 Irish populations at 27 °C, we determined their *IIL1* TTC/GAA repeat lengths by PCR (Fig 5A, B). Variations in repeat length were observed between these populations including several populations with dramatically expanded repeat

tracts (Fig 5B). A total of 10 samples representing seven geographic locations (At216, At249, At319, At322, At369, At370, At379, At434, At460-462 and At533) contained TTC/GAA tracts containing over 120 triplets, which has been identified as the threshold for displaying the *iil* phenotype (Sureshkumar, Todesco et al. 2009). A range in *iil* phenotype severity was also observed among the affected accessions (Fig 5C).

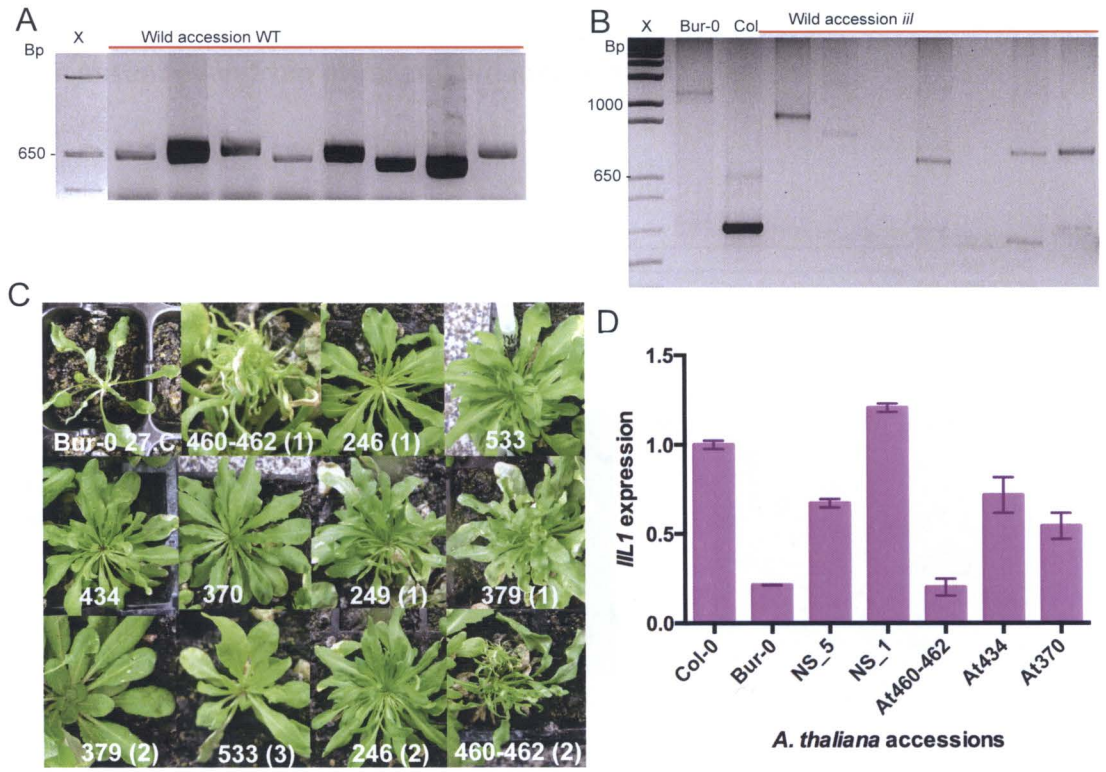



**Figure 5. Analysis of *IIL1* TTC/GAA triplet repeat variation. (A)** Representative PCR analysis showing variability in the *IIL1* triplet repeats within the non-expanded range among the Irish accessions. **(B)** PCR analysis showing variability of the *IIL1* repeat among accessions with dramatically expanded repeat tracts. Lanes 4, 5, 7, 9 and 10 are wild samples At460-462, 319, 369, 434 & 370 respectively. **(C)** Variability in the severity of the *iil* phenotype at 27 °C among wild Irish accessions. **(D)** qPCR expression analysis of the *IIL1* gene in Bur-0, NS_5 NS_1 and 3 wild accessions; At460-462, 434 and 370. Where NS_5 & NS_1 are natural suppressors, in Bur-0 background, of repeat tract length harbouring ~200 and ~60 repeats respectively. Results displayed as relative expression to Col-0 control. Error bars indicate standard error. qPCR primers, as in Appendix - Table S2.

In addition to the Irish strains, we observed the repeat expansion in two other populations. One accession from Scotland (At434), collected from Callander,

Scotland, UK also carried the expansion (Fig 5). Additionally we tested for repeat expansion in strains that are closer to Bur-0 (Cal-0, Mc-0, Ty-0 and Tamm-2) as defined through the 1001 genomes project (http://signal.salk.edu/Images/1001genomes.ppmcp2.png) and recovered the repeat expansion only in the Cal-0 strain (from Calver Fell, Derbyshire, Figure 6). Both the Scottish accession and Cal-0 also display typical *iil* phenotypes at higher temperatures (Fig 5C & 6). We conclude that the expanded TTC/GAA repeat at the *IIL1* locus is restricted mostly to Irish *A. thaliana* accessions, being found in several locations on the island of Ireland and possibly two locations in Britain, but is absent even from other related accessions.

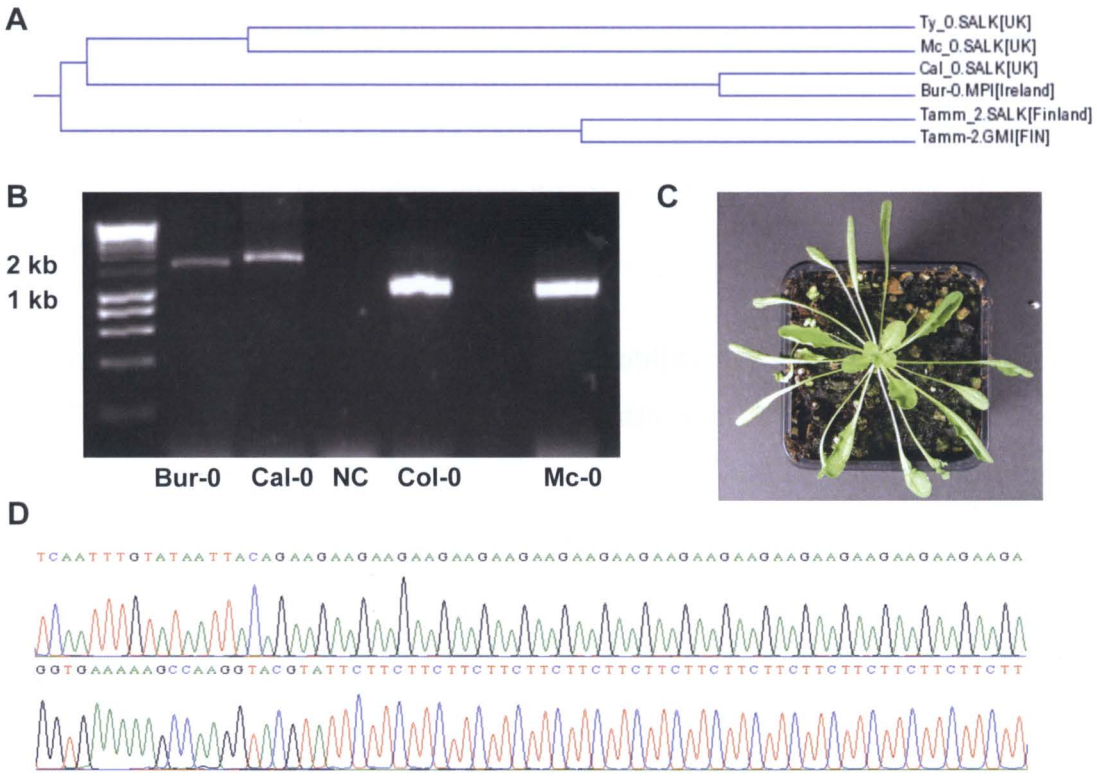



**Figure 6. Detection of repeat expansion in Cal-0. (A)** A phylogenetic tree showing the relationship of accessions with Bur-0 based on the 1001 genome project from http://signal.salk.edu/Images/1001genomes.ppmcp2.png). **(B)** TTC/GAA repeat PCR at the *IIL1* locus, showing expansions in Bur-0 and Cal-0 accessions, negative control (NC) and Col-0 & Mc-0 *A. thaliana* acessions without expansion. **(C)** Cal-0 accession displaying similar *iil* growth defects at elevated temperature. **(D)** Sequencing results of Cal-0 accession *IIL1* locus displaying TTC/GAA expansion. Forward direction showing start of expanded section (top) and reverse direction showing expansion end, no definite overlap possible (middle). (Image supplied by Dr Peter McKeown).

### 1.2.4 Irish populations display bimodal distribution of repeat length in the non-expanded range

Earlier studies have shown that the *IIL1* TTC/GAA repeat number varies between 0-36 in global populations and that the copy number is normally distributed. To analyse repeat length in the non-expanded range, we first analysed a subset of samples, At314 to At322, which were collected from a location approximately 4 km from the Burren. Sample At313, from County Fermanagh in Northern Ireland ~170 km from the Burren, was used as a geographically distal control. There is a noticeable variability in repeat length between samples At313 and At314-322 (Fig 5A), with samples At314-322 having slightly longer repeat tracts for their allele that lacks the repeat expansion. PCR products for the non-expanded alleles were excised from the gel and sequenced, which gave TTC/GAA repeat numbers of 11 for sample At313 and 26 for samples At314-At322 (Appendix - Table S1), suggesting a potential association between the higher copy number in the non-expanded alleles with repeat expansion.

To estimate the pattern of repeat length variability, we amplified and sequenced the TTC/GAA repeat from 131 Irish populations of *A. thaliana.* In contrast to global populations (Sureshkumar, Todesco et al. 2009), where a continuous normal distribution of repeats with copy numbers between 0-36 was observed, the Irish population displayed a distinct bimodal distribution in repeat number (Fig 7A). To identify patterns in the observed variation in repeat numbers we geographically mapped the *IIL1* TTC/GAA repeat lengths of each population (Fig 7B). While there is potentially some clustering of longer TTC/GAA repeats in the Burren region, there is no obvious pattern relating the TTC/GAA repeat lengths and geographical origin.

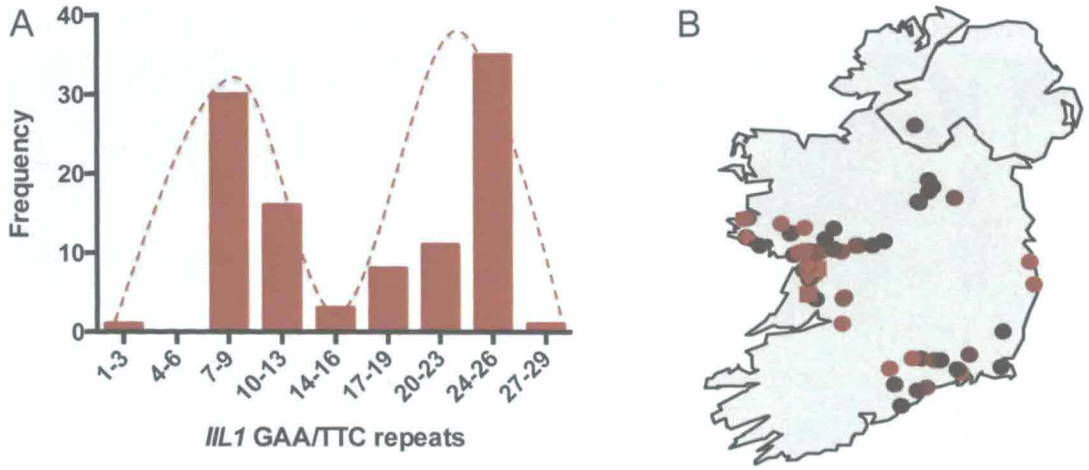



**Figure 7. TTC/GAA mutational dynamics in Irish *A. thaliana*.** **(A)** Frequency distribution of the copy number of the TTC/GAA repeats of the non-expanded alleles among wild Irish accessions. **(B)** Heat map of *IIL1* triplet repeat length overlaid onto geographic map. Black to Red indicates increasing repeat length from 3 repeats to 29 repeats.

There was also a trend observed in wild Irish accession with dramatically expanded TTC/GAA tracts at the *IIL1* locus, where they displayed mild *iil* phenotypic growth defects in temperatures as low as 23 °C in addition to elevated temperatures. The lab strain Bur-0 does not normally display the *iil* cryptic growth defects at 23 °C, despite having a larger TTC/GAA expansion than that observed in most *IIL1* expanded wild accessions.

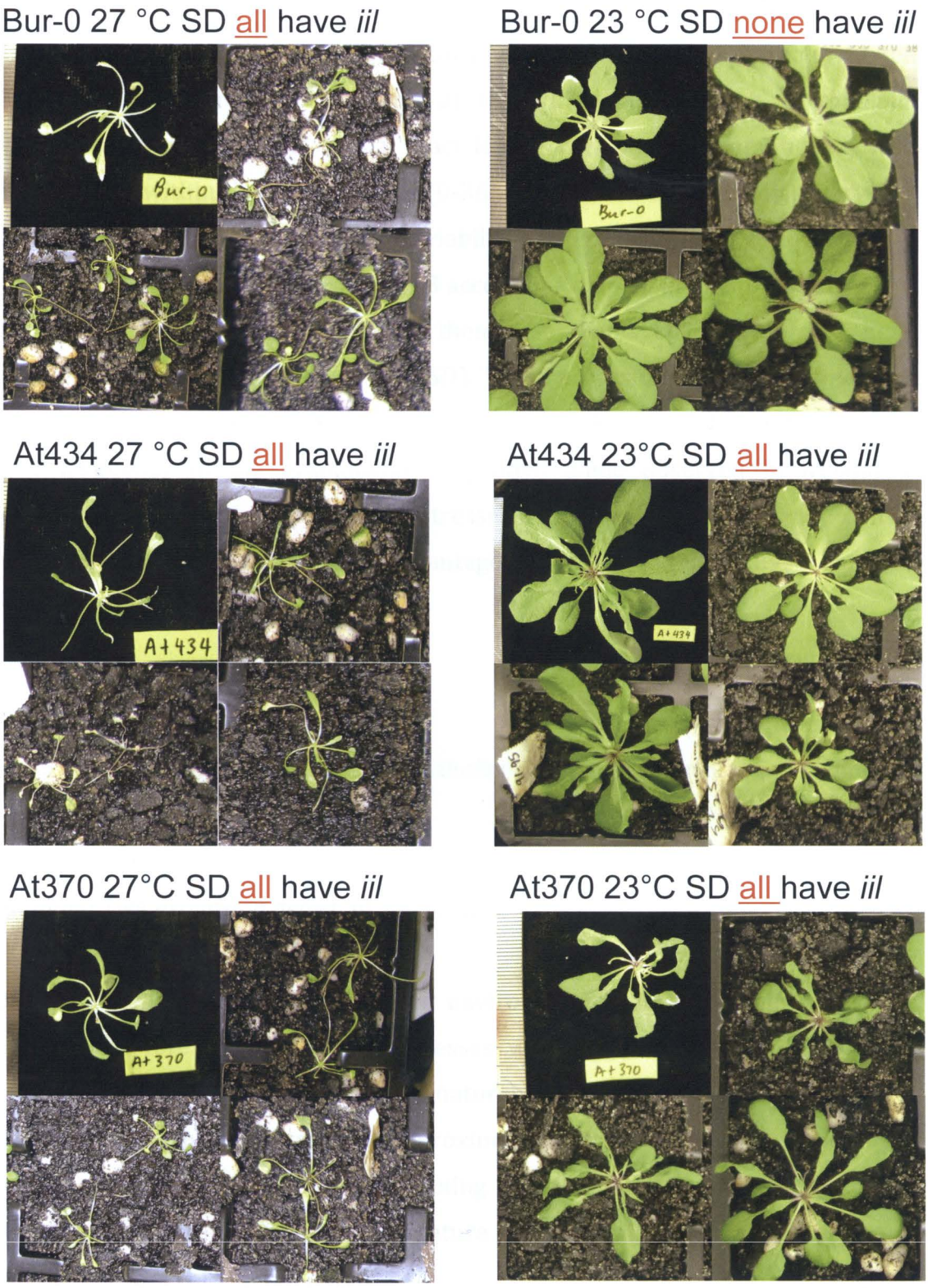



**Figure 8. Mild *iil* phenotype displayed in 23 °C wild accessions At434 & At370.** Bur-0 TTC/GAA expansion is greater than wild accessions At434 & At370. Bur-0 and wild accessions At434 & At370 all display *iil* phenotype at 27 °C. However At434 & At370 were consistently found to also display mild *iil* phenotypic growth defects at 23 °C. Where as Bur-0 generally did not. Thus it appears the wild accessions have a lower temperature threshold before the *iil* phenotype appears despite a less expanded TTC/GAA.

In summary, a population genetic study was undertaken to characterize the TTC/GAA triplet repeat variability at the *IIL1* locus in Irish *A. thaliana.* Considerable variability in repeat tract length was observed among the wild populations within the normal range (0-36 repeats) as well as the discovery of 8 accessions displaying extreme variability in repeat length, maintaining dramatically expanded alleles. These 8 accessions display *iil* phenotypes at 27 °C SD (of variable severity). Some of them also displayed reduced *IIL1* gene expression upon qPCR analysis (Fig 5D). These accessions are in a similar, yet distinctly different genetic background to Bur-0 as determined by SSLP analysis. The frequency of the dramatically expanded allele observed in Ireland is considerably high; perhaps specific stresses in local environment could confer this allele with an environmental advantage. This triggered us to investigate our second aim.

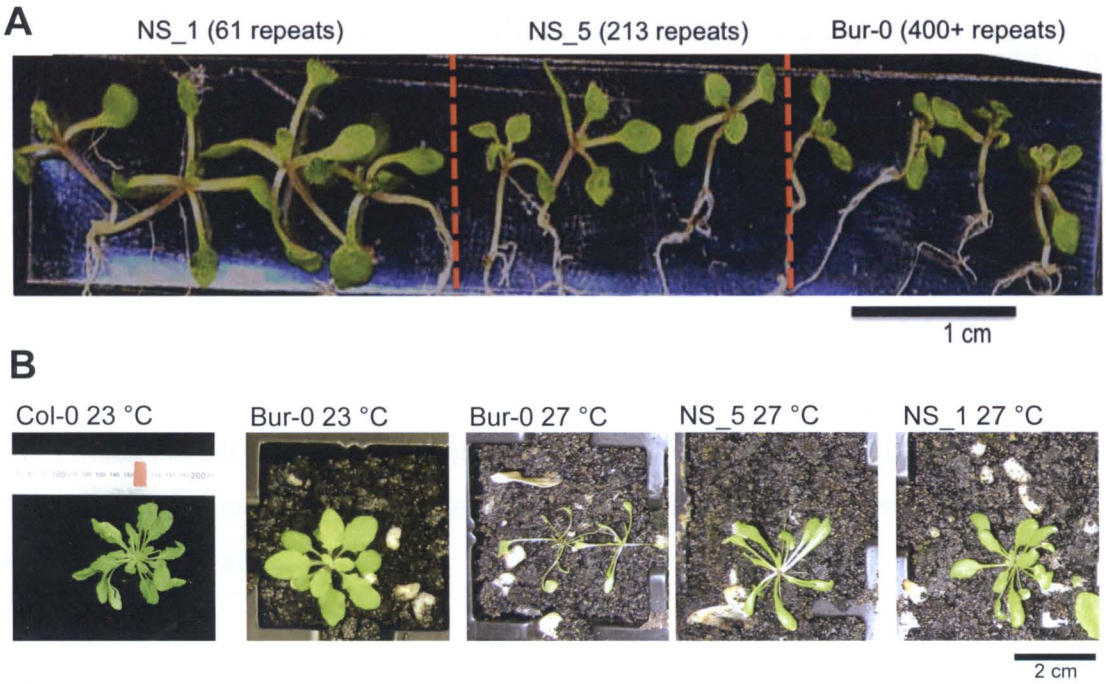
### 1.2.5 Influence of abiotic stresses on *A. thaliana* TNR expansion associated defect

***AIM 2:*** *To analyse the impacts of various abiotic environmental stresses on the A. thaliana TNR expansion defect 'iil' phenotype.*

Stresses used were designed to mimic environmental levels the plants may face in nature. More than 10 abiotic stresses were applied to *A. thaliana* Bur-0 accession (400 + repeats) and 2 natural suppressors (NS) in the Bur-0 background with repeat tracts of approximately 60 and 200 as described in the materials & methods. Extensive screening of these plants in soil and MS plates has found that only elevated temperature and to a mild extent UV-B exposure unmasks the cryptic *iil* phenotype.

Differences in plant size were observed in 7- day old Bur-0 and the natural suppressors when grown on MS plates (Fig 9A). Bur-0, NS_1 and NS_5 plants were observed visually and weighed and found to be the same size at 23 °C.

However, at 27 °C the Bur-0 plants were found to be smaller than the NS_1 plants. This revealed an anti-correlation with plant size and TTC/GAA expansion tract length (Fig 9A). Longer growth duration at 27 °C, in soil and on MS plates revealed the emergence of the *iil* phenotype at ~2.5-3 weeks growth. There was also a strong correlation with *iil* phenotype and TTC/GAA expansion tract length. That is, larger repeats were associated with a more severe *iil* phenotype (Fig 9B).




**C**

**Comparison of *IIL1* gene expression in Bur-0 accession *A. thaliana* & natural suppressors (NS) in Bur-0 TTC/GAA tract length at 23 °C & 27 °C**

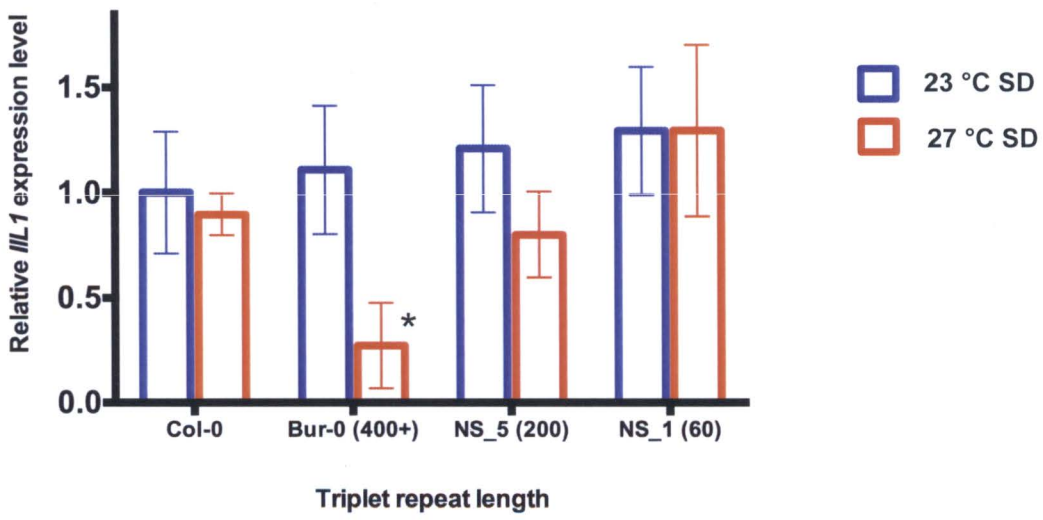


**Figure 9. Impacts of elevated temperature on *A. thaliana* TNR expansion defect. (A)** 1-week old seedlings of NS_1 (61 repeats), NS_5 (213 repeats) and Bur-0 (400+ repeats) grown in 27 °C short days. A noticeable anti-correlation of seedling size and TTC/GAA repeat tract length is apparent. (Note that no plants yet display an *iil* phenotype). **(B)** Representative images of Col-0, Bur-0 at 23 °C & 27 °C, NS_5 (~200 reps) at 27 °C and NS_1 (~60 reps) at 27 °C. Bur-0 & NS plants at 4-weeks growth. Bur-0 at 27 °C displays a strong *iil* growth defect that is absent in 23 °C. NS_5 plants typically display a lower frequency and milder *iil* phenotype compared to Bur-0. NS_1 plants typically do not display *iil* growth defects at 27 °C. **(C)** *IIL1* expression levels are similar between 23 °C & 27 °C for Col-0 control accession without expansion (23 repeats). A direct anti-correlation is observed between *IIL1* expression and TTC/GAA expansion length at 27 °C. * (asterisk) indicates significant decrease in *IIL1* expression in 27 °C Bur-0 plants compared to 23 °C Bur-0 plants as determined by students *t*-test, with 27 °C Bur-0 plants *IIL1* expression ~34% of 23 °C Bur-0 plant *IIL1*. All qPCR data used *TUB2* expression as a reference gene with results presented as relative expression to Col-0 at 23 °C. Students *t*-test was used to calculate significance as the difference in relative expression (*p*-value cut off <0.05) Error bars indicate standard error. Expression results for 23 °C & 27 °C treatments were derived from 8 biological replicates and 4 technical replicates for each accession.

The stong correlation between repeat tract length and *iil* phenotype (Fig 9B), was accompanied with a trend of decreased *IIL1* expression at 27 °C with increasing repeat tract length (Fig 9C). Standard deviations for all expression data were calculated from the average $^{\Delta\Delta}$Ct of technical replicates between biological replicates.

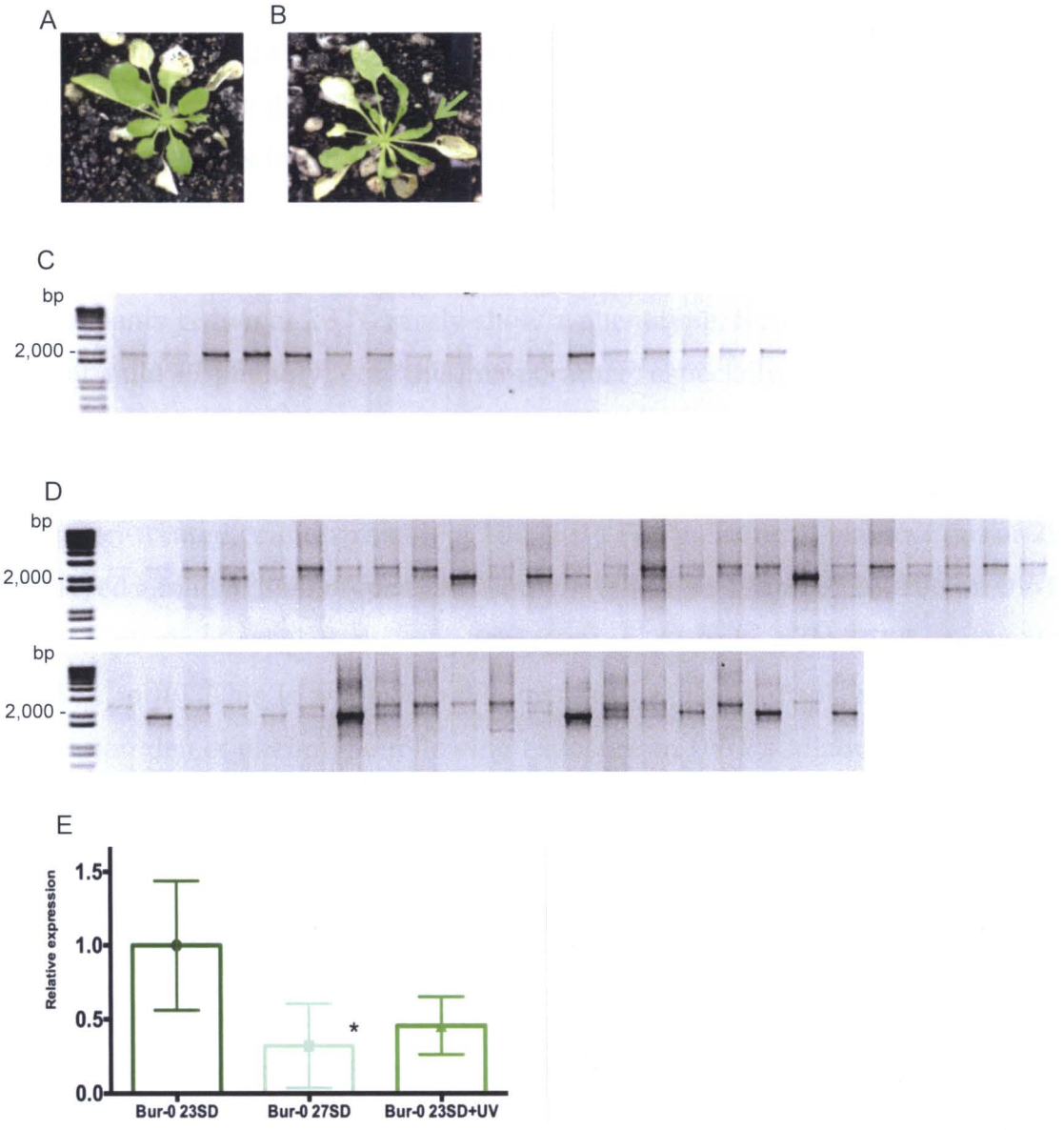



**Figure 10. Impacts of UV-B light on *A. thaliana* TNR expansion defect. (A)** 4-week old Bur-0 23 °C SD control plant displaying normal phenotype. **(B)** 4-week old Bur-0 23 °C SD + UV-B treated plant displaying *iil* phenotype. Green arrow points out emergence of narrowed tapered leaves characteristic of the *iil* phenotypic defect. **(C)** *IIL1* TTC/GAA repeat-spanning PCR on 4-week old Bur-0 23 °C SD control plants displaying presence of TTC/GAA expansion. **(D)** *IIL1* PCRs on 4-week old Bur-0 23 °C SD + UV-B treated plants displaying presence of extensive TTC/GAA expansion instability. **(E)** qPCR of *IIL1* expression data of Bur-0 23 °C SD, Bur-0 27 °C SD and Bur-0 23 °C SD + UV-B treatment samples. All genomic PCR and expression data were obtained from 4-week old plants. * (Asterisk) indicates significant decrease in *IIL1* expression in Bur-0 27 °C SD samples relative to Bur-0 23 °C SD ($p$-value <0.05) calculated using students $t$-test. Results displayed as relative expression to Bur-0 23 °C SD control. Error bars indicate standard error. Expression results for were derived from 11 biological replicates and 4 technical replicates in each category.

In addition to elevated temperature, UV stress also induced a mild *iil* phenotype in Bur-0 plants at 23 °C (Fig 10B), however this was not observed in either NS. The emergence of the *iil* phenotypic growth defects in UV-B exposed Bur-0 plants at 23 °C was found to be associated with massive instability of the repeat expansion (Fig 10D).

Bur-0 plants grown at 23 °C rarely show a phenotype. However some plants do exhibit mild *iil* phenotype at this temperature, especially if they are found to have larger expansions. An increase in the frequency of plants exhibiting *iil* growth defects was observed in UV-B treated Bur-0 plants at 23 °C relative to their non-treated counterparts (Fig 10C & D). For the control plants 2 out of 28 displayed a mild *iil* phenotype (>93% no *iil*). Where as 17 out of 26, 10 min UV-B treated plants displayed the *iil* phenotype at 2 weeks post UV-B treatment (~35% no *iil*). This is a considerable increase in plants displaying the *iil* TNR expansion defect phenotype following exposure to UV-B radiation (from 7% to 36%). Upon *IIL1* TTC/GAA tract PCR analysis of the 4-week old control (Fig 10C) and UV-B treated (Fig 10D) plants it was discovered that the genomic expansion displayed massive variability in UV-B treated plants (Fig 10D).

qPCR expression analysis also revealed a decrease in *IIL1* expression in the *iil*-phenotype-displaying UV-B treated plants (Fig 10E). Perhaps the *iil* phenotype is not directly due to UV. Rather, it may be an indirect affect due to the dramatic further increase in genomic expansion the UV induces.

Although the different stresses induced phenotypic responses from our plants, such as stunted growth and curled leaves (e.g. characteristic of NaCl stress) There were no conditions that caused differing phenotypes among the Bur-0 and NS plants, which would be suggestive of direct affects of the expansion, aside from elevated temperature and to an extent, UV exposure.

## 1.3 DISCUSSION

### 1.3.1 Expanded TTC/GAA allele at the *IIL1* locus maintained in Ireland at relatively high frequency

Of the ~530 samples collected we were able to recover large expansions, above the threshold length previously determined to be around 126-150 repeats (Chapter 3) in 8 populations. This gave a frequency of ~6.2% although collection was biased toward the Burrren area where Bur-0 was originally collected with at least 3 of the populations with expansions collected from that region. The expanded allele appears to be maintained in the wild some 55 y after it was originally collected in 1958. It is currently unknown why these repeat expansions occur at a higher frequency than otherwise expected for a seemingly deleterious mutation. However, the climatic conditions in Ireland are such that the cryptic deleterious phenotype would likely never appear and thus it is possible the mutation exists as a neutral allele. It is also possible that the repeat expansion may be conferring a yet unknown advantage to the Irish plants.

### 1.3.2 The *IIL1* TTC/GAA repeat expansion displays conditional neutrality

We have shown that the *IIL1* triplet repeat expansions are recoverable even after 50 years of original collection and occur at a moderate frequency in the Irish populations of *A. thaliana.* This suggests that this repeat expansion may possibly be either neutral or beneficial in the Irish populations. The phenotypic consequences of repeat expansion is possibly masked in the temperate Irish environment in which temperatures of 27 °C do not occur at times of the year when days are still short. However, it can be considered that it would still be disadvantageous to the fitness as loss of function of *IIL1* in the Col-0 background displays slow growth even under normal conditions (Knill, Reichelt et al. 2009, Sureshkumar, Todesco et al. 2009). The rarity of the *IIL1* TTC/GAA expansion in

the worldwide populations would support this. The fitness reduction would be even more marked in *A. thaliana* accessions from, for example, Mediterranean climates, where short days in spring and autumn may commonly be warm enough to reveal (unmask) the *iil* phenotype to natural selection. Alternatively, it is also possible that the repeat expansion may confer a yet unknown advantage to wild plants growing in the climatic conditions of the British Isles. For example, natural suppressors with reduced repeats in the Bur-0 background tend to flower early compared to Bur-0 (Appendix - Table S1).

It is not yet clear what effect early flowering would have in the environment of the Burren, where the repeat expansion appears to be most prevalent, but it could conceivably expose flowering plants to the risk of late frosts. In such a scenario, the seemingly detrimental allele may provide an advantage in selective environments in the form of antagonistic pleiotropy. There are currently very few examples of such conditional deleterious mutations known to display antagonistic pleiotropy. Recently, our lab has described loss-of-function alleles for the *ICARUS1* gene, which encodes a protein that is required for plant growth at higher temperatures, to be occurring at high frequencies in local populations, while being relatively rare in global populations, similar to the *IIL1* situation we describe here (Zhu et al, in press). The *IIL1* TTC/GAA polymorphism in *A. thaliana* thus provides an ideal model for analyzing the underlying mechanisms for such mutations and their role in environmental selection and evolutionary adaptation of natural species.

### 1.3.3 The occurrences of the *IIL1* repeat expansion outside Ireland.

Consistent with earlier findings, the *IIL1* repeat expansion appears to be infrequent even among the Irish accessions. Interestingly, we have also observed the repeat in expansion in a British accession outside Ireland (At434 from Callander, Scotland). We have also screened a closely related accession to Bur-0; Cal-0 collected from Calver Fell, UK and found that it also contained the

TTC/GAA expansion. However, there are some previously published results that raise doubts about the Cal-0 accession. In 2002, Cal-0 accession has been described as having a different leaf architecture when compared with Bur-0 (Perez-Perez, Serrano-Cartagena et al. 2002). However, publications after 2002 describe Bur-0 and Cal-0 to be identical for a variety of phenotypes (Lempe, Balasubramanian et al. 2005, Katori, Ikeda et al. 2010, Nam, Koh et al. 2011). In addition, unlike any other accession, Cal-0 appears to be unusually close to Bur-0 as shown through the 1001 genome project although Cal-0 and Bur-0 have not been flagged as identical (http://1001genomes.org) (Fig 6). Nevertheless, as we have found a strain from Scotland containing the expansion, it suggests that the repeat expansion is present mostly, but not exclusively in Ireland.

### 1.3.4 Variability of the non-expanded *IIL1* triplet repeats in Irish populations

Variability in TTC/GAA repeat length, within the normal range, was observed in this study. Excluding the dramatically expanded samples, TTC/GAA tracts in the Irish population spanned 1-28 repeats (Fig 7A). This range is only slightly narrower than what we have previously reported (0-36 repeats) across a global population set (Sureshkumar, Todesco et al. 2009). In addition, the Irish population exhibits a bimodal distribution in repeat lengths with peak frequencies of 9 and 26 repeats. This could indicate that the *IIL1* repeat in the Irish populations is comprised of two conserved genetic groups. Further investigation into the genotypic information of each accession is required to reveal the true extent of genetic diversity amongst the Irish *A. thaliana* (this is explored in Chapter 2). We determined that certain repeat lengths were favored over others. For example, of the 150 sequenced samples, 28 populations had nine repeats, 14 had eleven repeats and only 2 had ten repeats. It has been hypothesized that the number of repeats may reflect adaptations to the internal genetic environments (Undurraga, Press et al. 2012). While Undurraga *et* al's study on the *ELF3* gene (Undurraga, Press et al. 2012) demonstrated phenotypic impacts of polyQ variation in plants, our results indicate the repeat lengths of

noncoding trinucleotide repeats may also be confined in a background-dependent manner.

It is often suggested that there is a correlation between repeat length and the expression levels with the larger expansion resulting in severe reductions in expression levels (Ashley and Warren 1995, Kovtun and McMurray 2008). In our samples, we found this correlation is not perfect with several wild accessions with shorter TTC/GAA expansions e.g. At249 of repeat length ~260 expressing less *IIL1* compared to wild accessions with slightly larger repeat tracts e.g. At533 of >300 reps (results not shown) upon qPCR analysis. However, overall there does seem to be some correlation between repeat length and *iil* severity with wild accessions At434 (~200 reps) expressing more *IIL1* compared to At370 (~250 reps) and At460-462 (~350 reps) (Fig 5D) and wild accessions At460-462 displaying a more severe *iil* phenotype than At370 and At434 (Fig 5C). Therefore, there does seem to be a correlation between *iil* severity, *IIL1* expression & TTC/GAA expansion length.

It was also noticed that a lot of the wild accessions with dramatic TTC/GAA expansions exhibited mild *iil* phenotypes in 23 °C conditions as well as 27 °C, where as the lab strain Bur-0 typically only displays the *iil* phenotype in more than 23 °C growth conditions (Fig 8). Perhaps the collected wild accessions are more sensitive to 23 °C temperatures, as 23 °C itself could be a much higher temperature than they are used to in Ireland.

Overall, we have shown that even mutations with very strong phenotypes can be maintained in nature if their effects are conditional. The triplet repeat expansion, which causes severe growth impairments at high ambient temperatures, is maintained in the Irish population for at least the past 56 years. Our findings show that the repeat genotype is actually segregating at an appreciable frequency in the Irish population, that is, a frequency too high for the allele to be strongly deleterious in the Irish population. Though it is difficult to rule out balancing selection or fluctuating selection, data suggests that within Ireland, the allele might be neutral or beneficial. These Irish strains of *A. thaliana* along with

the natural suppressors provide an excellent resource to further analyze mechanisms that could explain the high frequency of the repeat expansion in the Irish populations, and indeed investigating this avenue is a future direction for this work.

### 1.3.5 *A. thaliana IIL1* TNR expansion defect is unmasked in elevated temperature and UV-B exposure

As a phenotype from the *IIL1* TNR expansion only emerged in response to elevated temperature, and to an extent UV-B exposure, we can conclude that the affects of the expansion are largely temperature specific and not a general stress response. Bur-0 plants at 23 °C, express *IIL1* to a level similar to that of other *A. thaliana* accessions such as Col-0 (Fig 9C). However, at 27 °C the Bur-0 plants express less *IIL1* relative to 23 °C. Bur-0 as well as other accessions at 27 °C including Col-0 (Fig 9C). This information gives us further insight into the nature of the molecular affects of the expansion; in the plants the level of gene expression is specifically due to temperature and to an extent UV.

The increase in UV-B treated Bur-0 plants at 23 °C displaying the *iil* defect may potentially be attributed to the increased TTC/GAA expansion instability favouring expansion (Fig 10D). However, whether the UV leads to an increase in *iil* severity indirectly through increasing the expansion tract or through a direct mechanism is difficult to say and further analysis into the affect of UV-B in Bur-0 is a future direction.

### 1.3.6 UV-B radiation stress dramatically increases TTC/GAA expansion instability.

UV-B light stress was found to induce the *iil* phenotype in Bur-0 plants. Upon repeat TTC/GAA PCR analysis, massive instability was observed (Fig 10D).

Potentially, rather than directly inducing the *iil* phenotype it is possible that the UV-B light stress induced DNA damage to the plants, which resulted in massive repeat expansion instability and subsequent to the repeat instability, the emergence of the *iil* phenotype.

There is mounting evidence in support of a contributing role of aberrant DNA repair and TNR expansion instability in TNRDs. This is explored in the Bur-0 model in Chapter 3. One hypothesis is: Exposure to UV light damages the plants DNA which in turn activates DNA repair pathways. Aberrant repair of the expansion region, due to its repetitive nature and secondary structure formation, leads to massive variability in repeat lengths, with both large expansions and contractions occurring (Fig 10D). The dramatic expansions are sufficient to induce the *iil* phenotype even at 21-23 °C. Individual 23 °C UV-treated plants displaying the *iil* phenotype were all found to have large instability favouring expansion over contraction in their TTC/GAA repeat at the *IIL1* locus.

## 1.4 MATERIALS & METHODS

### 1.4.1 Plant material

Plants were grown either in long days (LD, 16 h light/8 h dark) or short days (SD, 8 h light/16 h dark) at 23 °C or 27 °C in growth rooms or chambers (Percival Scientific) or in a greenhouse under long day conditions, as indicated. Maximum humidity was 65%. Seeds were sterilized in 1.5 ml tubes in 70% ethanol with 0.1% Triton X-100 for 3-5 min and washed in 95% ethanol for 1-2 min. Seeds were then either sown in soil or spotted onto 0.5 x MS media (pH 5.7), made from 2.17 g Murashige and Skoog salts (Sigma) and 7.0 g agar $L^{-1}$. Seeds from one representative individual from each distinct location were grown in 23 °C LD conditions to bulk seed stocks and ensure a homozygous phenotype in each line.

### 1.4.2 DNA and RNA extraction

DNA and RNA from wild Irish accessions and commercial *A. thaliana* strains were extracted from fresh rosette leaf tissue from 1 or 4 week-old plants (as indicated) after flash freezing in liquid nitrogen. For DNA extraction, 900 µl of warmed cetyl trimethyl ammonium bromide (CTAB) buffer (20 g CTAB and 81.83 g NaCl, 40 ml EDTA (0.5 M pH 8.0), 100 ml of Tris-HCl (1 M pH 8.0) in a 1 L solution) was added to the samples. The samples were vortexed and incubated at 65 °C for 40 min. 600 µl of Chloroform:phenol (24:1 v:v) was added to the tubes, which were inverted several times then centrifuged (14,000 rpm, RT, 20 min). The supernatant was transferred to a new microfuge tube containing 300 µl of chilled isopropanol and left to incubate at 4 °C for 15 min. The samples were then centrifuged again (14,000 rpm, RT, 10 min) to pellet the DNA. The supernatant was discarded and the pellet washed with 500 µl of wash buffer (76% v/v EtOH and 10 mM ammonium acetate). Washing was done by inverting the samples several times, centrifuging them (14,000 rpm, RT, 10 min) and discarding the supernatant. The pelleted DNA was left to air dry for 5-10 min prior to adding 60 µl of TE buffer (pH 8.0) with RNAse (1 µl/100 µl TE buffer) and incubating at 37 °C for 1 h to ensure all DNA had re-dissolved. DNA was stored at -20 °C.

For RNA extraction, the finely ground leaf tissue was incubated with 1 ml TRIzol reagent (Invitrogen) for 5 min, RT. 0.2 ml chloroform was added per 1 ml TRIzol and samples shaken vigorously by hand for 15 sec. Samples were then incubated for 10-15 min, RT and centrifuged (12,000 g, 10 min, 4 °C). Upper aqueous phase was transferred to a new microfuge tube containing 0.5 ml isopropanol to precipitate RNA. Samples were inverted and left to incubate for 10 min, RT then centrifuged (12,000 g, 10 min, 4°C). Supernatant was discarded and the pellet washed with 1 ml 75% ethanol, vortexed, and again centrifuged (9,500 g, 5 min, 4 °C). RNA pellets were left to dry for 5-10 min, RT then dissolved in 30-50 µl RNase free sterile Milli Q $H_2O$. [RNA] quality and quantity determined with NanoDrop (Thermo Scientific).

### 1.4.3 Polymerase chain reaction (PCR)

The concentration of all gDNA was determined using UV spectrophotometry (NanoDrop; Thermo Scientific). Polymerase chain reaction (PCR) was performed on the region containing the *IIL1* TTC/GAA repeat with MyTaq Red mix or ExTaq (Takara) used according to manufacturer's instructions.

*IIL1* locus

For PCR amplification of the *IIL1* locus containing the TTC/GAA repeat: 200 ng of gDNA was combined with 0.05 µl Extaq polymerase (Takara Bio), 1 µL of 10X Extaq extra buffer, 0.8 µL of 2.5 mM dNTPs, 6.65 µL of sterile milliQ $H_2O$ (Merck Millipore) and 0.5 µL of each *IIL1* primer (SKB_608 forward primer: 5'-GCCTGCAATGACCTTCTTGT -3' and SKB_561 reverse primer: 5'-GGGATGACAATGACGGAGAA -3) (Appendix - Table S2) were used.

*IIL1* PCR cycling:

Amplification for the *IIL1* loci: Initial denaturation 95 °C for 3 min, followed by 35 cycles of 95 °C for 10 sec, 72 °C for 30 sec and 72 °C for 2.5 min, followed by a final extension 72 °C for 10 min. PCR products were run on 1% agarose gels + EtBr in TAE buffer for 30 min at 100 V and visualized using a Carestream Molecular Imaging (Carestream MI) Software, version 5.0.

### 1.4.4 Sequencing

For sequencing the microsatellite repeats, PCR products were excised from gel and purified using Wizard SV Gel and PCR Clean-up system (Promega). PCR products were sequenced (Micromon, Monash University) and the sequences were aligned and analysed through Seqman (DNAStar-Lasergene). The repeat number was quantified for each accession (Appendix - Table S1).

### 1.4.5 cDNA synthesis: Reverse transcription

cDNA was synthesized using the Invitrogen SuperScript® III First Strand Synthesis System. Briefly, 10 µl of samples containing 1,000 µl total RNA were incubated for 30 min, 37 °C with 1 µl DNase and 1 µl 10X DNase buffer. 1 µl of 25 mM EDTA was added and samples incubated for 10 min, 65 °C. 1 ul Oligo (dT) 18 and 1 µl sterile milliQ $H_20$ were added and incubated for a further 10 min, 65 °C to prime samples. Samples were then immediately cooled on ice and cDNA synthesis set up containing: 1X cDNA synthesis buffer, 20 U inhibitor, 1 mM deoxynucleotide, 10 U enzyme and 1000 ng primed RNA. Reactions were run 60 min, 50 °C, then 5 min 85 °C.

### 1.4.6 Quantitative real-time PCR

RNA from 1 or 4-week old rosette leaves (as indicated) was extracted using Trizol, and 1 µg of total RNA was reverse transcribed using a Reverse Transcription kit (Invitrogen) as per manufacturer's instructions. Quantitative RT-PCR was performed using the SYBR Green system (Roche) on a Lightcycler 480 (Roche Diagnostics). qPCR reactions were set up as per manufacturer's instructions. Briefly, 2 µl cDNA (diluted 1:4), 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 7 µl sterile PCR grade nuclease-free $H_2O$, 10 µl Lightcycler 480 SYBR Green-1 Master (Roche). All samples were run in Biological triplicate unless otherwise stated. Expression data for *IIL1* gene were normalized to *TUB2* reference gene expression.

Reactions were run for 40 cycles and data analysis carried out using standard ΔΔCt methodology. Briefly, threshold cycles (cT) were based on a reaction reaching specific fluorescence intensity in the log-linear phase of the amplification curve. ΔcT was calculated as the difference in cT between *IIL1* and the reference gene *TUB2*. PCR efficiency was assumed to be the same and relative transcript abundance compared to 23 °C Col-0 was calculated as $2^{-\Delta\Delta cT}$.

The standard temperature profiles included an initial denaturation for 5 min at 95 °C, 40 cycles of amplification at 95 °C for 10 s, 72 °C for 15 s and 72 °C for 10 s, melt curve analysis for 5 s at 95 °C and 1 min at 65 °C, and a cooling phase for 30 s at 30 °C. Additionally, qPCR products were run on agarose gels to verify single products. (See Appendix - Table S2 for list of primers).

### 1.4.7 Statistical analyses

Gene expression data are presented as mean ± SEM of triplicate assays in at least three independent experiments, and the significance of the difference between groups evaluated with the Student's *t*-test (two-tailed) or Pearson's correlation coefficient. *p*-value < 0.05 was considered significant. Graphs were constructed using Prism 6 programme.

### 1.4.8 Abiotic stress

Bur-0 plants and 2 Bur-0 plants that were natural suppressors in the repeat tract, NS_1 (~60 repeats) & NS_5 (~200 repeats) were grown in 12 different mild stress environments. Abiotic stress conditions were designed to impose a stress on the plants to an extent that they may encounter in nature. I.e. the conditions were designed to put a strain on plant function but not great enough to kill the plants.

Stresses:

1. Drought (dehydration)
2. Rehydration
3. Salinity [NaCl] (150 mM)
4. Osmotic stress

5. Osmotic stress/mannitol (300 mM)
6. Abscisic acid (ABA) (Hormone treatment)
7. Oxidative stress (Methyl Viologen) (10 μM & 0.3 μM)
8. Wounding stress
9. UV-B light stress
10. Nutrient deprivation (Sucrose deficient MS)
11. pH variation (100 mM of NaHCO3 on MS plates)
12. High temperature (27 °C)

*A. thaliana* plants, Bur-0 strain, were germinated and grown in 15 cm pots filled with a 1:1 (v/v) mixture of perlite and vermiculite and watered every 3 d with 0.53 Hoagland solution under short day conditions (SD) (8 h light, 16 h dark regime) at approximately 2500 lux at 22.C for 2 weeks and 60% humidity. 50 Bur-0 seeds were used for each stress and 50 seeds for the control for each stress. All experiments were conducted on 2-week-old soil-grown plants and/or on MS agar plates as indicated.

For MS plate treatment, Seeds were sterilized in 1.5-ml tubes, in 70% EtOH with 0.1% Triton X-100 for 3-5 min and washed in 95% EtOH for 1-2 min. Twenty seeds for each accession were spotted onto 0.5 MS media (pH 5.7), made of 2.17g Murashige and Skoog salts (Sigma) and 7.0g Agar for 1 litre. The various stress-inducing compounds were added to the MS agar prior to making the petri dishes. Plates were placed in light and humidity controlled Percival chambers (Percival Inc, Canada).

#### *1.4.8.1 Drought (soil water deficit)*

Watering was stopped from 2 weeks after germination until the soil was dry, with relative water content of 5%, which typically took 5 d (Huang, Wu et al. 2008).

#### *1.4.8.2 Rehydration*

After this dehydration treatment, some plants were re-watered (Huang, Wu et al. 2008).

#### *1.4.8.3 Salinity (NaCl)*

For salinity stress, the soil was saturated with 150 mM NaCl solution, and about 100 mL of the solution was poured on the soil surface of each plant and allowed to drain. Salinity stress was also applied to MS plates by incorporating 150 mM NaCL to MS plates prior to planting. The effects of salt stress on iil phenotype were determined based on the literature (Galpaz and Reymond 2010). (Yokoi, Quintero et al. 2002), (Yamaguchi, Takahashi et al. 2006) (Kim, Bamba et al. 2007) (Denekamp and Smeekens 2003) (Nakashima, Kiyosue et al. 1997).

#### *1.4.8.4 Osmotic stress*

Increased concentrations of a solute around the plant cells causes water to leave the cells through osmosis. This also inhibits the transport of substrates and cofactors into the cell, causing an abiotic stress. Mannitol is typically used to induce osmotic stress in plants (Kreps, Wu et al. 2002).

For osmotic stress, the soil was saturated with 400 mM exogenous mannitol solution, and about 50 mL of the solution was poured on the soil surface of each plant and allowed to drain (Xiong and Zhu 2002, Kilian, Whitehead et al. 2007, Skirycz, Claeys et al. 2011). 300 mM mannitol was also added to MS plates for plate screening.

#### *1.4.8.5 Abscisic acid (ABA) stress*

The phytohormone abscisic acid (ABA) is produced under water deficit conditions and plays an important role in the stress response and tolerance of plants to drought and high salinity. When roots begin to sense a water shortage in their soil, ABA is released as an early response, to trigger closure of stomatal pores.

ABA binds to receptor proteins in the plasma membrane and cytosol of guard cells, which cause the concentration of free $Ca^{2+}$ to increase in the cytosol due to influx from outside the cell and release of $Ca^{2+}$ from internal stores such as the endoplasmic reticulum and vacuoles. This causes the subsequent activation of plasma membrane-localized anion channels stopping the uptake of any further $K^+$ into the cells causing a loss of $K^+$. The loss of these solutes causes an increase in water potential, which results in water diffusing back out of the cell by osmosis. This makes the cell flaccid, which results in the closing of the stomatal pores (Hamilton, Hills et al. 2000, Kohler, Hills et al. 2003). ABA also causes an increase in $H_2O_2$ production, which serves as a signaling intermediate to promote stomatal closure (Yamaguchi, Takahashi et al. 2006). Exogenous application of ABA induces a number of genes that respond to dehydration and cold stress (Zhu 2001, Shinozaki, Yamaguchi-Shinozaki et al. 2003).

Mixed isomers of ABA (Sigma) were first dissolved in a small volume of 0.1 M KOH and made up to 1 mM with distilled water. *A. thaliana* plants were treated with ABA when they were 2 weeks old. The ABA solution with 0.02% Tween-20 was sprayed to cover the foliage completely three times a day for 3 consecutive days (Xing, 2001). MS plates containing either 1 μM or 5 μM ABA were also used.

#### *1.4.8.6 Oxidative stress*

MS plates containing 10 μM and 0.3 μM Methyl viologen were made. Methyl viologen is a reducing compound with a strong redox potential that brings about oxidation. It is commonly used in higher doses as an herbicide.

#### *1.4.8.7 Wounding stress*

Wounding of the plants was performed by incision of rosette leaves using scalpel or forceps. Wounding was applied through cutting of leaf tips (resulting in ~20% of the leaf being damaged). DNA for PCR analysis was collected from leaf tissue around the wounding site (Denekamp and Smeekens 2003). Mechanical wounding was performed 4–6 h after dawn (Walley, Coughlan et al. 2007).

#### *1.4.8.8 UV-B light stress*

The lids of the growth boxes were removed and the boxes covered with a 3 mm quartz plate. The plants were irradiated for 5-10 min with UV-B light with a biologically effective quantity of 1.18 W $m^{-2}$. The UV-B light source consisted of six Philips TL40W/12 fluorescent tubes (kmax of310 nm, half- bandwidth of 40 nm. Afterwards the lids were closed and the boxes transferred back to the standard phytochamber percival until harvest. Under these conditions both the damaging short-wavelength and the photomorphogenic long-wavelength UV-B response pathways are induced in *A. thaliana* (Kilian, Whitehead et al. 2007). Pre-existing leaves were marked with a back sharpie to distinguish leaves that grew before and after UV treatment. 2 weeks post UV-B treatment, phenotypic data was recorded for the + & - UV exposed plants. 180 2-week old Bur-0 plants were exposed to UV-B radiation.

#### *1.4.8.9 Nutrient deprivation*

For nutrient deprivation stress, MS plates were made containing either no sucrose or ½ the amount of MS salts.

#### *1.4.8.10 pH variation*

MS plates typically pH 5.7, were altered with $NaHCO_3$ to make the agar pH 7. This slightly alkaline environment is a stress to optimal *A. thaliana* growth and mimics the pH of soil in Burren, Ireland.

#### *1.4.8.11 Elevated temperature*

For elevated temperature stress plants either on MS plates or in soil, were grown in 27 °C short day controlled growth chambers (Percival). Short day (SD) conditions consisted of 8 h of light followed by 16 h of darkness. Relative humidity was maintained at 65%. The soil-grown plants were watered periodically.

## 1.5 REFERENCES


Ahmed, M. and P. Liang (2012). "Transposable elements are a significant contributor to tandem repeats in the human genome." Comp Funct Genomics **2012**: 947089.

Ashizawa, T., J. R. Dubel and Y. Harati (1993). "Somatic instability of CTG repeat in myotonic dystrophy." Neurology **43**(12): 2674-2678.

Ashley, C. T., Jr. and S. T. Warren (1995). "Trinucleotide repeat expansion and human disease." Annu Rev Genet **29**: 703-728.

Balasubramanian, S., S. Sureshkumar, J. Lempe and D. Weigel (2006). "Potent induction of Arabidopsis thaliana flowering by elevated growth temperature." PLoS Genet **2**(7): e106.

Carpenter, N. J. (1994). "Genetic anticipation. Expanding tandem repeats." Neurol Clin **12**(4): 683-697.

Denekamp, M. and S. C. Smeekens (2003). "Integration of wounding and osmotic stress signals determines the expression of the AtMYB102 transcription factor gene." Plant Physiol **132**(3): 1415-1423.

Fondon, J. W., 3rd and H. R. Garner (2004). "Molecular origins of rapid and continuous morphological evolution." Proc Natl Acad Sci U S A **101**(52): 18058-18063.

Frieman, M. B. and B. P. Cormack (2004). "Multiple sequence signals determine the distribution of glycosylphosphatidylinositol proteins between the plasma membrane and cell wall in Saccharomyces cerevisiae." Microbiology **150**(Pt 10): 3105-3114.

Galpaz, N. and M. Reymond (2010). "Natural variation in Arabidopsis thaliana revealed a genetic network controlling germination under salt stress." PLoS One **5**(12): e15198.

Gong, L., S. Parikh, P. J. Rosenthal and B. Greenhouse (2013). "Biochemical and immunological mechanisms by which sickle cell trait protects against malaria." Malar J **12**: 317.

Grosse, S. D., I. Odame, H. K. Atrash, D. D. Amendah, F. B. Piel and T. N. Williams (2011). "Sickle Cell Disease in Africa: A Neglected Cause of Early Childhood Mortality." Am J Prev Med **41**(6): S398-405.

Hamilton, D. W., A. Hills, B. Kohler and M. R. Blatt (2000). "Ca2+ channels at the plasma membrane of stomatal guard cells are activated by hyperpolarization and abscisic acid." Proc Natl Acad Sci U S A **97**(9): 4967-4972.

Hammock, E. A., M. M. Lim, H. P. Nair and L. J. Young (2005). "Association of vasopressin 1a receptor levels with a regulatory microsatellite and behavior." Genes Brain Behav **4**(5): 289-301.

Hegde, M. V. and A. A. Saraph (2011). "Unstable genes unstable mind: beyond the central dogma of molecular biology." Med Hypotheses **77**(2): 165-170.

Hennessy, R., McNamara, M., & Hoctor, Z. (2010). Stone, Water & Ice: A Geology Trip Through the Burren. Burren Connect Project, pp. 64.

Huang, D., W. Wu, S. R. Abrams and A. J. Cutler (2008). "The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors." J Exp Bot **59**(11): 2991-3007.

Katori, T., A. Ikeda, S. Iuchi, M. Kobayashi, K. Shinozaki, K. Maehashi, Y. Sakata, S. Tanaka and T. Taji (2010). "Dissecting the genetic control of natural variation in salt tolerance of Arabidopsis thaliana accessions." J Exp Bot **61**(4): 1125-1138.

Kayser, M., L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Kruger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff, M. Stoneking and A. Sajantila (2000). "Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs." Am J Hum Genet **66**(5): 1580-1588.

Kennedy, L., E. Evans, C. M. Chen, L. Craven, P. J. Detloff, M. Ennis and P. F. Shelbourne (2003). "Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis." Hum Mol Genet **12**(24): 3359-3367.

Kilian, J., D. Whitehead, J. Horak, D. Wanke, S. Weinl, O. Batistic, C. D'Angelo, E. Bornberg-Bauer, J. Kudla and K. Harter (2007). "The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses." Plant J **50**(2): 347-363.

Kim, J. K., T. Bamba, K. Harada, E. Fukusaki and A. Kobayashi (2007). "Time-course metabolic profiling in Arabidopsis thaliana cell cultures after salt stress treatment." J Exp Bot **58**(3): 415-424.

Knill, T., M. Reichelt, C. Paetz, J. Gershenzon and S. Binder (2009). "Arabidopsis thaliana encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation." Plant Mol Biol **71**(3): 227-239.

Kohler, B., A. Hills and M. R. Blatt (2003). "Control of guard cell ion channels by hydrogen peroxide and abscisic acid indicates their action through alternate signaling pathways." Plant Physiol **131**(2): 385-388.

Kovtun, I. V. and C. T. McMurray (2008). "Features of trinucleotide repeat instability in vivo." Cell Res **18**(1): 198-213.

Kreps, J. A., Y. Wu, H. S. Chang, T. Zhu, X. Wang and J. F. Harper (2002). "Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress." Plant Physiol **130**(4): 2129-2141.

Lee, J., J. Nam, H. C. Park, G. Na, K. Miura, J. B. Jin, C. Y. Yoo, D. Baek, D. H. Kim, J. C. Jeong, D. Kim, S. Y. Lee, D. E. Salt, T. Mengiste, Q. Gong, S. Ma, H. J. Bohnert, S. S. Kwak, R. A. Bressan, P. M. Hasegawa and D. J. Yun (2007). "Salicylic acid-mediated innate immunity in Arabidopsis is regulated by SIZ1 SUMO E3 ligase." Plant J **49**(1): 79-90.

Lempe, J., S. Balasubramanian, S. Sureshkumar, A. Singh, M. Schmid and D. Weigel (2005). "Diversity of flowering responses in wild Arabidopsis thaliana strains." PLoS Genet **1**(1): 109-118.

Levdansky, E., J. Romano, Y. Shadkchan, H. Sharon, K. J. Verstrepen, G. R. Fink and N. Osherov (2007). "Coding tandem repeats generate diversity in Aspergillus fumigatus genes." Eukaryot Cell **6**(8): 1380-1391.

Ma, L., J. S. Jensen, M. Mancuso, R. Hamasuna, Q. Jia, C. L. McGowin and D. H. Martin (2012). "Variability of trinucleotide tandem repeats in the MgPa operon and its repetitive chromosomal elements in Mycoplasma genitalium." J Med Microbiol **61**(Pt 2): 191-197.

Mangiarini, L., K. Sathasivam, A. Mahal, R. Mott, M. Seller and G. P. Bates (1997). "Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation." Nat Genet **15**(2): 197-200.

Mirkin, S. M. (2007). "Expandable DNA repeats and human disease." Nature **447**(7147): 932-940.

Morales, F., J. M. Couto, C. F. Higham, G. Hogg, P. Cuenca, C. Braida, R. H. Wilson, B. Adam, G. del Valle, R. Brian, M. Sittenfeld, T. Ashizawa, A. Wilcox, D. E. Wilcox and D. G. Monckton (2012). "Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity." Hum Mol Genet **21**(16): 3558-3567.

Moxon, E. R., P. B. Rainey, M. A. Nowak and R. E. Lenski (1994). "Adaptive evolution of highly mutable loci in pathogenic bacteria." Curr Biol **4**(1): 24-33.

Nakashima, K., T. Kiyosue, K. Yamaguchi-Shinozaki and K. Shinozaki (1997). "A nuclear gene, erd1, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally up-regulated during senescence in Arabidopsis thaliana." Plant J **12**(4): 851-861.

Nam, M., S. Koh, S. U. Kim, L. L. Domier, J. H. Jeon, H. G. Kim, S. H. Lee, A. F. Bent and J. S. Moon (2011). "Arabidopsis TTR1 causes LRR-dependent lethal systemic necrosis, rather than systemic acquired resistance, to Tobacco ringspot virus." Mol Cells **32**(5): 421-429.

Perez-Perez, J. M., J. Serrano-Cartagena and J. L. Micol (2002). "Genetic analysis of natural variations in the architecture of Arabidopsis thaliana vegetative leaves." Genetics **162**(2): 893-915.

Scannell, D. A. W. M. J. P. (1983). Flora of Connemara and The Burren, Royal Dublin Society & Cambridge University Press.

Shinozaki, K., K. Yamaguchi-Shinozaki and M. Seki (2003). "Regulatory network of gene expression in the drought and cold stress responses." Curr Opin Plant Biol **6**(5): 410-417.

Skirycz, A., H. Claeys, S. De Bodt, A. Oikawa, S. Shinoda, M. Andriankaja, K. Maleux, N. B. Eloy, F. Coppens, S. D. Yoo, K. Saito and D. Inze (2011). "Pause-and-Stop: The Effects of Osmotic Stress on Cell Proliferation during Early Leaf Development in Arabidopsis and a Role for Ethylene Signaling in Cell Cycle Arrest." Plant Cell **23**(5): 1876-1888.

Stern, D. L. and V. Orgogozo (2008). "The loci of evolution: how predictable is genetic evolution?" Evolution **62**(9): 2155-2177.

Sureshkumar, S., M. Todesco, K. Schneeberger, R. Harilal, S. Balasubramanian and D. Weigel (2009). "A genetic defect caused by a triplet repeat expansion in Arabidopsis thaliana." Science **323**(5917): 1060-1063.

Undurraga, S. F., M. O. Press, M. Legendre, N. Bujdoso, J. Bale, H. Wang, S. J. Davis, K. J. Verstrepen and C. Queitsch (2012). "Background-dependent effects of polyglutamine variation in the Arabidopsis thaliana gene ELF3." Proc Natl Acad Sci U S A **109**(47): 19363-19367.

Walley, J. W., S. Coughlan, M. E. Hudson, M. F. Covington, R. Kaspi, G. Banu, S. L. Harmer and K. Dehesh (2007). "Mechanical stress induces biotic and abiotic stress responses via a novel cis-element." PLoS Genet **3**(10): 1800-1812.

Webb, D. A. (1961). Noteworthy Plants of the Burren. G. Section B: Biological, and Chemical Science. Proceedings of the Royal Irish Academy.

Xiong, L. and J. K. Zhu (2002). "Molecular and genetic aspects of plant responses to osmotic stress." Plant Cell Environ **25**(2): 131-139.

Yamaguchi, K., Y. Takahashi, T. Berberich, A. Imai, A. Miyazaki, T. Takahashi, A. Michael and T. Kusano (2006). "The polyamine spermine protects against high salt stress in Arabidopsis thaliana." FEBS Lett **580**(30): 6783-6788.

Yokoi, S., F. J. Quintero, B. Cubero, M. T. Ruiz, R. A. Bressan, P. M. Hasegawa and J. M. Pardo (2002). "Differential expression and function of Arabidopsis thaliana NHX Na+/H+ antiporters in the salt stress response." Plant J **30**(5): 529-539.

Zhu, J. K. (2001). "Plant salt tolerance." Trends Plant Sci **6**(2): 66-71.

# CHAPTER 2. Population genetic analyses of wild Irish *A. thaliana* using DArT sequencing SNP data

## 2.1 INTRODUCTION

Regional populations of *A. thaliana* have been shown to be complementary to the global collection of accessions often in answering questions that are more specific to their regions and in analyzing forces that shape evolution at a regional/local level (Le Corre, Roux et al. 2002, Pico, Mendez-Vigo et al. 2008, Bomblies, Yant et al. 2010, Long, Rabanal et al. 2013). Therefore, we reasoned that the analysis of Irish populations would also provide an opportunity to address the population genetic signatures of an island population in a model plant, as a complement to such genetic studies performed in continental populations (Le Corre, Roux et al. 2002, Aranzana, Kim et al. 2005, Pico, Mendez-Vigo et al. 2008, Bomblies, Yant et al. 2010, Lewandowska-Sabat, Fjellheim et al. 2010, Yin, Kang et al. 2010, Long, Rabanal et al. 2013). We therefore widened our study of the TTC/GAA repeat variability at the *IIL1* locus, and investigated the broader genetic architecture of *A. thaliana* populations from the vicinity of the Burren, and around Ireland.

In this chapter we take advantage of the 131 individual *A. thaliana* accessions collected from different locations across the island of Ireland (Chapter 1 Fig 1) to help characterize and assess both the phenotypic and genotypic diversity through genome-wide genotyping with DArT-SEQ technology and extensive analysis of many phenotypes.

### 2.1.1 Genome-wide genotyping in *A. thaliana*

*A. thaliana* is a model genetic organism not only for use in molecular biology, but also for ecological and evolutionary genetics by revealing the geographical structure of genetic variation (Borevitz and Nordborg 2003). By identifying genetic differences in natural populations we can relate this to the functional significance of those traits with respect to the organism's adaptive evolution to its particular ecological niche, thus providing insights into such phenomenon as adaptation and speciation (Borevitz and Nordborg 2003). *A. thaliana* is ideal for understanding natural variation, owing to its highly 'selfing' nature, wide geographic distribution, small well-characterized genome and considerable sequence resources available. *A. thaliana* also shows strong evidence of local adaptation (Fournier-Level, Korte et al. 2011, Hancock, Brachi et al. 2011) making it an ideal model for population genetic studies to advance understanding of diversity and evolution- areas of high scientific relevance.

Population studies have been undertaken on wild *A. thaliana* from Sweden, Norway, Morocco, Spain's Iberia peninsula and China (He, Kang et al. 2007, Pico, Mendez-Vigo et al. 2008, Lewandowska-Sabat, Fjellheim et al. 2010, Long, Rabanal et al. 2013). These studies have revealed to some extent, the evolutionary trajectory of the species, genes selected for in the local environments, potential routes of introduction as well as when and from where the species was likely introduced (Aranzana, Kim et al. 2005, Pico, Mendez-Vigo et al. 2008, Yin, Kang et al. 2010). To date, no study on the genetics of natural variation has been undertaken on wild *A. thaliana* populations in Ireland. This study contributes to elucidating the ultimate goal of obtaining a comprehensive view of genetic variation in populations by analyzing variation in a single isolated region and possessing the potential of comparing the findings to other similar studies from different geographical regions.

Here, we take advantage of the 131 natural Irish populations collected as in Chapter 1, to undertake a phenotypic analysis to reveal natural variation in phenotypes in the context of the local environment. Thirteen demographic and phenotypic traits including: latitude and longitude, altitude, *FRIGIDA* deletion, annual rainfall, degree of leaf serration, *IIL1* TTC/GAA repeat number, root

length at 23 °C SD, hypocotyl elongation at 23 °C SD, days to flower (DTF) at 23 °C SD & 23 °C LD as well as germination at 23 °C SD & 23 °C LD were analysed in this study (Appendix - Table S1). The responses of each accession to the various phenotypes can be related to the local environment as a genotype by environment (GXE) interaction, to identify genetic variants underlying natural variation within the phenotypes.

**Aim:** To address the population genetic signatures of the island Ireland population in the model plant, *A. thaliana*, as a complement to similar genetic studies performed in continental populations.

## 2.2 RESULTS

### 2.2.1 Genetic Diversity in Irish *A. thaliana*

In order to determine the wild accession genetic signatures, we analysed the genetic diversity among Irish accessions. To investigate the genetic diversity, we performed low coverage whole genome DArT sequencing (DArT-SEQ) on 150 samples from 131 distinct populations across Ireland (Appleby, Edwards et al. 2009). The DArT-SEQ sequencing produced diversity data for 11,242 SNPs and 12,651 presence-absence variants across the *A. thaliana* genome (Raw data will be submitted to the short reads archive upon publication). Using a set of 5,369 SNP markers for which genotypic information is available for more than 90% of the accessions, we analysed the genetic relatedness across the Irish accessions.

| Sample size | Total SNPs | Seg Sites | π.Bp-1 | θ.Bp-1 | Haplotypes | TajimaD |
|---|---|---|---|---|---|---|
| 131 | 5,369 | 5,043 | 0.21105 | 0.20174 | 67 | 0.74865 |

**Table 1. Summary of genetic diversity in wild Irish *A. thaliana*.** Average pairwise divergence (π), segregating sites, and θ estimates (4Nμ) calculated from DArT SNP sequencing data to

represent genetic diversity within the samples collected. Without heterozygous sites being set to missing data, as described in the materials & methods, genotyping section, 0.99% of the SNP loci were polymorphic. SNP data had a mean number of alleles over loci of 2.77 and a mean heterozygosity (*HO*) over loci of 0.033. Pi ($\pi$) estimates were obtained as the average of the values for all pairwise comparisons of theta ($\theta$) = $4N\mu$ were for an autosomal gene of a diploid organism (N and $\mu$ are the effective population size and the mutation rate per nucleotide site per generation, respectively). Values given per base pair.

### 2.2.2 Linkage Disequilibrium (LD), LD Decay, Population Structure, and Kinship Analysis

The genetic relationship among the wild *A. thaliana* samples was analysed by TASSEL software to generate a genetic distance tree plot (Fig 1A), based solely on diversity within SNP data. Neighbor-joining analysis of the 131 genotypes revealed phylogenetic clustering of individuals originating from the same or similar geographical locations suggestive of isolation by distance (Fig 1A). STRUCTURE analysis (Pritchard, Stephens et al. 2000) using the admixture model for the multilocus genotype data with the subpopulation number selection criteria of Rosenberg et al. (Rosenberg, Pritchard et al. 2002) suggested that the wild Irish *A. thaliana* populations consisted of three main genetic clusters (K=3), which were evenly distributed across the island of Ireland (Fig. 1B, C). The proportion of each cluster present in each individual is shown visually in figure 1B and then geographically in Figure 1C.

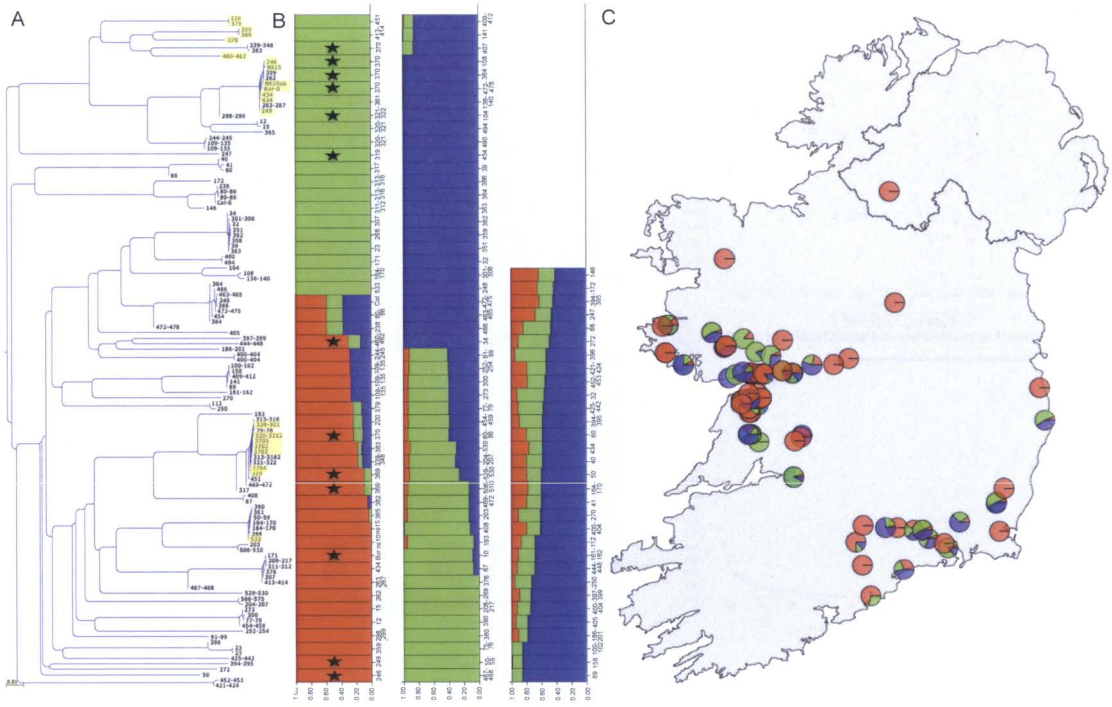



**Figure 1. Genetic diversity among Irish *A. thaliana* accessions. (A)** One of the neighbor-joining phylogenetic trees constructed using simple parsimony substitution models on SNP data in TASSEL software. Accessions with the *IIL1* triplet expansion are highlighted. **(B)** STRUCTURE bar plot showing the three genetic clusters identified by the analysis. Numbers 12-533 indicate the individual Irish accessions used in this study. The black stars represent samples that contain the repeat expansion. **(C)** Pie charts representing proportion of clusters 1 (red), 2 (green) and 3 (blue) present in each accession genotyped, as from (B), overlayed onto a geographic map of Ireland.

This was also confirmed by principal component analysis of the covariate data, in which a total of 40% of the variance was explained by three principal components (Where, 18.4, 12.2 and 9.6% are the variances explained by the first to third components respectively) (Fig 2), again suggesting that the genetic relatedness is associated with these three major clades observed in the Irish populations of *A. thaliana.* We found accessions that carried the repeat expansion fell into at least two of the 3 distinct clusters, suggesting that the repeat expansion has spread across Ireland in different genetic backgrounds. However, there is still some clustering with one of the clades completely lacking any samples that contained the repeat expansion (Fig 1B).

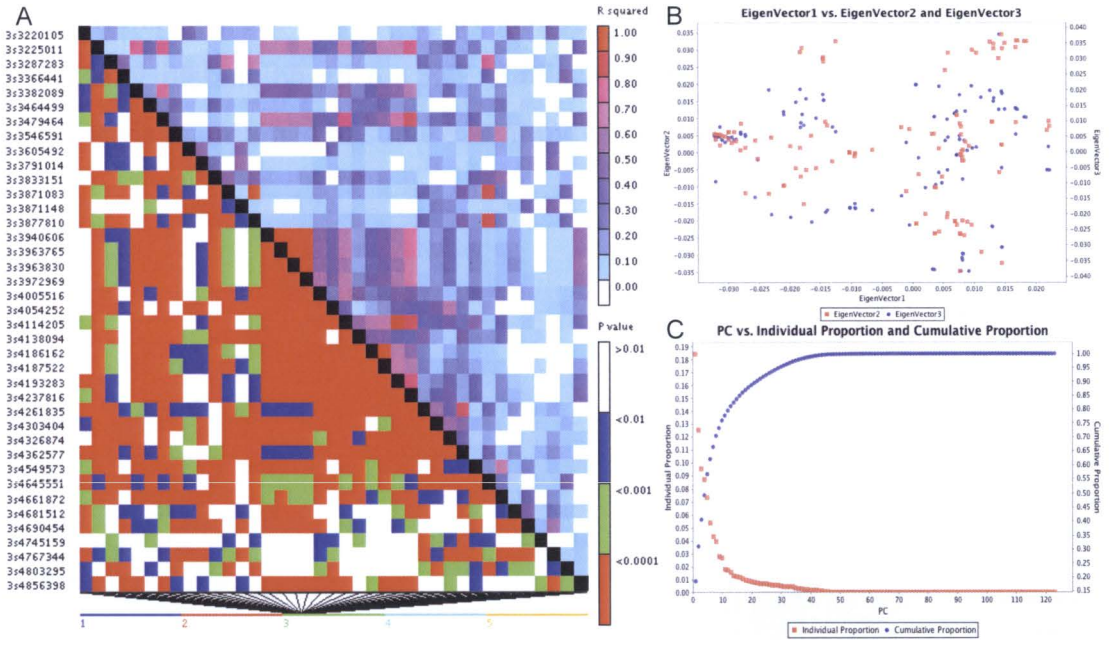



**Figure 2. Linkage disequilibrium (LD) hap map and Principal Component Analysis (PCA)** **(A)** Representative LD block on Chr3 showing a region of strong LD. Upper triangle presents $r^2$ values, with a colour gradient of red-white indicating strong ($r^2$ = 1) to weak ($r^2$ = 0) LD. The bottom triangle presents the associated *P*-value. **(B, C)** Principal component analysis (PCA). The relationship between variables is presented visually. **(B)** Scatter plot showing clustering of PC2 and PC3 relative to PC1. **(C)** Scree Plot graphing the eigenvalue against the component number. There are 3 PCs follow the 'elbow' in the scree plot. Therefore, 3 PCs were taken to be the point at which the remaining eigenvalues are relatively small.

Linkage disequilibrium (LD) was calculated with TASSEL software between homozygous SNP marker loci. A nonlinear regression model that estimates the decay of LD with distance was implemented using a pairwise analysis for all 5,369 SNP loci. $r^2$ values were regressed on their physical distances (Fig 3A). LD decay analysis revealed that the average decay of LD in terms of physical distance declined to $r^2 < 0.1$ at 6-7 Mb (~18 cM) (Fig 3A). LD analysis presented a well distributed weak to strong LD between markers indicating there were no major large haplotypes present across the Irish *A. thaliana* populations overall. Nevertheless, there were some regions showing strong LD (Fig 2A).

Next we plotted the genetic similarity for each accession, against their geographical distances from each other in order to visualise the genetic decay over distance (Fig 3B). As expected in established populations, *A. thaliana*

populations that are further apart geographically were also found to be more genetically distant based on SNP similarity matrix (Fig 3B). Thus population structure is maintained to some extent in Irish *A. thaliana*.

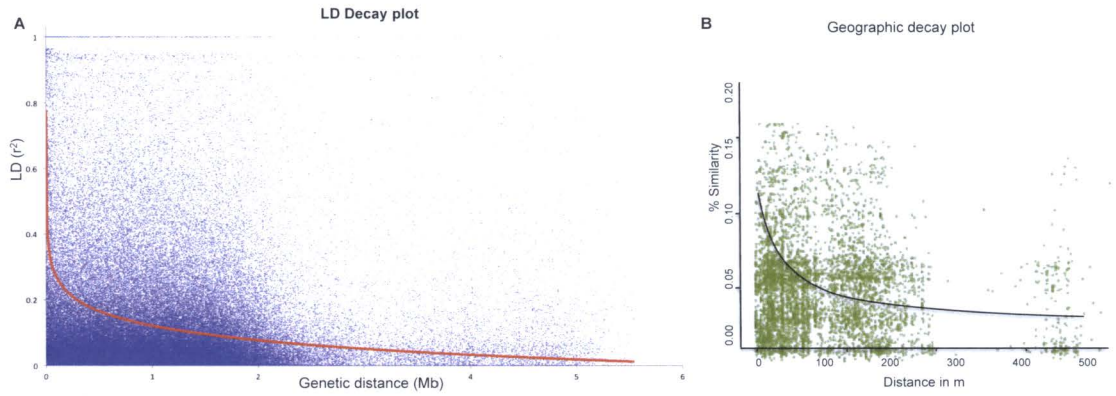



**Figure 3. Linkage disequilibrium decay (LD) analysis of Irish *A. thaliana* & genetic isolation by distance. (A)** LD decay analysis of Irish *A. thaliana*. LD decay plot determined by squared correlations of allele frequencies ($r^2$) against genetic distance between polymorphic sites amongst wild *A. thaliana* accessions. Inner trend line is a nonlinear logarithmic regression curve of genetic distance between polymorphic loci. Genetic similarity over geographical distance. **(B)** Isolation by distance plot for the Irish accessions. Genetic similarity is derived from pairwise SNP similarity between all markers for each taxa. Similarity matrix is generated by a neighbour-joining simple parsimony model using TASSEL software and plotted against geographical distance in m. Black trend line is generated through nonlinear logarithmic regression.

### 2.2.3 Phenotypic variation in local Irish accessions

To determine whether the genetic diversity determined through DArT-SEQ is reflected at the phenotypic level, we undertook phenotypic analysis of the genotyped Irish collection for several developmental traits. We measured seed germination, root length, hypocotyl elongation, degree of leaf serration, as well as flowering time (Appendix - Table S1). There was substantial phenotypic varaition across all traits similar to that of global populations. Furthermore, all the measured traits displayed high broad-sense heritability with the $H^2$ ranging from 78-91% suggesting that most of the phenotypic variation is genetically controlled (Table 2) .

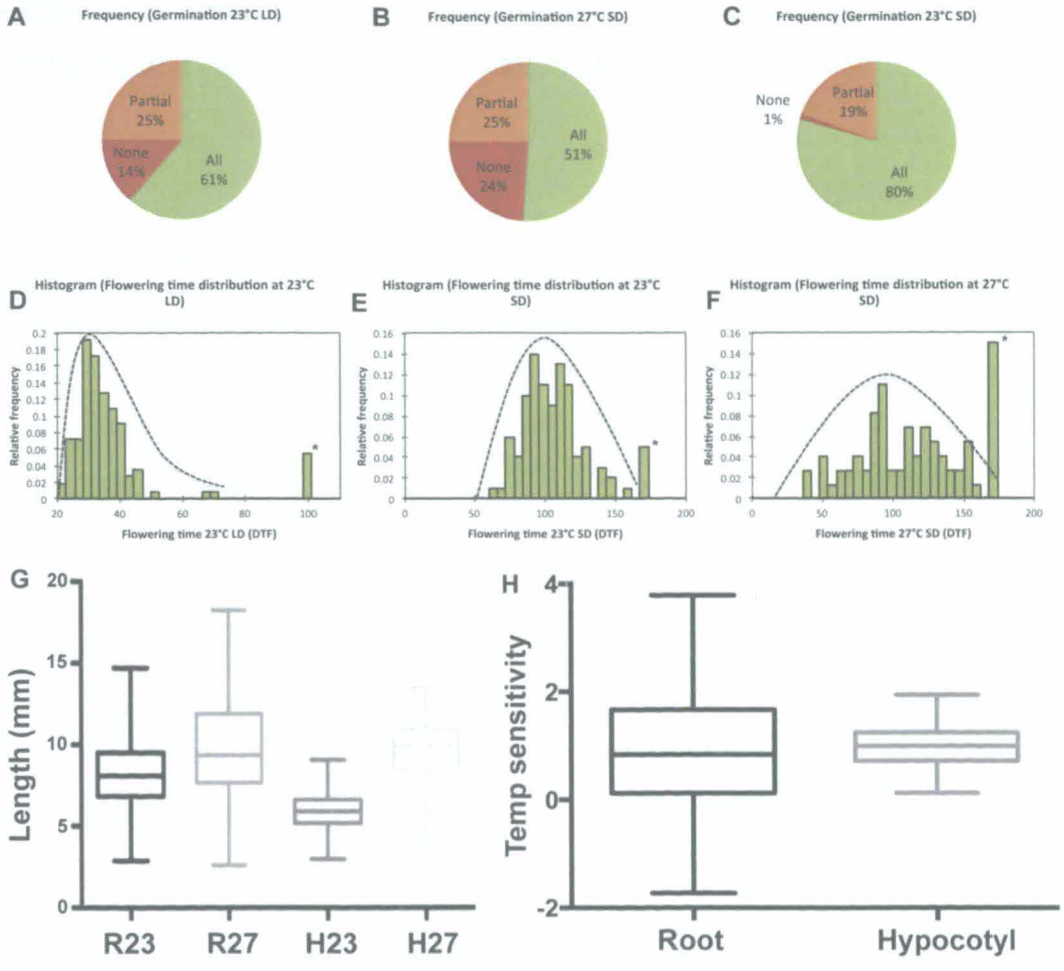



**Figure 4. Phenotypic variation in wild Irish *A. thaliana*. (A-C)** Variation in germination responses in wild Irish *A. thaliana*. Where (A, B & C) display percentage germination in 23 °C LD, 23 °C SD and 27 °C SD respectively. **(D-F)** Distribution of flowering times expressed as DTF (days to flower) among wild accessions in various conditions. Where (D, E & F) are 23 °C LD, 23 °C SD and 27 °C SD conditions respectively. * (Asterisk) indicates accessions extremely late flowering (>5 months, presumably winter annuals). **(G)** Box plots for root length 'R' and hypocotyl elongation 'H' in 23 °C SD and 27 °C SD, expressed in mm. **(H)** Temperature sensitivity in root length and hypocotyl elongation among wild accessions were calculated from regression of hypocotyl/root length onto temperature means (23SD vs. 27SD). Boxplots display the slope of correlation between 23 °C SD and 27 °C SD.

### 2.2.4 Irish *A. thaliana* variation in germination responses

Greatest germination occurred in 23 °C short day (SD) with only 1% of the 131 wild accessions not germinating. At 27 °C SD only approximately half of the accessions (51%) completely germinated, while nearly a quarter (24%)

exhibited partial germination. Germination in 23 °C long day (LD) was somewhat in between that observed in 23 °C SD and 27 °C SD (Fig 4A-C).

### 2.2.5 Variation of flowering responses in wild Irish *A. thaliana*

Flowering time was measured in 150 natural Irish *A. thaliana* accessions under three different environmental conditions: 23 °C LD, 23 °C SD and 27 °C SD. Days to flowering (DTF) was measured for all 150 wild accessions along with the common laboratory strains Col-0, Bur-0 and L*er* as well as the Cvi-0 strain and two natural suppressors in the Bur-0 background (NS_1 and NS_5).

In general, the distribution of flowering times reflected the trends observed in global populations (Lempe, Balasubramanian et al. 2005) (Werner, Borevitz et al. 2005, Shindo, Lister et al. 2006). Typically, both long days and elevated temperatures accelerate flowering, and the wild accessions flowered, on average, earlier in 23 °C LD compared to 23 °C SD (Fig 4D, E). There were a considerable number of accessions very late flowering, especially at 27 °C SD (Fig 4F). In contrast to the prominent peak of early-flowering accessions observed in 23 °C LD, flowering time was more broadly distributed in 23 °C SD and 27 °C SD (Fig 4D, E & F).

### 2.2.6 Variation in hypocotyl elongation and root lengths among Irish *A. thaliana*

Since the original Bur-0 accession displayed the *iil* phenotype in a temperature-dependent manner, we wondered whether the temperature sensitivity of these accessions differs. Therefore, next we analysed the response of root length and hypocotyl elongation of each accession between 23 °C and 27 °C in short days. Root length and hypocotyl elongation can be used to measure temperature sensitivity, with both hypocotyl and root length typically increasing at higher temperatures. The degree by which the root lengthens and hypocotyl extends,

are both measures of the sensitivity of a particular accession to the temperature. In Figure 4G, the distribution of root length and hypocotyl elongation at both 23 °C SD and 27 °C SD among the wild accessions is presented on the same scale to depict the shift and distribution for each parameter.

The root length range, for 23 °C SD and 27 °C SD, is 4.5-17.0 mm and 5.6-18.3 mm respectively with a change of 0.0-13.7 cm. The hypocotyl elongation ranges from 2.5-9.0 mm for 23 °C SD and 3.3-13.4 mm for 27 °C SD. Each hypocotyl and root parameter was normally distributed, with the greatest shift in length, occurring in the hyocotyls. The variation in sensitivity to different environmental temperatures, 23SD and 27SD, was obvious when hypocotyl and root lengths were regressed on their environmental means (Fig4H). Root length temperature sensity ranged from approximately -1.8 to 3.9 and temperature sensitivity of hypocoyl elongation ranged from approximately 0.5 to 2.

### 2.2.7 Variation in temperature response of hypocotyl elongation, root length and flowering time in wild Irish *A. thaliana*

There is a strong phenotypic correlation between hypocotl lengths at different temperatures. The same trend was also observed with root lengths although the correlation was higher in hypocotyl elongation at different temperatures. The range of differences in hypocotyl elongation lengths between different temperatures ranged from 0-12.0 mm and the range of differences in root lengths between different temperatures ranged from -2.6-13.7 mm. In the Irish population, root length gives a more linear measure of correlation ($R^2$ = 0.51831) over hypocotyl elongation ($R^2$ = 0.29622) (Fig 5A, B).

Flowering time correlation between different temperatures, 23 °C and 27 °C SD, was measured, for each accession, as their shift in DTF between 23 °C SD and 27 °C SD conditions. *A. thaliana* use ambient temperature as an environmental cue to either initiate or delay flowering. Typically plants will flower earlier in 27 °C

than in 23 °C. Here, the flowering time correlation between different temperatures was not very dramatic with only an $R^2$ = 0.1142 (Fig 5C). Two wild accessions, At161-162 & At204-207, were identified as having the opposite temperature response than otherwise expected with early flowering at 23 °C and very late flowering at 27 °C.

On average hypocotyl length and root length increased by 43.6% and 18.5% respectively with an increase in temperature. As the correlations between different temperatures were lower for flowering than for hypocotyl and root elongation, probably there is more GXE influence in flowering.

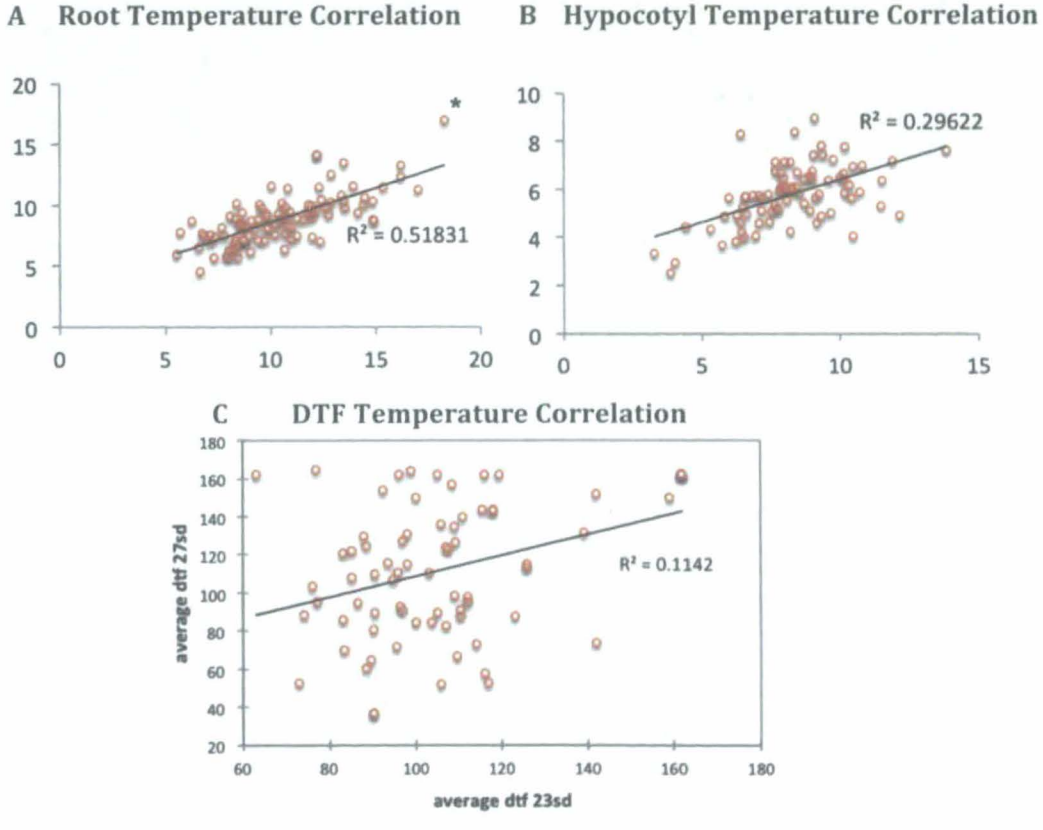



**Figure 5. Root length, hypocotyl elongation and flowering time temperature correlation among wild accessions. (A)** Correlation analysis of root length between 23 °C and 27 °C SD of wild accessions. **(B)** Correlation analysis of hypocotyl elongation between 23 °C and 27 °C SD of wild accessions. [NB: *= wild accession At112, which does not show a typical temperature response, with unusually long hypocotyls at both 23SD and 27SD]. **(C)** Correlation analysis of flowering time (measured as DTF) between 23 °C and 27 °C SD of wild accessions. Flowering time, measured as average DTF in 27 °C is plotted against average DTF in 23 °C. $R^2$ = 0.1142. [NB: *161-162 & *204-207 = early flowering at 23 °C and very late flowering at 27 °C].

In addition to the measured traits, broad phenotypic variation was observed across the Irish population in general (Fig 6) which is consistent with the identified genotypic diversity.

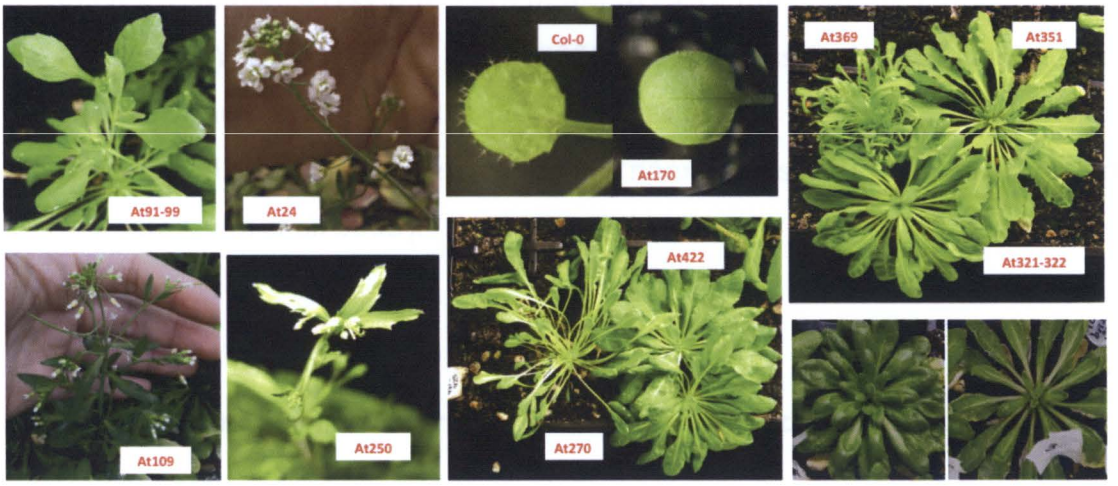



**Figure 6. Representation of the broad phenotypic variability amongst wild Irish *A. thaliana*.** Variations in leaf morphology such as between samples At91-99 & At250 were observed. Sample At24 is an apparent agamous mutant. Sample At422 is a trichomeless mutant. Sample At109 represents a discovered population of dwarf *A. thaliana*. Sample At422 displays a wild accession seemingly insensitive to elevated temperature, as it does not display a typical phenotypic response to high temperature such as the characteristic cotyledon elongation present in sample At270.

### 2.2.8 Analysis of the *FRIGIDA* genotype among wild Irish *A. thaliana* accessions.

All 131 wild accessions used in this study were also genotyped for the two known deletions at the *FRI* locus. Different allelic variants of the *FRIGIDA* (*FRI*) gene (Col or L*er*-type deletions) are a major indicator of whether the plant will be early or late flowering. Our wild accessions were screened for the presence/absence of Col-type and L*er*-type *FRI* deletions (Appendix - Table S1). We analysed by PCR analysis, all 131 wild accessions for both mutations. Nine out of 131 wild accessions carried the Col-type *FRI* deletion. They were; At146, At317, At379, At467, At490, At161-162, At339-348, At 80-86 and At460-462 all

of which flowered early in 23LD. That is these wild accessions averaged 33.2 DTF while Col-0, in this study, averaged 22 DTF and Bur-0 68 DTF. Col-0 is known to flower early and Bur-0 is known to flower late. None of the wild accessions possessed L*er*-type *FRI* deletions (Appendix - Table S1). This pattern is quite distinct from global populations where approximately 34% of global *A. thaliana* accessions were found to carry the Ler-type *FRI* deletion and approximately 56% of global *A. thaliana* accessions were found to carry the Col-type *FRI* deletion at the *FRI* locus (Werner, Borevitz et al. 2005). Nevertheless, it should be noted that the original Bur-0 accession also lacks either of these deletions, but carries a modified *FLC* allele, which appears to confer nonresponsiveness to vernalisation (Werner, Borevitz et al. 2005).

### 2.2.9 Natural variation in leaf serration among wild Irish accessions of *A. thaliana*.

Degree of leaf serration was graded using an internal scale from 1-9 of increasing serration (Appendix - Figure S1). A slightly bimodal distribution was observed in the Irish population, with the largest peak containing moderately serrated leaves and a minor peak containing more severely serrated leaves (Fig 7A). Geographic distribution of leaf serration indicates this trait is fairly evenly distributed across the country. Though, there is potentially a slight clustering of more severely serrated leaves in the southeastern region of the country (Fig 7B).

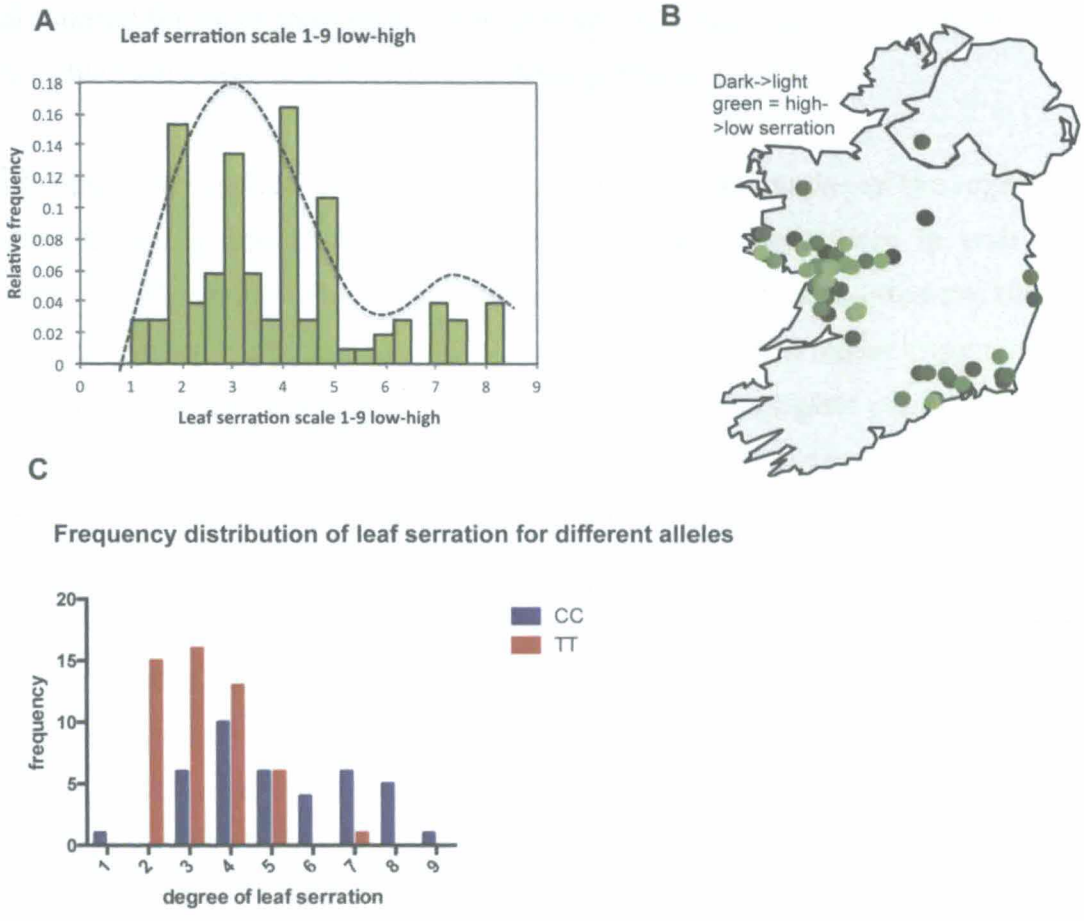



**Figure 7. Leaf serration in wild Irish *A. thaliana*.** **(A)** Frequency histogram depicting slight bi-modal distribution in leaf serration among the 131 samples. The majority of accessions had a low-moderate level of leaf serration. **(B)** Geographic distribution of leaf serration across Ireland where, light green to dark green represents low to high serration. **(C)** Differences in the distributions of leaf serration among Irish accessions with the genotype at 7.85Mb in chromosome 4 (DArT-Seq marker ID : 100017906).

### 2.2.10 Variation in triplet repeats and its association with phenotypes.

Since variation in triplet repeats has been shown to be associated with phenotypic differences in other species, we analsed whether any such association exists in our populations. To assess whether phenotypic variation can be explained by triplet repeat variation at the *IIL1* locus, we explored a model with the phenotypes as response and the copy number of the triplet repeats as a factor (Fig 7C). Consistent, with earlier findings, we found a significant association with the *iil* phenotype and the triplet repeats, which

accounted for more than 60% of the phenotypic differences, suggesting that we are able to discover true associations through this approach.

We observed significant associations between the copy number of the repeats at the *IIL1* locus and leaf serration and temperature sensitivity in root and hypocotyl elongation. The copy number of triplet repeats accounted for 10% of variation in leaf serration ($p<0.0059$) with the plants with higher copy number displaying somewhat reduced leaf serrations. Similarly, higher copy number of the *IIL1* repeat was associated with reduced sensitivity to temperature both in root and hypocotyl elongation, which accounted for 8% ($p<0.0172$) or 11% ($p<0.0084$) for the traits respectively. Consistent with this, analysis of DART-Seq markers around (7.7-7.9Mb) the *IIL1* locus (7.8 Mb), revealed that a marker at position 7.85 Mb (Marker ID: 100017906) was strongly associated with leaf serration, explaining 30% of the variation ($p<0.0001$). Thus, we found it appears that the *IIL1* tri-nucleotide repeat variation even within the normal range may be associated with phenotypic variation. However, it is likely that this association reflects a potential LD with another gene, which may modulate phenotypic responses in leaf shape.

| Phenotypic trait | Variance | Mean | Range | STDev | Heritability |
|---|---|---|---|---|---|
| DTF (23LD) | 175.9685 | 35.629623 | 22-100 | 13.26531 | 78% |
| Root length (23SD) | 4.171006 | 8.9832248 | 4.544761-17.03192 | 2.042304 | 91% |
| Hypocotyl elongation (23SD) | 1.325227633 | 5.771875428 | 2.967523-8.99073 | 1.151185 | 86% |
| Leaf Serration | 3.193282067 | 3.829326923 | 1-8.25 | 1.786975 | n/a |
| *IIL1* TTC/GAA repeat number | 9571.068646 | 46.81818182 | 3-400+ | 97.83183 | n/a |

**Table 2.** Summary statisics for phenotypes of the Irish *A. thaliana* population analysed in this study.

## 2.3 DISCUSSION

### 2.3.1 Genetic diversity of Irish *A. thaliana*

The diversity among the Irish population is extensive both in terms of genotype (SNP derived) and phenotype. I have shown that the Irish population falls into 3 main phylogenic branches. There are many wild accessions that are, in terms of their genetic background more similar to the lab strains of Col-0, Cvi and L*er* than they are to each other or to Bur-0, the only accession previously collected from Ireland. This indicates that there is quite a broad genetic variance among the Irish accessions. This notion is supported by the measured phenotypic traits of flowering time, hypocotyl & root length, temerature response, range of *IIL1* TTC/GAA repeats, degree of leaf serration and germination patterns at various temperatures. All of these phenotypic parameters span response ranges similar to that observed from other geographical regions (Le Corre, Roux et al. 2002, Pico, Mendez-Vigo et al. 2008, Bomblies, Yant et al. 2010, Long, Rabanal et al. 2013, Brennan, Mendez-Vigo et al. 2014). Thus, multiple colonization events (from Britain or continental Europe) are likely. Further genetic analysis will be required to resolve this question.

A previous study looking at LD decays in a global sample of 19 *A. thaliana* accessions, chosen for maximum genetic diversity, had an average LD decay of about 10 Kb (Kim, Plagnol et al. 2007). This is a fairly steep decay in LD similar to that seen in humans (Kim, Plagnol et al. 2007), which might be surprising given the 'selfing' nature of *A. thlaiana*. But it reflects the fact that outcrossing is common in *Arabdopsis* (Kim, Plagnol et al. 2007). More inbread populations desplay slower decaying LD (Kim, Plagnol et al. 2007). Even after repeated inter-crossing, RIL populations still suffer from slowly decaying linkage disequilibrium (LD) (Keurentjes, Bentsink et al. 2007). Our population displayed an LD decay $r^2$ >0.1 at approximately 7-8 Mb, and a decrease in $r^2$ by 50% at <5 Kb. These results are consistent with similar studies on global populations (Kim, Plagnol et al. 2007), but still indicates that the Irish population is quite genetically diverse

in terms of its haplotype breakdown and LD decay over genetic distance and geographical distance. Given that the there is only a fairly weak correlation between genetic and geographical distance (Fig 3B), it indicates that the *A. thaliana* subpopulations are not very well established and the different genotypes are distributed around the country frequently probably through human interference as the vast majority of collection sites were well populated areas.

### 2.3.2 Diversity within phenotypic traits of Irish *A. thaliana*

Distribution of flowering times among wild accessions at 23 °C LD followed the same distribution curve as reported in ealier studies, with only a slightly narrower range which could be due to a smaller sample size (Johanson, West et al. 2000, Lempe, Balasubramanian et al. 2005) (Werner, Borevitz et al. 2005, Shindo, Lister et al. 2006). Heritability ranged from 78-91%, which is comparable to that observed in global *A. thaliana* flowering response studies under the same environmental conditions (Lempe, Balasubramanian et al. 2005).

The change in root and hypocotyl length observed among the wild accessions with the increase in temperature is consistant with *A. thaliana* from the northern hemisphere where temperature changes considerably from summer-winter. Plants from these regions have develpoed the ability to detect changes in temerature as environmental cues to know when to comit to critical life decissions such as when to germinate and flower. Thus, it is an adaptive advantage for such plants to be able to respond sharply to changes in temperature. For example C*vi* is an accession of *A. thaliana* that originates from the tropics where temperature varies little between different seasons of the year, this particular accesson has adapted to its local environment and as a result is largely temperature insensitive (Zhu and Balasubramanian, personal communication).

### 2.3.3 Temperature sensitivity of Irish *A. thaliana*

The general trend among the Irish population is: As temperature increases from 23 °C-27 °C, flowering time decreases while root length and hypocotyl elongation increase. This is as expected, however, there are some individuals within the Irish population, that that do not conform to this general trend. Out of the 131 accessions used in this study, approximately 18 did not show a typical response to temperature in terms of flowering time. That is, they flowered noticably later at 27SD than 23SD (they were, At24, At104, At141, At171, At300, At351, At369, At397, At15, At204-207, At239-243, At320-321, At321-322, At370, At400-404, At425-442, At454-459 and At463-465). There were 8 accessions (At300, At321-322, At358, At361, At364, At370, At407 & At409-412) that did not have significantly longer roots at 27 °C SD than 23 °C SD and 5 accessions that exhibited the opposite trend, where root length at 23 °C SD was longer than root length at 27 °C SD (At77-79, At239-243, At319, At359 & At451). These accessions are interesting candidate's for further analysis of their GXE response.

For hypocotyl elongation, which is the most sensitive response to temperature, all measured accessions had an increase in hypocotyl length to some extent. The change in hypocotyl length ranged from 0.0-7.5 mm. Several accessions had only a small response (<1.5 mm) including 12 of the accessions that responded poorly to DTF and root length temperature changes (At 24, At319, At 351, At359, At369, At397, At407, At451, At320-321, At321-322, At370, At425-442 and At454-459). Typically, the same individuals that did not respond to temperature change in any one of the temperature related traits were not responding well in the others either. This suggests that those accessions have decreased temperature sensitivity. If an accession responded abnormally in one temperature responsive trait but normally in others, this would suggest that that accession has a problem with the specific trait itself rather than temperature response. (e.g. if an accession responded in a typical fashion to temperature change in DTF and root length, but then showed no response in hypocotyl, this would suggest that the

problem lies within that accessions ability to extend its hypocotyl rather than its ability to detect temperature). Wild accessions (At 24, At319, At 351, At359, At369, At397, At407, At451, At320-321, At321-322, At370, At425-442 and At454-459) displayed across-the-board decreased temperature sensitivity. These wild accessions are candidates for further analysis into temperature response pathways which is one avenue of future direction for this work. Using these accessions we aim to investigate and unearth genes and biological processes involved in temperature response, which is another very important area of research.

Of these 13 temperature insensitive accessions, 4 were collected from the same location from different plants and at different time points i.e, one year appart. As only 1 representative individual plant from each distinct location (in general) was used in this phenotypic analysis, other individuals from the same location should also be analysed for their temperature response to DTF, root and hypocotyl responses. Plants collected from the same population are expected to be genetically similar (almost clones) as *A. thaliana* has such a high self fertilization frequency.

Our studies also revealed some interesting associations between the copy number of the *IIL1* intronic triplet repeat with leaf serration and temperature sensitivity (Fig 7D). Copy number variation in the non-expanded range has been shown to be associated with phenotypic variation in other systems, although most of these associations are with repeats that occur within the coding region. Therefore, it is likely that the observed associations reflect and LD with another locus somewhere in the nearby region. The stronger association with the DArT-Seq marker (100026931), which appears to explain more variation, supports this hypothesis. This is also consistent with the observed bimodal distribution of repeats that we observed in the Irish accessions. Nevertheless, further analysis is required to explore this association at the functional level.

### 2.3.4 Limitations of our study

We used STRUCTURE to analyse the population structure of the Irish accessions. Genetic relationships can be simple or complex. Statistical methods that interpret relationships do not always describe the actual relationship among populations e.g. if populations have an isolation-by-distance population structure, an unweighted pair group method of genetic structure would provide misleading depiction of genetic structure (Kalinowski 2011).

Another limitation with our dataset is that due to cost restraints we were unable to sequence the whole genomes of the 131 wild accession. Whole genome sequences as opposed to SNP data, would have allowed much more indepth genetic analyses such as, investigation into evolution, ansestery and phlogeny of the Irish *A. thaliana* population. However, the primary aim of this Irish *A. thaliana* collection was to analyse the TTC/GAA triplet repeat variability with the potential of recovering the dramatically expanded allele if it were still present in the wild, which was achieved. However, as DNA, seeds, plant tissue and extensive phenotypic data are available for all collected wild Irish accessions, this allows for future in-depth whole genome analysis of the Irish population. Another future direction for this work is to compare the Irish population to a global population of *A. thaliana* which would allow us to characterise this isand population in a more meaningful context.

### 2.3.5 A new collection of Irish *A. thaliana*

This study has generated a new collection of genotyped strains from an island population that is suitable for studying natural variation in *A. thaliana.* We have also provided evidence that the Irish population displays both considerable genotypic and phenotypic diversity, likely comparable to that of other regional populations (Le Corre, Roux et al. 2002, Pico, Mendez-Vigo et al. 2008, Bomblies, Yant et al. 2010, Long, Rabanal et al. 2013, Brennan, Mendez-Vigo et al. 2014).

This Irish collection will also serve as a rich resource for further genetic analysis into the history of the post-glacial colonization of Ireland by *A. thaliana*. Native to Eurasia and Northern Africa, *A. thaliana* is believed to have colonized Ireland in the postglacial period (Francois, Blum et al. 2008). However, the actual time period and route of introduction to Ireland is unknown. Overall, our phylogenetic analysis, population structure analysis and Principle Component Analysis results further indicate there may be at least 3 distinct genetic backgrounds of *A. thaliana* occupying Ireland, which may be remnants of successive colonization processes.

## 2.4 MATERIALS & METHODS

The collection of plant material has been described in Chapter 1.

### 2.4.1 Phenotyping

***Growth conditions & Plant culture:*** All plants were grown in controlled growth rooms with a temperature variability of about +/- 0.1 °C under long day conditions (16 h of light) or short day conditions (8 h of light). Maximal humidity was 65%.

***Germination & Flowering time:*** Six seeds per accession were sown in a completely randomized design in individually numbered positions of multiple-well flats filled with soil. Germination was scored for each accession under all conditions: 23 °C SD, 27 °C SD and 23 °C LD. Germination was scored as either 'All', 'Partial' or 'None'. If all 6 plants germinated, germination was classified as 'All', if between seven and one plant germinated, it was classified as 'Partial' and if none of the seeds germinated, 'None'. Flowering time was measured as both toal leaf number (TLN) and days to flower (DTF) for all accessions.

***Root length & Hypocotyl elongation***: For analysis of root length and hypocotyl elongation, seeds from all 131 accessions were sterilized in 1.5 ml tubes, in 70% EtOH with 0.1% Triton X-100 for 3-5 min and washed in 95% EtOH for 1-2 min. Seeds were then spotted onto 0.5 x MS media (PH 5.7), made from 2.17g Murashige and Skoog salts (Sigma) and 7.0g Agar for 1 litre. Seeds were stratified for four days in the dark at 4 °C then moved to light and humidity controlled Percival chambers (Percival Inc, Canada) of 23 °C SD, 27 °C SD or 23 °C LD conditions. Root length was measured 5 days post germination and hypocotyl elongation measured 10 days post germination from photographic images, using ImageJ64 software for Mac (National Institutes of Health, Bethesda, Maryland, USA). A completely randomized design of two repetitions of 20 plants per accession for each condition was used.

***Demographic traits & Leaf Serration***: Demographic traits including latitude & longitude, altitude, soil type were recorded at the collection site. Annual rainfall data was obtained from The Irish Metrological service online, MET éireann. Degree of leaf serration was based on a relative scale generated from within the dataset, ranging from 1-9 of increasing serration. Scale in Appendix - Figure S1. A single representative image of each of the 131 accessions was taken of 4-week old plants and 3 independent volunteers were asked to score the degree of serration based on the provided scale. The average score was taken for each accession.

### 2.4.2 Genotyping

Genotyping of 150 individuals was carried out by Diversity Arrays Technology Pty Ltd (DArT P/L, Canberra, Australia) as described previously (Sansaloni, Petroli et al. 2010). All samples with >10% missing data were excluded. Minor allele frequency (MAF) was set to 0.05% and all taxa markers having missing data in more than 20% of their loci were excluded. This filtering resulted in a dataset of 5,369 SNP markers from which genetic analysis was undertaken.

Generation of genetic trees, principal component analysis (PCA) and Linkage disequilibrium (LD) were analysed using TASSEL version 4.3.5 (Trait Analysis by aSSociation, Evolution and Linkage) (Bradbury, Zhang et al. 2007).

### 2.4.3 Data analysis

A subset of 5,369 DArT SNP markers was selected for the analysis based on quality parameters out of the 11,242 included on the genotyping array. The selected DArT markers had Call Rate (percentage of samples that could be scored as '0' or '1') > 80%, Reproducibility (reproducibility of scoring between replicated samples) > 97% and were polymorphic with frequencies of samples scored as '0' or '1' ranging between 0.95 and 0.05. Level of polymorphism for each marker was calculated by the expected heterozigosity: $H_e = 1\text{-}\Sigma P_i^2$ where $P_i$ is the frequency of the *i*th allele at the DArT marker locus. Because DArT markers are dominant, the allele frequencies were estimated making the frequency of '0' genotypes equivalent to $P_i^2$, assuming Hardy-Weinberg Equilibrium.

The 5,369 marker data were used to determine both population structure and pair-wise kinship coefficients among the 131 individuals, using the model-based Bayesian clustering algorithm implemented by STRUCTURE v.2.3.4 (Pritchard, Stephens et al. 2000) and the kinship procedure of Hardy (Hardy 2003) using the software package SPAGeDi v.2.1 (Hardy OJ), respectively. STRUCTURE analyses were performed assuming an admixture model with default settings (no informative priors were used). STRUCTURE was run from 1 to 20 inferred clusters (*K*) with 10 independent runs for each *K*, each run starting with a burn-in period of 10,000 steps followed by 10,000 Markov Chain Monte Carlo iterations. The most probable value of *K* was selected according to the Evanno method (Evanno, Regnaut et al. 2005). Negative kinship values were set to zero following Yu *et al.* (Yu, Pressoir et al. 2006). Pair-wise linkage disequilibrium (LD) between individual DArT markers was calculated by the square allele frequency correlation coefficient ($r^2$) implemented in the program TASSEL v.4.3.5 (Bradbury, Zhang et al. 2007) and their statistical significance was

computed by 1,000 permutations using the two-sided Fisher's Exact test (Weir 1996). Mean $r^2$ values were calculated separately for unlinked loci and for loci on the same chromosome. The 95$^{th}$ percentile of the square root transforming of $r^2$ distribution was taken as a population-specific critical value of $r^2$ (Breseghello and Sorrells 2006), beyond which LD was likely to be caused by genetic linkage.

### 2.4.4 Statistics

To evaluate environmental sensitivity, root and hypocotyl measurements were regressed onto the environmental mean, [log(root/hypocotyl$_{23SD}$) - log(root/hypocotyl$_{27SD}$)]/[log(mean root/hypocotyl$_{23SD}$) - log(mean root/hypocotyl$_{27SD}$)].

***Pairwise Linkage disequilibrium (LD) and LD decay***: The extent of LD was estimated from the DArTseq SNP data, as the squared allele frequency correlation ($r^2$) for each accession using TASSEL software. This produces two values $D'$ and $r^2$. $D'$ is the standardized disequilibrium coefficient (a statistic used for determining whether recombination or homoplasy has occurred between a pair of alleles) and $r^2$ represents the correlation between alleles at two loci. When multiple alleles are present at a particular locus, a weighted average of $D'$ or $r^2$ is calculated between the two loci (Farnir, Coppieters et al. 2000).

This weighted average is determined by calculating $D'$ or $r^2$ for all possible combinations of alleles, and then weighting them according to the alleles frequency. When only two alleles are present at both loci a two-sided Fisher's Exact test is calculated to determine *P*-values. If more than two alleles are present, permutations are used to calculate the proportion of permuted gamete distributions that are less probable then the observed gamete distribution under the null hypothesis ($H_o$) of independence (Weir 1996). To ensure an unbiased *P*-value, the option of employing rapid permutations was selected. This algorithm runs permutations until 10 are found that are more significant than the observed P-value, thus slightly reducing the *P*-value.

LD decay graphs were plotted with genetic (Mb) or geographic distance (Km) vs. $r^2$ for each marker pair locus located on the same chromosome, using nonlinear regression as described by Remington et al. (Remington, Thornsberry et al. 2001). The expected decay of LD was estimated according to the following equation:

$$E(r^2) = [10 + pd)/(11 + pd)](1 + \{(3 + pd)\,[12 + 12pd + p^2d^2]\}/[n(2 + pd)(11 + pd)])$$

Where, $n$ denotes the number of sequences, $p = 4N_e c$ between adjacent sites, $d$ is the distance between the two sites of a pairwise comparison, and $c$ is the recombination rate (Hill and Weir 1988).

Genetic distance was also plotted against geographic distance (Km). Genetic distance was based on pairwise SNP similarities between all accessions. The genetic distance matrix was generated by a neighbor-joining simple parsimony model using TASSEL. The geographic distance matrix was a pairwise matrix of distances between accessions, calculated by the Haversine formula (Sinnott 1984) using the latitude and longitude for each accession.

***Population Structure and Kinship***: Estimation of population structure ($Q$) and kinship (K) relationships were derived using only loci that had pairwise $r^2$ values < 0.5 for all possible combinations. From the $R^2$ data 344,933 markers from the 500,825 SNP marker-pair comparisons had $R^2$ < 0.5 among all pairwise comparisons. 999 randomly chosen markers from this marker subset were used as input data for SPAGeDi analysis for calculation of $Q$ and K. Population structure was first characterized using STRUCTURE to estimate subpopulation membership of each line in these two populations individually (Pritchard, Stephens et al. 2000). The admixture model with correlated allele frequencies was used with a burn-in of 10,000 and 10,000 MCMC iterations for subpopulations numbers ranging from 1 to 15. Ten runs for each $K$ value were performed, and the STRUCTURE HARVESTER software (implementing the

Evanno method) was used to select the optimum number of *K*. The optimum number of subpopulations was determined by the Wilcoxon two sample *t* test as described by Rosenberg et al. (2002) by comparing the posterior probability for successive adjacent subpopulations (K2 vs. K3, K3 vs. K4, and so on) (Rosenberg, Pritchard et al. 2002). The smaller *K* value in a pairwise comparison for the first nonsignificant Wilcoxson test was chosen as the best number of subpopulations.

Principal component analysis (PCA) was also used to control for population structure using TASSEL software. The number of principal components (eigenvectors per combination of SNP markers) that collectively explained 40% of the variation was selected for the analysis.

A pairwise kinship coefficient matrix (**K**-matrix) that estimates the probability of recent co-ancestory between genotypes (Bette A. Loiselle 1995) was determined using SPAGeDi (Hardy OJ). A randomly chosen subset of 999 markers was used as input into the SPAGeDi software. The formula is:

$$F_{ij} = (Q_{ij} - Q_m)/(1 - Q_m) \approx r_{ij}$$

Where $r_{ij}$ is the pairwise kinship coefficient, $F_{ij}$ is an estimator of the coefficient, $Q_{ij}$ is the probability of the identity by state between random loci for genotypes *i* and *j*, and $Q_m$ is the average probability of identity by state for loci from random genotypes in the population used to draw *i* and *j*. The $F_{ij}$ was calculated for all pairwise combinations in each of the two populations. Negative values in the kinship matrix was set to zero as described by Yu et al. (Yu, Pressoir et al. 2006)

## 2.5 REFERENCES


Appleby, N., D. Edwards and J. Batley (2009). "New technologies for ultra-high throughput genotyping in plants." Methods Mol Biol **513**: 19-39.

Aranzana, M. J., S. Kim, K. Zhao, E. Bakker, M. Horton, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. Tang, C. Toomajian, B. Traw, H. Zheng, J. Bergelson, C. Dean, P. Marjoram and M. Nordborg (2005). "Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes." PLoS Genet **1**(5): e60.

Bette A. Loiselle, V. L. S., John Nason and Catherine Graham (1995). "Spatial Genetic Structure of a Tropical Understory Shrub, Psychotria officinalis (Rubiaceae)." American Journal of Botany . Published by: Botanical Society of America **82**(11): 25.

Bomblies, K., L. Yant, R. A. Laitinen, S. T. Kim, J. D. Hollister, N. Warthmann, J. Fitz and D. Weigel (2010). "Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of Arabidopsis thaliana." PLoS Genet **6**(3): e1000890.

Borevitz, J. O. and M. Nordborg (2003). "The impact of genomics on the study of natural variation in Arabidopsis." Plant Physiol **132**(2): 718-725.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss and E. S. Buckler (2007). "TASSEL: software for association mapping of complex traits in diverse samples." Bioinformatics **23**(19): 2633-2635.

Brennan, A. C., B. Mendez-Vigo, A. Haddioui, J. M. Martinez-Zapater, F. X. Pico and C. Alonso-Blanco (2014). "The genetic structure of Arabidopsis thaliana in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe." BMC Plant Biol **14**: 17.

Breseghello, F. and M. E. Sorrells (2006). "Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars." Genetics **172**(2): 1165-1177.

Evanno, G., S. Regnaut and J. Goudet (2005). "Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study." Mol Ecol **14**(8): 2611-2620.

Farnir, F., W. Coppieters, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar

and M. Georges (2000). "Extensive genome-wide linkage disequilibrium in cattle." Genome Res **10**(2): 220-227.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt and A. M. Wilczek (2011). "A map of local adaptation in Arabidopsis thaliana." Science **334**(6052): 86-89.

Francois, O., M. G. Blum, M. Jakobsson and N. A. Rosenberg (2008). "Demographic history of european populations of Arabidopsis thaliana." PLoS Genet **4**(5): e1000075.

Hancock, A. M., B. Brachi, N. Faure, M. W. Horton, L. B. Jarymowycz, F. G. Sperone, C. Toomajian, F. Roux and J. Bergelson (2011). "Adaptation to climate across the Arabidopsis thaliana genome." Science **334**(6052): 83-86.

Hardy, O. J. (2003). "Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers." Mol Ecol **12**(6): 1577-1588.

Hardy OJ, V. X. "(2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618–620. doi: 10.1046/j.1471-8286.2002.00305.x."

He, F., D. Kang, Y. Ren, L. J. Qu, Y. Zhen and H. Gu (2007). "Genetic diversity of the natural populations of Arabidopsis thaliana in China." Heredity (Edinb) **99**(4): 423-431.

Hill, W. G. and B. S. Weir (1988). "Variances and covariances of squared linkage disequilibria in finite populations." Theor Popul Biol **33**(1): 54-78.

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino and C. Dean (2000). "Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time." Science **290**(5490): 344-347.

Kalinowski, S. T. (2011). "The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure." Heredity (Edinb) **106**(4): 625-632.

Keurentjes, J. J., L. Bentsink, C. Alonso-Blanco, C. J. Hanhart, H. Blankestijn-De Vries, S. Effgen, D. Vreugdenhil and M. Koornneef (2007). "Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population." Genetics **175**(2): 891-905.

Kim, S., V. Plagnol, T. T. Hu, C. Toomajian, R. M. Clark, S. Ossowski, J. R. Ecker, D. Weigel and M. Nordborg (2007). "Recombination and linkage disequilibrium in Arabidopsis thaliana." Nat Genet **39**(9): 1151-1155.

Le Corre, V., F. Roux and X. Reboud (2002). "DNA polymorphism at the FRIGIDA gene in Arabidopsis thaliana: extensive nonsynonymous variation is consistent with local selection for flowering time." Mol Biol Evol **19**(8): 1261-1271.

Lempe, J., S. Balasubramanian, S. Sureshkumar, A. Singh, M. Schmid and D. Weigel (2005). "Diversity of flowering responses in wild Arabidopsis thaliana strains." PLoS Genet **1**(1): 109-118.

Lewandowska-Sabat, A. M., S. Fjellheim and O. A. Rognli (2010). "Extremely low genetic variability and highly structured local populations of Arabidopsis thaliana at higher latitudes." Mol Ecol **19**(21): 4753-4764.

Long, Q., F. A. Rabanal, D. Meng, C. D. Huber, A. Farlow, A. Platzer, Q. Zhang, B. J. Vilhjalmsson, A. Korte, V. Nizhynska, V. Voronin, P. Korte, L. Sedman, T. Mandakova, M. A. Lysak, U. Seren, I. Hellmann and M. Nordborg (2013). "Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden." Nat Genet **45**(8): 884-890.

Pico, F. X., B. Mendez-Vigo, J. M. Martinez-Zapater and C. Alonso-Blanco (2008). "Natural genetic variation of Arabidopsis thaliana is geographically structured in the Iberian peninsula." Genetics **180**(2): 1009-1021.

Pritchard, J. K., M. Stephens and P. Donnelly (2000). "Inference of population structure using multilocus genotype data." Genetics **155**(2): 945-959.

Remington, D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt, J. Doebley, S. Kresovich, M. M. Goodman and E. S. t. Buckler (2001). "Structure of linkage disequilibrium and phenotypic associations in the maize genome." Proc Natl Acad Sci U S A **98**(20): 11479-11484.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky and M. W. Feldman (2002). "Genetic structure of human populations." Science **298**(5602): 2381-2385.

Sansaloni, C. P., C. D. Petroli, J. Carling, C. J. Hudson, D. A. Steane, A. A. Myburg, D. Grattapaglia, R. E. Vaillancourt and A. Kilian (2010). "A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus." Plant Methods **6**: 16.

Sinnott, R. W. (1984). "Virtues of the Haversine." Sky and Telescope **68**(2): 159.

Weir, B. S. (1996). Genetic Data Analysis II : Methods for Discrete Population Genetic Data by Bruce S. Weir

Werner, J. D., J. O. Borevitz, N. H. Uhlenhaut, J. R. Ecker, J. Chory and D. Weigel (2005). "FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions." Genetics **170**(3): 1197-1207.

Yin, P., J. Kang, F. He, L. J. Qu and H. Gu (2010). "The origin of populations of Arabidopsis thaliana in China, based on the chloroplast DNA sequences." BMC Plant Biol **10**: 22.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich and E. S. Buckler (2006). "A unified mixed-model method for association mapping that accounts for multiple levels of relatedness." Nat Genet **38**(2): 203-208.

# CHAPTER 3. Investigating molecular mechanisms underlying the TTC/GAA repeat expansion associated genetic defect in the Bur-0 accession of *A. thaliana*

## 3.1 INTRODUCTION

### 3.1.1 Common properties underlying TNR expansion

TNRDs though vastly different in pathology, display common underlying properties in terms of repeat instability, genetic anticipation and the involvement of DNA repair. In general, there is a strong correlation between increasing TNR lengths and repeat instability in TNRDs (Ashizawa, Dubel et al. 1993, De Biase, Rasmussen et al. 2007, Liu and Leffak 2012). In terms of LOF TNRDs, such as friedreich's ataxia (FRDA), Myotonic dystrophy (DM) and Spinal and bulbar muscular atrophy (SBMA) there is similarly a strong correlation between degree of protein deficiency and repeat instability (Ashizawa, Dubel et al. 1993, Rusmini, Simonini et al. 2011, Morales, Couto et al. 2012). Genetic anticipation, the phenomenon of increased disease severity when inherited over successive generations, with the exception of poly alanine disorders, is hallmark to all TNRDs (Carpenter 1994, He and Todd 2011).

Commonalities with DNA repair also exist amongst most TNRDs (Schweitzer and Livingston 1997, Ezzatizadeh, Pinto et al. 2012). This was first discovered when studying CAG/CTG TNR length in the Huntington (*HTT*) gene, where a drastic expansion in TNR length was found in cells of specifically affected regions in the brain cortex and striatum (Mangiarini, Sathasivam et al. 1997, Kennedy, Evans et al. 2003). These cells were non-proliferating thus replicative mechanisms could not explain the high levels of expansion observed and the DNA repair pathways, base excision repair (BER) and mismatch repair (MMR) subsequently found to

contribute to the somatic cell TNR expansion instability in HD (Watanabe, Tanaka et al. 2000, Hashida, Goto et al. 2001, Kovtun and McMurray 2008, Mollersen, Rowe et al. 2010, Du, Campau et al. 2012, Ezzatizadeh, Pinto et al. 2012). Since then, similar DNA repair pathway involvement has been observed in TNRDs including Myotonic dystrophy type 1 (DM1), Myotonic dystrophy type 2 (DM2), Dentatorubral-pallidoluysian atropy (DRPLA), Fragile X syndrome (FXS) and Friedreich's ataxia (FRDA) (Spiro, Pelletier et al. 1999, Henricksen, Tom et al. 2000, Liu, Zhang et al. 2004, Owen, Yang et al. 2005, Kovtun, Liu et al. 2007, Jarem, Wilson et al. 2011, Kantartzis, Williams et al. 2012).

### 3.1.2 Current status of model systems for triplet expansion research

#### *3.1.2.1 Frataxin knock out models*

The development and use of animal and cellular models of FRDA are essential requirements for the understanding of the disease mechanisms and the investigation of potential therapeutic strategies. Mammalian models, such as mice, are critical in FRDA research. Fundamentally, an accepted prerequisite for clinical trials of a compound in humans is the successful alleviation of the disease in animal models. To date there are two types of mice models used to study FRDA. The first are full frataxin deletion models. Since complete deletion of frataxin in mice is not viable, resulting in day 6.5 embryonic lethality (Cossee, Puccio et al. 2000), there are mouse models with tissue-specific knock out of *FXN* (e.g., nervous system, heart and pancreas). These models reproduce most of the characteristic features of the disease, including hypertrophic cardiomyopathy, progressive spinocerebellar, and sensory ataxia, and some aspects of the diabetes. There are also smallRNA methods that target and greatly reduce, but not totally abolish, *FXN* expression. These models show FRDA pathophysiology but totally lack the underlying genomic mutation. The frataxin deletion models can be used to evaluate new therapies that treat the downstream effects of

frataxin deficit, such as oxidative stress (idebenone) (Soriano, Llorens et al. 2013) and mitochondrial iron accumulation (deferiprone) (Soriano, Llorens et al. 2013). However, these models do not contain a GAA/TTC expansion and thus research into the pathological mechanisms of the GAA/TTC expansion are not applicable to such systems nor are research of therapeutic approaches that may function to increase *FXN* expression specifically in FRDA.

#### *3.1.2.2 FRDA models with engineered GAA/TTC expansions*

The second approach is by engineering GAA/TTC expansions in the mouse *FXN* gene or by knocking-out mouse *FXN* and expressing human *FXN* transgenes harbouring GAA/TTC expansions derived from FRDA patients. KIKI, KIKO, YG8R, and YG22R are GAA/TTC repeat expansion based mouse models. KIKI mice contain homozygous KI $(GAA)_{230}$ repeat expansion in their 1st intron at *FXN* locus. They express 66-83% WT *FXN* (Perdomini, Hick et al. 2013). KIKO mice are KIKI crossed with a full KO *FXN* allele and produce 25-36% *FXN* (Perdomini, Hick et al. 2013). These mice have only mild FRDA symptoms suggesting they are still producing enough frataxin, perhaps if less than 25% WT frataxin was produced they would exhibit clinical symptoms. The lack of FRDA clinical pathophysiology in these mice is a limitation. YG8 mice contain human *FXN* transgenes in a mouse *FXN* null/KO. These mice have two GAA/TTC sequences of 90 and 190 repeats. Similarly, there is a YG22 line with one sequence of 190 repeats. When these are crossed with total *FXN* KO (crossed with heterozygotes and screened for offspring that inherit the null allele) they are known as YG8R and YG22R respectively. They contain GAA/TTC repeat expansions and express only human frataxin (Al-Mahdawi, Pinto et al. 2006). These mice possess most of the clinical aspects of FRDA but not the cardiac phenotype. Mice models containing the GAA/TTC expansion allow for pharmacological approaches that target the genomic impact of the expansion directly such as histone deacetylase inhibitors (HDACi) that change the genomic region of the expansion mutation to make it more accessible to transcription thus increasing *FXN* expression. A limitation with these humanized mice models are that they lack the genetic

context. That is, the repeat expansion is not in the native *FXN* gene. On the other hand, the KIKI and KIKO mice recapitulate the genomic *FXN* expansion, but only have minor phenotype (likely due to the still high *FXN* expression). Thus there is no current murine model with large *FXN* GAA/TTC repeat expansions that induce FRDA symptoms, though efforts to generate such mice are ongoing (Martelli, Napierala et al. 2012). It's important that mouse models not only exhibit similar clinical signs and features associated with the biochemical and cellular features of FRDA. But as the genetic basis plays an essential role in FRDA, a model that captures the GAA/TTC repeat instability is also important. To date no single model meets these criteria.

#### *3.1.2.3 Cellular models for FRDA*

Other models for FRDA research include FRDA patient-derived cells: fibroblasts, lymphoblasts, induced pluripotent stem cells (iPSC)s, and iPSC-derived cells. Cells derived from FRDA patients carry the complete *FXN* locus together with GAA/TTC repeat expansions and regulatory sequences, they are therefore the most relevant models. Cell types such as lymphoblasts and fibroblasts are not pathological tissues in FRDA and therefore do not exhibit the complex biochemical phenotypes associated with FRDA, despite having reduced levels of frataxin (Rotig, de Lonlay et al. 1997, Sturm, Bistrich et al. 2005). These cells could be useful for biomarker research and to study therapeutic candidates able to prevent cellular damage related to oxidative stress or to modulate frataxin levels (Martelli, Napierala et al. 2012).

The most interesting cells to study remain neurons and cardiomyocytes, which are particularly affected in FRDA, but not accessible from patients. The poor availability of such cell types can be circumvented by generation of induced pluripotent stem cells (iPSCs) from primary FRDA fibroblasts. This process re-programs the fibroblasts into embryonic-like cells able to be differentiated into FRDA pathological tissues such as neurones and cardiomyocytes. These FRDA iPSC models have demonstrated retention of the pathological GAA/TTC repeat

expansions and decreased levels of *FXN* mRNA and frataxin protein compared to controls (Martelli, Napierala et al. 2012). Neurons and cardiomyocytes derived from FRDA iPSCs present different levels of disruption of mitochondrial homeostasis; these data parallel the mitochondrial damage found in FRDA patients and mouse models (Puccio, Simon et al. 2001). Drawbacks with the iPSC models include lack of specificity of the differentiation protocols (the yield of differentiation into cardiomyocytes is less than 1%). They also do not allow for observations of FRDA pathophysiology on the whole-organism level. However, a major benefit to using iPSC cells is that they exhibit FRDA-like pathological phenotypes as well as the GAA/TTC expansions allowing for study of both the somatic instability and expansion dynamics in relation to a pathological phenotype.

#### *3.1.2.4 The Bur-0 repeat expansion as a model for FRDA*

Given the absence of a natural occurrence of a repeat expansion in model organisms, the Bur-0 system provides an attractive avenue to explore similarities. Given the genetic power and tool kit in *Arabidopsis,* this system opens up novel avenues for exploring the mechanisms at the fundamental level, which may be of relevance to several human genetic diseases that occur due to repeat expansions. The repeat expansion in Bur-0 is the same GAA/TTC repeat expansion. Although the Bur-0 accession contains a TTC/GAA expansion in a different gene to FRDA and in the opposite direction when compared to FRDA, the nature of the expansion itself, genic location as well as similar effects at the basic molecular level influencing gene expression suggests that this system can be exploited to understand the fundamental aspects of FRDA mutation. Since there were fundamental similarities, in this chapter, we explored whether additional similarities could be found between the triplet expansions in plants and what is known in human diseases.

### 3.1.3 Genetic anticipation in plants

Genetic anticipation is defined by increasing severity and earlier onset of a disease as it is inherited through consecutive generations. Genetic anticipation is commonly associated with dynamic mutations in DNA (Carpenter 1994, Ashley and Warren 1995, Delatycki, Williamson et al. 2000, Kovtun and McMurray 2008, Evans-Galea, Lockhart et al. 2014). Unlike static mutations, which are retained in somatic tissues and stably transmitted to offspring, triplet repeat expansions are dynamic and continue to mutate within tissues and across generations. Longer tracts are more likely to undergo an expansion mutation than shorter tracts. As repeat tract length correlates with disease severity and age of onset, this phenomenon leads to genetic anticipation, a hallmark of most TNRDs.

One hypothesis is that as the tract length increases, the integrity of the repeating sequence decreases and when a certain threshold length is reached, the intermolecular properties of the repeating sequence change and that local region is able to slip out of its normal double helical structure and form different, more thermodynamically favourable, local DNA secondary structures (Sakamoto, Chastain et al. 1999, Schweitzer and Livingston 1999, Kennedy, Evans et al. 2003, Ruggiero and Topal 2004, Mirkin 2007, Liu and Leffak 2012). When this happens it may interfere with normal DNA replication as cells possess a plethora of mechanisms that function to protect against DNA abnormalities (Schweitzer and Livingston 1997, Ura and Hayes 2002, Christmann, Tomicic et al. 2003, Jiricny 2006, Wyman and Kanaar 2006, Du, Campau et al. 2012, Ezzatizadeh, Pinto et al. 2012, Kim and Wilson 2012, Budden and Bowden 2013). There are DNA repair systems that specifically recognise these mutations in DNA. Their influences and attempts to resolve the abnormal structures may be futile and their disordered affects may in fact induce the repeats to expand further.

In order to have a deleterious effect, the number of repeats must cross a certain threshold length, which differs amongst TNRDs and depends on the nature of the repeating nucleotides (Delatycki, Williamson et al. 2000, McMurray 2010). In TNRDs, the increased disease severity and early onset of symptoms (genetic

anticipation) is directly correlated with expansion length. That is, genetic anticipation is largely due to a further expansion in the next generation. However, that is not always the case and in some instances the genetic anticipation observed in successive generations, is not associated with a greater repeat expansion (Delatycki and Corben 2012, Evans-Galea, Carrodus et al. 2012). In such instances there must be other mechanisms causing the genetic anticipation e.g. inheritance of epigenetic abnormalities such as aberrant methylation, this could be known as 'epigenetic anticipation' (with the definition of: Increasing severity and earlier onset of a disease as it is inherited through consecutive generations due to epigenetic factors). The inheritance of epigenetic changes associated with TNR mutant alleles as they are passed over generations is an area, to my knowledge unexplored.

The mechanisms that govern genetic anticipation are thus far unknown (Mirkin 2007). If the phenomenon of genetic anticipation in FRDA is due to the affects of the intrinsic nature of the GAA/TTC repeat expansion itself, 27 °C Bur-0 would be an ideal model to study all levels of 'genetic anticipation', given the shorter generational times in *Arabidopsis*. However, in order to investigate why and how, on the molecular level, the GAA/TTC expansions are influencing genetic anticipation we first need to determine if genetic anticipation is actually occurring in the Bur-0 model. If so, the Bur-0 model will hold exciting potential for research into genetic anticipation and TNRDs in general.

### 3.1.4 DNA repair pathways contribute to TNR instability

In light of the mounting evidence of the roles various DNA repair pathways play in triplet repeat instability in many TNRDs, we set out to investigate the involvement of DNA repair mechanisms in TNR instability in the Bur-0 model. DNA is constantly at risk of mutation by endogenous stressors such as reactive oxygen species and alkylating agents as well as exogenous stressors such as UV light (Christmann, Tomicic et al. 2003). Organisms have evolved DNA repair

mechanisms to minimize mutations and preserve the integrity of their genetic material (Christmann, Tomicic et al. 2003). DNA repair pathways are highly conserved from prokaryotes to humans, signifying the importance these genes played during evolution (Ura and Hayes 2002) and as such many of the DNA repair genes found in *A. thaliana* have orthologous counterparts in mice and humans.

In humans there are 5 main DNA repair pathways that have been described. They are; the Base Excision Repair pathway (BER), which removes short oligonucleotides that have been damaged by ionizing radiation, alkylating agents or oxidation (Kim and Wilson 2012), the Nucleotide Excision Repair pathway (NER), which removes pyramidine dimers that are caused by UV-B radiation (Budden and Bowden 2013), the MisMatch Repair pathway (MMR), which repairs small base mismatches, and insertions/deletions that occur during DNA replication (Jiricny 2006), the Homologous Recombination Repair pathway (HRR), which repairs double strand breaks by using a homologous DNA template to repair the break caused by ionizing radiation or reactive oxygen species (Wyman and Kanaar 2006) and the Non-Homologous End-Joining pathway (NHEJ), which also repairs double strand breaks, however it does not require an homologous DNA template during repair (Wyman and Kanaar 2006). Out of these DNA repair pathways, only one so far, the MMR pathway, has been shown to be involved in the FRDA GAA/TTC instability (Ezzatizadeh, Pinto et al. 2012). While we attempted to systematically test individual pathways, due to time constrains only the MMR pathway has been explored through transgenics and those results are presented in this chapter.

Overall, this chapter explores the cross-species similarities of the TNR expansion in 27 °C Bur-0 with those in human TNRDs, particularly FRDA. Here we use the Bur-0 model to explore the molecular nature of the *IIL1* TTC/GAA expansion. We first aim to determine whether the phenomenon of genetic anticipation, hallmark in TNRDs, is also occurring in the Bur-0 model. In this chapter, we demonstrate that there are signs of genetic anticipation in the Bur-0 model and we were able to determine a potential threshold value after which the genetic

instability appears to increase resulting in genetic anticipation. We have then explored whether DNA repair pathways are contributing to the TTC/GAA expansion associated phenotypes and show that compromising the mismatch repair pathway suppresses the repeat expansion associated *iil* phenotype in *A. thaliana,* providing a potential pathway which can be considered for further analysis in FRDA.

## 3.2 RESULTS

***Aim 1.*** *To determine if the phenomenon of 'genetic anticipation', hallmark in TNRDs, is also occurring in the Bur-0 model. If so, we aim to determine the repeat threshold for expansion. I.e., the tract length that causes the repeat to become unstable and expand dramatically in the next generation. We also investigate repeat instability in plants exposed at elevated temperature.*

### 3.2.1 Genetic anticipation in *A. thaliana* Bur-0

The Bur-0 repeat expansion $(TTC/GAA)_{400+}$ can spontaneously contract or enlarge further due to the high instability of repeat expansions in general, over generations. Large repeat expansions are highly dynamic mutations with instability increasing as repeat length increases. This instability is usually observed as small increases or decreases in expansion tract length of the same mutant allele when it is inherited over generations as well as within different tissues of a single individual. This instability, which can be observed, is shown in figure 5. This instability has been previously exploited to isolate spontaneous suppressor lines, which allowed the establishment of a series of suppressors with varying TTC/GAA repeats lengths (Sureshkumar, Todesco et al. 2009). Here we take advantage of these natural suppressor lines, which range from ~61 to ~264 repeats, to study the repeat dynamics of TTC/GAA expansion tracts in

Bur-0.

### 3.2.2 Trans-generational increases in repeat lengths are observed in Bur-0 at elevated temperature conditions.

In order to determine if genetic anticipation was occurring, Bur-0 with 400+ repeats and natural suppressors NS1, NS2, NS3, NS4, NS5 and NS6 with repeat tracts of ~ 60, 126, 157, 181, 213 & 264 respectively, were grown over 1 generation for 4 weeks at 27 °C temperature, where the *iil* phenotype is visible. The TTC/GAA repeat expansion at the *IIL1* locus was amplified and analysed through PCR. From these analyses, it can be seen that the repeat lengths in suppressors NS_3 (~157 repeats) to NS_6 (~264 repeats) increased in the second generation but the same trend was not observed in other suppressors with ~126 repeats or less. Genetic anticipation represented as an increase in frequency of progeny with *iil* phenotype, was not observed in Bur-0 plants as all individuals in both generations exhibited the defect phenotype due to the expansion being much larger than the required threshold length (Fig 1).

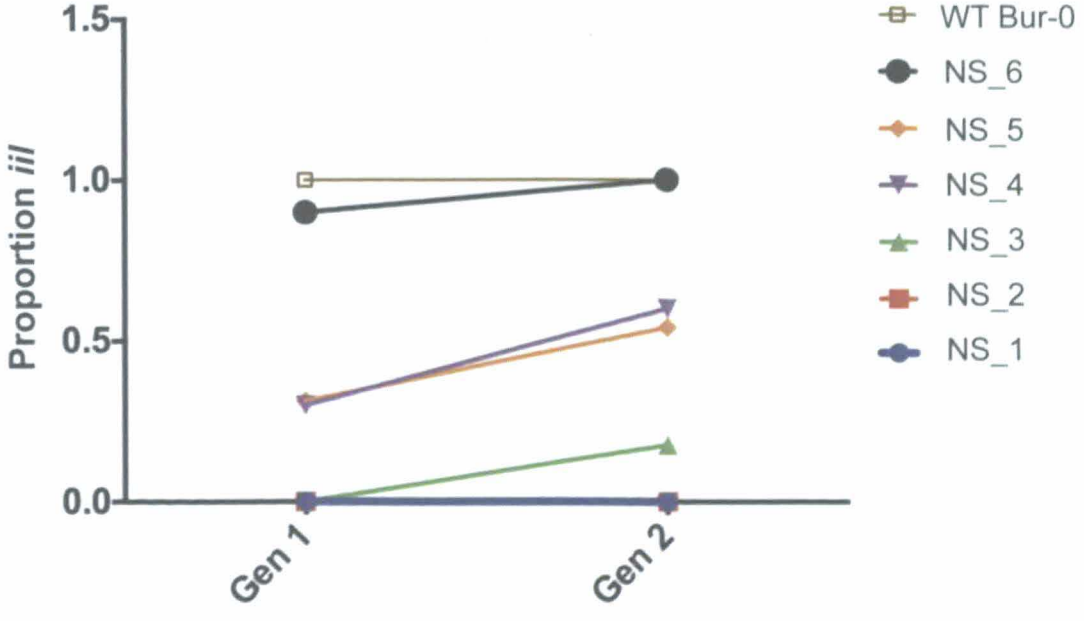


**Figure 1. Proportion of plants displaying the *iil* phenotype at 27 °C across a generation.** Proportion of individuals displaying TNR defect is displayed as % with *iil*. Here natural suppressors (NS) 1 (~61 repeats) and 2 (~126 repeats) (NS_1 & NS_2) displayed no *iil* phenotypes in either generation 1 or 2. NS_3 (~157 repeats) had no individuals display an *iil* phenotype in generation 1, however ~18% (9 out of 51) of individuals in generation 2 displayed a mild *iil* phenotype. NS_4 (~181 repeats) had ~30% (12 out of 40) individuals displaying an *iil* phenotype in generation 1 and ~60% (21 out of 35) displaying an *iil* phenotype in generation 2. NS_5 (~213 repeats) had ~32% (15 out of 48) individuals displaying an *iil* phenotype in generation 1 and ~60% (16 out of 29) displaying an *iil* phenotype in generation 2. NS_6 (~264 repeats) had ~90% (27 out of 30) individuals displaying an *iil* phenotype in generation 1 and 100% (25 out of 25) displaying an *iil* phenotype in generation 2. 100% of Bur-0 (400+ repeats) individuals displayed an *iil* phenotype in both generation 1 and 2.

In terms of repeat tract instability, NS_1 (~61 repeats) did not display repeat instability or any change in tract length across generation 1 or 2 for both 23 °C & 27 °C temperature conditions (Fig 2A & B). There was also no repeat instability detected or distinguishable change in repeat length of generation 1 or 2 NS_2 (~126 repeats) plants when grown at 23 °C. However some minor variability in tract length was observed in generation 1 & 2 in plants grown at elevated temperature (27 °C) (Fig 2C & D). Thus, results indicate that perhaps repeats of ~126 reps begin to lose their integrity at the DNA level when exposed to higher temperatures. Though expansions of that length were not associated with a disease phenotype. NS_3 plants (~157 repeats) display a slightly more unstable repeat tract tending towards expansion in generation 2 (Fig 2E & F). Instability was also observed in both generation 1 and 2 NS_3 plants grown at 27 °C, to a greater extent than NS_2 (Fig 2D & F).

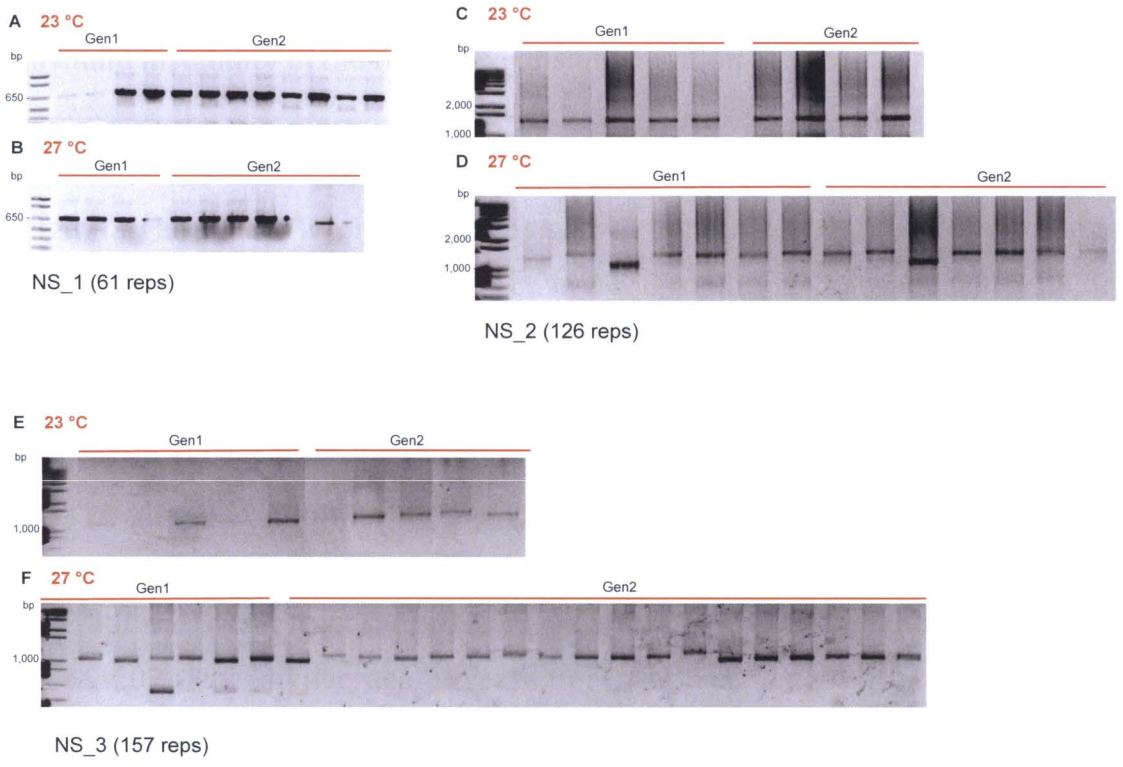



**Figure 2. Mutational dynamics of the TTC/GAA expansion at the *IIL1* locus in NS_1, NS_2 and NS_3 plants in 23 °C & 27 °C conditions. (A)** NS_1 genomic TTC/GAA tract over 2 generations in 23°C. **(B)** NS_1 genomic TTC/GAA tract over 2 generations in plants exposed to 27 °C *iil* phenotype inducing conditions. **(C)** NS_2 genomic TTC/GAA tract over 2 generations in 23 °C. **(D)** NS_2 genomic TTC/GAA tract over 2 generations in plants exposed to 27 °C *iil* phenotype inducing conditions. **(E)** NS_3 genomic TTC/GAA tract over 2 generations in 23 °C. **(F)** NS_3 genomic TTC/GAA tract over 2 generations in plants exposed to 27 °C *iil* phenotype inducing conditions. NS_1 PCR used SKB_608 & 561 primers, thus expected product size for NS_1 = 787 bp. NS_2 used SKB_608 & 609 primers, thus expected product size = 1,471 bp. NS_3 used SKB_608 & 561 primers, thus expected product size = 1,075 bp. (Appendix - Table S2).

NS_1 expected product size = 787 bp. Generation 1 & 2 repeat tracts in both 23 °C and 27 °C conditions did not display variations in tract length. NS_2 showed no variability in generation 1 & 2 in 23 °C conditions and slight variability ranging from ~1,100-1,700 bp. Suggesting repeat tracts of ~2-202, for both generation 1 and 2 in 27 °C conditions. NS_3 showed no variability in generation 1 and slight variability ranging from 1,100-1,600 bp in 27 °C conditions. Suggesting repeat tracts of ~165-332. Generation 1 in 27 °C conditions displayed variation of ~550-1,100 bp suggesting repeat tracts of ~0-165. Generation 2 in 27 °C conditions displayed variation of ~900-1,400 bp, suggesting repeat tracts

of ~99-265. Thus in Bur-0 NS with <120 TTC/GAA repeats there seems to be only mild instability in the repeat expansion mutation observed.

However, NS_4 with ~ 181 repeats, displayed a clear increase in genomic repeat length in generation 2 for both plants grown in both 23 °C & 27 °C conditions (Fig 3). NS_4 generation 1 repeat tract, in 23 °C conditions, ranged from ~1,630-1,650 bp, suggesting repeat tracts of ~179-183. Generation 2 repeat tract ranged from ~1,700-2,000 bp, suggesting repeat tracts of ~202-302. Thus an average increase in repeat tract length by ~ 70 repeats in the second generation was observed in 23 °C growth conditions. In 27 °C conditions, generation 1 repeat tract ranged from ~1,200-1,550 bp. Suggesting repeat tracts of ~36-153. Generation 2 repeat tract ranged from ~1,300-1,700 bp. Suggesting repeat tracts of ~69-202. Thus an average increase in repeat tract length by ~ 40 repeats in the second generation was observed in 27 °C growth conditions. Plants in 27 °C conditions displayed grater variability in their expansion tract length than in 23 °C.

Therefore our results indicate that a repeat tract of >150<180 appears to be the threshold length at which point the repeat expansions become considerably unstable and tend toward expansion when the allele is passed on to the next generation. This identified repeat threshold length for expansion correlates with the sudden increase in frequency of individuals exhibiting the *iil* phenotype in Figure 1.

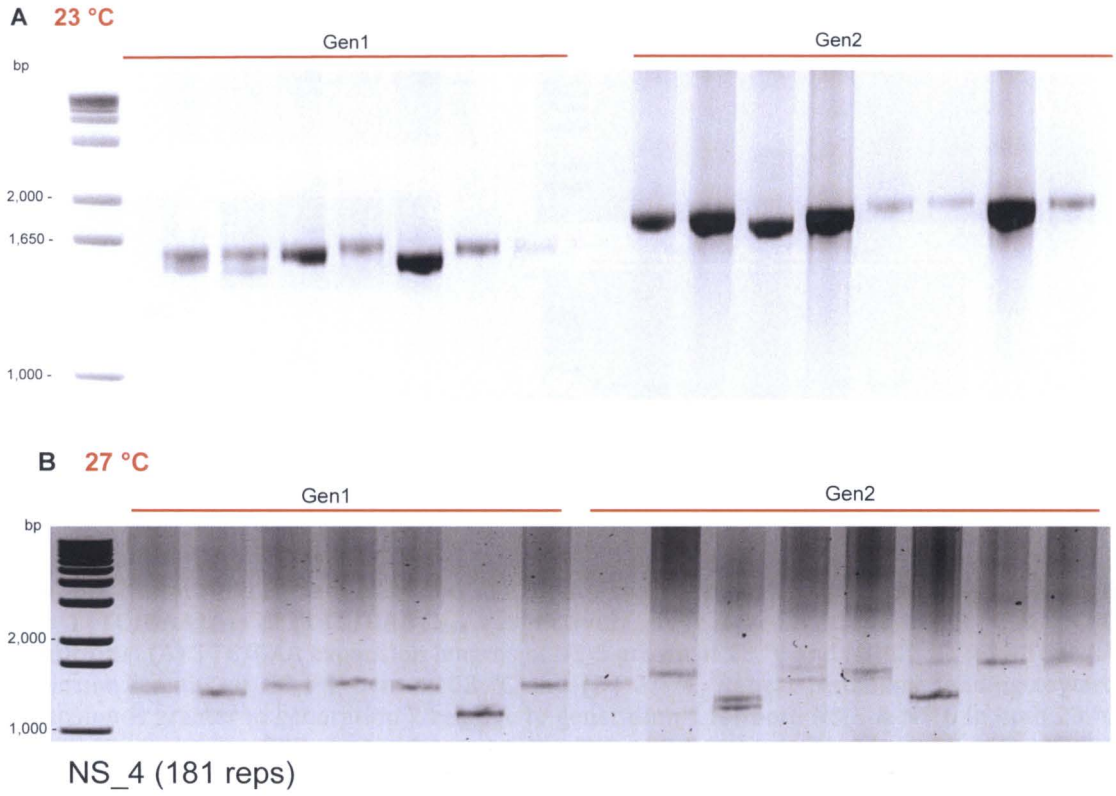



**Figure 3. Repeat expansion dynamics in generations 1 & 2, for natural suppressor 4 (NS_4) (TTC/GAA)$_{\sim 181}$.** In **(A)** 23 °C & **(B)** 27 °C growth conditions. Repeat tract displays increased instability, tending towards expansion over contraction in generation 2 (Gen2) in both 23 °C & 27 °C conditions. [NB: the higher bands at ~3 & 4 Kb in generation 2 individuals could potentially represent dramatic expansions]. PCR was performed using SKB_608 &609 primers (Appendix - Table S2). Thus expected product size = 1,636 bp.

NS_5 & 6 also displayed an increase in repeat instability, tending towards expansion in generation 2, relative to generation 1, as well as increased repeat tract instability in general, potentially more so in 27 °C conditions (Fig 4).

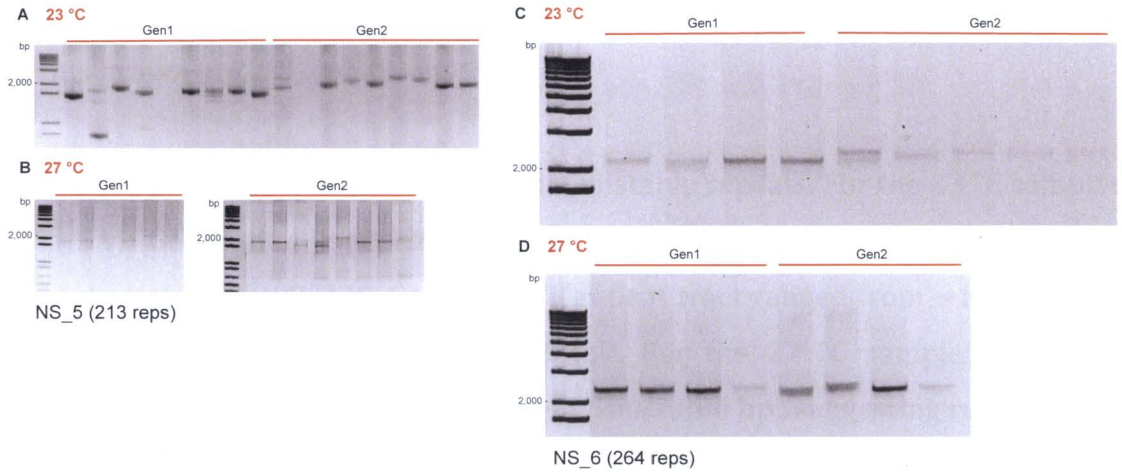



**Figure 4. Mutational dynamics of the TTC/GAA expansion at the *IIL1* locus of NS_5 & NS_6 with (TTC/GAA)$_{\sim 213}$ & (TTC/GAA)$_{\sim 264}$ respectively, over 2 generations in 23 °C & 27 °C conditions. (A)** TTC/GAA expansion lengths of NS_5 grown at 23 °C and **(B)** 27 °C. **(C)** TTC/GAA expansion lengths of NS_6 grown at 23 °C and **(D)** 27 °C. Repeat instability tending toward expansion is greater in generation 2 relative to generation 1 for both NS_5 & NS_6 in both 23 °C and 27 °C conditions. PCR was performed using SKB_608 &609 primers (Appendix - Table S2). Expected PCR product size for NS_5 = 1,732 bp. Expected product size for NS_6 = 1,885 bp.

For NS_5, the generation 1 repeat tract in 23 °C conditions ranged from ~1,600-1,900 bp, suggesting repeat tracts of ~170-269. Generation 2 repeat tract ranged from ~1,650-2,200 bp, suggesting repeat tracts of ~186-370. For the 27 °C samples, generation 1 repeat tract ranged from ~1,600-2,200 bp, suggesting repeat tracts of ~169-370. Generation 2 repeat tract ranged from ~1,500-2,300 bp, suggesting repeat tracts of ~136-276. Thus the average repeat length increased by 58 from approximately 220 to 278 repeats in 23 °C conditions. A similar level of instability was observed between generation 1 and 2 for NS_5 plants grown in 27 °C conditions.

For NS_6, the generation 1 repeat tract in 23 °C conditions ranged from ~2,000-2,200 bp, suggesting repeat tracts of ~302-369. Generation 2 repeat tract ranged from ~2,100-2,400 bp, suggesting repeat tracts of ~335-436. For the 27 °C samples, generation 1 repeat tract ranged from ~2,400-2,500 bp, suggesting repeat tracts of ~436-469. Generation 2 repeat tract ranged from ~2,200-2,600 bp, suggesting repeat tracts of ~369-502. Thus the average repeat length increased by 50 from approximately 336 to 386 repeats in 23 °C conditions. A

similar level of instability was observed between generation 1 and 2 for NS_6 plants grown in 27 °C conditions.

Bur-0 (400+ repeats) displayed the most unstable repeats. For the 23 °C samples (Fig 5A): Generation 1 repeat tract ranged from ~2,000-2,100 bp, suggesting repeat tracts of ~302-336. Generation 2 repeat tract ranged from ~1,400-2,200 bp, suggesting repeat tracts of ~102-339. For the 27 °C samples (Fig 5B): Generation 1 repeat tract ranged from ~1,500-2,800 bp., suggesting repeat tracts of ~136-570. Generation 2 repeat tract ranged from ~1,400-2,500 bp, suggesting repeat tracts of ~102-470. Thus a change in expansion instability was not clear between generations (1 & 2) in both 23 °C & 27 °C conditions (Fig 5), potentially due to the instability being so high in both instances. This high level of instability may be due to the intrinsic nature of larger expansions. However, a dramatic increase in repeat instability was observed in plants from 27 °C conditions relative to 23 °C conditions, regardless of the generation (1 or 2) (Fig 5B).

Hence these results indicate that there is greater instability tending towards expansion for Bur-0 plants exposed to the *iil* phenotype inducing conditions of elevated temperature. This suggests a direct role of elevated temperature influencing repeat instability favoring expansion.

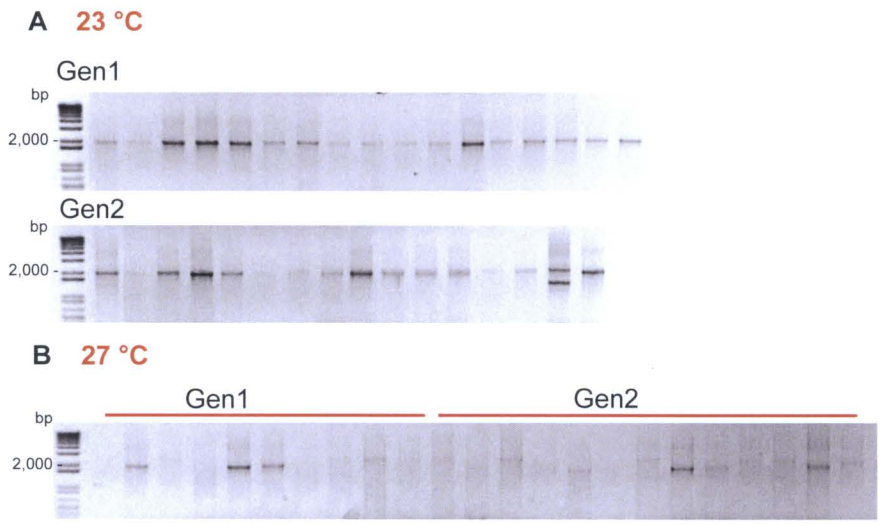


**Figure 5. Mutational dynamics of the TTC/GAA expansion at the *IIL1* locus of Bur-0 with $(TTC/GAA)_{400+}$ over 2 generations in 23 °C & 27 °C conditions. (A)** TTC/GAA expansion lengths of Bur-0 grown at 23 °C. **(B)** TTC/GAA expansion lengths of Bur-0 exposed to the *iil* phenotype-inducing conditions of 27 °C. PCR used SKB_608 & 609 primers, thus expected product size = 2,293 bp. (Appendix - Table S2).

Our results also show a direct correlation between TTC/GAA tract length and instability (Fig 6). The range of variability in expansion lengths for Bur-0 and each of the Bur-0 NSs across generations 1 & 2 for both 23 °C & 27 °C conditions is displayed in Figure 6. An increase in instability favoring expansion is displayed for strains with >126 reps in 23 °C and for strains with >60 repeats in 27 °C. The expansion instability is much higher across both generations, for plants exposed to 27 °C elevated temperature conditions and thus a strong correlation of increased repeat length in the second generation was lost among strains with larger expansions and indeed not observed in Bur-0. However, a direct correlation in level of further repeat tract expansion across the generation was observed in strains with >120 repeats in 23 °C. These results of observed increase in expansion tract length in the second generation correlate with the observed genetic anticipation in Figure 1. Thus the genetic anticipation observed in the plants can be attributed, at least in part, to an increase in expansion tract length.

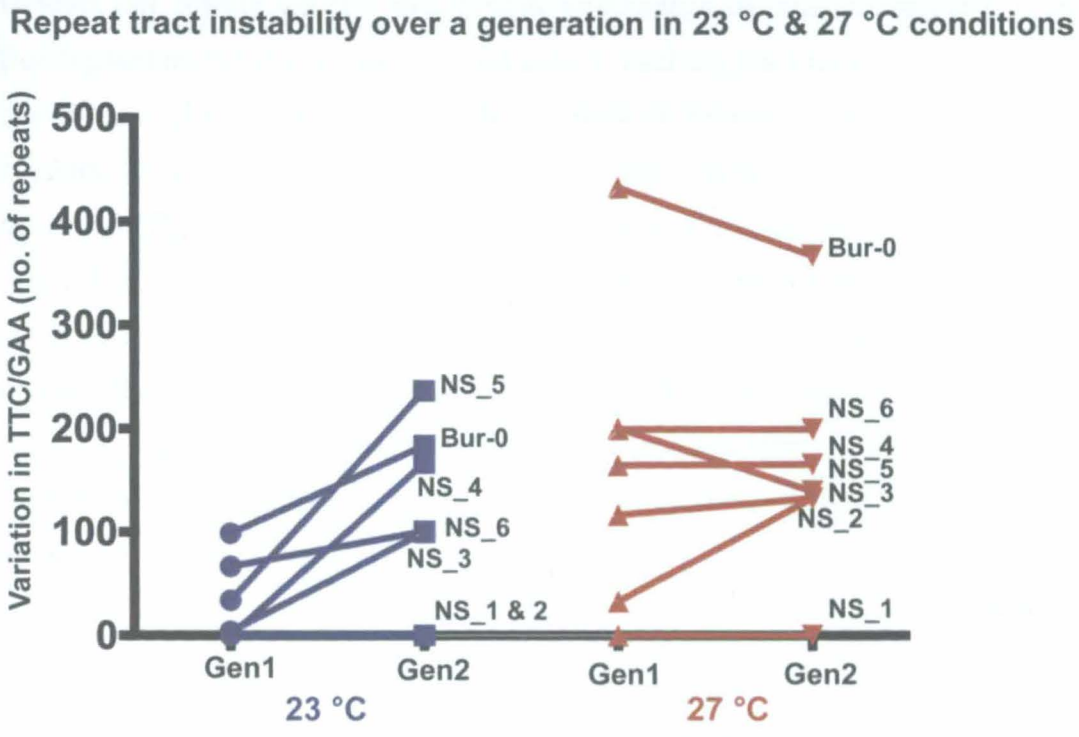



**Figure 6. Mutational dynamics of TTC/GAA expansion over a generation in 23 °C & 27 °C conditions.** Here repeat tract instability was measured as the average expansion tract lengths as determined through PCR analysis, among individual plants for each generation. For most strains at 23 °C with >120 repeats, there is a distinct increase in repeat tract length in the second generation, with a direct correlation of repeat length and level of further expansion in generation 2.

| TNRD | Gene | Codon | WT | Pathogenic |
|---|---|---|---|---|
| Fragile X syndrome | *FMR1* | CGG/CCG | 6-53 | 230+ |
| Myotonic dystrophy | *DMPK* | CTG/CAG | 5-37 | 50+ |
| Spinocerebellar ataxia | *SCA8* | CTG/CAG | 16-37 | 110-250 |
| Friedreich's ataxia | *FXN* | GAA/TTC | 7-34 | 100+ |
| Iregularly impaired leaves (*iil*) phenotype | *IIL1* | TTC/GAA | 0-35 | ~180+ |

**Table 1. Comparison of experimentally determined threshold for expansion and *iil* disease pathology in Bur-0 plants with various non-coding human TNRDs including FRDA.** Here WT indicates the normal range in repeat length. 0-35 repeats in Bur-0 was determined by Sureshkumar et al (Sureshkumar, Todesco et al. 2009) through a global analysis of >95 *A. thaliana* accessions. *iil* pathogenic range was determined in this study on genetic anticipation. Note the similarities between the WT range in repeat length and pathogenic repeat length between disorders.

Overall, our results indicate that genetic anticipation is indeed observed in the Bur-0 plant model of a TNR expansion defect. That the tract length for expansion and disease phenotype is a pre-mutation allele of between >157<181 TTC/GAA repeats. This is the tract length at which point genetic anticipation is first observed (Fig 1) and substantial instability observed in the genomic expansion (Fig 2 & 3). Our results also find that presence of the disease phenotype (27 °C) is directly correlated with level of repeat instability in expanded alleles of greater than 157 TTC/GAA repeats. Thus, regarding the expansion mutational dynamic impact on disorder phenotype/symptoms, striking similarities between human non-coding TNRDs and the Bur-0 TNR expansion defect were observed (Table 1).

### 3.2.3 DNA MMR pathway is involved in contributing to Bur-0 *iil* TNR expansion associated growth defect

***Aim 2.*** *To investigate whether DNA repair pathways influence the TTC/GAA expansion tract length at the IIL1 locus and if so, whether the DNA repair pathways are influencing TNR instability and IIL1 expression.*

To analyse the effect of the mismatch repair pathway on the *iil* phenotype, we generated two different artificial microRNAs against *MSH2* (At3g18524) under the *35S* CaMV promoter *(35S::amiRMSH2-C1* and *35S::amiRMSH2-C2*) and transformed them into the Bur-0 plants. 11 T1 plants with 7 independent T1 derived from 35S::*amiR-MSH2-C1* and 4 T1 derived from 35S::*amiR-MSH2-C2* were analysed at 27 °C compared to the untransformed Bur-0 plants. The T1s germinated and grown in 27 °C displayed a delayed onset of the *iil* phenotype. *iil* phenotypic growth defects begin to appear at ~2.5-3 weeks post germination in 27 °C Bur-0. The *35S::amiR-MSH2* T1s in the Bur-0 background consistently failed to display the *iil* phenotype until 5-6 weeks post germination in 27 °C. Figure 7A-C display the distinctive *iil* phenotype in the Bur-0 controls at 4-weeks

and absence of the *iil* phenotype in the 35S::*amiR-MSH2* T1s, 1 and 2, at 4-weeks respectively.

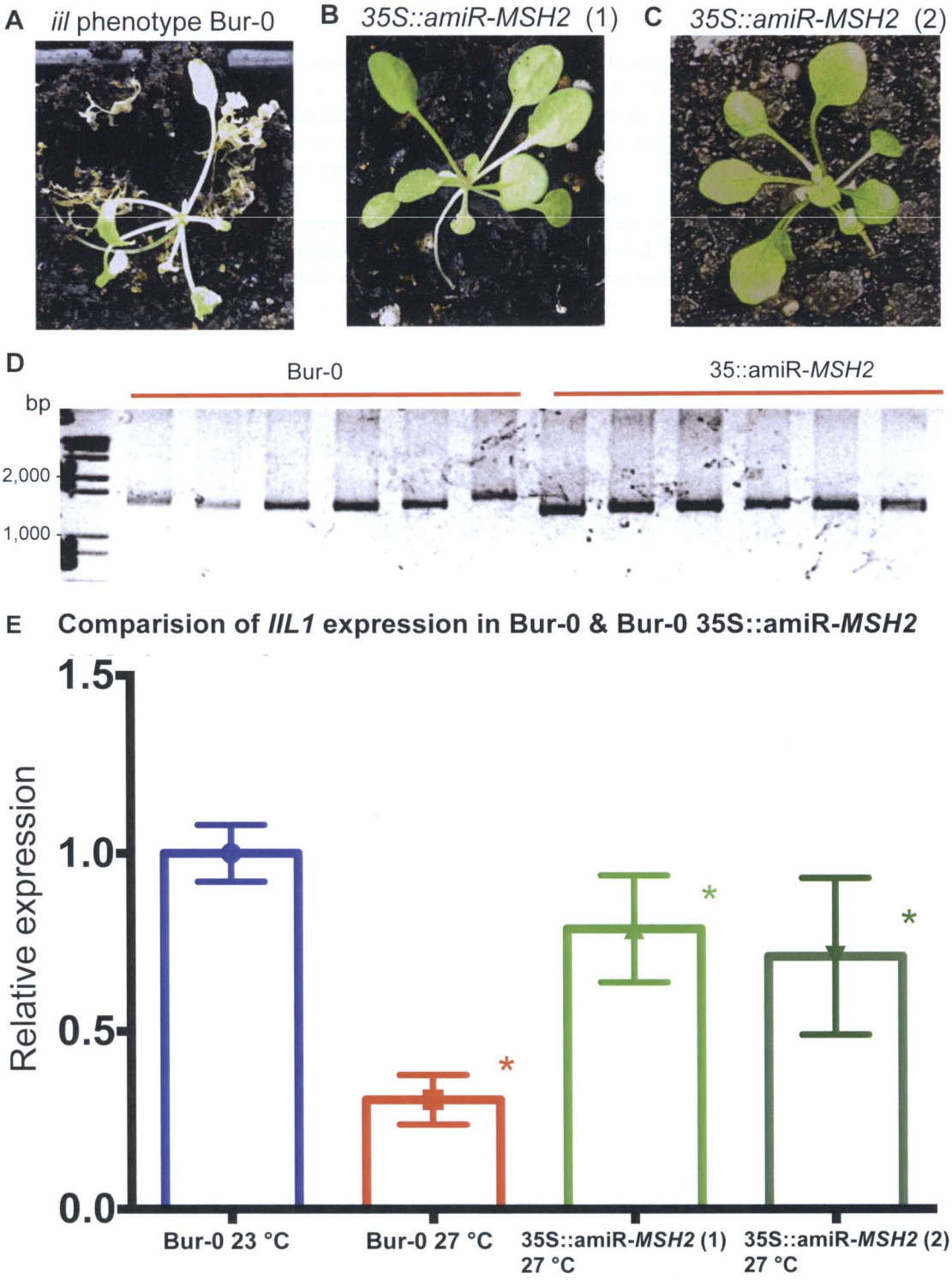


**Figure 7. Analysis of T1 *35S::amiR-MSH2* impacts on the *irregularly impaired leaves* (*iil*) defect.** **(A)** Representative individual of 4-week old post germination Bur-0 control plant at 27 °C. Plant is displaying the twisted, tapered leaf phenotypic growth defects characteristic in the *iil* defect. **(B)** Representative individual of 4-week old post germination 35S::*amiR-MSH2-C1* T1 line at 27 °C. **(C)** Representative individual of 4-week old post germination 35S::*amiR-MSH2-C2* T1 line at 27 °C. Note T1 with both amiRNA constructs display an identical phenotype. **(D)** PCR of genomic TTC/GAA expansion in 6 biological replicate Bur-0 control plants at 27 °C, and 6 samples from 35S::*amiR-MSH2* T1 transformants, comprised of 3 biological replicates from construct 1 followed by 3 biological replicates from construct 2 resectively, at 27 °C. PCR utilized SKB_608 & 561 primers (Appendix - Table S2). **(E)** q-RT-PCR analysis of *IIL1* gene expression levels in 4-week old Bur-0 controls grown in 23 °C conditions, Bur-0 controls grown in 27 °C conditions and Bur-0 35S::*amiR-MSH2-C1 or C2* T1 lines grown in 27 °C (note expression results for both 35S::amiR-*MSH2* constructs were not significantly different). *IIL1* expression levels were normalized to *TUB2* reference gene expression (Appendix - Table S2). Results displayed as expression relative to Bur-0 23 °C control. * (Asterisks) indicates significant change in expression levels as determined by students *t*-test with cut-off *p*-value <0.05. Red asterisk indicates significant decrease in *IIL1* expression (*p*-value 0.0383) in Bur-0 27 °C relative to the Bur-0 23°C. Green asterisks indicates significant increases in *IIL1* expression (*p*-values 0.0391 & 0.0487 respectively for C1 & C2) in Bur-0 35S::amiR-*MSH2* T1 27 °C relative to Bur-0 27°C. There was no significant difference in *IIL1* expression between Bur-0 23 °C and either 35S::*amiR-MSH2* T1 construct at 27 °C. q-RT-PCR primers used as in Appendix - Table S2.

In order to analyse the impacts of the *35S::amiR-MSH2* T1 on TTC/GAA expansion dynamics and *IIL1* expression, tissue was collected from 4-week old plants and both DNA and RNA extracted. PCRs were performed on DNA from 6 Bur-0 and 6 35S::*amiR-MSH2* T1 lines (3 from each independent amiR construct) plants (Fig 7D). There is no obvious difference in TTC/GAA expansion length between the different genotypes. However, there is potentially, slightly more instability in the repeat expansion in Bur-0 and it appears that compromising *MSH2* potentially reduces repeat instability (Fig 7D).

Extracted RNA from the 4-week old plants was converted into cDNA and *IIL1* expression levels analysed by q-RT-PCR (Fig 7E). A significant decrease in *IIL1* expression in Bur-0 27 °C plants relative to Bur-0 23 °C plants (*p*-value 0.0383) can be seen, with Bur-0 27 °C plants expression approximately 30% of Bur-0 23 °C *IIL1*. The 4-week old *35S::amiR-MSH2* T1 plants, that did not yet display an *iil* phenotype, exhibited an *IIL1* expression level similar to that of Bur-0 23 °C. However, the *IIL1* expression level between Bur-0 and *35S::amiR-MSH2* constructs in the Bur-0 background at 27 °C was significantly different (*p*-value s 0.0391 & 0.0487 respectively for C1 & C2) with the *35S::amiR-MSH2* plants expressing approximately 2.5-fold increase in *IIL1* expressions each.

Our results indicate that absence of MSH2 protein leads to a significant increases in Bur-0 27 °C *IIL1* expression in 4-week old plants, comparable to expression level observed at 23 °C, where the *iil* phenotype is not seen in *A. thaliana*. We also found a potential decrease in the level of genomic TTC/GAA expansion instability in *MSH2* mutants. The expansion was still present in plants carrying the artificial microRNAs against MSH2 (Fig 7D) indicating they are exhibiting a milder disease phenotype in spite of the expansion.

## 3.3 DISCUSSION

### 3.3.1 Genetic anticipation is occurring in *A. thaliana*

Our findings suggest that genetic anticipation is occurring in *A. thaliana* as plants with ≥157 TTC/GAA repeats exhibited increased disease severity and earlier onset of *iil* phenotype in generation 2 (Fig 1). Genetic anticipation is difficult to observe in the original Bur-0 strain as all individuals display the disease phenotype thus a difference in the phenotype is difficult to distinguish between the generations. However, by taking advantage of the natural suppressors we were able to show that there was an increase in disease severity over the generation. However, this was only observed in plants with ≥157 repeats (first observed in NS_3). This suggests that a repeat tract of <150 is not sufficient to cause the defect phenotype in any generation, consistent with earlier analysis (Sureshkumar, Todesco et al. 2009). All plants with >264 TTC/GAA repeats exhibited the *iil* phenotype in both generations and so genetic anticipation was not clearly observed at the phenotypic level.

### 3.3.2 Determining threshold TTC/GAA tract length for expansion

In terms of increased disease severity over the generations, there was an ~18% increase in plants displaying the *iil* phenotype in generation 2 of NS_3 plants, with no plants displaying the *iil* phenotype in generation 1 and 9/51 displaying a mild phenotype in generation 2. This together with the mild instability observed in NS_2 plants genomic tract length (in the 27 °C conditions) (Fig 2D) indicates that around 126-150 TTC/GAA repeats may be the tract length at which point the repeats begin to lose their integrity and thus could be considered an allelic pre-mutation length for the *IIL1* TNR expansion disorder.

In NS_4, with ~180 repeats, there is a considerable increase in plants exhibiting the defect phenotype in the second generation, with ~30% (12/40) individuals displaying an *iil* phenotype in generation 1 and ~60% (21/35) displaying an *iil* phenotype in generation 2 (Fig 1). This is a 100% increase in level of disease severity in the second generation. Repeat tracts of ~180 TTC/GAA were also associated with a distinct increase in repeat instability tending towards expansion, at the DNA level (Fig 3). This indicates that >150<180 is the threshold tract length for expansion and also the threshold length at which point the disease symptoms emerge.

Plants with longer repeat tracts are associated with a greater frequency of individuals displaying the *iil* phenotype. This trend suggests a direct correlation with genetic anticipation and TTC/GAA tract length. That is, the longer the initial repeat in generation 1 the greater the extent of genetic anticipation in generation 2. However, as the tract length approaches around 400 as seen in Bur-0 the majority of individuals in generation 1 exhibit the *iil* phenotype and thus a defect saturation affect is observed and genetic anticipation unobservable in Bur-0 and suppressors in which expansions are beyond the threshold length. Regardless, based on the observation of disorder phenotype in NS plants over the generations it does appear that a similar phenomenon of genetic anticipation is in fact occurring (Fig 1).

From repeat >126 (NS_2) to 400+ (Bur-0) there is a direct correlation between repeat tract length and repeat instability in with or without exposure to high temperature. That is, as repeat length increases, so does the repeat instability. This is an observation also observed in FRDA (Koeppen 2011) and many other TNRDs (McMurray 2010, Liu and Leffak 2012). The repeat instability in 27 °C is much greater than at lower temperature conditions (Fig 5). The *iil* phenotype only appears in elevated temperatures (27 °C) thus the presence of the defect phenotype appears to be correlated with repeat instability.

### 3.3.3 *A. thaliana iil* TNR expansion defect exhibits similarities to FRDA

Our results indicate striking similarities between the Bur-0 TNR expansion defect and observations found in non-coding human TNRDs such as Fragile X syndrome, Spinocerebellar ataxia and Friedreich's ataxia (Mirkin 2005, Mirkin 2007). Given these similarities in the repeat dynamics when inherited across generations and observation of genetic anticipation Bur-0 could well be used as a model for in depth investigation into the underlying mechanisms governing genetic anticipation at all levels.

### 3.3.4 Involvement of the DNA MMR pathway in Bur-0 TNR expansion defect

We have shown that compromising *MSH2* activity suppresses the TTC/GAA repeat expansion associated phenotypes in *A. thaliana.* Unlike the Bur-0 plants, 35S::*amiR-MSH2* plants displayed a delayed onset and milder form of the characteristic *iil* phenotypic growth defects associated with Bur-0. This milder and delayed *iil* onset was consistent across all *35S::amiR-MSH2* transformants, suggesting that the phenotypic consequences in those transformed plants were due to the influence of the *amiRMSH2* itself and not due to insertional disruption of another gene, which can potentially happen during *A. tumefaciens* transformation (Sugui, Chang et al. 2005).

An hypothesis as to why *35S::amiR-MSH2* plants display a delayed/milder phenotype could be due to the possibility of MSH2 proteins stabilizing the expansion-induced secondary structures. The function of DNA MMR pathway MSH2 and MSH3 proteins is to recognize and bind mismatched DNA base pairs as discussed above. As large TNR expansions are known to misalign and form hairpin secondary structures with many mismatched base pairs, it is plausible that the MSH2 proteins may recognize and bind to the TTC/GAA expansion thereby stabiliziging their local secondary structure. This would contribute to the transcriptional downregulation as RNA polymerase can only extend over and transcribe DNA in the appropriate B-form structure. This is consistent with our observations of elevated *IIL1* expression in *35S::amiR-MSH2* plants in spite of the genomic expansion (Fig 7D & E).

DNA repair is an essential component of every organism and accordingly is highly conserved across all species (Christmann, Tomicic et al. 2003). Our findings in *A. thaliana* thus are relevant to the human situation as DNA repair has already been shown to affect TNR instability and gene expression in mice and other organism models of various TNRDs including FRDA (Gonitel, Moffitt et al. 2008, Ezzatizadeh, Pinto et al. 2012). Of the various DNA repair mechanisms, the DNA mismatch repair (MMR) pathway has been shown to contribute to the pathology of FRDA in various models (Ezzatizadeh, Pinto et al. 2012). Our findings thus suggest that highly conserved fundamental pathways modulate repeat expansion associated effects in diverse species arguing for the use of *A. thaliana* as a suitable model for further defining the molecular mechanisms.

### 3.3.5 Future directions

Although only the DNA MMR pathway has been found to contribute to FRDA pathology in various models, other TNRDs have also found additional DNA repair pathways to be involved in the disorder pathology, thus a future direction would be to analyse the effects of all orthologous DNA repair pathways in the *A.*

*thaliana* model. While we have analysed only *amiRMSH2* in this chapter, we have constructed artificial microRNAs against several genes indispensible in 3 other DNA repair pathways; the Homologous recombination repair pathway (HRR) (genes *RAD54* and *ATM*), the Non-homologous end-joining pathway (NHEJ) (genes *KU70* and *KU80*), the Nucleotide excision repair pathway (NER) (gene *UVH1*), as well as the *MSH1* gene in the Mismatch repair pathway (MMR). These constructs could also be transformed into *A. thaliana* and a similar analysis to that presented here on the MMR pathway undertaken. Growing T2 and potentially T3 lines would allow analysis of the inter-generational TNR instability in absence of the MSH2 protein. Thus the transgenic lines generated through this study can have many future applications investigating the mechanisms of the MMR pathway in the Bur-0 TNR expansion defect and *MSH2* affects on TNR instability.

## 3.4 CONCLUSIONS

In conclusion we have shown that the Bur-0 triplet expansion displays trans-generational dynamics similar to repeat expansions observed in humans and it appears that the DNA MMR pathway protein MSH2 is contributing to the increased severity of the Bur-0 TNR expansion defect at 27 °C. The phenotypic consequences and affects on gene expression in *35S::amiRMSH2* plants are consistent with findings in other FRDA model systems. Thus we can draw another molecular similarity between FRDA and the *A. thaliana iil*. This provides a platform for study into the molecular mechanisms through which DNA repair pathways influence TNR expansion pathology. In the next chapter, I explore these similarities to further identify molecular mechanisms and potential targets for effective management of repeat expansion associated genetic defects.

## 3.5 MATERIALS & METHODS

### 3.5.1 Tissue samples, DNA & RNA extractions

Tissue from basal rosette leaves of 4-week old plants were collected in 1.5 ml microfuge tubes, snap frozen in liquid $N_2$ then finely ground using a mortar and pestle. DNA extractions, RNA extractions, cDNA synthesis and qPCR reactions were carried out as decribed in chapter 1.

### 3.5.2 Genetic anticipation analysis protocol

*A. thaliana* lines: Bur-0 with $(TTC.GAA)_{400+}$, NS_1 with $(TTC.GAA)_{61}$, NS_2 $(TTC.GAA)_{126}$, NS_3 $(TTC.GAA)_{157}$, NS_4 $(TTC.GAA)_{181}$, NS_5 $(TTC.GAA)_{213}$, NS_6 $(TTC.GAA)_{264}$ were used in this study. These repeat numbers were determined earlier in the lab through a combination of southern blots and PCRs (Bernal and Balasubramanian, personal communication).

Two flats each containing ~30 individual plants of each *A. thaliana* strain (Bur-0 & the 6 NS) were sown in soil and one of each flat maintained under 23 °C short day (SD) conditions and the other under 27 °C SD conditions in controlled temperature growth rooms. The flats maintained in both 23 °C SD and 27 °C SD were kept post germination for 4-weeks then transferred to 23 °C Long Day (LD) conditions.

Maintaining the plants from germination to 4 weeks is sufficient time to induce the *iil* growth defect phenotype in Bur-0 plants. Under the 27 °C temperature conditions Bur-0 plants are unable to flower and set viable seeds. In order to study the influence of TTC/GAA instability over the generations we need to be able to collect seeds. Therefore, after 4 weeks of growth at 27 °C the plants were transferred to 23 °C LD to recover and set seeds. If Bur-0 plants are maintained

under 27 °C conditions for >4 weeks they are generally unable to recover from the growth defects and do not recover normal growth. Therefore plants were exposed to the *iil* TNR expansion defect phenotype-inducing conditions for the maximum time possible. This time period is long enough to induce the *iil* phenotype (Fig 7A).

4 weeks post germination tissue samples were collected from individual plants and gDNA TTC/GAA expansion determined by PCR. All plants were left in 23 °C LD conditions until they set seeds. The mature seeds were then collected and kept for ~2 weeks in a dehumidifying chamber. Those seeds were designated generation 2 (Gen2). Once Gen2 seeds were obtained from all *A. thaliana* lines (Bur-0 & the 6 NS) from both temperature conditions, they were stratified and again planted as for Gen1. Tissue samples from Gen2 plants were collected after 4 weeks and gDNA TTC/GAA tract analysed by PCR. This process was successfully continued over several generations. This allowed for us to observe changed in TTC/GAA tract length over the generation +/- *iil* defect-inducing conditions for Bur-0 and each of the NS lines. Changes in expansion instability and tract length as well as plant phenotypes over generations were analysed.

23 °C & 27 °C


### 3.5.3 Artificial miRNA *(35S::amiR-MSH2)* cloning protocol

In attempt to inhibit the DNA mismatch repair (MMR) pathway, Two alternative artificial microRNAs against *MSH2* (At3g18524), a gene indispensible for the function of the MMR pathway, were designed using the Web MicroRNA Designer http://wmd3.weigelworld.org/cgi-bin/webapp.cgi as described by Schwab et al (Schwab, Ossowski et al. 2006) by a previous honours student Mr Aleksej

Stevanovic. Two independent artificial microRNAs were designed against *MSH2* to improve the chances of a successful knock-out (Appendix - Table S4 & Appendix - Figure S7).

Artificial microRNAs (amiRNA) were initially generated by site directed mutagenesis using the *pRS300* vector as a template (Appendix - Figure S2). This created 3 different PCR products for each construct (Fig 8A). The DNA fragments contained overlapping DNA sequences, which were then used to subsequently fuse the 3 PCR products into 1 contiguous chimeric amiRNA structure for each construct (Fig 8B). This fusion PCR generated the 699 bp amiRNA backbones. The amiRNA backbones were then ligated into *pGEM-T easy* vectors and transformed into *E. coli* to propagate the construct. Restriction digestion was performed on the transformed *E. coli* cells to confirm presence of the insert (Fig 8C). This restriction digestion resulted in 2 bands, the top at 3.1 kb is the *pGEM-T easy* vector (Appendix - Figure S3) and the lower band is a 572 bp fragment of the amiRNA backbone (Fig 8C).

Following the restriction digestion the constructs were sequenced to confirm the correct amiRNA backbones were present. amiRNA backbones were then transferred into a *pJLBlue reverse* vector (Appendix - Figure S4). Restriction digestion was again performed to confirm the presence of the insert (Fig 8D). The restriction digestion resulted in a 436 bp fragment of the amiRNA backbone and 2305 bp fragment of *pJLBlue reverse* vector (Fig 8D). *pJLBlue* constructs were then recombined into the *pKF210* binary vectors (Appendix - Figure S5), with the help of a *pSoup* plasmid (Appendix - Figure S6) and transformed into *E. coli*. Again, restriction digestion confirmed presence of the insert releasing a 436 bp amiRNA backbone, a 1032 bp 35S promoter and 5214 bp of the remaining vector (Fig 8E).

Constructs were sequenced to confirm no mutations had occurred during the cloning process, and *pKF210*-amiRNA construct transformed into *Agrobacterium tumefaciens*. Colony PCR confirmed presence of the correct amiRNA backbone (Fig 8F), prior to transformation of the Bur-0 plants. Seeds from the transformed

plants were collected and treated with BASTA solution to select for plants transformed with the *pKF210* vector. To confirm those transformants (T1) contained the amiRNA construct, amiRNA construct specific PCRs were preformed (Fig 8G). The PCR amplified a 1215 bp fragment confirming the presence of the amiRNA construct.

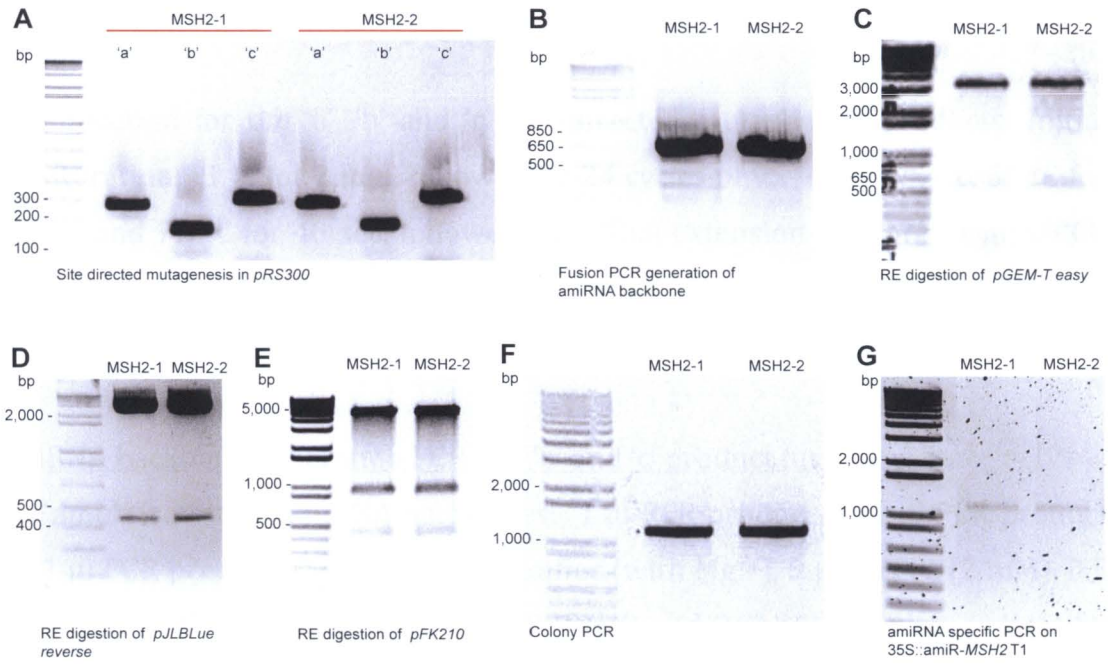



**Figure 8. *MSH2*-amiRNA cloning process. (A)** Site directed mutagenesis on *pRS300* vector, generation of 'a', 'b' and 'c' fragments for generating *amiRNA-MSH2-1* and amiRNA-MSH2-2 backbones. PCRs generated products of 271 bp, 170 bp and 298 bp from left to right. Primers used as in Appendix - Table S2. **(B)** Fusion PCR producing amiRNA backbones. Primers used as in Appendix - Table S3. **(C)** Restriction digestion of *pGEM-T easy*. 3.1 kb band is *pGEM-T easy* vector and 572 bp band is a fragement of the amiRNA backbone. **(D)** Restriction digestion of *pJLBLue reverse* vectors containing amiRNA backbones. 2305 bp band is *pJLBLue reverse* vector and 436 bp band is a fragement of the amiRNA backbone. **(E)** Restriction digestion of *pFK210* constructs containing amiRNA backbones. 5214 bp band is a portion of *pKF210* vector, 1032 bp band is the 35S promoter and 436 bp band is the amiRNA backbone. **(F)** Colony PCR on *A. tumeaciens* cells containing the *pFK210* vectors with amiRNA backbones. amiRNA specific primers were used (Appendix - Table S4 & S5) to generate 1215 bp PCR products. **(G)** PCR products confirming the presence of the amiRNAs in transformed T1 plants. PCRs used amiRNA specific primers (Appendix - Table S4 & S5) to generate 1215 bp DNA fragments. All PCR products and RE digests were run on1.2% agarose gels for 30 min-1 h at 100 V. 1kb+ DNA ladders were used (Invitrogen, USA).

### 3.5.4 Site directed mutagenesis of *pRS300* vector

For amplification of 'a', 'b' and 'c': 2 µl of *pRS300* plasmid DNA (1:100), 5 µl 10X PCR buffer (with $Mg^{++}$), 5 µl dNTPs (2 mM), Phusion DNA polymerase, 33.5 µl MilliQ $H_2O$ and 2 µl of each oligo primer (10 µM) were used. (see Appendix - Table S4 for list of primers ).

'a', 'b' and 'c' PCR cycling:

Amplification for the 'a', 'b' and 'c' site-directed mutagenesis products: Initial denaturation 95 °C for 2 min, followed by 24 cycles of 95 °C for 30 sec, 55 °C for 30 sec and 72 °C for 40 sec, followed by a final extension 72 °C for 7 min. PCR products were run on 1.2% agarose gels + EtBr in TAE buffer for 30 min at 100 V.

amiRNA backbone generation PCR ('a', 'b' and 'c' product fusion)

For amplification of amiRNA backbones: 1 µl PCR product 'a', 1 µl PCR product 'b', 1 µl PCR product 'c', 5 µl 10X PCR buffer (with $Mg^{++}$), 5 µl dNTPs (2 mM), 0.5 µl Phusion DNA polymerase, 33 µl MilliQ $H_2O$ and 2 µl of each oligo primer 'A' and 'B' (10 µM) were used. (SKB_792 5'-CTGCAAGGCGATTAAGTTGGGTAAC-3' & SKB_793 5'-GCGGATAACAATTTCACACAGGAAACAG-3') (Appendix - Table S4).

'a', 'b' and 'c' fusion PCR cycling:

Amplification of the 'a', 'b' and 'c' fusion products: Initial denaturation 95 °C for 2 min, followed by 24 cycles of 95 °C for 30 sec, 55 °C for 30 sec and 72 °C for 90 sec, followed by a final extension 72 °C for 7 min. PCR products were run on 1.2% agarose gels + EtBr in TAE buffer for 30 min at 100 V.

PCR to confirm presence of insert in *pFK210* constructs

Each of the 2 MSH2-amiRNA constructs used different specific primers (Appendix - Table S4 & S5). For amplification of specific amiRNAs: 2 µl DNA, 2 µl 10X PCR buffer, 2 µl dNTPs (2 mM), 0.5 µl Taq polymerase Thermopol (NEB),

11.5 µl MilliQ $H_2O$ and 1 µl of each amiRNA specific primer (Appendix - Table S4 & S5) were used. PCR cycling conditions are the same as for 'a', 'b', and 'c' fusion PCR.

### 3.5.5 Poly-A tailing

6.5 µl purified amiRNA backbone DNA, 0.5 µl Taq DNA polymerase (Thermopol by New England Biolabs, USA), 1 µl of 2 mM dATPs, 1 µl of 10 X PCR Buffer (NEB) and 1 µl of 12 mM $MgCl_2$ were incubated at 72 °C for 30 min.

*Liquid LB media*

The liquid LB media was used to incubate *E. coli* and *A. tumefaciens* cells in order to obtain a large amount of cells. The liquid LB media was prepared by adding 10 g of Tryptone, 5 g of yeast extract and 10 g of NaCl to a 1 L Erlenmeyer flask containing 1 L of deionized $H_2O$. The pH was adjusted to 7.5 before autoclaving.

*LB agar plates*

Plates were prepared by adding 10 g Tryptone, 5 g yeast extract, 10 g NaCl and 15 g agar to a 1 L Erlenmeyer flask containing 1 L of deionized $H_2O$. The pH was adjusted to 7.5 before autoclaving and distributed into petri dishes in a laminar flow cabinet.

### 3.5.6 Construction of amiRNA backbones

The web-based tool designed 4 primers for each amiRNA, which were then used to produce amiRNA backbones. In the 1st step, the *pRS300* vector (Appendix - Figure S2 & S7) was used as a PCR template and 3 separate PCR reactions performed (Fig 8A), which produced 3 different PCR products 'a', 'b' and 'c' for each amiRNA. Each of the 2 amiRNAs were cloned using different primers (Appendix - Table S4 & S5). The PCRs were designed to replicate the endogenous amiRNA* and amiRNA (Appendix - Figure S7) sequences, found on the *pRS300* vector, with gene specific sequences through a type of site-directed mutagenesis.

This altered the short (21 bp) amiRNA* and amiRNA sequences, while keeping the rest of the plasmid sequence unchanged allowing the creation of gene specific amiRNAs. The 'a', 'b' and 'c' PCR products were then separated on an electrophoretic gel (2% agarose, 100 V, 60 min), visualized, then excised from the gel, and purified using the Wizard ® SV Gel and PCR Clean-Up System (Promega, USA) following the manufacturer's instructions. Those purified 'a', 'b' and 'c' PCR products were then used as a template for a fusion PCR designed to fuse the 3 PCR products into a 699 bp chimeric DNA structure (Fig 8B). Primers used for the fusion PCR were the same for both amiRNA constructs (Appendix - Table S5). The fusion PCR products were separated by gel electrophoreses (2% agarose, 100 V, 60 min) and again visualized, excised from the gel and purified using the Wizard ® SV Gel and PCR Clean-Up System (Promega, USA). These steps constructed the 2 anti-*MSH2* amiRNA backbones.

The amiRNA backbones were then poly-A tailed and ligated into a *pGEM-T easy* vector (Appendix - Figure S3).

### 3.5.7 Subcloning the amiRNA backbones into pGEM-T

*pGEM-T easy* is a 3015 bp linear vector containing overhanging poly-T nucleotides at both ends that facilitate binding to poly A nucleotides present on the amiRNA backbones. Ligation was performed using the pGEM®-T Easy Vector System I (Promega, USA) following the manufacturer's instructions. Then transformed into *E. coli*. Transformation was achieved through heat shock, briefly 3 µl of the ligated DNA was added to 50 µl of competent DH5α *E. coli* cells and incubated at 42 °C for 90 sec, immediately following this samples were placed on ice for 2 min then incubated in 1 ml LB media at 37 °C and 180 rpm for 90 min. 200 µl of the sample in LB media was then spread on LB agar plates containing 100 ng/ml Ampicillin, IPTG and X-Gal and incubated O/N at 37 °C.

The purpose of subcloning into *pGEM-T easy* was to allow for easy blue/white

selection of amiRNA backbones in *E. coli* cells facilitated by beta-galactosidease (β-Gal) gene. If the *pGEM-T easy* self ligates it activates the β-Gal gene, which produces a blue pigment in the presence of X-Gal and IPTG. If the amiRNA backbones are ligated into *pGEM-T easy* this causes insertional inactivation of β-Gal and such colonies appear white. Thus colonies containing the insert appear white, while colonies containing the self-ligated vector are blue. *pGEM-T easy* also contains an *Ampicillin resistance* (*Amp*$^R$) gene, which allows for positive selection of colonies that have successfully taken up the vector.

Following the O/N incubation white colonies were selected and incubated in 3 ml liquid LB media (containing 100 ng/ml Ampicillin) for 24 h at 180 rpm and 37 °C in order to obtain a large amount of cells. Subsequently the vectors were extracted from the *E. coli* using a QIAprep ® Spin Miniprep kit (Qiagen, Netherlands) according to the manufacturer's instructions.

### 3.5.8 Restriction digestions

Following the miniprep, the presence of the *pGEM-T easy* vector was determined by restriction digestion (Fig 8C). The restriction enzyme (RE) pairs used, were selected such that they only cut once, to release part of the amiRNA backbone. The digests were then electrophoretically separated (1.2% agarose gel, 100 V, 45 min) to confirm the presence of the insert. At this stage the amiRNA backbones were sequenced (Micromon, Monash) in order to determine that no mutations occurred during the subcloning process.

**Restriction digestion of *pGEM-T easy* vector:** 2 µl DNA, 1 µl RNAse, 2 µl 10X Buffer n°4 (New England Biolabs), 14.6 µl sterile $H_2O$ and 0.2 µl EcoRI-HF 20,000 U/µl (New England Biolabs) & 0.2 µl BamHI-HF U/ul (New England Biolabs) REs.

**Restriction digestion of *pGEM-T* construct:** 10 µl *pGEM-T* construct, 2 µl 10X Buffer n°4 (New England Biolabs), 7 µl sterile $H_2O$ and 0.5 µl HindIII-HF 20,000

U/µl (New England Biolabs) & 0.5 µl BamHI-HF U/µl (New England Biolabs) REs.

**Restriction digestion of *pJLBlue reverse* vector:** 7 µl *pJLBlue reverse* vector, 2 µl 10X Buffer n°4 (New England Biolabs), 10 µl sterile $H_2O$ and 0.5 µl HindIII-HF 20,000 U/µl (New England Biolabs) & 0.5 µl BamHI-HF U/µl (New England Biolabs) REs.

**Restriction digestion of *pJLBlue reverse* construct:** 2 µl *pJLBlue reverse* construct, 2 µl 10X Buffer n°4 (New England Biolabs), 15.6 µl sterile $H_2O$ and 0.2 µl HindIII-HF 20,000 U/µl (New England Biolabs) & 0.2 µl BamHI-HF U/µl (New England Biolabs) REs.

**Restriction digestion of *pFK210* construct:** 4 µl *pKF210* construct, 2 µl 10X Buffer n°4 (New England Biolabs), 13.6 µl sterile $H_2O$ and 0.2 µl HindIII-HF 20,000 U/µl (New England Biolabs) & 0.2 µl BamHI-HF U/µl (New England Biolabs) REs.

All restriction digestions were incubated at 37 °C for 1 h.

### 3.5.9 DNA sequencing

All sequencing was performed by Micromon™ (Monash, VIC, Australia). The plasmids sent for sequencing were extracted using the QIAprep® Spin Miniprep Kit (Qiagen, Netherlands) according to the manufacturer's instructions. Following the extraction, the DNA concentrations were determined using the NanoDrop™ 1000 spectrophotometer (Thermo Scientific, USA). The DNA was prepared by diluting the concentration of vectors to 90 ng/µl. Then adding 11 µl of DNA (90 ng/µl) with 1 µl of primer (0.8 pmol/µl). All products were sequenced using the *pGEM-T easy* specific forward and reverse primers (Appendix - Table S5). The sequencing results were retrieved using SeqMan 8.0.2 (DNASTAR, USA) and the sequences assembled and compared against the

reference DNA using the same software.

### 3.5.10 Transferring the insert from pGEM-T easy into pJLBlue revrese

Before the amiRNA backbones could be transferred into the *pFK210* vector, they were transferred into the *pJLBlue* reverse vector (Appendix - Figure S4). This step facilitated the subsequent transfer into the *pFK210* vector. The *pJLBlue* reverse vector is 2335 bp long and contains 2 important features. The 1$^{st}$ one is the *Kanamycin resistance* gene, which allows for positive selection of cells containing the vector, and the second are the attL1 and attL2 sites. These sites allow recombination between vectors containing the attL1/2 sites, and vectors containing the attR1/2 sites e.g the *pFK210* vector (Appendix - Figure S5).

In order to transfer the inserts from *pGEM-T easy* into *pJLBlue reverse*, both vectors were designed using the same RE pair. The RE pair was selected so that each enzyme cuts only once in each vector, in order to release the amiRNA backbones from *pGEM-T easy* and create sticky ends on the *pJLBlue reverse* vector. Following restriction digestion, the products were separated by gel electrophoresis (1.2% agarose gel, 100 V, 45 min), visualized and the amiRNA backbones and *pJLBlue reverse* vectors excised from the gel and purified using the Wizard® SV Gel and PCR Clean-Up System kit (Promega, USA) following the manufacturer's instructions. The REs used for the restriction digest created compatible 'stick ends', which allowed for the amiRNA backbones to be ligated into *pJLBlue reverse* by keeping the correct directionality.

Following ligation, 4 µl of the ligand product was transformed into 50 µl of competent DH5α *E. coli* cells by heat shock. The cells were transferred onto LB agar plates containing 50 ng/µl Kanamycin in order to select for cells containing the *pJLBlue reverse*. Following the transformation the cells were grown at 37 °C, O/N.

Following the O/N incubation, colonies were picked from each LB agar plate and grown in 3 ml of liquid LB media (containing 50 ng/µl Kanamycin), 180 rpm, 37 °C, O/N. Subsequently the vectors were extracted from *E. coli* cells using a QIAprep® Spin Miniprep Kit (Qiagen, Netherlands). After the extraction, the presence of the vector was confirmed by restriction digestion (Fig 7D). The products were resolved by electrophoresis (1.2% agarose gel, 100 V, 45 min) and visualized to determine the presence of the inserts.

### 3.5.11 LR clonase recombination

Once the presence of the inserts were determined, they were transferred into the *pFK210* vector by a recombination reaction. The *pFK210* vector is a gateway vector and it is particularly useful as it contains origins of replication that allow expression in both bacteria and plants. The recombination between *pJLBlue reverse* and *pFK210* was possible due to the attL1 and attL2 sites present on *pJLBlue reverse* (Appendix - Figure S4). and the attR1 and attR2 sites present on *pFK210* (Appendix - Figure S5). When the recombinase enzyme is added to the 2 vectors, the 2 sites recombine exchanging the DNA flanking those 2 sites (amiRNA backbones). The recombination reaction was performed using the Gateway® Spin Miniprep Kit (Qiagen, Netherlands). The presence of the *pFK210* vector and the amiRNA backbone was determined by restriction digestion (Fig 8E). Once the presences of inserts were confirmed, the vectors were sequenced to again make sure no mutations had occurred along the way. Sequencing was performed as above, except with *pFK210* specific forward and reverse primers. Once sequencing confirmed the presence of the correct sequence, the vectors were transformed into *Agrobacterium tumefaciens.*

### 3.5.12 Agrobacterium tumefaciens transformation

*A. tumefaciens* have been extensively used to introduce foreign DNA into *A.*

*thaliana* plants (Clough and Bent 1998), therefore *A. tumefaciens* cells were used to transfer the *pFK210* vectors into Bur-0 plants. The *pFK210* vector has the ability to replicate inside both *E. coli* however it cannot replicate inside *A. tumefaciens* without the presence of the *pSoup* helper plasmid (Appendix - Figure S6). *pSoup* provides the replicase function, which allows the expression of the *pFK210* vector in *A. tumefaciens.*

The transformation of *A. tumefaciens* was performed by electroporation using the Electroporator ECM 630 (Fisher Biotech). 50 µl of electrocompetent *A. tumefaciens* GV3001 cells were transferred to pre-chilled 100 µl electroporation cuvettes (Eppendorf, Germany). Subsequently, 2 µl of the desired *pFK210* vector and 2 µl *pSOUP* of helper plasmid were added to the cuvettes and electroporated under the following settings: Capacitance = 25 µF, Voltage = 2.4 kV and Resistance = 200 Ω. Immediately after electroporation, 1 ml liquid LB media was added to the cuvettes, the cells transferred to a 1.5 ml microfuge tube and incubated at 28 °C, 130 rpm for 4 h. Following the incubation, cells were transferred onto LB agar plates (containing Gentinomycin (10 µg/ml), Rifampicin (50 µg/ml), Spectinomycin (100 µg/ml) and Tetracycline (10 µg/ml)). The plates were subsequently incubated in a MS Vortex Incubator (MS major science, USA) at 28 °C for 48 h. After 48 h, colonies were selected from each plate and grown in 3 ml of liquid LB media (containing Gentinomycin (10 µg/ml), Rifampicin (50 µg/ml), Spectinomycin (100 µg/ml) and Tetracycline (10 µg/ml)) for 24 h at 28 °C, 130 rpm. Following the incubation, the presences of the vectors were verified by colony PCR.

### 3.5.13 Agrobacterium tumefaciens colony PCR

In order to determine if *A. tumesfaciens* cells have successfully up taken the vectors, 200 µl of *A. tumesfaciens* were transferred to a 1.5 ml microfuge tube and incubated at 100 °C for 10 min in order to lyse the cell walls and release the vectors. The cells were then centrifuged at 8,000 rpm for 2 min to remove

cellular debris and the supernatant transferred to another microfuge tube and used as a PCR template. Each *pFK210* construct used a different set of primers (Appendix - Table S4 & S5) in order to confirm the presence of the correct amiRNA construct. Once the presence of the insert was verified, the *pFK210* vectors were transformed into *A. thaliana* by infecting the plants with the transgenic *A. tumefaciens* cells.

### 3.5.14 Transformation

Floral dipping was performed to transfer the vectors into Bur-0 *A. thaliana* plants. The flowers of the plants were dipped into a solution containing transgenic *A. tumefaciens* cells, in order to transform the modified *pFK210* vectors into *A. thaliana* plants. A large number of *A. tumefaciens* cells were grown by inoculating 2 ml of transgenic *A. tumefaciens* cells into 500 ml of liquid LB media (containing Gentinomycin (10 µg/ml), Rifampicin (50 µg/ml), Spectinomycin (100 µg/ml) and Tetracycline (10 µg/ml)). The media was incubated in Erlenmeyer flasks at 28 °C and 130 rpm O/N to allow bacteria to grow. The following day, the cells were transferred into 1 L centrifuge tubes and centrifuged in a JAIO rotor at 4,000 rpm for 15 min to pellet the cells. The supernatant was discarded and pellet resuspended in transformation solution (prepared by adding 500 ml deionized $H_2O$, 25 g sucrose and 187.5 µl SilWet).

*A. thaliana* plants aged between 35-40 days (in the stages of early flowering) were dipped into the solution, ensuring the flowers and flower buds were submerged in the solution for a few seconds. Following floral dipping the plants were left lying on the side and covered with a dome for 24 h in plant growth chambers 23 °C, Long day conditions (16 h of light and 8 h of dark). After 24 h the plants were placed in an upright position and left to grow and the seeds to mature. In order to maximize the chances of a successful transformation, many *A. thaliana* plants were transformed with the same construct several times.

### 3.5.15 Selecting transgenic plants

Once mature seeds of the transformed plants were collected, they were left in a dehumidifying chamber for 1-2 weeks, stratified in 0.1% agarose solution for 3 days at 4 °C, planted on soil and left to grow for 4 days in 23 °C LD conditions. The plants were watered with 2 L of the strong herbicide BASTA at 30 µg/ml. After germination the seeds were transferred to 27 °C short day conditions (8 h light and 16 h dark) (the *iil* phenotype-inducing conditions). On days 5, 7, 9 and 11 the plants were sprayed with BASTA solution (30 µg/ml). BASTA selected for plants containing the *pFK210* vector as this vector also a BAR gene (Appendix - Figure S5) which neutralizes the effect of the herbicide. All untransformed seedlings died.

### 3.5.16 Verifying the presence of the insert

The presence of the amiRNA constructs in *A. thaliana* plants were determined by PCR using specific primers for each amiRNA construct (Fig 8G).

### 3.5.17 Analysis of transformants

The TNR expansion defect phenotype, *iil*, in the 27 °C Bur-0 transformats was observed and compared to the untransformed plants. The genomic TTC/GAA expansion at the *IIL1* locus of each transformant was analysed by PCR and their *IIL1* gene expression levels analysed by q-RT-PCR.

## 3.6 REFERENCES


Al-Mahdawi, S., R. M. Pinto, O. Ismail, D. Varshney, S. Lymperi, C. Sandi, D. Trabzuni and M. Pook (2008). "The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues." Hum Mol Genet **17**(5): 735-746.

Al-Mahdawi, S., R. M. Pinto, D. Varshney, L. Lawrence, M. B. Lowrie, S. Hughes, Z. Webster, J. Blake, J. M. Cooper, R. King and M. A. Pook (2006). "GAA repeat expansion mutation mouse models of Friedreich ataxia exhibit oxidative stress leading to progressive neuronal and cardiac pathology." Genomics **88**(5): 580-590.

Ashley, C. T., Jr. and S. T. Warren (1995). "Trinucleotide repeat expansion and human disease." Annu Rev Genet **29**: 703-728.

Aziz, N. A., C. K. Jurgens, G. B. Landwehrmeyer, W. M. van Roon-Mom, G. J. van Ommen, T. Stijnen and R. A. Roos (2009). "Normal and mutant HTT interact to affect clinical severity and progression in Huntington disease." Neurology **73**(16): 1280-1285.

Budden, T. and N. A. Bowden (2013). "The Role of Altered Nucleotide Excision Repair and UVB-Induced DNA Damage in Melanomagenesis." Int J Mol Sci **14**(1): 1132-1151.

Carpenter, N. J. (1994). "Genetic anticipation. Expanding tandem repeats." Neurol Clin **12**(4): 683-697.

Cattaneo, E., C. Zuccato and M. Tartari (2005). "Normal huntingtin function: an alternative approach to Huntington's disease." Nat Rev Neurosci **6**(12): 919-930.

Christmann, M., M. T. Tomicic, W. P. Roos and B. Kaina (2003). "Mechanisms of human DNA repair: an update." Toxicology **193**(1-2): 3-34.

Clough, S. J. and A. F. Bent (1998). "Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana." Plant J **16**(6): 735-743.

Cossee, M., H. Puccio, A. Gansmuller, H. Koutnikova, A. Dierich, M. LeMeur, K. Fischbeck, P. Dolle and M. Koenig (2000). "Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation." Hum Mol Genet **9**(8): 1219-1226.

Delatycki, M., R. Williamson and S. Forrest (2000). "Friedreich ataxia: an overview." J Med Genet **37**(1): 1-8.

Delatycki, M. B. and L. A. Corben (2012). "Clinical Features of Friedreich Ataxia." J Child Neurol **27**(9): 1133-1137.

Djousse, L., B. Knowlton, M. Hayden, E. W. Almqvist, R. Brinkman, C. Ross, R. Margolis, A. Rosenblatt, A. Durr, C. Dode, P. J. Morrison, A. Novelletto, M. Frontali, R. J. Trent, E. McCusker, E. Gomez-Tortosa, D. Mayo, R. Jones, A. Zanko, M. Nance, R. Abramson, O. Suchowersky, J. Paulsen, M. Harrison, Q. Yang, L. A. Cupples, J. F. Gusella, M. E. MacDonald and R. H. Myers (2003). "Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease." Am J Med Genet A **119A**(3): 279-282.

Du, J., E. Campau, E. Soragni, S. Ku, J. W. Puckett, P. B. Dervan and J. M. Gottesfeld (2012). "Role of mismatch repair enzymes in GAA/TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells." J Biol Chem **287**(35): 29861-29872.

Evans-Galea, M. V., N. Carrodus, S. M. Rowley, L. A. Corben, G. Tai, R. Saffery, J. C. Galati, N. C. Wong, J. M. Craig, D. R. Lynch, S. R. Regner, A. F. Brocht, S. L. Perlman, K. O. Bushara, C. M. Gomez, G. R. Wilmot, L. Li, E. Varley, M. B. Delatycki and J. P. Sarsero (2012). "FXN methylation predicts expression and clinical outcome in Friedreich ataxia." Ann Neurol **71**(4): 487-497.

Evans-Galea, M. V., P. J. Lockhart, C. A. Galea, A. J. Hannan and M. B. Delatycki (2014). "Beyond loss of frataxin: the complex molecular pathology of Friedreich ataxia." Discov Med **17**(91): 25-35.

Ezzatizadeh, V., R. M. Pinto, C. Sandi, M. Sandi, S. Al-Mahdawi, H. Te Riele and M. A. Pook (2012). "The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model." Neurobiol Dis **46**(1): 165-171.

Gomes-Pereira, M., L. Foiry, A. Nicole, A. Huguet, C. Junien, A. Munnich and G. Gourdon (2007). "CTG trinucleotide repeat "big jumps": large expansions, small mice." PLoS Genet **3**(4): e52.

Gonitel, R., H. Moffitt, K. Sathasivam, B. Woodman, P. J. Detloff, R. L. Faull and G. P. Bates (2008). "DNA instability in postmitotic neurons." Proc Natl Acad Sci U S A **105**(9): 3467-3472.

Gray, M., D. I. Shirasaki, C. Cepeda, V. M. Andre, B. Wilburn, X. H. Lu, J. Tao, I. Yamazaki, S. H. Li, Y. E. Sun, X. J. Li, M. S. Levine and X. W. Yang (2008). "Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit

progressive and selective neuropathogenesis in BACHD mice." J Neurosci **28**(24): 6182-6195.

Hick, A., M. Wattenhofer-Donze, S. Chintawar, P. Tropel, J. P. Simard, N. Vaucamps, D. Gall, L. Lambot, C. Andre, L. Reutenauer, M. Rai, M. Teletin, N. Messaddeq, S. N. Schiffmann, S. Viville, C. E. Pearson, M. Pandolfo and H. Puccio (2013). "Neurons and cardiomyocytes derived from induced pluripotent stem cells as a model for mitochondrial defects in Friedreich's ataxia." Dis Model Mech **6**(3): 608-621.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." Nat Rev Mol Cell Biol **7**(5): 335-346.

Jones, J., A. Estirado, C. Redondo and S. Martinez (2013). "Stem cells from wildtype and Friedreich's ataxia mice present similar neuroprotective properties in dorsal root ganglia cells." PLoS ONE **8**(5): e62807.

Kennedy, L., E. Evans, C. M. Chen, L. Craven, P. J. Detloff, M. Ennis and P. F. Shelbourne (2003). "Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis." Hum Mol Genet **12**(24): 3359-3367.

Kim, Y. J. and D. M. Wilson, 3rd (2012). "Overview of base excision repair biochemistry." Curr Mol Pharmacol **5**(1): 3-13.

Koeppen, A. H. (2011). "Friedreich's ataxia: pathology, pathogenesis, and molecular genetics." J Neurol Sci **303**(1-2): 1-12.

Kovtun, I. V. and C. T. McMurray (2008). "Features of trinucleotide repeat instability in vivo." Cell Res **18**(1): 198-213.

Liu, G. and M. Leffak (2012). "Instability of (CTG)n*(CAG)n trinucleotide repeats and DNA synthesis." Cell Biosci **2**(1): 7.

Lufino, M. M., A. M. Silva, A. H. Nemeth, J. Alegre-Abarrategui, A. J. Russell and R. Wade-Martins (2013). "A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds." Hum Mol Genet **22**(25): 5173-5187.

Martelli, A., M. Napierala and H. Puccio (2012). "Understanding the genetic and molecular pathogenesis of Friedreich's ataxia through animal and cellular models." Dis Model Mech **5**(2): 165-176.

McMurray, C. T. (2010). "Mechanisms of trinucleotide repeat instability during human development." Nat Rev Genet **11**(11): 786-799.

Miranda, C. J., M. M. Santos, K. Ohshima, J. Smith, L. Li, M. Bunting, M. Cossee, M. Koenig, J. Sequeiros, J. Kaplan and M. Pandolfo (2002). "Frataxin knockin mouse." FEBS Lett **512**(1-3): 291-297.

Mirkin, S. M. (2005). "Toward a unified theory for repeat expansions." Nat Struct Mol Biol **12**(8): 635-637.

Mirkin, S. M. (2007). "Expandable DNA repeats and human disease." Nature **447**(7147): 932-940.

Perdomini, M., A. Hick, H. Puccio and M. A. Pook (2013). "Animal and cellular models of Friedreich ataxia." J Neurochem **126 Suppl 1**: 65-79.

Pouladi, M. A., L. M. Stanek, Y. Xie, S. Franciosi, A. L. Southwell, Y. Deng, S. Butland, W. Zhang, S. H. Cheng, L. S. Shihabuddin and M. R. Hayden (2012). "Marked differences in neurochemistry and aggregates despite similar behavioural and neuropathological features of Huntington disease in the full-length BACHD and YAC128 mice." Hum Mol Genet **21**(10): 2219-2232.

Puccio, H. (2009). "Multicellular models of Friedreich ataxia." J Neurol **256 Suppl 1**: 18-24.

Puccio, H., D. Simon, M. Cossee, P. Criqui-Filipe, F. Tiziano, J. Melki, C. Hindelang, R. Matyas, P. Rustin and M. Koenig (2001). "Mouse models for Friedreich ataxia exhibit cardiomyopathy, sensory nerve defect and Fe-S enzyme deficiency followed by intramitochondrial iron deposits." Nat Genet **27**(2): 181-186.

Rotig, A., P. de Lonlay, D. Chretien, F. Foury, M. Koenig, D. Sidi, A. Munnich and P. Rustin (1997). "Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich ataxia." Nat Genet **17**(2): 215-217.

Ruggiero, B. L. and M. D. Topal (2004). "Triplet repeat expansion generated by DNA slippage is suppressed by human flap endonuclease 1." J Biol Chem **279**(22): 23088-23097.

Sakamoto, N., P. D. Chastain, P. Parniewski, K. Ohshima, M. Pandolfo, J. D. Griffith and R. D. Wells (1999). "Sticky DNA: self-association properties of long GAA/TTC repeats in R.R.Y triplex structures from Friedreich's ataxia." Mol Cell **3**(4): 465-475.

Schwab, R., S. Ossowski, M. Riester, N. Warthmann and D. Weigel (2006). "Highly specific gene silencing by artificial microRNAs in Arabidopsis." Plant Cell **18**(5): 1121-1133.

Schweitzer, J. K. and D. M. Livingston (1997). "Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast." Hum Mol Genet **6**(3): 349-355.

Schweitzer, J. K. and D. M. Livingston (1999). "The effect of DNA replication mutations on CAG tract stability in yeast." Genetics **152**(3): 953-963.

Shishkin, A. A., I. Voineagu, R. Matera, N. Cherng, B. T. Chernet, M. M. Krasilnikova, V. Narayanan, K. S. Lobachev and S. M. Mirkin (2009). "Large-scale expansions of Friedreich's ataxia GAA repeats in yeast." Mol Cell **35**(1): 82-92.

Soriano, S., J. V. Llorens, L. Blanco-Sobero, L. Gutierrez, P. Calap-Quintana, M. P. Morales, M. D. Molto and M. J. Martinez-Sebastian (2013). "Deferiprone and idebenone rescue frataxin depletion phenotypes in a Drosophila model of Friedreich's ataxia." Gene **521**(2): 274-281.

Sturm, B., U. Bistrich, M. Schranzhofer, J. P. Sarsero, U. Rauen, B. Scheiber-Mojdehkar, H. de Groot, P. Ioannou and F. Petrat (2005). "Friedreich's ataxia, no changes in mitochondrial labile iron in human lymphoblasts and fibroblasts: a decrease in antioxidative capacity?" J Biol Chem **280**(8): 6701-6708.

Sugui, J. A., Y. C. Chang and K. J. Kwon-Chung (2005). "Agrobacterium tumefaciens-mediated transformation of Aspergillus fumigatus: an efficient tool for insertional mutagenesis and targeted gene disruption." Appl Environ Microbiol **71**(4): 1798-1802.

Sureshkumar, S., M. Todesco, K. Schneeberger, R. Harilal, S. Balasubramanian and D. Weigel (2009). "A genetic defect caused by a triplet repeat expansion in Arabidopsis thaliana." Science **323**(5917): 1060-1063.

Ura, K. and J. J. Hayes (2002). "Nucleotide excision repair and chromatin remodeling." Eur J Biochem **269**(9): 2288-2293.

Wyman, C. and R. Kanaar (2006). "DNA double-strand break repair: all's well that ends well." Annu Rev Genet **40**: 363-383.

Zhang, Y., A. A. Shishkin, Y. Nishida, D. Marcinkowski-Desmond, N. Saini, K. V. Volkov, S. M. Mirkin and K. S. Lobachev (2012). "Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells." Mol Cell **48**(2): 254-265.

# CHAPTER 4. Short interfering RNAs (siRNA)s play a role in transcriptional down regulation of *FXN* seen in Friedreich's ataxia

## 4.1 INTRODUCTION

### 4.1.1 FRDA molecular pathology

The pathology of FRDA is ultimately caused by a decrease in frataxin protein expression. The level of frataxin expression is inversely correlated with the repeat expansion tract length, that is larger repeat expansions typically reduce the expression level more than shorter repeat expansions (Patel and Isaya 2001, Evans-Galea, Lockhart et al. 2014). In the *FXN* gene a GAA repeat of 7-34 is considered normal (Wild type, WT) and repeat tracts of more than 100 considered pathogenic (Patel and Isaya 2001). In individuals with severe forms of FRDA repeats tracts can reach more than 1000 (Ditch, Sammarco et al. 2009). Generally, the longer the repeat tract within the pathogenic range, the greater the decrease in *FXN* expression and concomitantly the more severe the symptoms (Koeppen 2011, Metz, Coppard et al. 2013). Pathogenic repeat tracts generally decrease *FXN* expression up to 4-29% of WT (Campuzano, Montermini et al. 1997). It is worth mentioning that individuals heterozygous for the expansion are asymptomatic and express ~half the amount of frataxin as individuals who carry WT alleles in a homozygous state (Chapdelaine, Coulombe et al. 2013). Thus, it is an example of Haplosufficiency and tells us that an increase in *FXN* expression to just half that of WT is likely sufficient to prevent symptoms.

FRDA was first described almost 150 years ago and almost 20 years ago it was first recognized that the causative mutation was due to a GAA/TTC trinucleotide repeat expansion in the first intron of the Frataxin gene (*FXN*) (Campuzano, Montermini et al. 1996). Translational blockage seems to be the main source of frataxin deficiency. In normal gene expression, introns are spliced from pre-mRNA, during mRNA processing. Hence, if the *FXN* mRNA transcripts were present the expansion-bearing intron would be removed and a functionally normal *FXN* mRNA and protein result. Unfortunately in FRDA, although the small amount of frataxin produced is functionally normal, the expansion seems to interfere with transcription contributing to the gene Loss of function (LOF) effect. This is in contrast to other TNRDs such as Huntington's disease, in which the expansion is exonic leading to a Gain of function (GOF) protein.

### 4.1.2 Factors contributing to frataxin insufficiency in FRDA

Substantial research has gone into investigating how the expansion affects *FXN* gene expression, and it appears that there are many factors that contribute to the decreased amount of frataxin (Patel and Isaya 2001, Evans-Galea, Lockhart et al. 2014). For example, it has been shown that at the DNA level, the long expansion tracts can slip out of the normal double helical structure of DNA and form different DNA secondary structures preventing RNA polymerase extension over the gene resulting in decreased mRNA transcripts produced. The expanded repeats seem to form unusual DNA structures known as "sticky DNA" (Sakamoto, Chastain et al. 1999, Wells 2009). *In vitro* studies clearly demonstrate the presence of "sticky DNA" (Sinden and Wells 1992, Wells 2009). These unusual structures interfere with both transcription and replication (Ohshima, Montermini et al. 1998, Wu, Chow et al. 1998). The GAA/TTC repeats interfere with transcription in a position and orientation dependent manner (Bidichandani, Ashizawa et al. 1998, Ohshima, Montermini et al. 1998, Mariappan, Catasti et al. 1999, Grabczyk and Usdin 2000, Jain, Rajeswari et al. 2002). There is also evidence that these secondary structures attract certain

proteins that can stabilize their structure, further preventing the amount of mRNA transcripts produced (Ezzatizadeh, Pinto et al. 2012).

As it appears that a transcriptional defect of the *FXN* gene is the main cause of frataxin deficiency, with the coding sequence unaltered and the mature mRNA successfully generated usually progressing through translation and protein formation as normal (Xia, Cao et al. 2012), pharmacological up-regulation of *FXN* expression may restore frataxin to therapeutic levels. Thus, one major avenue for the treatment for FRDA would be approaches that increase *FXN* transcription.

#### 4.1.3 RNA interference (RNAi)

RNA interference (RNAi) is an epigenetic process in which small double-stranded RNA molecules (dsRNA) induce sequence-specific gene silencing (Meister and Tuschl 2004). Micro RNAs (miRNA) and short interfering RNAs (siRNA) are 21–24 nt non-coding smallRNA sequences that regulate gene expression by targeting specific mRNA transcripts. The endogenous RNAi pathway intersects with the miRNA machinery, which in mammals regulates gene expression by inhibiting protein synthesis and inducing mRNA decay (Filipowicz, Bhattacharyya et al. 2008, Guo, Ingolia et al. 2010).

miRNAs are gene encoded and arise from a single-stranded pre-miRNA. Unlike miRNA, siRNAs are non-gene encoded and derived from dsRNA molecules. These dsRNA molecules can arise from an exogenous source (exo-siRNA), such as viral genomes or endogenous sources (endo-siRNA) such as from repetitive elements like transposons and long stretches of repetitive DNA sequences (Gaynor, Campbell et al. 2010). Endo- and exo-siRNAs and miRNAs are processed in the cytoplasm by the RNase III enzyme Dicer and loaded to Argonaut (Ago) proteins to form effector complexes (microRNP or RISC). Endo-siRNAs modulate innate immunity in plants (Blevins, Rajeswaran et al. 2006, Stram and Kuzntzova 2006, Guo, Ingolia et al. 2010), *Drosophila* (Wang, Aliyari et al. 2006, Zambon, Vakharia et al. 2006) and *C. elegans* (Kawli and Tan 2008, Lu,

Yigit et al. 2009). The function of mammalian endo-siRNAs is not well understood, though it has been found that they operate primarily in oocytes (Tam, Aravin et al. 2008, Watanabe, Totoki et al. 2008, Suh, Baehner et al. 2010). It is worth noting that the large inter-generational expansions observed as 'Genetic anticipation' have been found to generally occur during gametogenesis (Ashley and Warren 1995).

Somatic miRNAs in mammals are known to regulate numerous biological processes, and their disordered expression has been linked to a plethora of human diseases including cancer (Croce 2009) and diabetes (Guay and Regazzi 2013) as well as several TNRDs (Galloway and Nelson 2009, Osborne, Lin et al. 2009, Lee, Chu et al. 2011, Jovicic, Zaldivar Jolissaint et al. 2013), including FRDA (Mahishi, Hart et al. 2012). For example a resent screen has found that the miRNA miR-886-3p was upregulated in FRDA patient cells and when this miRNA was silenced by an anti-*miR* it led to the elevation of *FXN* message and protein levels (Mahishi, Hart et al. 2012). Indicating that this particular miRNA is contributing to decreased *FXN* expression in FRDA. Thus it appears that the epigenetic processes of DNA methylation, histone deacetylase activation and RNAi are contributing to frataxin deficiency in addition to the genetic barriers (Orr 2009).

In contrast to the known involvement of miRNA dysregulation in human diseases the function of endo-siRNAs has been largely unexplored and remains thus far elusive. There is however evidence to suggest that siRNAs may be acting to decrease expression of genes with repeat expansions in other organisms. Our lab has discovered that TTC/GAA repeat expansions can trigger production of 24 nt siRNAs in the case of *A. thaliana IIL1* gene (Eimer and Balasubramanian, personal communication). The siRNAs appear to localize at and around the site of the TTC/GAA repeat expansion, at a significantly higher level, in Bur-0 compared to natural suppressors, which differ from Bur-0 only by the repeat expansion in an otherwise identical genomic background. This tells us that the TTC/GAA repeat expansion can lead to siRNA production. It was also found that compromising the production of siRNAs by knocking down *DICKERLIKE 3*

(*DCL3*), which is required for the production of 24 nt-siRNAs, suppressed the *iil* phenotype and increased *IIL1* expression, indicating the potential for this machinery as a target for therapeutic intervention (Eimer and Balasubramanian, personal communication). The striking molecular similarities between the TTC/GAA expansion in Bur-0 *IIL1* and in FRDA *FXN* triggered us to investigate whether a similar situation was occurring in FRDA.

Although siRNAs have not been shown so far in the case of FRDA, dysregulated siRNAs have been shown in other repeat expansion diseases. In the presence of large GAA/TTC expansions, such as in FRDA, there is not only a decrease in *FXN* transcript but also an increase in *FXN* anti-sense transcript (De Biase, Chutake et al. 2009). A study in Myotonic dystrophy type 1 (DM1) *Drosophila* model, found that flies made to express anti-sense transcript of the *DM1* gene as well as the sense transcript were shown to have increased toxicity/disorder severity compared to flies that just expressed the protein encoding transcript. It has been hypothesized that the anti-sense transcript may be binding to the sense transcript and generating siRNA through the RNAi pathway (Yu, Teng et al. 2011). Furthermore, it was demonstrated that Dicer was able to recognize this sense:anti-sense dsRNA complex and generate 21 nt triplet-repeat derived siRNAs. Therefore, we reasoned to test whether siRNAs associated with repeat expansion are present in FRDA cell lines. Thus, we hypothesized that similar smallRNA dysregulation, if acting through a common TNR expansion intrinsic property, may also influence *FXN* expression.

### 4.1.4 Screens for molecules to increase gene expression in presence of repeat expansion

In recent times, there is accumulating evidence to suggest that methodologies that could result in increased expression of frataxin have a high potential for therapeutics of FRDA (Sandi, Al-Mahdawi et al. 2013). This is coupled with increases in our understanding of the molecular mechanisms of FRDA, although

several questions still remain unresolved (Wells and Ashiuzawa 2006, Koeppen 2011, Marmolino 2011, Payne 2011, Yandim, Natisvili et al. 2013). However, screening for molecules that increase frataxin levels has not been an easy one with only few reports so far, with the largest screen involving the screening of only 88 compounds selected from an *in silico* screening of 25,000 compounds (Li, Voullaire et al. 2013, Lufino, Silva et al. 2013).

Our lab has utilized the *A. thaliana* model to screen for compounds that increase gene expression in spite of the repeat expansion (Tasset and Balasubramanian, personal communication). The TTC/GAA expansion in the intron of the *ISOPROPYL MALATE ISOMERASE LARGE SUBUNIT 1 (IIL1)* gene results in the reduction of the expression of *IIL1* leading to the *iil* phenotype, which can be easily screened for visibly (Sureshkumar, Todesco et al. 2009, Willadsen, Cao et al. 2013). This finding opened unprecedented opportunities allowing for a large-scale drug screen, which was carried out by other members from our lab. Using the plant-based screening, a large-scale phenotypic suppression analysis was carried out for the *iil* phenotype to identify small molecules. Three different small molecule libraries (Natural Products Library (MS Discovery), LOPAC library (Sigma) and Diverset Library (Chembridge) representing a total of 4000 compounds were used to screen for their ability to suppress the *iil* phenotype. By screening 4000 compounds, more than 65 compounds were identified that suppress or modify the *iil* phenotype. Phenotypic suppression/modification could occur either by the compound affecting gene expression or by acting at a different level downstream in the pathway. Additional analysis revealed that several of these compounds increase *IIL1* expression in spite of repeat expansion, suggesting the potential for these compounds to be tested on FRDA cells for their ability to increase *FXN* expression (Tasset and Balasubramanian, personal communication). Thus, we hypothesized that such compounds, if acting through a common TNR expansion intrinsic property, may also influence *FXN* expression.

Based on the above, we set out to take a translational approach to screen for compounds and pathways, which can be used for FRDA therapeutics. To achieve this goal, we set out two distinct objectives.

1. To analyse whether siRNA mediated transcriptional downregulation observed in plants can also be seen in FRDA.
2. To test the efficacy pre-screened candidate molecules to increase FXN expression

In this chapter I show that smallRNAs are dysergulated in FRDA and 24 nt siRNAs that map to FXN are increased in patient cells compared to the lymphoblasts of unaffected individuals. I also demonstrate that blocking smallRNA production through chemical inhibitors have a positive effect on FXN expression. By analyzing the pre-screened candidate molecules in FRDA cells, I show that at least two additional compounds have the capability to increase FXN expression in spite of the repeat expansion.

## 4.2 RESULTS

### 4.2.1 SmallRNAs mapping to *FXN* are increased in presence of repeat expansion.

A total of 3,040 reads from the smallRNA sequencing of our normal and FRDA affected cell lines were mapped to the *FXN* locus, among which 51.95% were 22 nt long. The majority of reads were 22 nt, with the FRDA samples also displaying a slight peak at 18 nt. This was slightly different from the plant system, where the majority of the reads were 24 nt long. We found an increase in smallRNAs mapping to the *FXN* gene in affected cells, which was approximately 3.5 fold when compared with normal cells (Table 1). This data is consistent with what was found in the plant model, which also exhibited a dramatic increase in smallRNAs mapping to the *IIL1* locus in presence of the repeat expansion (Eimer

and Balasubramanian, personal communication).

**Normalised smallRNA levels**

| | Plant system | Human cells |
|---|---|---|
| Normal cells | 17.72 | 89.06 |
| Affected cells | 201.92 | 316.88 |

**Table 1. GAA/TTC repeat expansion is associated with increased smallRNA production.** Comparison of smallRNAs that map to *IIL1* (in plants) and *FXN* (in humans) between normal and affected individuals identified through next generation sequencing of smallRNAs. Human cell data is obtained from single replicate. The plant data is from Eimer and Balasubramanian, personal communication.

The specific regions of the *FXN* gene to which the smallRNAs map are represented visually in Figure 1. Major increases in smallRNA abundance occur on Chr9 positions 71667502- 71667520 & 71670843-71670861 and again at the 3' end at position 71709755- 71709773. These 3 highly abundant smallRNA are 19 nt (Fig 1A).

Critical *FXN* regulation is known to occur at the 5' end (Fig 1B). In healthy individuals, the region indicated by * adjacent to the upstream of the repeat are hypomethylated, while regions downstream are hypermethylated making this region of the gene overall in euchromatin. This region is aberrantly methylated in individuals with pathologically expanded GAA/TTC tracts, such that the region upstream the repeat is now hypermethylated and the region following the repeat hypomethylated. This induces heterochomatin, which can prevent transcription from occurring. Thus contributing to the frataxin deficiency in FRDA (De Biase, Chutake et al. 2009, Evans-Galea, Carrodus et al. 2012, Evans-Galea, Lockhart et al. 2014).

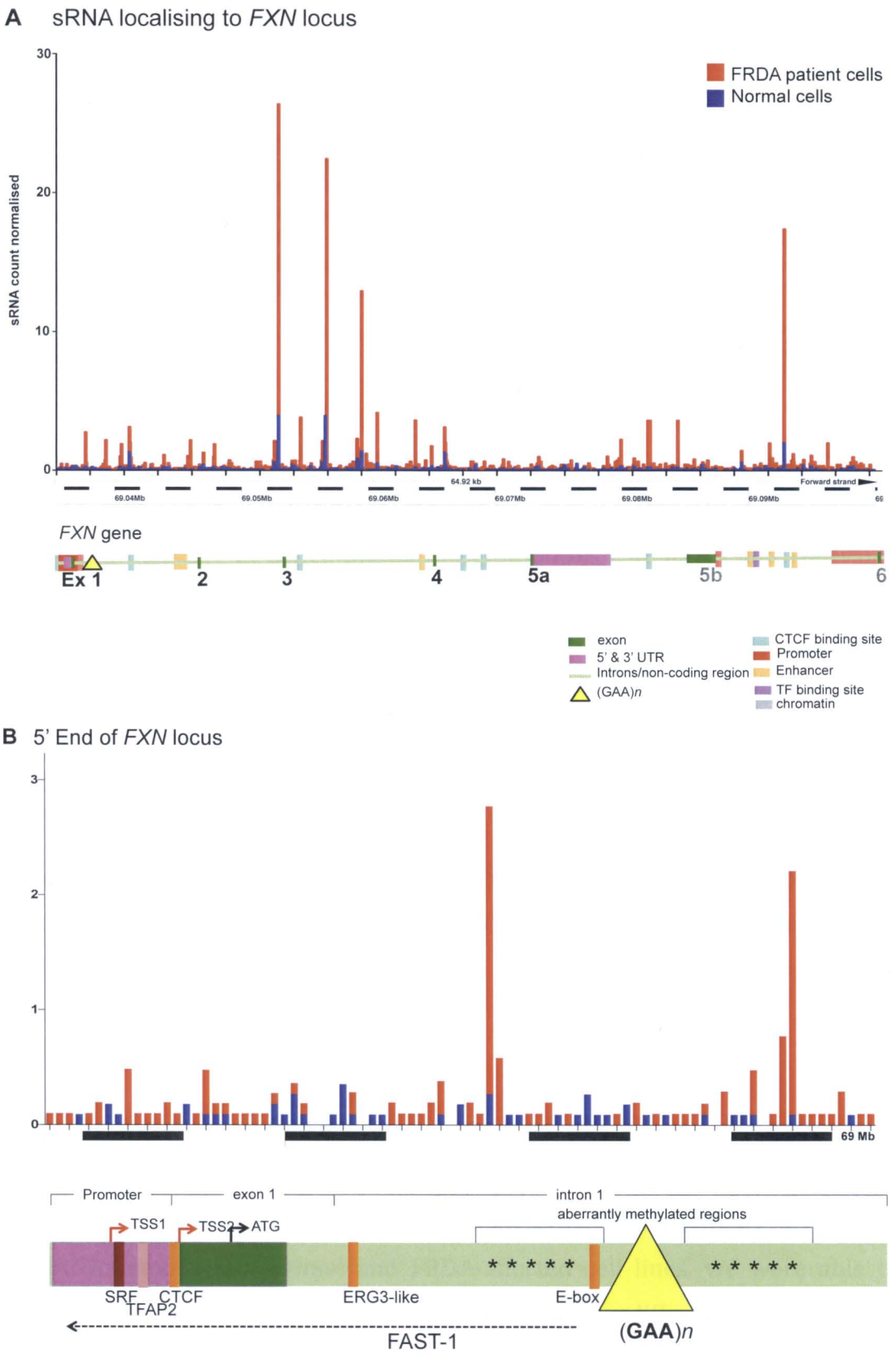



**Figure 1. Schematic representation of smallRNA distribution across the *FXN* locus in normal (blue) and FRDA-affected (red) cell lines. (A)** Distribution of smallRNA across the entire *FXN* gene relative to various genetic regions e.g. UTR, exon-intron and various gene regulatory sites. [NB: exons 5b and 6 are thought to be only present in a very minor alternative

isoform. For the most part human frataxin has 5 exons as indicated.] **(B)** Representation of the 5' region of *FXN* including the promoter, exon 1, and part of intron 1. Transcription start sites (TSS) are indicated by red arrows. The translation start site (ATG) is indicated by a black arrow. Several protein binding sites (labeled) and epigenetic changes known to be important for *FXN* regulation and silencing in Friedreich ataxia (FRDA) are also indicated. These include sites for the transcription factors serum response factor (SRF), transcription factor activator protein 2 (TFAP2), an early growth response protein 3-like factor (EGR3-like), and the chromatin insulator protein CCCTC-binding factor (CTCF) and a putative E-box binding protein. Dashed arrow represents *FXN* antisense transcript FAST-1 (which is dramatically increased in expanded *FXN* alleles). The two red peaks depict dramatic increases in the normalised smallRNA counts in regions immediately either side of the GAA/TTC repeat indicated by the yellow triangle at positions 71652807- 71652825 and 71654118- 71654135

We identified 2 regions flanking the repeat expansion at positions 71650666-71650686 and 71654118- 71654135 (5'-3') that have an ~10 and 21-fold increase respectively in siRNA targeting those regions (Fig 1B). Herman et al reported a decrease in active marks (acetylation of histone H3 and H4 lysine residues) and increase in repressive marks (di- and tri-methylation of histone H3 lysine 9) upstream and downstream of the GAA/TTC expansion respectively in primary lymphocytes from individuals with FRDA (Herman, Jenssen et al. 2006). These regions are indicated by * in Figure 1B and coincide with the genomic regions mapped by our upregulated smallRNA.

Aside from the overall increase in smallRNA specifically mapping to the *FXN* gene, there were very similar smallRNA profiles between the normal and FRDA-affected cell lines.

### 4.2.2 Differentially expressed miRNAs in FRDA cells

Since we carried out genome-wide analysis of differentially expressed smallRNAs among the normal and FRDA-affected cell lines, we were able to capture not only siRNAs, but also miRNAs that are differentially regulated. Therefore, first we analyzed whether known miRNAs that were previously found to be dysregulated were captured in our analysis. Consistent with previous findings (Mahishi, Hart et al. 2012), we found miR-886-3p was dysregulated in

our FRDA-affected cells, which was previously known to be associated with an increase in FXN deficiency (Table 3).

We then took a blind approach to identify miRNAs specifically modulated in FRDA cells as well as their putative targets, which were identified through computational analysis. Six of the most highly upregulated miRNAs in FRDA cells were selected for further analysis (Table 2). Interestingly, the specific genes identified as targets for the miRNAs in Table 2, are all associated with neuronal function implying a functional relevance for this altered smallRNA profiles.

| miRNA | log2Ratio (FA/cont) | Probability | target gene with miTG Score >0.99 | Target gene function | Phenotype | Phenotype in orthologue |
|---|---|---|---|---|---|---|
| hsa-miR-4286 | 3.94 | 0.96 | C2ORF21 | unknown | | |
| hsa-miR-181a-2-3p | 2.66 | 0.88 | SLC39A10 | May act as a zinc-influx transporter. | unknown | |
| hsa-miR-20b-3p | 2.62 | 0.84 | n/a | n/a | | |
| hsa-miR-3656 | 2.76 | 0.82 | MARK4 | kinase involved in microtubule bundle formation, nervous system devlopment and apoptosis | unknown | |
| hsa-miR-1260b | 3.95 | 0.95 | GDI1 | Regulates the GDP/GTP exchange reaction of most Rab proteins by inhibiting the dissociation of GDP from them, and the subsequent binding of GTP to them.Brain; predominant in neural and sensory tissues. Known to be involved in FXMR | MENTAL RETARDATION, X-LINKED 41, X-linked non-syndromic intellectual deficit, mental retardation X-linked type 41 (MRX41), mental retardation X-linked type 48 (MRX48) | |
| | | | GOLGA8F | unknown, potential product of pesudogene | unknown | |
| | | | TARDBP | DNA and RNA-binding protein which regulates transcription and splicing. May also be involved in microRNA biogenesis, apoptosis and cell division.Stabilizes the low molecular weight neurofilament (NFL) mRNA through a direct interaction with the 3' UTR. | AMYOTROPHIC LATERAL SCLEROSIS 10, WITH OR WITHOUT FRONTOTEMPORAL DEMENTIA, Amyotrophic Frontotemporal dementia with motor neuron disease | many neurological effects in zebrafish |
| | | | GOLGA8G | unknown whether protein encoding | unknown | |
| | | | ATF6B | Transcriptional factor that acts in the unfolded protein response (UPR) pathway by activating UPR target genes induced during ER stress. | unknown | |
| hsa-miR-619-5p | 2.63 | 0.83 | FOXG1 | Transcription repression factor which plays an important role in the establishment of the telencephalon. tissue specificity, regional subdivision of the developing brain and in the development of the Expression is restricted to the neurons of the developing telencephalon. | RETT SYNDROME, CONGENITAL VARIANT, 14q11.2 microduplication syndrome, 14q12 microdeletion syndrome, Atypical Rett syndrome, congenital variant of Rett syndrome (RTTCV) | many neurological effects in zebrafish |
| | | | DAPK1 | protein kinase. Mediates apoptosis especially in neurones | unknown | |

**Table: 2. microRNAs upregulated in FRDA-affected cells, their predicted target genes and target gene function.** Differential expression (DE) was calculated as log2Ratio (FA/cont). Six of the most highly DE miRNAs are displayed and represent an upregulation in miRNA of >2.5-fold in FRDA cells compared to control cells, with a probability of >80%.

To assess the biological impacts of the differential expression miRNAs we analysed the expression levels for their target genes. Expression of genes identified as the most likely targets for the differentially expressed (DE) miRNAs (Table 2) were quantified by q-RT-PCR analysis (Figure 2).

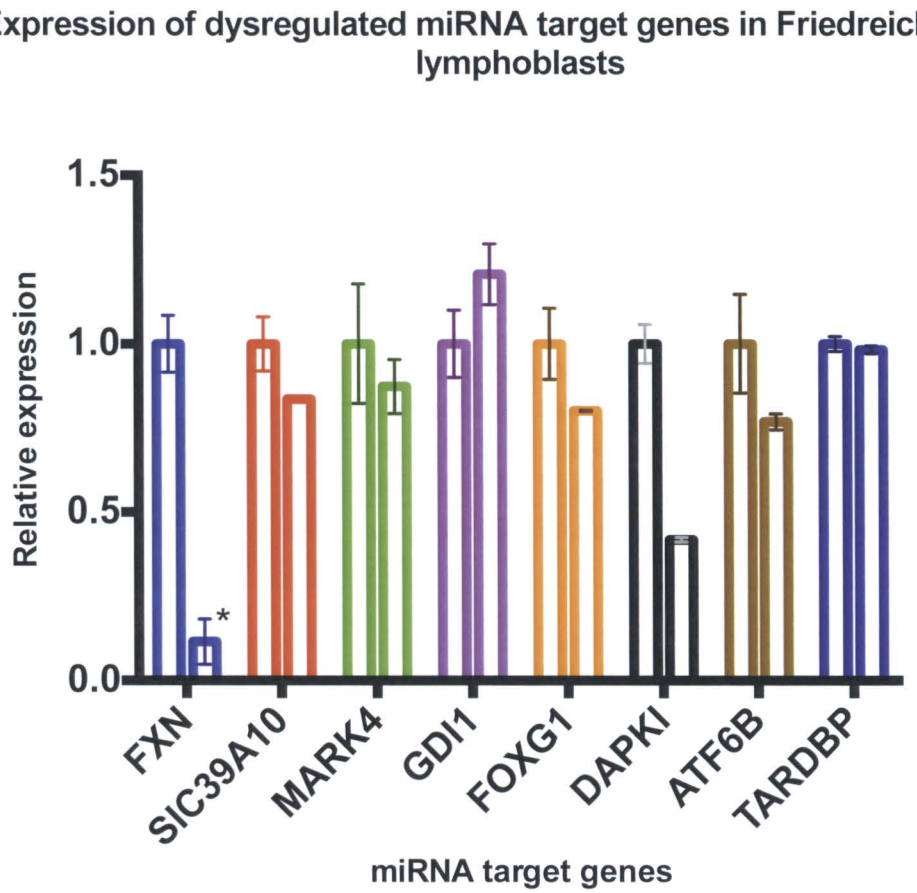



**Figure 2. Quantitative RT-PCR expression analysis of genes likely targeted, by miRNAs identified as highly upregulated in FRDA patient cells relative to control cell lines.** qPCR expression analysis on *FXN* and six other potential target genes for upregulated miRNAs. Data is grouped in pairs with control sample and FRDA-affected samples from left to right. Expressions of all samples were normalized to expression of *TUB2* reference gene.* Indicates a significant difference in expression of indicated gene between control and FRDA-affected samples. Error bars indicate standard error. All target genes, bar GDI1, display anti-correlation in their expression levels.

The difference in *FXN* expression between control and FRDA-affected patient cells was significant (*p*-value 0.0017). *FXN* expression in our FRDA-affected cells was found to be 10-20% that of control cells (Fig 2). Expression of *SIC39A10*, *MARK4*, *FOXG1*, *DAPKI*, *ATF6B* and *TARDBP* all exhibited decreased expression in FRDA cells compared to control. Although the decrease in expression was not found to be significant in our dataset (*p*-value <0.05), the trend in decreased expression in patient cells should be noted as decreased expression could suggest that the miRNAs upregulated in FRDA may be targeting those genes contributing to their decreased expression. *GDI1* was found to be much more highly expressed in both the control and patient cells, to a similar level as *FXN* in control cells whereas expression of all other genes are comparatively low.

| | |
|---|---|
| **miRNA highly differentially expressed** | hsa-miR-20b, hsa-miR-181a |
| **miRNA differentially expressed** | hsa-miR-886-3p*, hsa-miR-363, hsa-miR-21, hsa-miR-124, hsa-miR-130a, hsa-miR-10a, hsa-miR-505, hsa-miR-643, hsa-miR-155, hsa-miR-628-3p, hsa-miR-194, hsa-miR-192, hsa-miR-9, hsa-miR-212, hsa-miR-126, hsa-miR-335, hsa-miR-125b, hsa-miR-9, hsa-miR-132, hsa-miR-944, hsa-miR-500 |
| **miRNA not differentially expressed** | hsa-miR-886-5p, hsa-miR-923, hsa-miR-518f, hsa-miR-614 |

**Table 3. Previously identified dysregulated miRNA in Friedreich's ataxia**. Table displays the 27 miRNA identified as dysregulated in Friederich's ataxia affected cells by Rajiv R. Ratan et al. It can be seen that all bar 4 of those miRNA were also found to be dysregulated in our smallRNA seq data. 2 of which were found to be highly differentially expressed in our analysis and their biological impacts investigated further with their predicted target gene expression levels analysed. * indicates the miRNA found to contribute to FRDA pathology. [NB: Rajiv R. Ratan et al. Identified those miRNA using TaqMan Human MicroRNA Array Set v2.0 (Applied Biosystems by Life Technologies), which enables quantitation of 377 known human miRNAs. Our analysis identified 339164 smallRNA in control cell lines and 454322 smallRNA in FRDA cell lines]

Our results indicate a dysregulation in both miRNA and siRNA in FRDA. Our deep-smallRNA sequencing results are consistent with the literature (Mahishi, Hart et al. 2012) in that many of the miRNAs previously identified as dysregulated in FRDA were also found to be dysregulated in our study (Table 3).

Specifically, 27 out of the 30 miRNAs previously identified as disregulated in FRDA-affected cells by Mahishi, Hart et al, were also identified in our analysis (Table 3). Furthermore, additional miRNAs not previously identified as dyregulated, were uncovered in our study including; hsa-miR-4286, hsa-miR-3656, hsa-miR-1260b and hsa-miR-619-5p. All of these miRNA have target genes with functions known to be involved in the nervous system, oxidative stress and other TNRDs (except hsa-miR-4286 which has a currently unknown function)(Table 2). The functions of these target genes indicate a role in the biological pathways mediating FRDA pathology.

### 4.2.3 Compromising smallRNA production in FRDA cells leads to an increase in *FXN* expression.

In the plant system, there was found to be siRNAs, derived from the *IIL1* repeat expansion that appeared to play a role in reducing that gene's expression level (Eimer and Balasubramanian, personal communication). Those siRNAs were 24 nt. In *Arabidopsis*, DICER-LIKE 3 (DCL3) protein is responsible for production specifically of 24 nt smallRNAs. Thus, *DCL3* was targeted for silencing to specifically prevent the production 24 nt siRNAs without compromising the activity of other smallRNAs such as miRNAs and siRNAs of other lengths. Targeting *DCL3* through artificial microRNAs against *DCL3 (35S::amiRDCL3)* resulted in increased *IIL1* mRNA and protein expression and the suppression of the *iil* phenotype (Eimer and Balasubramanian, personal communication). In *Arabidopsis*, there are several Dicers responsible for processing miRNAs and siRNAs (Schauer, Jacobsen et al. 2002, Kurihara and Watanabe 2004, Lee, Nakahara et al. 2004). In humans however, there is only a single Dicer that processes both pre-miRNAs and pre-siRNAs. Thus blocking Dicer activity will prevent production of all miRNAs and siRNAs. Therefore, we took a chemical inhibition approach as opposed to RNAi mediated knocking down of the *DICER*, which will have relatively fewer unwanted effects on the growth of the cells. In humans, the ability to produce siRNAs and miRNAs of different lengths is reliant

on certain partner proteins that interact with Dicer. However, little is known about these accessory proteins (Noland and Doudna 2013). Therefore, we considered small molecules, which could inhibit the smallRNA machinery in general.

Recent studies have identified chemical modulators of miRNAs and siRNAs (Gumireddy, Young et al. 2008, Henn, Joachimi et al. 2008, Shan, Li et al. 2008, Deiters 2010, Watashi, Yeung et al. 2010, Young, Connelly et al. 2010) utilizing cell-based assays. One screen identified small molecules that inhibited loading of RNAs in to the RISC complex in cultured human cells, which included Aurintricarboxylic acid (ATA), Oxidopamine HCL (dHCL), and Suramin NaCl, with low IC50 values of 0.47, 1.61, and 0.69 μM respectively (Tan, Chiu et al. 2012). The compound ATA inhibits RNA binding to endogenous ARGONAUTE 2 (Ago2) and *de novo* RISC assembly, without disturbing preformed complexes of Ago2 with endogenous smallRNAs or inhibiting the catalytic activity of Ago2. ATA is a generic nuclease inhibitor, with potentially deleterious effects at high concentrations. However, in the relatively low concentration of 25 μM, ATA selectively inhibits loading of siRNAs to Ago2 without affecting other functions of the protein. Previous studies have shown that ATA is associated with low toxicity in tissue cultures (Hashem, Flaman et al. 2009, Tan, Chiu et al. 2012) and is well tolerated *in vivo* (Matsuno, Kozawa et al. 1998, Tan, Chiu et al. 2012).

Aurintricarboxylic acid (ATA) as described above was our main drug candidate for inhibition of siRNA production in our FRDA cell lines. As it has been identified as specifically inhibiting siRNA production in *H. Sapiens* (Tan, Chiu et al. 2012). We also tested Oxidopamine HCL (dHCL) and Suramin small molecule inhibitors of RISC loading. FRDA-affected cells treated with 25 μM ATA for 72 h were consistently found to increase *FXN* mRNA expression (Fig 3) and frataxin protein (Fig 4). Suramin and dHCL also gave a trend of increase in *FXN* expression upon treatment with Suramin and dHCL, though their levels of significance was $0.1>P<0.05$ (Fig 3).

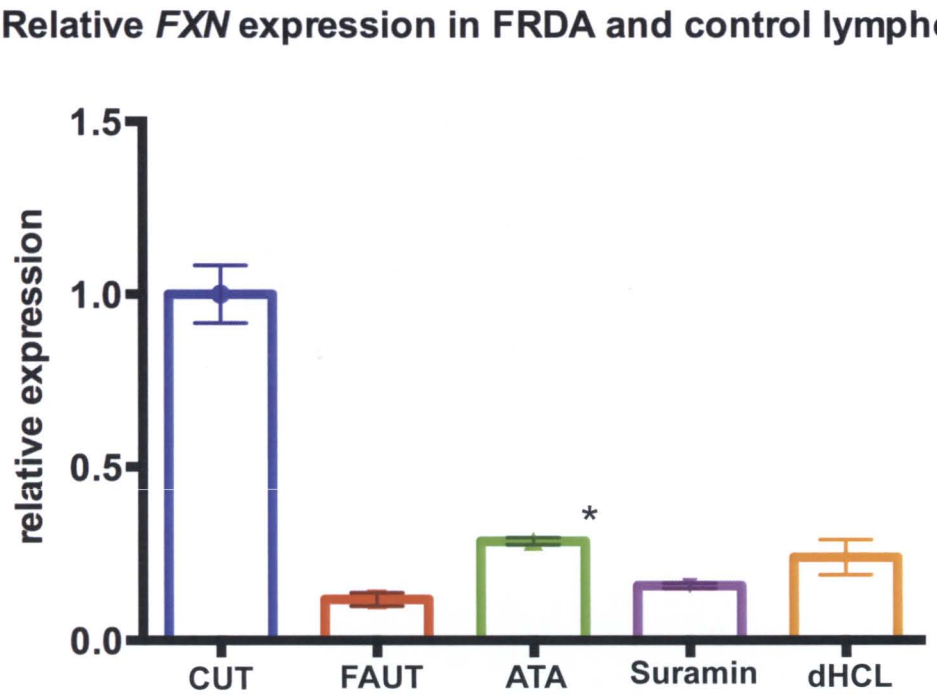



**Figure 3. Effects of Aurintricarboxylic acid (ATA), Suramin and Oxidopamine hydrochloride (dHCL) on *FXN* gene expression in FRDA-affected lymphoblasts *in vitro*.** *FXN* qPCR expression analysis. 21 biological replicates of control untreated cells (CUT) (0.1% DMSO treated) (blue), 20 biological replicates of untreated (0.1% DMSO treated) FRDA cells (FAUT) (red) and 31 biological replicates of FRDA-affected patient cells treated with 25 μM ATA, Suramin or dHCL for 72 h (green) were analysed for affects on *FXN* expression. * Indicates significant increase in *FXN* expression *p*-value <0.05. Error bars indicate standard error.

Here the FRDA untreated patient cells expressed 11.87% of the control cell frataxin where as the ATA treated samples expressed 28.86% of control cell frataxin. This is a 16.99% increase in *FXN* expression (*p*-value 0.0045). Suramin (*p*-value 0.103234) and dHCL (*p*-value 0.089) showed a trend of increase in *FXN* expression although they were not statistically significant.

Western blot analysis was carried out on control cells, FRDA-affected cells and FRDA-affected cells treated with 25 μM ATA for 72 h to test in a semi-quantitative manner, whether the observed increase in *FXN* mRNA expression translated to a detectable increase in frataxin protein production.

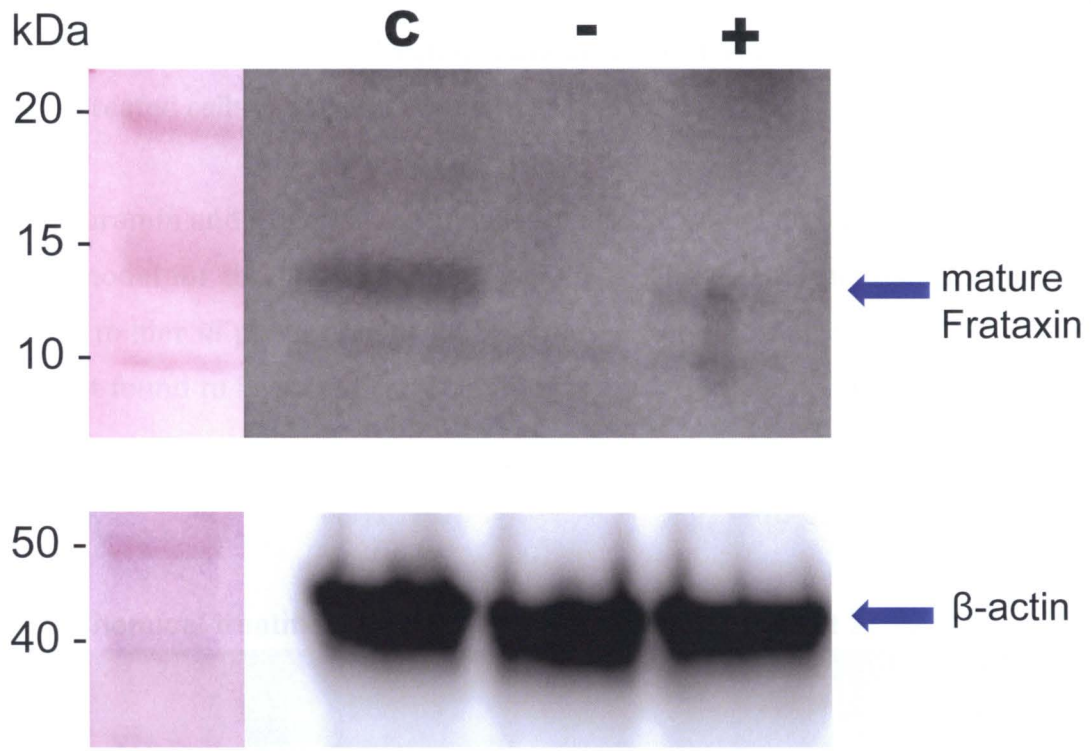



**Figure 4. Frataxin protein levels in Friedreich's ataxia patient cells treated with Aurintricarboxylic acid (ATA): Western blot analysis.** Whole-cell lysates of control lymphoblasts (Coriell #GM15851) and FRDA-affected lymphoblasts (Coriell #GM15850) +/- ATA treatment. For frataxin, specific bands at approximately 14 kDa was believed to represent mature frataxin, weaker bands at approximately 17 kDa are thought to represent larger frataxin isoforms. Specific band for β-actin protein control at approximately 45 kDa. [NB: With regard to the literature (Rufini, Fortuni et al. 2011) the second band corresponds to the mature frataxin].

Consistent with q-RT-PCR results (Fig 3), we found a potential increase in frataxin protein levels upon ATA treatment although the increase was relatively minor. Nevertheless, this highlighted in potential for ATA for further analysis and the potential for exploiting this pathway as a therapeutic target (Fig 4). Based on *FXN* expression results (Fig 3), Suramin and dHCL were also identified as potential compounds for FRDA therapeutics as both displayed increases in *FXN* expression with p-values ≤0.1. The presence of the repeat expansion was confirmed to ensure that the increased *FXN* production is despite the GAA/TTC repeat expansion (Fig 8).

Western blot analysis was also undertaken on cells treated with Suramin and dHCL but was not found to display a noticeable difference in frataxin relative to the untreated cells (results not shown).

ATA, Suramin and dHCL compounds were also screened on our Bur-0 *A. thaliana* plant model for the *iil* phenotype (Fig 5) The ATA treated plants did appear to have a milder *iil* phenotype at 27 °C although the difference in *IIL1* expression was not found to be statistically significant upon qPCR expression analaysis (Fig 5C).

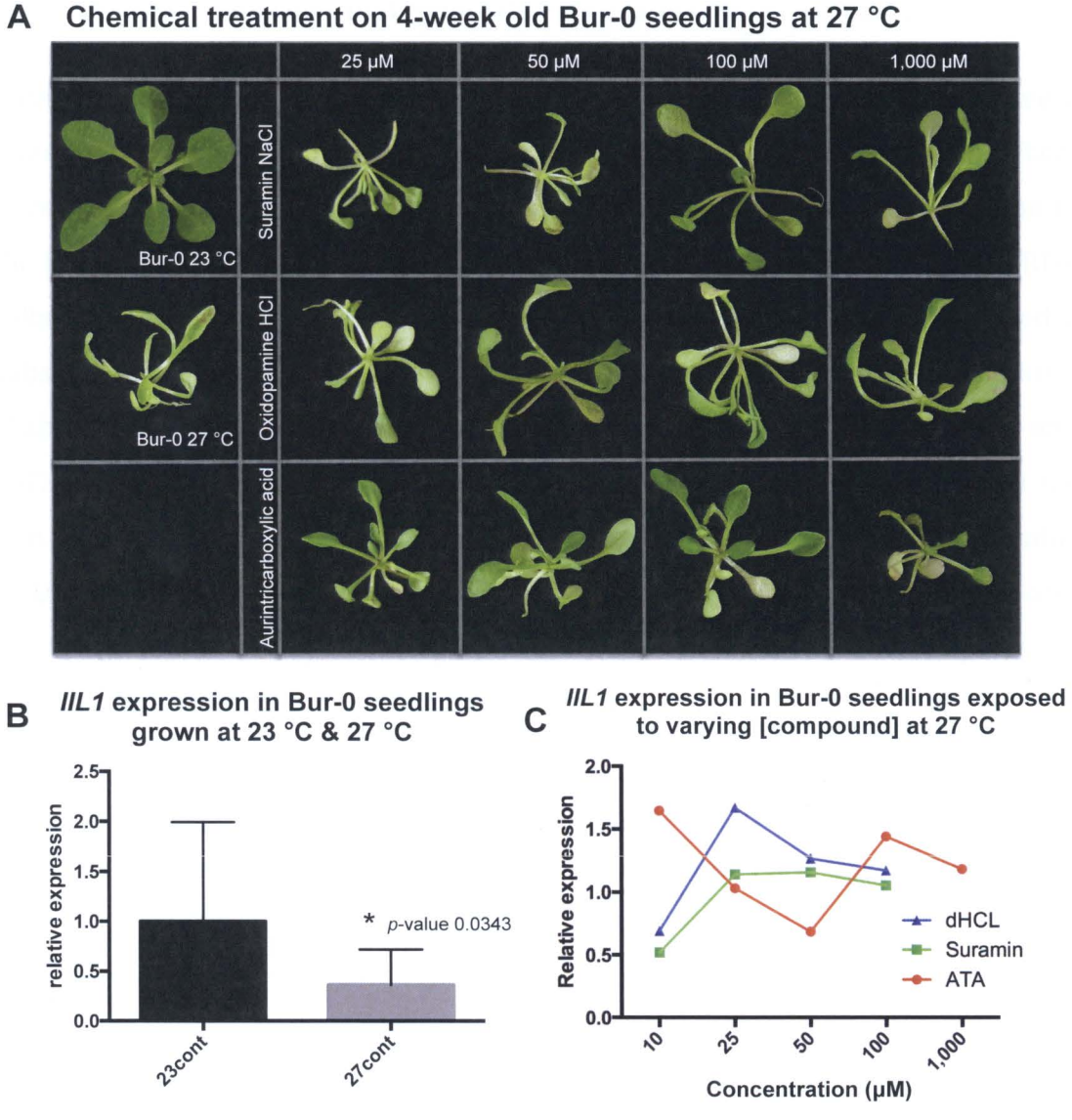


**Figure 5. SmallRNA inhibiting compound screen on 27 °C Bur-0 seedlings. (A)** Phenotypes of representative Bur-0 seedlings from each treatment group. Plants were grown in presence of Aurintricarboxylic acid (ATA), Oxidopamine hydrochloride (dHCL) and Suramin sodium salt of 25, 50, 100 and 1,000 μM concentrations. Bur-0 27 °C shows the *iregularly impaired leaves* (*iil*) phenotypic growth defects manifested as tapered, twisted leaves, which is caused by a TTC/GAA intronic expansion at the *IIL1* locus. Suramin NaCl appears to have a recovered phenotype at higher concentrations (100-1,000 μM). dHCL appears to have a potentially less severe phenotype across all concentrations and ATA appears to have only a very mild *iil* phenotype across all concentrations, evident as less twisted, more rounded leaves, especially in the younger leaves. **(B)** *IIL1* gene expression levels in untreated Bur-0 seedlings at 23 °C and 27 °C. At 27 °C, Bur-0 has a relative decrease in expression ~3-fold (*p*-value 0.0343). **(C)** Relative *IIL1* expression in small molecule treated Bur-0 at 27 °C across [10-1,000 μM]. Expression was normalized to *TUB2* gene and results graphed relative to the Bur-0 27 °C untreated samples. A statistically significant difference in *IIL1* expression in any of the treated plants relative to the 27 °C Bur-0 untreated control was not found. Error bars indicate standard error.

### 4.2.4 Additional molecules that modulate *FXN* expression in FRDA cells

Having discovered ATA can change *FXN* expression levels and seem to have a phenotypic impact in plants, we sought to look for additional compounds that have the potential to increase *FXN* expression. Previously a screen was done in the lab, which had identified potential candidate molecules for analysis in FRDA cells (Tasset and Balasubramanian, personal communication). We analysed a subset of 18 compounds for their affects on *FXN* expression in the FRDA cell lines (Table 1). Initially 10 μM of each compound were cultured with normal and FRDA cells and expression level quantified by qRT-PCR (Fig 6). Here we can see that 2 of the compounds, Visnagin (*p*-value <0.027) and Sphondin (*p*-value <0.01) as well as the previously identified compound Aurintricarboxylic acid (ATA) (*p*-value <0.03) all display significant increase in *FXN* expression. Conversely, compounds Fumarprotocetraric acid (*p*-value <0.016) & Carylophyllene oxide (*p*-value <0.022) display significant decreases in *FXN* expression relative to the FRDA untreated cells (FAUT). The remaining compounds were found to have no significant impact on *FXN* expression.

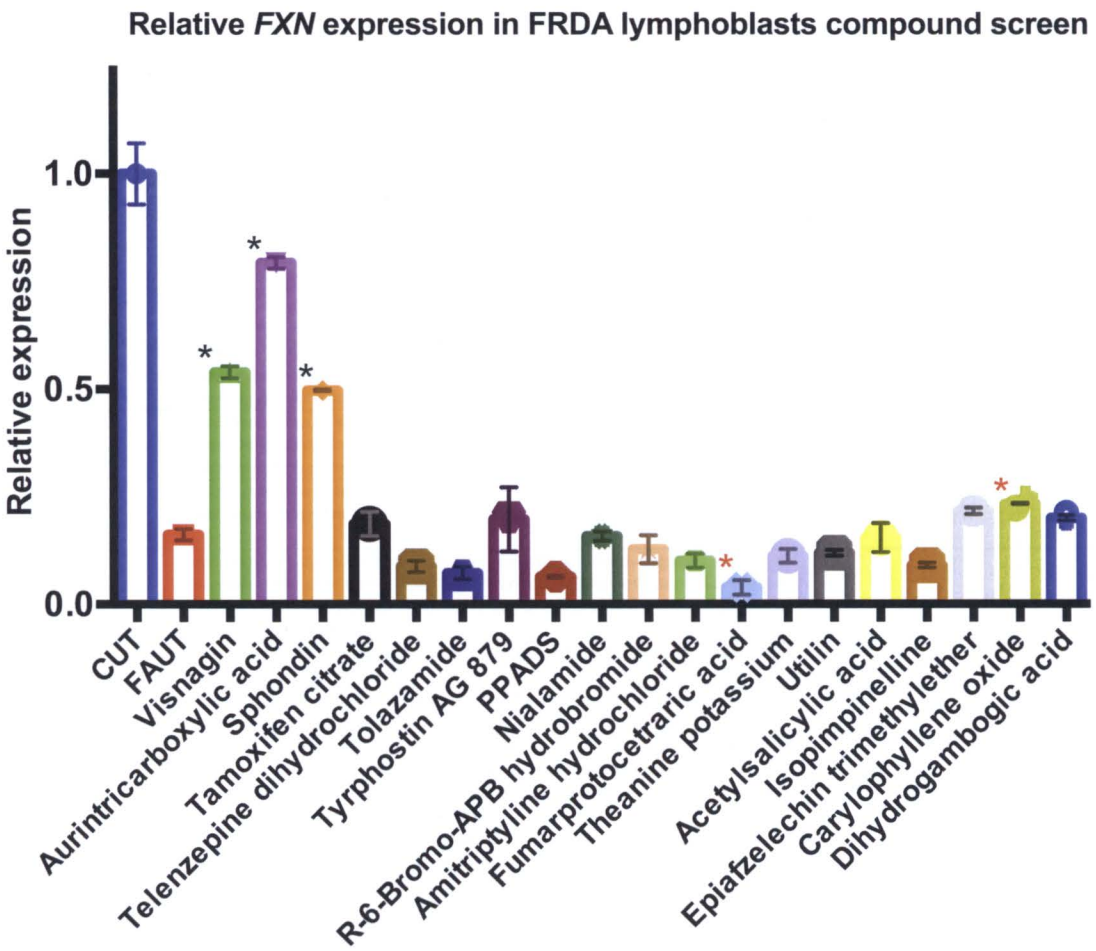



**Figure 6. Effects of various chemical compounds on *FXN* gene expression in FRDA-affected lymphoblastoid cells.** *FXN* expression was measured for each compound relative to the control untreated cells in their particular assay and qPCR run. With these particular cell lines the Friederich's ataxia untreated (FAUT) samples had an expression of 16.19% that of control untreated (CUT) *p*-value 7.3E-08. 48 biological samples of CUT and FAUT were used to generate this value with a SEM 0.0713 & 0.0128 respectively. Here, 19 compounds are presented. * Indicates compounds that produced a significant increase in *FXN* expression in FRDA cells. * (red asterisk) Indicates compounds that produced a significant decrease in *FXN* expression in FRDA cells. Student's two sided *t*-tests were used to determine if datasets were significantly different from each other. Error bars indicate standard error.

No significant changes were observed in control treated cells for any compound except for Visnagin (*p*-value <0. 026), Aurintricarboxylic acid (*p*-value <0. 0054), & Sphondin (*p*-value 0. 048) which also produced significant increases in *FXN* expression for control treated cells relative to control untreated cells (Results not shown).

| Compound | Biological activity/properties | plant system phenotype (+/-) | plant system expression (+/-) | human system expression (+/-) |
|---|---|---|---|---|
| Aurintricarboxylic acid | Readily polymerizes in aqueous solution, forming a stable free radical that inhibits protein-nucleic acid interactions. Clinical use inhibits viral reproduction (Hashem, Flaman et al. 2009) | + | nsd | + |
| Oxidopamine HCL | Neurotoxic compound selectively destroy dopaminergic neurons (Tiurenkov, Volotova et al. 2014). | + | nsd | + |
| Suramin sodium salt | Broad-spectrum antagonist of ATP receptors. Agonist of calcium channels in various excitable tissues e.g. muscles and neurons (Menezes, Barbosa et al. 2011). Activity against activity against RNA viruses (Ruprecht, Rossoni et al. 1985) | + | nsd | + |
| Visnagin | Vasodilator, reduces blood pressure by inhibiting calcium influx into cells. Protects against doxorubicin-induced cardiomyopathy through modulation of mitochondrial malate dehydrogenase (Lee, Jung et al. 2010) | + | + | + |
| Dihydrogambogic acid | Displays cytotoxicity towards a wide variety of tumor cells and has been shown to affect many important cell-signaling pathways (Felth, Lesiak-Mieczkowska et al. 2013) | + | + | nsd |
| Carylophyllene oxide | Constituent of essential oil of hemp Cannabis sativa. Selective agonist of cannabinoid receptor type-2. Exert significant cannabimimetic antiinflammatory effects, antinociceptive, neuroprotective and antidepressant (Amenta, Jallo et al. 2014). Approved by FDA. | + | + | - |
| Epiafzelechin trimethylether | Inhibitors of EP2 Prostaglandin E2 Receptor and sodium fluorescein uptake (Pang, Abeysinghe et al. 2013) | + | + | nsd |
| Isopimpinelline | Naturally occurring coumarin (product found in Apiaceae). Some evidence for inhibition in skin and breast cancers (Prince, Campbell et al. 2006). | - | - | nsd |
| Theanine potassium | Herbal extract from green tea. May have activity in modulating the metabolism of cancer chemotherapeutics agents. Cognitive enhancer and displays neuroprotective effects *in vivo*. Promotes self-renewal of human embryonic stem cells (hESC) (Atkinson, Lako et al. 2013). | + | + | nsd |
| Utilin | Stimulant of immune system. Stimulates T-lymphocyte population to increased release of lymphokines (Wang and Sun 2007). | + | + | nsd |
| Fumarprotocetraric acid | Antioxident properties. Found in some Lichens (de Barros Alves, de Sousa Maia et al. 2014). | + | + | - |
| Sphondin | Furanocoumarin derivative may have the therapeutic potential as an anti-inflammatory drug on airway inflammation by inhibiting IL–1β-induced COX–2 expression (Yang, Liang et al. 2002) | - | - | + |
| Acetylsalicylic acid | AKA Aspirin- salicylate drug, often used as an analgesic to relieve minor aches and pains, as an antipyretic to reduce fever, and as an anti-inflammatory medication (Lei, Zhou et al. 2014). FDA approved | + | + | nsd |
| Amitriptyline hydrochloride | Inhibits presynaptic reuptake of norepinephrine and serotonin in CNS. Management of chronic pain associated with migraine, tension headache, phantom limb pain, tic douloureux, diabetic neuropathy, peripheral neuropathy, cancer or arthritis; treatment of panic and eating disorders (Javed, Petropoulos et al. 2015). | + | - | nsd |
| R-6-Bromo-APB Hydrobromide | D1 Dopamine receptor agonist; potent enantiomer (Azaryan, Clock et al. 1996). | + | + | nsd |
| Nialamide | Non-selective, irreversible monoamine oxidase inhibitor. Mood-stimulating effects, previous use as antidepressant (Bilkei-Gorzo, Michel et al. 2007), withdrawn due to hepatotoxicity effects. | + | nsd | nsd |
| PPADS | Selective purinergic P2X antagonist. Inhibits excitatory postsynaptic responses including fast transmission at central synapses, contraction of smooth muscle cells, macrophage activation and apoptosis. Similar activity to Suramin (Menezes, Barbosa et al. 2011) | + | nsd | nsd |
| Tyrophostin AG 879 | Protein tyrosine kinase inhibitor. Inhibits nerve growth factor (NGF)- dependent TrkA tyrosine phosphorylation in PC-12 cells (Arias, Valle-Leija et al. 2014). Antitumoral effects on breast cancer cells | + | + | nsd |
| Tolazamide | Oral blood glucose lowering drug used for treatment of Type 2 diabetes. Lowers blood glucose acutely by stimulating the release of insulin from the pancreas by binding to ATP-sensitive potassium-channel receptors on the pancreatic cell surface, reducing potassium conductance and causing depolarization of the membrane. Depolarization stimulates calcium ion influx through voltage-sensitive calcium channels, raising intracellular concentrations of calcium ions, which induces the secretion, or exocytosis, of insulin (Schmitt and Johns 1995). | + | + | nsd |
| Tamoxifen citrate | Antagonist of the estrogen receptor in breast tissue via its active metabolite, 4-hydroxytamoxifen (Vachon, Schaid et al. 2015). Uses include- anticancer drug used to treat breast cancer. Improve fertility, treat premature puberty, bipolar disorder and thyroid conditions | + | - | nsd |
| Telenzepine $H_2Cl$ | Potent and selective muscarnic receptor antagonist reported to bind to mAChR M1. Inhibits gastric acid secretion (Kita, Ago et al. 2014). | + | nsd | nsd |

**Table 4. List of 21 candidate compounds screened on Bur-0 plants at 27 °C and FRDA-affected human cell lines.** Where + indicates an improvement in TTC/GAA expansion-associated phenotype (*iil* growth defects) / significant increase in gene expression. While – indicates a worsening of disease phenotype / significant decrease in gene expression. Nsd indicates 'no statistical difference' based on *t*-test analysis

Our results indicate that compounds Visnagin, Sphondin, ATA, dHCL and Suramin can significantly increase *FXN* gene expression in the FRDA cells. Visnagin was found to have the greatest affect in the plant system increasing *IIL1* expression and suppression of the *iil* phenotype. This compound was also identified as producing the greatest increase in *FXN* expression out of all 18 compounds screened on the FRDA cells (second only to ATA) (Fig 6). Conversely, Sphondin was identified through the plant screen as noticeably worsening the TTC/GAA expansion-associated growth defects and was associated with a significant decrease in *IIL1* expression. PCR analysis also revealed increases in genomic TTC/GAA expansion tract lengths and increased repeat instability (results not shown). However, Sphondin treatment on FRDA cells revealed a significant increase in *FXN* expression, with no detectable change in TTC/GAA length (Table 5 & Fig 8).

Carylophyllene oxide and Fumarprotocetraric acid were identified as improving Bur-0 phenotype and increasing gene expression. However, these compounds were identified as significantly decreasing *FXN* expression in the human cells. Whereas, Isopimpinelline noticeably worsened the *iil* growth defects and displayed decreased *IIL1* gene expression in plants as well as decreasing *FXN* gene expression in the FRDA cells. However, the decrease in expression was not significant in the human system.

Dihydrogambogic acid, Epiafzelechin trimethylether, Acetylsalicylic acid and Tyrophostin AG 879 all resulted in the suppression of the *iil* phenotype and increased gene expression in the plant system. Our results indicate they also increase *FXN* expression (Fig 6 & Table 4) in the human system, however the increase was not deemed significant with *t*-test analysis. Compounds Theanine potassium, Utilin, R-6-Bromo-APB hydrobromide and Tolazamide displayed an

in increase in *IIL1* expression in the plant system but a decrease in *FXN* expression in the cell lines, though this difference was again not significant.

Amitriptyline hydrochloride, Nialamide, PPADS, Tamoxifen citrate and Telenzepine dihydrochloride all exhibited an improved phenotype in the plant system but no significant difference in gene expression or in the case of Amitriptyline hydrochloride and Tamoxifen citrate, a decrease in *IIL1* expression. Treatment of the FRDA cells with these compounds similarly produced no significant difference in *FXN* expression.

The biological function of the compounds screened provides an insight into the pathways through which they may act to alter *IIL1* and *FXN* gene expression. Suramin and ATA both have activity against viral replication, specifically by recognizing viral dsRNAs. Most of the other drugs inhibit cell-signaling pathways. Potentially aspects of those pathways are important in *IIL1* and *FXN* expression (Table 1).

#### 4.2.5 Compound optimization *in vitro* assays

Given the promising effects ATA, Visnagin and Sphondin have on *FXN* expression levels in FRDA cells we next sort to optimize the concentration and duration of chemical treatment. With the intitial assays 10 μM of each compound was used and treatment was left for 72 h. There is potential that slightly higher or lower concentration and duration of treatment could have a more significant affect on *FXN* expression. Finding the concentration at which the compounds induce greatest increases in *FXN* expression with minimal adverse affects on cells is important when determining optimal concentration for further progression of candidate drugs in, *in vivo* pharmaceutical animal trials.

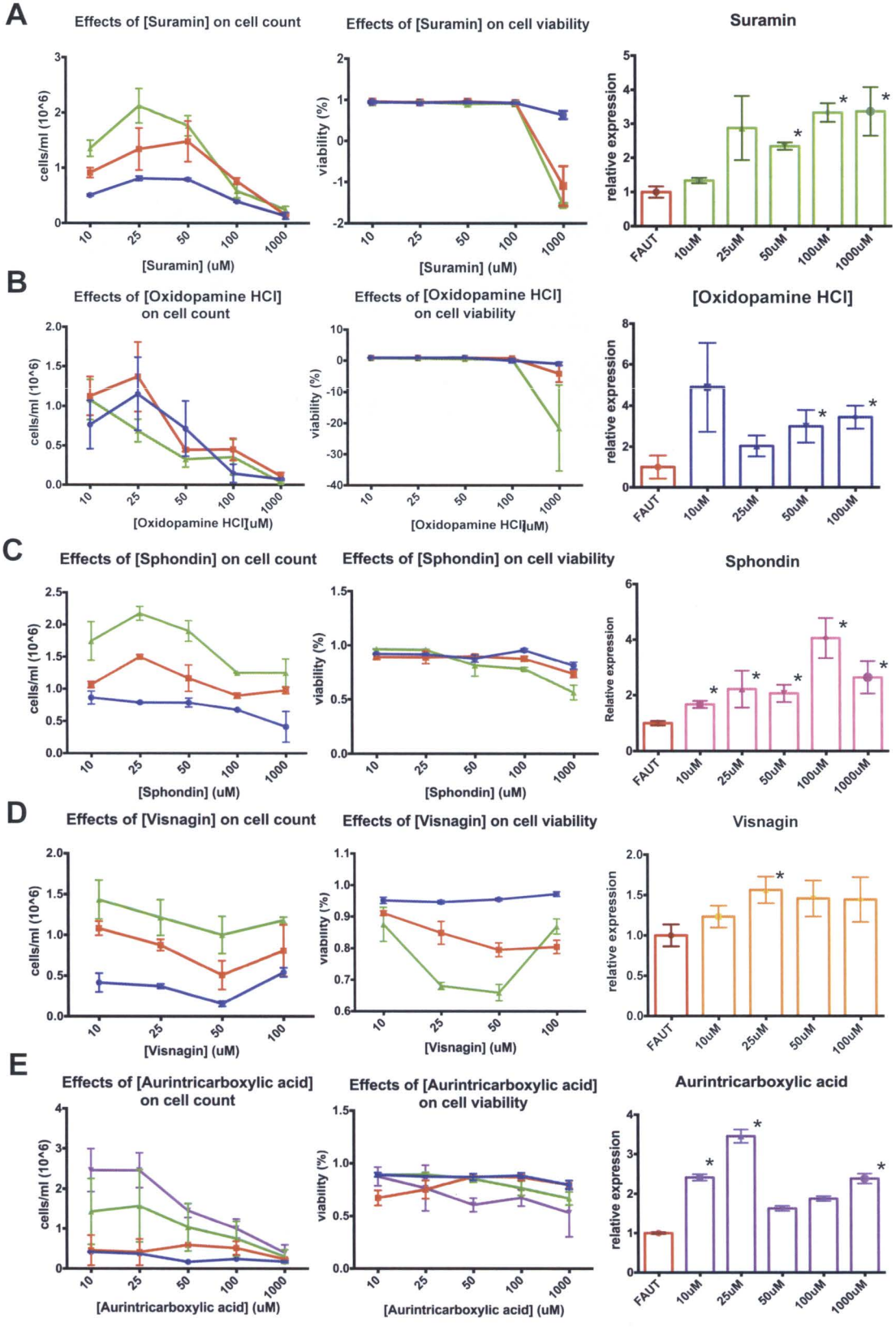



**Figure 7. Compound optimization assay.** Figure displays the affects on cell count/ proliferation, cell viability and *FXN* gene expression at various compound concentrations (y-axis) and treatment duration time points (coloured as indicated) for 5 compounds. Where **(A)** Suramin, **(B)** Oxidopamine HCl,(dHCL) **(C)** Sphondin, **(D)** Visnagin and **(E)** Aurintricarboxylic

acid (ATA). * indicates significant increase in *FXN* expression levels relative to Friedreich'a ataxia untreated (FAUT) cells as determined by *p*-value >0.05 with student-two-sided t-test.

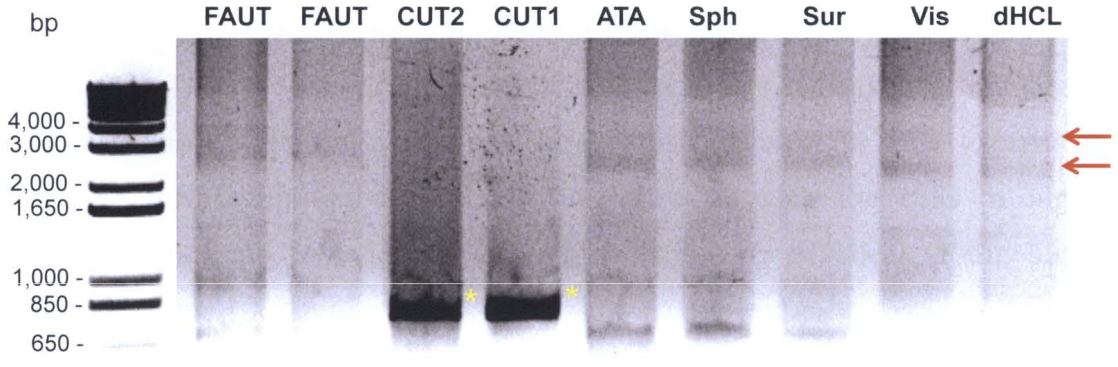



**Figure 8. *FXN* GAA/TTC repeat tract lengths of compound-treated FRDA cells.** Here PCR products from Friederich's ataxia (FRDA) untreated cells (FAUT), Control untreated cells and FRDA cells treated with 25 μM Aurintricarboxylic acid (ATA), 100 μM Sphondin (Sph), 100 μM Suramin (Sur), 25 μM Visnagin (Vis) and 10 μM Oxidopamine hydrochloride (dHCl) from left to right respectively. Red arrows indicate 2 bands of expected GAA/TTC expansion lengths in FRDA cells. PCR amplification of GAA/TTC expansion tracts revealed no detectable change in expansion tract length between FAUT and treated cells amongst any of the compounds. * (Yellow asterisks) indicate control sample PCR products with GAA/TTC repeat tracts within the normal range. [NB: PCR analysis was done on the same samples from which *FXN* expression data was obtained thus excluding possible variability amongst biological replicates.]

#### *4.2.5.1 Suramin*

No significant difference was found between *FXN* expression levels for different durations of Suramin chemical treatment. Thus, expression data presented in Fig 7A is constructed from pooled time points. Fig 7A cell count presents a steady decrease in cell proliferation at the higher compound concentrations of 100 μM and 1,000 μM for all treatment durations and an accompanying drop in cell viability at 1,000 μM especially for the longer duration periods. This suggests that at higher concentrations of Suramin (100-1,000 μM) there in a steady decline in cell proliferation and if cells are left in the presence of these higher concentrations for long time periods, namely 96 h and to a greater extent 120 h, there is a dramatic affect on their viability with < ½ of the cells non-viable.

Treatment of FRDA samples with Suramin at 50 μM (*p*-value 0.0509), 100 μM (*p*-value 0.0178) and 1,000 μM (*p*-value 0.0227) concentrations gave significant increases in *FXN* expression relative to FAUT. The expression level was increased by ~ 2.3 fold for 50 μM, and ~3.3 fold for both 100 μM and 1000 μM Suramin treatments. Thus these experiments indicate that in this cell based system higher concentrations of Suramin induce the greatest increase in *FXN* expression. However this compound appears to be more toxic to the cells at higher concentrations and the viability of cells may not be practical at the higher concentrations.

#### *4.2.5.2 Oxidopamine hydrochloride (dHCl)*

dHCL also exhibited no significant difference in *FXN* expression between treatment time points. Thus 72 h, 96 h and 120 h expression data was combined for each concentration. No expression data was obtainable from 100 μM treatments at 96 h and 120 h time points or from 1,000 μM treatments at any time point as higher concentrations of dHCL proved to be quite toxic to the cells and there was insufficient quality & quantity RNA extracted for expression analysis as most cells did not grow and died at higher concentrations (Fig 7B). Although there is a steady decline in cell proliferation as [dHCL] increases, there appears to be a buffering zone from concentrations 10-100 μM where the viability is not noticeably affected. That is within that concentration range the cells decrease in their proliferation but do not start to die until 1,000 μM at which concentration a notable death in the cells can be seen, especially at 120 h treatment duration.

dHCL at [10 μM] gave the greatest increase in *FXN* expression with an ~4.9-fold increase, though due to variability within biological replicates it had an significance of *p*-value 0.0959. 25 μM displayed no significant difference, 50 μM

exhibited a significance of *p*-value 0.0531 and 100 μM was strongly significant (*p*-value 0.0137). No values were obtained from 1,000 μM.

#### *4.2.5.3 Sphondin*

No significant difference was found between *FXN* expression levels for different durations of Sphondin chemical treatment. Thus expression data was combined in the associated graph. All Sphondin concentrations resulted in an increase in *FXN* expression from the untreated FRDA cells. 10 μM Sphondin (*p*-value 0.0014), 25 uM (*p*-value 0.0476), 50 μM (*p*-value 0.008), 100 μM (*p*-value 0.0032) and 1,000 μM (*p*-value 0.0352). The Sphondin treated cells at [100 μM] also exhibited a significant increase in *FXN* expression over [50 μM] and [10 μM] samples. This suggests that over the concentration range tested sphondin concentrations of 100-1,000 μM produce the greatest increase in expression.

Sphondin was noticeably milder on the cells with cell viability and cell count not statistically affected by Sphondin relative to the untreated cells, in the concentration range used. Thus as the increasing concentration does not seem to have a negative impact on cell proliferation the higher concentrations would be a good staring point for further analysis. In this optimization analysis Sphondin treatment at 10 μM resulted in an increase in *FXN* expression by ~1.6 fold whereas Sphondin treatment at 100 μM resulted in an increase in *FXN* expression by ~4-fold.

#### *4.2.5.4 Visnagin*

No significant difference was found between *FXN* expression levels for different durations of Visnagin treatment and as such expression data was combined for all time points. Visnagin at different concentrations had no significant affects on cell proliferation or viability. It also had the most mild affect on *FXN* expression

with 10 μM (*p*-value 0.181) displaying an insignificant affect on expression, 25 μM (*p*-value 0.008) displaying a significant increase in *FXN* expression relative to FAUT, and 50 μM and 100 μM treatments displaying *p*-values of 0.061 and 0.115 respectively. Additionally, Visnagin had a much smaller increase in *FXN* expression ~0.5-fold relative to untreated cells for 25 μM treatment.

#### *4.2.5.5 Aurintricarboxylic acid (ATA)*

It can be seen in Fig 7E that as the duration of treatment increases and the concentration of ATA compound increases the cell proliferation rate declines. However, the cell viability does not seem to be affected by the higher concentrations until the duration of treatment exceeds 144 h (4 days).

In this assay lower concentrations of ATA, in the range 10-25 μM, had the most significant increases in *FXN* expression with ATA 10 μM (*p*-value 0.019) and 25 μM (*p*-value 0.045), with accompanying ~2.5 and ~3.5-fold increases in expression respectively, relative to FAUT. ATA of 50 μM (*p*-value 0.055), 100 μM (*p*-value 0.06) and 1,000 μM (*p*-value 0.043) also displayed statistically significant increases in expression but with only ~0.5-1-fold increases in expression level.

With this further information regarding the affects on cell proliferation and viability as well as *FXN* expression levels at different concentrations for each of the compounds, we were able to determine the optimal concentration of each compound that leads to the greatest increase in *FXN* expression with minimal negative effects on cell proliferation and viability (Table 2).

| Compound | Concentration (μM) | Fold increase in *FXN* expression | *p*-value |
|---|---|---|---|
| Sphondin | 100 | 3.89 | 0.003220718 |
| Suramin | 100 | 3.34 | 0.017880634 |
| Oxidopamine HCl | 10 | 4.93 | 0.095986594* |
| Visnagin | 25 | 0.64 | 0.008965761 |
| Aurintricarboxylic acid | 25 | 3.51 | 0.045574509 |

**Table 5. Optimized compound treatments on Friedreich's ataxia (FRDA) cells.** Displaying optimal concentration for treatment with associated fold increase in *FXN* expression and *p*-value. Treatment duration had no affect for any compound at optimal concentration determined. * Oxidopamine HCl of *p*-value <0.1 significance.

## 4.3 DISCUSSION

### 4.3.1 Short interfering RNAs (siRNA)s play a role in transcriptional down regulation of *FXN* seen in FRDA

We have presented for the first time, a comprehensive overview of differentially expressed smallRNAs in FRDA-affected cell lines, including known and novel miRNAs (Tables 2 & 3) that may be contributing to FRDA pathology. We have shown an increase in smallRNAs, which is associated with the repeat expansion similar to the plant system, suggesting that repeat expansion can trigger smallRNA production, which in turn may contribute to the downregulation of *FXN* gene expression. Our results have found a ~3.5 fold increase in smallRNAs mapping to the *FXN* gene in FRDA cells (Table 1 & Fig 1). This enrichment of smallRNAs could well be contributing to the epigenetic silencing of *FXN* expression.

SmallRNAs are key epigenetic regulators of gene expression and disruptions in their expression have been linked to countless disorders including many TNRDs (Watanabe, Totoki et al. 2008, De Biase, Chutake et al. 2009, He and Todd 2011, Kumari and Usdin 2012). The 5' end of *FXN* is an important region for regulation

of expression. In this region we found ~10 and 21-fold increase in smallRNAs localizing to the adjacent regions upstream and downstream to the GAA/TTC repeat respectively (Fig 1B). In FRDA regions on either side of the GAA/TTC repeat are aberrantly methylated which induces a previously euchromatin state to a heterochromatin state forming an epigenetic block to transcription at the loci upstream proximal to the GAA/TTC repeat. One hypothesis is that the smallRNAs localizing to those same regions known to have aberrant methylation, may be acting through a siRNA-directed DNA methylation pathway to epigenetically silence *FXN* expression.

However, it's worth noting that our analysis was done on lymphoblast cells, which are not a pathological tissue in FRDA. There may be tissue specific toxic smallRNAs present in major sites of disease pathology e.g. the central and peripheral nervous system and/or the heart. As epigenetic regulation is largely tissue-specific (Lau, Hanel et al. 2004, Xia, Cao et al. 2012) it would be interesting to analyse the smallRNA profiles of clinically affected tissues. This is a future direction that could be done through the same deep-smallRNA sequencing approach.

### 4.3.2 Manipulating smallRNA pathways can help overcome FXN deficiency

It is currently unclear how these smallRNAs may be contributing to the problems in FRDA. However, irrespective of the mechanism, if they are indeed contributing to the *FXN* deficiency, targeting those aberrant siRNAs may have potential as a new type of gene therapeutic agent. Our results with Aurintricarboxylic acid (ATA) as well as small molecule inhibitors of mi-RISC Oxidopamine HCl (dHCl) and Suramin suggest a significant increase in *FXN* gene expression in FRDA lymphoblasts upon treatment arguing for the potential of this approach (Fig 4). As mentioned previously, FRDA exhibits haplosufficiency, with carriers of the mutation asymptomatic and producing ~50% the Frataxin as individuals carrying homozygous non-expanded alleles. FRDA affected individuals produce

4-29% frataxin as normal individuals (depending largely on repeat length). Therefore, increasing frataxin expression by just 25-50% would theoretically be sufficient to eliminate symptoms. Given this, even if epigenetic dysregulation plays only a minor role in contributing to FRDA pathology, if by manipulating the epigenetic role we are able to increase frataxin expression by even a small amount it may still be significant as a therapeutic for FRDA as even small increases in expression are beneficial to FRDA sufferers in terms of reducing symptoms and prolonging life (increasing life quality and duration). Thus, mainipulating smallRNAs holds a considerable potential as a treatment of FRDA.

Of the three compounds tested, ATA gave the greatest increase in *FXN* expression (Fig 3) and the only detectable increase in frataxin protein (Fig 4). Here we assumed the increase in *FXN* expression was due to ATA's ability to inhibit siRNA activity. Deep smallRNA sequencing on ATA treated FRDA cells would be required to determine whether the ATA was actually acting by decreasing smallRNA production. This is a future experiment that could be done.

A conflicting point worth mentioning is that, as mentioned above, ATA is also known to be a potent inhibitor of human flap endonuclease inhibitor 1 (FEN1). FEN1 acts on specialized DNA structures that occur during the processes of DNA replication, repair and recombination. FEN1 is a structure-specific nuclease in that it specifically removes the 5′ unannealed flap in a branch structure resulting from strand displacement synthesis. FEN1 is critical in maintaining genome stability. It has been shown that Human flap endonuclease 1 (h-FEN1) mutations have dramatic effects on repeat instability (Ruggiero and Topal 2004). FEN1 plays an important role in maintaining the integrity of repeated sequences. Studies *in vivo* and *in vitro* have shown the *FEN1*Δ mutant significantly enhanced triplet repeat expansion (Liu, Zhang et al. 2004). It is thought that FEN1 endonuclease efficiently removes a TNR flap before it can form an intermediate that leads to expansion upon ligation or chromosome breakage (Ruggiero and Topal 2004). How FEN1 employs its endonuclease activity to resolve a triplet repeat flap is unknown.

However, there is other evidence to suggest that FEN1 loss of function and triplet instability/expansion effect is species dependent and in humans FEN1 loss may not affect triplet repeat instability (Moe, Sorbo et al. 2008). Thus our results may not necessarily contradict FEN1 results in human cell lines. Therefore, something that inhibits hFEN1 function such as potentially ATA, may actually contribute to FRDA pathology in one sense as well as improving it in another. A future direction could be to look into synthesizing drugs and synthetic compounds (mimetics) similar in structure to ATA but varying slightly. This has the potential to reveal compounds that act through the same mechanism as ATA but with improved potency. Perhaps a synthetic compound that inhibits siRNA production but not FEN1 endonuclease could be generated and this may be a more efficacious drug for treatment of FRDA.

We have found several miRNAs to be differentially expressed between patient cells and normal cells, including miR-886-3p which has been previously identified to play a contributing role to FRDA pathology (Tables 2 & 3) (Mahishi, Hart et al. 2012). We identified the potential target genes for these miRNAs and tested whether the expression of these target genes is altered in FRDA cells.

As suspected, anti-correlation between the miRNAs and their potential target genes was observed. Five of the genes identified by (miRSearch 3.0) as being likely targets for the upregulated miRNAs in FRDA cells (Table 2), were displaying a trend of decreased gene expression in the FRDA cells although it turned out to be not significant. The anti-correlation trend however does supports the idea that those miRNAs may actually be having a biological affect and further experiments are required to test these preliminary results, since the loss of the target gene appear to contribute to the disorder symptoms (Table 2). Almost all of the identified target genes are expressed in neuronal cells and are thus potentially represent pathways that are modulated by the smallRNA machinery in FRDA. Therefore, another future direction, would be exploring the transcriptome of the FRDA cells in order to define the biological pathways of therapeutic significance.

### 4.3.3 Small molecule translational screen from *Arabidopsis* to Friedreich's ataxia

Through this translational approach from *A. thaliana* to human cell lines, we have identified 5 compounds that appear to have significant impacts on gene expression in both the plant and human systems. Any compounds that influence gene expression in both plant and human systems are likely to act through a fundamentally conserved mechanism to increase gene expression.

Of the 18 compounds identified in the plant screen and subsequently screened on the human cells, 4 compounds significantly altered *FXN* gene expression. The 2 compounds identified as significantly and consistently increasing *FXN* expression were Visnagin and Sphondin. Visnagin also produced an improvement in both Bur-0 phenotype and *IIL1* expression. Visnagin is known to act as a vasodilator reducing blood pressure by inhibiting calcium influx into cells. It protects against doxorubicin-induced cardiomyopathy through modulation of mitochondrial malate dehydrogenase (Lee, Jung et al. 2010) (Table 1). However, how these findings relate to its effect on increasing *FXN* expression still remains unknown. Sphondin was found to worsen the plant *iil* phenotypic defect and decrease *IIL1* gene expression. Upon PCR analysis it was also found to increase TTC/GAA repeat instability displaying even longer repeat expansion tracts. Sphondin is a furanocoumarin derivative that may have therapeutic potential as an anti-inflammatory drug on airway inflammation by inhibiting expression of IL-1β-induced COX-2 (Yang, Liang et al. 2002). Interestingly, Sphondin was found to increase *FXN* expression. Starting from the known biological functions of Visnagin and Sphondin, further investigation into how they mediate their affects on gene expression in *A. thaliana* and Human cells is required.

Two other compounds Fumarprotocetraric acid & Carylophyllene oxide significantly decreased *FXN* expression despite increasing *IIL1* expression. These compounds would obviously not be considered as a treatment for FRDA. However, given that they are influencing both *IIL1* and *FXN* expression, albeit in

opposite ways, it still tells us that they are likely able to influence gene expression through a mechanism common to both systems and is therefore relevant. Further analysis into the mechanistic actions of those 2 compounds may uncover molecular processes contributing to the frataxin deficiency.

### 4.3.4 Limitations of our findings.

This translatory screen enabled us to identify 2 (Visnagin & Sphondin) compounds that increase *FXN* expression in frataxin deficient human cells through specifically a mechanism intrinsic to the expansion itself. It has also identified 2 other compounds (Fumarprotocetraric acid & Carylophyllene oxide) that are negatively affecting gene expression through a mechanism intrinsic to the expansion itself. In addition, the compounds ATA, dHCL and Suramin were also identified as increasing *FXN* expression.

It's worth mentioning that not all compounds discovered as potential drugs to treat FRDA have practical applications. Just because a compound gives an increase in *FXN* gene expression in a cell-based assay does not necessarily correlate with an increase in gene expression in human trials. For example the drug Resveratrol was found to increase *FXN* expression in FRDA mouse models but this did not translate to significant increase in *FXN* expression *in vivo* despite an improvement in clinical outcome (Li, Voullaire et al. 2013). Other FRDA drugs in phase 1 and 2 clinical trials, show an in vivo increase in *FXN* expression, but are not associated with an increase in clinical outcome. And some have been found to result in a significant worsening of clinical outcome for FRDA patients. Thus, drugs that show potential as therapeutics do not always translate into suitable therapies at the whole organism level (Delatycki and Corben 2012).

### 4.3.5 Future Directions

This work opens up several avenues for future work. It would be interesting to test whether our identified siRNAs are inducing DNA methylation in the FRDA lymphoblasts, (which has already previously been shown to be hypermethylated). a future step might be to specifically blocking the *FXN* promoter localizing siRNA by anti-mIR generation, then measuring the effects on methylation status of the *FXN* promoter (by bisulfide sequencing) and gene expression status of the treated cells by q-RT-PCR.

Given that target genes for most of the differentially expressed miRNAs are expressed in neuronal cells, it would be interesting to explore the transcriptome of the FRDA cells could help identify downstream pathways and identify additional targets for therapeutic intervention. In order to further characterize the targets of miRNAs differentially regulated in FRDA cells a future direction would be to assess the impacts of the gene targets at the cellular level through biochemical experiments such as use of mimetics or sponges to knockdown the specific miRNAs and then assess the impact on these genes at the transcriptional level. Alternatively, development of anti-*miR* against the desired miRNAs could be used in attempt to increase the expression of the targeted genes and observe the effect on FRDA symptoms and pathology. These analyses would allow further characterization of the molecular signatures of the downstream pathways affected in FRDA.

A conflicting aspect regarding the Bur-0 model is, how could there be an increase in repeat-derived smallRNAs if the disease is caused by an *IIL1* transcriptional deficiency and those *IIL1* transcripts are the pool from which the smallRNAs are derived? This may seem counterintuitive, however as mentioned previously, FRDA is associated with a large increase in *FXN* antisense transcript (FAST-1) (Delatycki, Williamson et al. 2000) . These antisense transcripts could potentially be the endogenous pool of the rasiRNA. Though future experiments to confirm this could be done.

With the identification of the 5 drug candidates as having potential for treatment of FRDA, the future directions for this project are immense. It would be

interesting to understand the molecular basis for this impact of *FXN* gene expression. Several possibilities exist, which can be explored. First, one could test whether the small molecules affect the formation of "sticky DNA/RNA" which could be analysed using biophysical techniques. Second, we could assess whether any of the differentially regulated siRNAs/miRNAs discovered in our study are modulated in presence of the small molecules. Generating mimetics and similar compounds to those drugs identified and screening for greater increase in *FXN* expression could also be done. Transcriptome sequencing of FRDA compound treated cells could identify downstream pathways influencing FRDA frataxin deficiency and could potentially uncover additional targets for therapeutic intervention.

To progress the identified 5 drug candidates through to potentially clinical trials, Further animal trials are required. However, at present there are no suitable animal models of FRDA that contain GAA/TTC expansions in their *FXN* gene. Most of the pharmaceutical screening on mice is done utilizing models such as humanized YG8R FRDA mice, which do not contain an expansion in their *FXN* gene (Li, Voullaire et al. 2013). Pharmaceutical screens used to identify candidate drugs that increase *FXN* expression in such mice, must utilize mechanisms that induce gene transcription (e.g. act on the *FXN* promoter to drive transcription) and those drug candidates would not act through mechanisms related to the TTC/GAA expansion mutation itself. Through our *A. thaliana*-FRDA human cell screen we have selectively identified compounds that must have an affect on the intrinsic properties of the GAA/TTC expansion itself and not specifically the *FXN* gene.

Another analysis that could be done as a future experiment is to analyse *FXN* expression levels following treatment on FRDA induced pluripotent stem cells (iPSC). Such model iPCS cell lines are already in existence and our drugs may produce more specific affects in cell types affected by FRDA such as neurons and cardiomyocytes (Hick, Wattenhofer-Donze et al. 2013). For example frataxin deficiency causes oxidative stress and subsequent cell death in dorsal root

ganglia and cardiac tissue (Arias, Valle-Leija et al. 2014). By treating iPSC cells with our drug candidates the affects on such cellular pathological aspects of FRDA such as iron accumulation and oxidative stress could be analysed.

## 4.4 MATERIALS & METHODS

### 4.4.1 Cell Lines

Deep smallRNA sequencing was carried out on lymphoblast cell lines derived from normal (Coriell #GM15851) as well as FRDA clinically affected lymphoblast individuals (Coriell #GM15850). Cells were cultured under identical conditions (37 °C +/-0.1 °C, 5% $CO_2$). Two biological replicates of each cell line were sent for deep sequencing of smallRNAs (BGI- Beijing Genomics Institute, Shenzhen, China) followed by bioinformatics analysis. The cell lines were derived from an FRDA patient (affected cells) and his brother, who was homozygous WT at the *FXN* locus, to minimize differences due to gender and genetic background.

Catalog No. GM15851 Lot. D Product. Cell culture (B-Lymphocyte/Epstein-Barr Virus) Sample Description. FRIEDREICH ATAXIA 1; FRDA

Catalog No. GM15850 Lot. G Product. Cell culture (B-Lymphocyte/Epstein-Barr Virus) Sample Description. FRIEDREICH ATAXIA 1; FRDA

Lymphoblastoid cell lines obtained from the Coriell cell repository (Camden, New Jersey, USA) were established by Epstein- Barr virus transformation of peripheral blood mononuclear cells using phytohemagluttinin as a mitogen. Cells were free from bacterial, fungal and mycoplasma contamination.

Patient lymphoblast cell lines used: FRDA affected individual: GM15850 &

homozygous normal individual: GM15851. GM15850 cells are derived from a clinically affected 13 yr Caucasian male, homozygous for the GAA/TTC expansion in the *FXN* gene with alleles of approximately 650 and 1030 repeats. GM15851 cells are derived from a clinically unaffected 14 yr Caucasian male, who has two FRDA alleles in the normal range of GAA trinucleotide repeats. GM15850 and GM15851 cell lines are derived from brothers of a similar age to reduce biological differences that may arise due to gender, age and genetic backgrounds.

### 4.4.2 Lymphoblast cell culture

Cell culture conditions: Roswell Park Memorial Institute (RPMI) 1640 cell culture media (Life Technologies), 2 mM L-glutamine, 15% FBS- non-heat inactivated, 2% pen/strep (Sigma). Cells were maintained in T25 tissue culture flasks with 10ml medium in an upright position with vented caps in a SANYO MCO-20AIC cell culture incubator at 37 °C & 5% $CO_2$.

Lymphoblast cultures were shipped in T25 tissue culture flasks filled to capacity with CO2-equillibrated medium. Upon receipt, cell cultures flasks were incubated for 12 h @ 37 °C unopened. Lymphoblast cultures were then counted and split if sufficient growth had occurred, to yield a cell density of 200,000-500,000 viable cells/ml.

Lymphoblastoid cell lines grow in suspension culture with cells clumped in loose aggregates. Cultures were seeded at an initial concentration of ≥ 200,000 viable cells/ml. The plateau growth level is ~1,000,000 cells/ml and was reached after 3-5 days of sub-culturing (with pre-warmed fresh media). The pH of cultures become quite acidic at this point and require sub-culturing (the pH indicator, phenol red in the RPMI media causes the culture to become yellow).

### 4.4.3 Cell storage in liquid $N_2$

To maintain integrity, cell cultures were preserved as frozen 'seed stocks'. For freeze storage, 1 x $10^7$ cellls were pooled, Usually a fully confluent (density 1 x $10^6$ cells/ml) 10 ml flask was used. As lymphoblasts grow in suspension as loose aggregates, cultures were homogonised by gently agitating the culture and trituration with pipette. Cell viability was calculated by trypan blue staining and haemocytometer counting. Breify, a 10 µl sample of culture to be counted was placed in a microcentifuge tube with 10 ml of 0.4% trypan blue dye in PBS for a 1:1 ratio. Samples were loaded in a hemocytometer and counted. Cultures were centrifuged for 10 min at 100 (+/- 20) X g at 4-10 °C. The pelleted cells were resuspended in 1ml of cold (4-10 °C) freeze medium (RPMI 1640 with 30% FBS and 5% DMSO) to yield approximately 5-7 million viable cells/ml. The 1 ml freeze medium cell suspension were aliquoted into plastic cryovials (Sigma) and placed in control freezing chambers (Thermo Scientific), which freeze the samples at -1 °C /min to -80 °C. Cryovials were then transferred to liquid $N_2$ storage. To recover cell lines for further culture, cryovials were thawed rapidly in a 37 °C water bath and re-suspended in fresh culture medium to a concentration of ~0.5 x $10^6$ cells/ml.

### 4.4.4 Harvesting cells

10 ml culture samles (~$1^7$ cells) were gently triturated, collected in 10 ml BD Falcon tubes and washed twice in phosphate-buffered saline (PBS). Briefly, samples were centrifuged (1,500 rpm/450 g, 5 min, RT). Supernatant was discarded and cell pellets washed by resuspension in 1 ml PBS, in microfuge tubes, then centrifuged (1,500 rpm/450 g, 5 min, RT). Washed cell pellets were then used to extract DNA, RNA or protein. [NB: Protocols were modified to accommodate different initial cell concentrations. E.g. for 2.5 ml samples (~$2.5^6$ cells, protocol was quartered].

### 4.4.5 DNA Isolation

For cell pellet of ~$1^7$ cells, 0.5 ml of lysis buffer (50 mM Tris (pH 8.0), 100 mM $Na_2EDTA$, 0.5% SDS) and 10 µl of 20 mg/ml Proteinase K were added and samples incubated at 55 °C for 2-3 hrs with occasional gentle mixing/inversion (until cells have lysed). 0.5 ml of phenol/chloroform/isoamyl alcohol (25:24:1) solution was then added to each tube followed by gentle inversion 10-20 X. Samples were then centrifuged (13,000 rpm, 5 min, RT) and upper aqueous phase transferred to new microfuge tube containing 0.8 ml of 100% ethanol and 16 µl of 5 M NaCl. Tubes were gently inverted several times and centrifuged (13,000 rpm, 5 min, RT). Precipitated DNA pellets were washed with 1 ml of 70% ethanol (13,000 rpm, 5 min, RT). Supernatant was removed and DNA pellets left to air dry for 5-10 min followed by re-suspension in 20-400 ul of sterile milli Q $H_2O$. [DNA] quality and quantity determined with NaoDrop (Thermo Scientific).

### 4.4.6 RNA Isolation and cDNA synthesis

For sRNA deep sequencing analysis, total RNA was extracted from 10 ml (~$1^7$ cells) samples using the Qiagen RNeasy Plus Mini Kit (Qiagen) and Qiagen QIAshredder (Qiagen) according to the manufacturer's instructions. RNA requirements of sample quality ≥ 10 ug, sample concentration ≥ 400 ng/ul and sample purity OD260/280 = 1.8-2.2 were met. RNA was isolated from cell pellets through the TRIzol method described in chapter 1, and cDNA synthesized using the Invitrogen SuperScript® III First Strand Synthesis System also as described in Chapter 1.

#### 4.4.7 Protein Isolation

Cells were lysed using 100 µL Laemmli buffer (2X) (10% (w/v) SDS, 1 M Tris-HCL (pH 6.8) and 0.02% bromophenol blue in $H_2O$) per sample, and incubated for 3 min at 95 °C. Cellular debris was pelleted by centrifugation at 16,000 rpm at RT for 2 min. Protein concentration was estimated based on cell count/sample.

#### 4.4.8 PCR

The concentration of all gDNA was determined using UV spectrophotometry (Nanodrop; Thermo Scientific). 200 ng of gDNA was combined with 0.05 µL Extaq polymerase (Takara Bio), 1 µL of 10X Extaq extra buffer, 0.8 µL of 2.5 mM dNTPs (Thermo Scientific), 6.65 µL of sterile milli Q $H_2O$ (Merck Millipore) and 0.5 µL of each *FXN* primer (Suppl. Table S1) were used to amplify the GAA/TTC tract loci (Suppl. Figure S2). Amplification for the *FXN* loci: Initial denaturation 94 °C for 5 min, followed by 10 cycles of 94 °C for 20 sec, 58 °C for 30 sec and 68 °C for 5 min, followed by a further 20 cycles of 94 °C for 20 sec, 63 °C for 30 sec and 68 °C for 5 min + 20 sec per cycle. Final extension 68 °C for 10 min. PCR products were run on 0.8% agarose gels + ethidium bromide in TAE buffer for 30 min at 100 V.

#### 4.4.9 smallRNA seq data analysis

*Differential Expression:*

Raw counts for miRNA were identified from the small sequence reads. For smallRNA differential expression, unique smallRNA sequences and their abundance counts were extracted from all smallRNA sequences perfectly

matching the reference genome. siRNA and miRNA that had a (raw, unnormalised) count of less than five in both case and control were excluded from the analysis, due to low power and information. siRNA and miRNA that were present in one set but not the other, and had a count of greater than four, were *not* excluded.

TMM normalisation was used. edgeR was used to determine differential expression using the standard 'non-generalised' linear model approach. MDS plots suggest dataset similarity between biological replicates (case/control replicates one/two) indicating no batch effect present. Differental expresion was calculated as logFC (logfold change), logCPM (log of counts-per-million: a relative estimation of count), P-value (raw *p*-value), FDR (Benjamini-Hochberg adjusted P value).

*smallRNAs matching FXN gene:*

All smallRNAs matching the *FXN* gene were pulled out from the data and the normalised count (normalised count= ((raw count / library size) * 1000000) and length of sequences calculated.

#### 4.4.10 Quantitative real-time PCR

Quantitative RT-PCR was performed using both The SYBR Green system (Roche) and TaqMan® Gene Expression system (Applied Biosystems) on a Lightcycler 480 (Roche Diagnostics). SYBR Green qPCR reactions were set up as described in Chapter 1. For TaqMan assays, 20 µl reactions containing 1 µl cDNA and 19 µl 1:20 diluted TaqMan probe were run in triplicate. TaqMan probe: specific for the human *FXN* gene (Assay ID: Hs00175940_m1) and Human *GAPDH* reference gene (Assay ID: Hs03929097_g1). All samples were run in Biological triplicate unless otherwise stated. Expression data for all human genes were normalized to *GAPDH* reference gene expression and all *A. thaliana* genes were normalized to

*TUB2*. Expression data is presented as relative difference to either untreated control cells or untreated FRDA-affected cells as stated (or relative to Bur-0 untreated at 27 °C in *A. thaliana*).

Reactions were run for 40 cycles and data analysis carried out using standard ΔΔCt methodology as described in chapter 1. Expression results between both systems were indistinguishable.

Statistical analyses was determined through student's *t*-test (two-tailed) or Pearson's correlation coefficient as described in chapter 1. $p$-value < 0.05 was considered significant. Graphs were constructed using Prism 6 programme.

#### 4.4.11 Western blot analysis

For Western blotting 40-50 μg protein was added per lane of 4-12% Bis-Tris gels (Invitrogen). Nitrocellulose membranes (Highbond) were stained with Ponceau S dye (Invitrogen) for protein detection prior to western blotting. Membranes were blocked with 5% (w/v) skim milk powder in TBS-T buffer (10 ml Tris HCl pH 7.5 1 M, 8.7 g NaCl & 1 ml Tween) for 1 h at RT with rotation. Following blocking, membranes were incubated with primary antibodies diluted (1:5,000) in TBS-T and left o/n at RT with rotation. HRP-conjugated anti-mouse secondary antibody was used at 1:10,000 dilution with TBS-T and left for 1 h at RT. Membranes were washed 3X with TBS-T following primary and secondary antibody incubations. Luminol reagent was added as a HRP substrate, and used to detect and quantify the protein signal. Membranes were imaged using a LI-COR Odyssey.

Antibodies used:

Monoclonal Anti-FXN antibody produced in mouse. Clone 3G9, purified immunoglobulin (#WH0002395M3, Sigma-Aldrich)

Monoclonal Anti-β-Actin antibody produced in mouse. Clone AC-74, purified immunoglobulin (#A2228, Sigma-Aldrich)

Monoclonal anti-mouse HRP–conjugated $IgG_{2a}$. HA-probe f-7 (#sc-7392, Santa Cruz Biotechnology)

#### 4.4.12 Small molecule Screening Assay

For pharmacological compound analysis *in vitro*; FRDA-affected lymphoblasts (Coriell #GM15850) and control lymphoblasts (Coriell #15851) were seeded into sterile 6-well culture plates at an initial density of 3 x $10^5$ cells/ml in 3 ml complete RPMI media containing 3 µl compound to be screened. For a list of compounds, see Table 1, Chapter 5. Compounds were screened in biological triplicates for FRDA cells and biological duplicate for control cells unless otherwise specified. 0.1% DMSO was added as a mock treatment. Following treatment of 72 h (or indicated duration in optimization assay), each 3ml sample was split into 3 microfuge tubes, containing 1ml each. Thus 3 aliquotes of washed cell pellets were obtained for each sample and used for DNA, RNA and protein extraction separately. All assays were performed on triplicate cultures in at least three independent experiments.

For chemical treatment of Bur-0 plants, 100 µl of compound at the desired concentration or 0.1% DMSO as a mock treatment, was spread a top solidified MS-media in 6-well sterile plates and left for 10 min in a laminar flow cabinet to absorb into the media. Stratified Bur-0 seeds were then placed in the wells, the lid sealed with parafilm and plates incubated in 27 °C short day Percival chambers for 4-weeks. Photographs of seedling phenotype were then taken and

tissue from each individual plant analysed for genomic TTC/GAA tract size (PCR) and *IIL1* gene expression level.

#### 4.4.13 Optimization assay

In the optimization assays, FRDA and control cells were incubated with 5 compounds ATA, Suramin, Sphondin, Oxidopamine HCl (dHCL) and Visnagin, at 10 μM, 25 μM, 50 μM, 100 μM and 1,000 μM concentrations for 72 h, 96 h and 120 h (and 144 h for ATA) timeframes. For each time point and compound concentration, cells were counted for analysis of affects on cell proliferation and cell viability. Viability was measured to correct for variations in cell number from well to well due to possible toxicity of different test compounds. Viability was determined as % living/dead cells to determine treatment affects on viability and q-RT-PCR analysis undertaken to determine affects on *FXN* expression levels. (Protein concentration in samples was made uniform for Western blot analysis based on viable cell count in each sample).

## 4.5 REFERENCES


Ahmed, M. and P. Liang (2012). "Transposable elements are a significant contributor to tandem repeats in the human genome." Comp Funct Genomics **2012**: 947089.

Al-Mahdawi, S., R. M. Pinto, D. Varshney, L. Lawrence, M. B. Lowrie, S. Hughes, Z. Webster, J. Blake, J. M. Cooper, R. King and M. A. Pook (2006). "GAA repeat expansion mutation mouse models of Friedreich ataxia exhibit oxidative stress leading to progressive neuronal and cardiac pathology." Genomics **88**(5): 580-590.

Amenta, P. S., J. I. Jallo, R. F. Tuma, D. C. Hooper and M. B. Elliott (2014). "Cannabinoid receptor type-2 stimulation, blockade, and deletion alters the vascular inflammatory responses to traumatic brain injury." J Neuroinflammation **11**(1): 191.

Appleby, N., D. Edwards and J. Batley (2009). "New technologies for ultra-high throughput genotyping in plants." Methods Mol Biol **513**: 19-39.

Aranzana, M. J., S. Kim, K. Zhao, E. Bakker, M. Horton, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. Tang, C. Toomajian, B. Traw, H. Zheng, J. Bergelson, C. Dean, P. Marjoram and M. Nordborg (2005). "Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes." PLoS Genet **1**(5): e60.

Arias, E. R., P. Valle-Leija, M. A. Morales and F. Cifuentes (2014). "Differential contribution of BDNF and NGF to long-term potentiation in the superior cervical ganglion of the rat." Neuropharmacology **81**: 206-214.

Ashizawa, T., J. R. Dubel and Y. Harati (1993). "Somatic instability of CTG repeat in myotonic dystrophy." Neurology **43**(12): 2674-2678.

Ashley, C. T., Jr. and S. T. Warren (1995). "Trinucleotide repeat expansion and human disease." Annu Rev Genet **29**: 703-728.

Atkinson, S. P., M. Lako and L. Armstrong (2013). "Potential for pharmacological manipulation of human embryonic stem cells." Br J Pharmacol **169**(2): 269-289.

Azaryan, A. V., B. J. Clock and B. M. Cox (1996). "Mu opioid receptor mRNA in nucleus accumbens is elevated following dopamine receptor activation." Neurochem Res **21**(11): 1411-1415.

Bette A. Loiselle, V. L. S., John Nason and Catherine Graham (1995). "Spatial Genetic Structure of a Tropical Understory Shrub, Psychotria officinalis (Rubiaceae)." American Journal of Botany . Published by: Botanical Society of America **82**(11): 25.

Bidichandani, S. I., T. Ashizawa and P. I. Patel (1998). "The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure." Am J Hum Genet **62**(1): 111-121.

Bilkei-Gorzo, A., K. Michel, F. Noble, B. P. Roques and A. Zimmer (2007). "Preproenkephalin knockout mice show no depression-related phenotype." Neuropsychopharmacology **32**(11): 2330-2337.

Blevins, T., R. Rajeswaran, P. V. Shivaprasad, D. Beknazariants, A. Si-Ammour, H. S. Park, F. Vazquez, D. Robertson, F. Meins, Jr., T. Hohn and M. M. Pooggin (2006). "Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing." Nucleic Acids Res **34**(21): 6233-6246.

Boland, C. R. and A. Goel (2010). "Microsatellite instability in colorectal cancer." Gastroenterology **138**(6): 2073-2087 e2073.

Bomblies, K., L. Yant, R. A. Laitinen, S. T. Kim, J. D. Hollister, N. Warthmann, J. Fitz and D. Weigel (2010). "Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of Arabidopsis thaliana." PLoS Genet **6**(3): e1000890.

Borevitz, J. O. and M. Nordborg (2003). "The impact of genomics on the study of natural variation in Arabidopsis." Plant Physiol **132**(2): 718-725.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss and E. S. Buckler (2007). "TASSEL: software for association mapping of complex traits in diverse samples." Bioinformatics **23**(19): 2633-2635.

Brennan, A. C., B. Mendez-Vigo, A. Haddioui, J. M. Martinez-Zapater, F. X. Pico and C. Alonso-Blanco (2014). "The genetic structure of Arabidopsis thaliana in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe." BMC Plant Biol **14**: 17.

Breseghello, F. and M. E. Sorrells (2006). "Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars." Genetics **172**(2): 1165-1177.

Budden, T. and N. A. Bowden (2013). "The Role of Altered Nucleotide Excision Repair and UVB-Induced DNA Damage in Melanomagenesis." Int J Mol Sci **14**(1): 1132-1151.

Campuzano, V., L. Montermini, Y. Lutz, L. Cova, C. Hindelang, S. Jiralerspong, Y. Trottier, S. J. Kish, B. Faucheux, P. Trouillas, F. J. Authier, A. Durr, J. L. Mandel, A. Vescovi, M. Pandolfo and M. Koenig (1997). "Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes." Hum Mol Genet **6**(11): 1771-1780.

Campuzano, V., L. Montermini, M. D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Canizares, H. Koutnikova, S. I. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P. I. Patel, S. Di Donato, J. L. Mandel, S. Cocozza, M. Koenig and M.

Pandolfo (1996). "Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion." Science **271**(5254): 1423-1427.

Carpenter, N. J. (1994). "Genetic anticipation. Expanding tandem repeats." Neurol Clin **12**(4): 683-697.

Chapdelaine, P., Z. Coulombe, A. Chikh, C. Gerard and J. P. Tremblay (2013). "A Potential New Therapeutic Approach for Friedreich Ataxia: Induction of Frataxin Expression With TALE Proteins." Mol Ther Nucleic Acids **2**: e119.

Christmann, M., M. T. Tomicic, W. P. Roos and B. Kaina (2003). "Mechanisms of human DNA repair: an update." Toxicology **193**(1-2): 3-34.

Clough, S. J. and A. F. Bent (1998). "Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana." Plant J **16**(6): 735-743.

Cossee, M., H. Puccio, A. Gansmuller, H. Koutnikova, A. Dierich, M. LeMeur, K. Fischbeck, P. Dolle and M. Koenig (2000). "Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation." Hum Mol Genet **9**(8): 1219-1226.

Croce, C. M. (2009). "Causes and consequences of microRNA dysregulation in cancer." Nat Rev Genet **10**(10): 704-714.

de Barros Alves, G. M., M. B. de Sousa Maia, E. de Souza Franco, A. M. Galvao, T. G. da Silva, R. M. Gomes, M. B. Martins, E. P. da Silva Falcao, C. M. de Castro and N. H. da Silva (2014). "Expectorant and antioxidant activities of purified fumarprotocetraric acid from Cladonia verticillaris lichen in mice." Pulm Pharmacol Ther **27**(2): 139-143.

De Biase, I., Y. K. Chutake, P. M. Rindler and S. I. Bidichandani (2009). "Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription." PLoS One **4**(11): e7914.

De Biase, I., A. Rasmussen, A. Monticelli, S. Al-Mahdawi, M. Pook, S. Cocozza and S. I. Bidichandani (2007). "Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life." Genomics **90**(1): 1-5.

Deiters, A. (2010). "Small molecule modifiers of the microRNA and RNA interference pathway." AAPS J **12**(1): 51-60.

Delatycki, M., R. Williamson and S. Forrest (2000). "Friedreich ataxia: an overview." J Med Genet **37**(1): 1-8.

Delatycki, M. B. and L. A. Corben (2012). "Clinical Features of Friedreich Ataxia." J Child Neurol **27**(9): 1133-1137.

Denekamp, M. and S. C. Smeekens (2003). "Integration of wounding and osmotic stress signals determines the expression of the AtMYB102 transcription factor gene." Plant Physiol **132**(3): 1415-1423.

Ditch, S., M. C. Sammarco, A. Banerjee and E. Grabczyk (2009). "Progressive GAA.TTC repeat expansion in human cell lines." PLoS Genet **5**(10): e1000704.

Du, J., E. Campau, E. Soragni, S. Ku, J. W. Puckett, P. B. Dervan and J. M. Gottesfeld (2012). "Role of mismatch repair enzymes in GAA.TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells." J Biol Chem **287**(35): 29861-29872.

Du, J., E. Campau, E. Soragni, S. Ku, J. W. Puckett, P. B. Dervan and J. M. Gottesfeld (2012). "Role of Mismatch Repair Enzymes in GAA{middle dot}TTC Triplet-repeat Expansion in Friedreich Ataxia Induced Pluripotent Stem Cells." J Biol Chem **287**(35): 29861-29872.

Dubrova, Y. E., M. Plumb, J. Brown and A. J. Jeffreys (1998). "Radiation-induced germline instability at minisatellite loci." Int J Radiat Biol **74**(6): 689-696.

Evanno, G., S. Regnaut and J. Goudet (2005). "Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study." Mol Ecol **14**(8): 2611-2620.

Evans-Galea, M. V., N. Carrodus, S. M. Rowley, L. A. Corben, G. Tai, R. Saffery, J. C. Galati, N. C. Wong, J. M. Craig, D. R. Lynch, S. R. Regner, A. F. Brocht, S. L. Perlman, K. O. Bushara, C. M. Gomez, G. R. Wilmot, L. Li, E. Varley, M. B. Delatycki and J. P. Sarsero (2012). "FXN methylation predicts expression and clinical outcome in Friedreich ataxia." Ann Neurol **71**(4): 487-497.

Evans-Galea, M. V., P. J. Lockhart, C. A. Galea, A. J. Hannan and M. B. Delatycki (2014). "Beyond loss of frataxin: the complex molecular pathology of Friedreich ataxia." Discov Med **17**(91): 25-35.

Ezzatizadeh, V., R. M. Pinto, C. Sandi, M. Sandi, S. Al-Mahdawi, H. Te Riele and M. A. Pook (2012). "The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model." Neurobiol Dis **46**(1): 165-171.

Farnir, F., W. Coppieters, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar and M. Georges (2000). "Extensive genome-wide linkage disequilibrium in cattle." Genome Res **10**(2): 220-227.

Felth, J., K. Lesiak-Mieczkowska, P. D'Arcy, C. Haglund, J. Gullbo, R. Larsson, S. Linder, L. Bohlin, M. Fryknas and L. Rickardson (2013). "Gambogic acid is cytotoxic to cancer cells through inhibition of the ubiquitin-proteasome system." Invest New Drugs **31**(3): 587-598.

Filipowicz, W., S. N. Bhattacharyya and N. Sonenberg (2008). "Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?" Nat Rev Genet **9**(2): 102-114.

Fondon, J. W., 3rd and H. R. Garner (2004). "Molecular origins of rapid and continuous morphological evolution." Proc Natl Acad Sci U S A **101**(52): 18058-18063.

Forgacs, E., J. D. Wren, C. Kamibayashi, M. Kondo, X. L. Xu, S. Markowitz, G. E. Tomlinson, C. Y. Muller, A. F. Gazdar, H. R. Garner and J. D. Minna (2001). "Searching for microsatellite mutations in coding regions in lung, breast, ovarian and colorectal cancers." Oncogene **20**(8): 1005-1009.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt and A. M. Wilczek (2011). "A map of local adaptation in Arabidopsis thaliana." Science **334**(6052): 86-89.

Francois, O., M. G. Blum, M. Jakobsson and N. A. Rosenberg (2008). "Demographic history of european populations of Arabidopsis thaliana." PLoS Genet **4**(5): e1000075.

Frieman, M. B. and B. P. Cormack (2004). "Multiple sequence signals determine the distribution of glycosylphosphatidylinositol proteins between the plasma membrane and cell wall in Saccharomyces cerevisiae." Microbiology **150**(Pt 10): 3105-3114.

Galloway, J. N. and D. L. Nelson (2009). "Evidence for RNA-mediated toxicity in the fragile X-associated tremor/ataxia syndrome." Future Neurol **4**(6): 785.

Galpaz, N. and M. Reymond (2010). "Natural variation in Arabidopsis thaliana revealed a genetic network controlling germination under salt stress." PLoS One **5**(12): e15198.

Gaynor, J. W., B. J. Campbell and R. Cosstick (2010). "RNA interference: a chemist's perspective." Chem Soc Rev **39**(11): 4169-4184.

Gong, L., S. Parikh, P. J. Rosenthal and B. Greenhouse (2013). "Biochemical and immunological mechanisms by which sickle cell trait protects against malaria." Malar J **12**: 317.

Gonitel, R., H. Moffitt, K. Sathasivam, B. Woodman, P. J. Detloff, R. L. Faull and G. P. Bates (2008). "DNA instability in postmitotic neurons." Proc Natl Acad Sci U S A **105**(9): 3467-3472.

Gordenin, D. A., T. A. Kunkel and M. A. Resnick (1997). "Repeat expansion--all in a flap?" Nat Genet **16**(2): 116-118.

Grabczyk, E., M. Mancuso and M. C. Sammarco (2007). "A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro." Nucleic Acids Res **35**(16): 5351-5359.

Grabczyk, E. and K. Usdin (2000). "The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner." Nucleic Acids Res **28**(14): 2815-2822.

Grosse, S. D., I. Odame, H. K. Atrash, D. D. Amendah, F. B. Piel and T. N. Williams (2011). "Sickle Cell Disease in Africa: A Neglected Cause of Early Childhood Mortality." Am J Prev Med **41**(6): S398-405.

Guay, C. and R. Regazzi (2013). "Circulating microRNAs as novel biomarkers for diabetes mellitus." Nat Rev Endocrinol **9**(9): 513-521.

Gumireddy, K., D. D. Young, X. Xiong, J. B. Hogenesch, Q. Huang and A. Deiters (2008). "Small-molecule inhibitors of microrna miR-21 function." Angew Chem Int Ed Engl **47**(39): 7482-7484.

Guo, H., N. T. Ingolia, J. S. Weissman and D. P. Bartel (2010). "Mammalian microRNAs predominantly act to decrease target mRNA levels." Nature **466**(7308): 835-840.

Hamilton, D. W., A. Hills, B. Kohler and M. R. Blatt (2000). "Ca2+ channels at the plasma membrane of stomatal guard cells are activated by hyperpolarization and abscisic acid." Proc Natl Acad Sci U S A **97**(9): 4967-4972.

Hammock, E. A., M. M. Lim, H. P. Nair and L. J. Young (2005). "Association of vasopressin 1a receptor levels with a regulatory microsatellite and behavior." Genes Brain Behav **4**(5): 289-301.

Hancock, A. M., B. Brachi, N. Faure, M. W. Horton, L. B. Jarymowycz, F. G. Sperone, C. Toomajian, F. Roux and J. Bergelson (2011). "Adaptation to climate across the Arabidopsis thaliana genome." Science **334**(6052): 83-86.

Hardy, O. J. (2003). "Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers." Mol Ecol **12**(6): 1577-1588.

Hardy OJ, V. X. "(2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618–620. doi: 10.1046/j.1471-8286.2002.00305.x."

Hashem, A. M., A. S. Flaman, A. Farnsworth, E. G. Brown, G. Van Domselaar, R. He and X. Li (2009). "Aurintricarboxylic acid is a potent inhibitor of influenza A and B virus neuraminidases." PLoS One **4**(12): e8350.

Hashida, H., J. Goto, T. Suzuki, S. Jeong, N. Masuda, T. Ooie, Y. Tachiiri, H. Tsuchiya and I. Kanazawa (2001). "Single cell analysis of CAG repeat in brains of dentatorubral-pallidoluysian atrophy (DRPLA)." J Neurol Sci **190**(1-2): 87-93.

He, F., D. Kang, Y. Ren, L. J. Qu, Y. Zhen and H. Gu (2007). "Genetic diversity of the natural populations of Arabidopsis thaliana in China." Heredity (Edinb) **99**(4): 423-431.

He, F. and P. K. Todd (2011). "Epigenetics in nucleotide repeat expansion disorders." Semin Neurol **31**(5): 470-483.

Henn, A., A. Joachimi, D. P. Goncalves, D. Monchaud, M. P. Teulade-Fichou, J. K. Sanders and J. S. Hartig (2008). "Inhibition of dicing of guanosine-rich shRNAs by quadruplex-binding compounds." Chembiochem **9**(16): 2722-2729.

Hennessy, R., McNamara, M., & Hoctor, Z. (2010). Stone, Water & Ice: A Geology Trip Through the Burren. Burren Connect Project, pp. 64.

Henricksen, L. A., S. Tom, Y. Liu and R. A. Bambara (2000). "Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion." J Biol Chem **275**(22): 16420-16427.

Herman, D., K. Jenssen, R. Burnett, E. Soragni, S. L. Perlman and J. M. Gottesfeld (2006). "Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia." Nat Chem Biol **2**(10): 551-558.

Hick, A., M. Wattenhofer-Donze, S. Chintawar, P. Tropel, J. P. Simard, N. Vaucamps, D. Gall, L. Lambot, C. Andre, L. Reutenauer, M. Rai, M. Teletin, N. Messaddeq, S. N. Schiffmann, S. Viville, C. E. Pearson, M. Pandolfo and H. Puccio (2013). "Neurons and cardiomyocytes derived from induced pluripotent stem cells as a model for mitochondrial defects in Friedreich's ataxia." Dis Model Mech **6**(3): 608-621.

Hill, W. G. and B. S. Weir (1988). "Variances and covariances of squared linkage disequilibria in finite populations." Theor Popul Biol **33**(1): 54-78.

Huang, D., W. Wu, S. R. Abrams and A. J. Cutler (2008). "The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors." J Exp Bot **59**(11): 2991-3007.

Jain, A., M. R. Rajeswari and F. Ahmed (2002). "Formation and thermodynamic stability of intermolecular (R*R*Y) DNA triplex in GAA/TTC repeats associated with Freidreich's ataxia." J Biomol Struct Dyn **19**(4): 691-699.

Jarem, D. A., N. R. Wilson, K. M. Schermerhorn and S. Delaney (2011). "Incidence and persistence of 8-oxo-7,8-dihydroguanine within a hairpin intermediate exacerbates a toxic oxidation cycle associated with trinucleotide repeat expansion." DNA Repair (Amst) **10**(8): 887-896.

Javed, S., I. N. Petropoulos, U. Alam and R. A. Malik (2015). "Treatment of painful diabetic neuropathy." Ther Adv Chronic Dis **6**(1): 15-28.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." Nat Rev Mol Cell Biol **7**(5): 335-346.

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino and C. Dean (2000). "Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time." Science **290**(5490): 344-347.

Jovicic, A., J. F. Zaldivar Jolissaint, R. Moser, F. Silva Santos Mde and R. Luthi-Carter (2013). "MicroRNA-22 (miR-22) overexpression is neuroprotective via

general anti-apoptotic effects and may also target specific Huntington's disease-related mechanisms." PLoS One **8**(1): e54222.

Kalinowski, S. T. (2011). "The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure." Heredity (Edinb) **106**(4): 625-632.

Kantartzis, A., G. M. Williams, L. Balakrishnan, R. L. Roberts, J. A. Surtees and R. A. Bambara (2012). "Msh2-Msh3 Interferes with Okazaki Fragment Processing to Promote Trinucleotide Repeat Expansions." Cell Reports **2**(2): 216-222.

Katori, T., A. Ikeda, S. Iuchi, M. Kobayashi, K. Shinozaki, K. Maehashi, Y. Sakata, S. Tanaka and T. Taji (2010). "Dissecting the genetic control of natural variation in salt tolerance of Arabidopsis thaliana accessions." J Exp Bot **61**(4): 1125-1138.

Kawli, T. and M. W. Tan (2008). "Neuroendocrine signals modulate the innate immunity of Caenorhabditis elegans through insulin signaling." Nat Immunol **9**(12): 1415-1424.

Kayser, M., L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Kruger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff, M. Stoneking and A. Sajantila (2000). "Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs." Am J Hum Genet **66**(5): 1580-1588.

Kennedy, L., E. Evans, C. M. Chen, L. Craven, P. J. Detloff, M. Ennis and P. F. Shelbourne (2003). "Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis." Hum Mol Genet **12**(24): 3359-3367.

Kerrest, A., R. P. Anand, R. Sundararajan, R. Bermejo, G. Liberi, B. Dujon, C. H. Freudenreich and G. F. Richard (2009). "SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination." Nat Struct Mol Biol **16**(2): 159-167.

Keurentjes, J. J., L. Bentsink, C. Alonso-Blanco, C. J. Hanhart, H. Blankestijn-De Vries, S. Effgen, D. Vreugdenhil and M. Koornneef (2007). "Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population." Genetics **175**(2): 891-905.

Kilian, J., D. Whitehead, J. Horak, D. Wanke, S. Weinl, O. Batistic, C. D'Angelo, E. Bornberg-Bauer, J. Kudla and K. Harter (2007). "The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses." Plant J **50**(2): 347-363.

Kim, J. K., T. Bamba, K. Harada, E. Fukusaki and A. Kobayashi (2007). "Time-course metabolic profiling in Arabidopsis thaliana cell cultures after salt stress treatment." J Exp Bot **58**(3): 415-424.

Kim, S., V. Plagnol, T. T. Hu, C. Toomajian, R. M. Clark, S. Ossowski, J. R. Ecker, D. Weigel and M. Nordborg (2007). "Recombination and linkage disequilibrium in Arabidopsis thaliana." Nat Genet **39**(9): 1151-1155.

Kim, Y. J. and D. M. Wilson, 3rd (2012). "Overview of base excision repair biochemistry." Curr Mol Pharmacol **5**(1): 3-13.

Kita, Y., Y. Ago, K. Higashino, K. Asada, E. Takano, K. Takuma and T. Matsuda (2014). "Galantamine promotes adult hippocampal neurogenesis via M(1) muscarinic and alpha7 nicotinic receptors in mice." Int J Neuropsychopharmacol **17**(12): 1957-1968.

Knill, T., M. Reichelt, C. Paetz, J. Gershenzon and S. Binder (2009). "Arabidopsis thaliana encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation." Plant Mol Biol **71**(3): 227-239.

Koeppen, A. H. (2011). "Friedreich's ataxia: pathology, pathogenesis, and molecular genetics." J Neurol Sci **303**(1-2): 1-12.

Kohler, B., A. Hills and M. R. Blatt (2003). "Control of guard cell ion channels by hydrogen peroxide and abscisic acid indicates their action through alternate signaling pathways." Plant Physiol **131**(2): 385-388.

Kovtun, I. V., Y. Liu, M. Bjoras, A. Klungland, S. H. Wilson and C. T. McMurray (2007). "OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells." Nature **447**(7143): 447-452.

Kovtun, I. V. and C. T. McMurray (2001). "Trinucleotide expansion in haploid germ cells by gap repair." Nat Genet **27**(4): 407-411.

Kovtun, I. V. and C. T. McMurray (2008). "Features of trinucleotide repeat instability in vivo." Cell Res **18**(1): 198-213.

Kreps, J. A., Y. Wu, H. S. Chang, T. Zhu, X. Wang and J. F. Harper (2002). "Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress." Plant Physiol **130**(4): 2129-2141.

Kumari, D. and K. Usdin (2012). "Is Friedreich ataxia an epigenetic disorder?" Clin Epigenetics **4**(1): 2.

Kunkel, T. A. (1993). "Nucleotide repeats. Slippery DNA and diseases." Nature **365**(6443): 207-208.

Kurihara, Y. and Y. Watanabe (2004). "Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions." Proc Natl Acad Sci U S A **101**(34): 12753-12758.

Lahue, R. S. and D. L. Slater (2003). "DNA repair and trinucleotide repeat instability." Front Biosci **8**: s653-665.

Lau, J. C., M. L. Hanel and R. Wevrick (2004). "Tissue-specific and imprinted epigenetic modifications of the human NDN gene." Nucleic Acids Res **32**(11): 3376-3382.

Le Corre, V., F. Roux and X. Reboud (2002). "DNA polymorphism at the FRIGIDA gene in Arabidopsis thaliana: extensive nonsynonymous variation is consistent with local selection for flowering time." Mol Biol Evol **19**(8): 1261-1271.

Lee, J. K., J. S. Jung, S. H. Park, S. H. Park, Y. B. Sim, S. M. Kim, T. S. Ha and H. W. Suh (2010). "Anti-inflammatory effect of visnagin in lipopolysaccharide-stimulated BV-2 microglial cells." Arch Pharm Res **33**(11): 1843-1850.

Lee, S. T., K. Chu, W. S. Im, H. J. Yoon, J. Y. Im, J. E. Park, K. H. Park, K. H. Jung, S. K. Lee, M. Kim and J. K. Roh (2011). "Altered microRNA regulation in Huntington's disease models." Exp Neurol **227**(1): 172-179.

Lee, Y. S., K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer and R. W. Carthew (2004). "Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways." Cell **117**(1): 69-81.

Lei, J., Y. Zhou, D. Xie and Y. Zhang (2014). "Mechanistic Insights into a Classic Wonder Drug-Aspirin." J Am Chem Soc.

Lempe, J., S. Balasubramanian, S. Sureshkumar, A. Singh, M. Schmid and D. Weigel (2005). "Diversity of flowering responses in wild Arabidopsis thaliana strains." PLoS Genet **1**(1): 109-118.

Levdansky, E., J. Romano, Y. Shadkchan, H. Sharon, K. J. Verstrepen, G. R. Fink and N. Osherov (2007). "Coding tandem repeats generate diversity in Aspergillus fumigatus genes." Eukaryot Cell **6**(8): 1380-1391.

Lewandowska-Sabat, A. M., S. Fjellheim and O. A. Rognli (2010). "Extremely low genetic variability and highly structured local populations of Arabidopsis thaliana at higher latitudes." Mol Ecol **19**(21): 4753-4764.

Li, L., L. Voullaire, C. Sandi, M. A. Pook, P. A. Ioannou, M. B. Delatycki and J. P. Sarsero (2013). "Pharmacological screening using an FXN-EGFP cellular genomic reporter assay for the therapy of Friedreich ataxia." PLoS One **8**(2): e55940.

Lin, Y., L. Hubert, Jr. and J. H. Wilson (2009). "Transcription destabilizes triplet repeats." Mol Carcinog **48**(4): 350-361.

Liu, G. and M. Leffak (2012). "Instability of (CTG)n*(CAG)n trinucleotide repeats and DNA synthesis." Cell Biosci **2**(1): 7.

Liu, Y., H. Zhang, J. Veeraraghavan, R. A. Bambara and C. H. Freudenreich (2004). "Saccharomyces cerevisiae flap endonuclease 1 uses flap equilibration to maintain triplet repeat stability." Mol Cell Biol **24**(9): 4049-4064.

Long, Q., F. A. Rabanal, D. Meng, C. D. Huber, A. Farlow, A. Platzer, Q. Zhang, B. J. Vilhjalmsson, A. Korte, V. Nizhynska, V. Voronin, P. Korte, L. Sedman, T.

Mandakova, M. A. Lysak, U. Seren, I. Hellmann and M. Nordborg (2013). "Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden." Nat Genet **45**(8): 884-890.

Lu, R., E. Yigit, W. X. Li and S. W. Ding (2009). "An RIG-I-Like RNA helicase mediates antiviral RNAi downstream of viral siRNA biogenesis in Caenorhabditis elegans." PLoS Pathog **5**(2): e1000286.

Lufino, M. M., A. M. Silva, A. H. Nemeth, J. Alegre-Abarrategui, A. J. Russell and R. Wade-Martins (2013). "A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds." Hum Mol Genet **22**(25): 5173-5187.

Ma, L., J. S. Jensen, M. Mancuso, R. Hamasuna, Q. Jia, C. L. McGowin and D. H. Martin (2012). "Variability of trinucleotide tandem repeats in the MgPa operon and its repetitive chromosomal elements in Mycoplasma genitalium." J Med Microbiol **61**(Pt 2): 191-197.

Mahishi, L. H., R. P. Hart, D. R. Lynch and R. R. Ratan (2012). "miR-886-3p levels are elevated in Friedreich ataxia." J Neurosci **32**(27): 9369-9373.

Malinina, L. (2005). "Possible involvement of the RNAi pathway in trinucleotide repeat expansion diseases." J Biomol Struct Dyn **23**(3): 233-235.

Mangiarini, L., K. Sathasivam, A. Mahal, R. Mott, M. Seller and G. P. Bates (1997). "Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation." Nat Genet **15**(2): 197-200.

Mariappan, S. V., P. Catasti, L. A. Silks, 3rd, E. M. Bradbury and G. Gupta (1999). "The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia." J Mol Biol **285**(5): 2035-2052.

Marmolino, D. (2011). "Friedreich's ataxia: past, present and future." Brain Res Rev **67**(1-2): 311-330.

Martelli, A., M. Napierala and H. Puccio (2012). "Understanding the genetic and molecular pathogenesis of Friedreich's ataxia through animal and cellular models." Dis Model Mech **5**(2): 165-176.

Matsuno, H., O. Kozawa, M. Niwa, K. Tanabe, K. Ichimaru, Y. Takiguchi, M. Yokota, H. Hayashi and T. Uematsu (1998). "Multiple inhibition of platelet activation by aurintricarboxylic acid prevents vascular stenosis after endothelial injury in hamster carotid artery." Thromb Haemost **79**(4): 865-871.

McMurray, C. T. (1999). "DNA secondary structure: a common and causative factor for expansion in human disease." Proc Natl Acad Sci U S A **96**(5): 1823-1825.

McMurray, C. T. (2010). "Mechanisms of trinucleotide repeat instability during human development." Nat Rev Genet **11**(11): 786-799.

Meister, G. and T. Tuschl (2004). "Mechanisms of gene silencing by double-stranded RNA." Nature **431**(7006): 343-349.

Menezes, M. F., S. P. Barbosa, C. A. De Andrade, J. V. Menani and P. M. De Paula (2011). "Purinergic mechanisms of lateral parabrachial nucleus facilitate sodium depletion-induced NaCl intake." Brain Res **1372**: 49-58.

Metz, G., N. Coppard, J. M. Cooper, M. B. Delatycki, A. Durr, N. A. Di Prospero, P. Giunti, D. R. Lynch, J. B. Schulz, C. Rummey and T. Meier (2013). "Rating disease progression of Friedreich's ataxia by the International Cooperative Ataxia Rating Scale: analysis of a 603-patient database." Brain **136**(Pt 1): 259-268.

Miret, J. J., L. Pessoa-Brandao and R. S. Lahue (1998). "Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **95**(21): 12438-12443.

Mirkin, S. M. (2005). "Toward a unified theory for repeat expansions." Nat Struct Mol Biol **12**(8): 635-637.

Mirkin, S. M. (2007). "Expandable DNA repeats and human disease." Nature **447**(7147): 932-940.

Moe, S. E., J. G. Sorbo and T. Holen (2008). "Huntingtin triplet-repeat locus is stable under long-term Fen1 knockdown in human cells." J Neurosci Methods **171**(2): 233-238.

Mollersen, L., A. D. Rowe, E. Larsen, T. Rognes and A. Klungland (2010). "Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice." PLoS Genet **6**(12): e1001242.

Morales, F., J. M. Couto, C. F. Higham, G. Hogg, P. Cuenca, C. Braida, R. H. Wilson, B. Adam, G. del Valle, R. Brian, M. Sittenfeld, T. Ashizawa, A. Wilcox, D. E. Wilcox and D. G. Monckton (2012). "Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity." Hum Mol Genet **21**(16): 3558-3567.

Moxon, E. R., P. B. Rainey, M. A. Nowak and R. E. Lenski (1994). "Adaptive evolution of highly mutable loci in pathogenic bacteria." Curr Biol **4**(1): 24-33.

Nakashima, K., T. Kiyosue, K. Yamaguchi-Shinozaki and K. Shinozaki (1997). "A nuclear gene, erd1, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally up-regulated during senescence in Arabidopsis thaliana." Plant J **12**(4): 851-861.

Nam, M., S. Koh, S. U. Kim, L. L. Domier, J. H. Jeon, H. G. Kim, S. H. Lee, A. F. Bent and J. S. Moon (2011). "Arabidopsis TTR1 causes LRR-dependent lethal systemic necrosis, rather than systemic acquired resistance, to Tobacco ringspot virus." Mol Cells **32**(5): 421-429.

Neumann, R., V. E. Lawson and A. J. Jeffreys (2010). "Dynamics and processes of copy number instability in human gamma-globin genes." Proc Natl Acad Sci U S A **107**(18): 8304-8309.

Noland, C. L. and J. A. Doudna (2013). "Multiple sensors ensure guide strand selection in human RNAi pathways." RNA **19**(5): 639-648.

Ohshima, K., L. Montermini, R. D. Wells and M. Pandolfo (1998). "Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo." J Biol Chem **273**(23): 14588-14595.

Orr, H. T. (2009). "Unstable nucleotide repeat minireview series: a molecular biography of unstable repeat disorders." J Biol Chem **284**(12): 7405.

Osborne, R. J., X. Lin, S. Welle, K. Sobczak, J. R. O'Rourke, M. S. Swanson and C. A. Thornton (2009). "Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy." Hum Mol Genet **18**(8): 1471-1481.

Owen, B. A., Z. Yang, M. Lai, M. Gajec, J. D. Badger, 2nd, J. J. Hayes, W. Edelmann, R. Kucherlapati, T. M. Wilson and C. T. McMurray (2005). "(CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition." Nat Struct Mol Biol **12**(8): 663-670.

Pang, Y., I. S. B. Abeysinghe, J. He, X. He, D. Huhman, K. M. Mewan, L. W. Sumner, J. Yun and R. A. Dixon (2013). "Functional Characterization of Proanthocyanidin Pathway Enzymes from Tea and Their Application for Metabolic Engineering1[W][OA]." Plant Physiol **161**(3): 1103-1116.

Patel, P. I. and G. Isaya (2001). "Friedreich ataxia: from GAA triplet-repeat expansion to frataxin deficiency." Am J Hum Genet **69**(1): 15-24.

Payne, R. M. (2011). "The Heart in Friedreich's Ataxia: Basic Findings and Clinical Implications." Prog Pediatr Cardiol **31**(2): 103-109.

Pearson, C. E. and R. R. Sinden (1998). "Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA." Curr Opin Struct Biol **8**(3): 321-330.

Perdomini, M., A. Hick, H. Puccio and M. A. Pook (2013). "Animal and cellular models of Friedreich ataxia." J Neurochem **126 Suppl 1**: 65-79.

Perez-Perez, J. M., J. Serrano-Cartagena and J. L. Micol (2002). "Genetic analysis of natural variations in the architecture of Arabidopsis thaliana vegetative leaves." Genetics **162**(2): 893-915.

Pico, F. X., B. Mendez-Vigo, J. M. Martinez-Zapater and C. Alonso-Blanco (2008). "Natural genetic variation of Arabidopsis thaliana is geographically structured in the Iberian peninsula." Genetics **180**(2): 1009-1021.

Prince, M., C. T. Campbell, T. A. Robertson, A. J. Wells and H. E. Kleiner (2006). "Naturally occurring coumarins inhibit 7,12-dimethylbenz[a]anthracene DNA adduct formation in mouse mammary gland." Carcinogenesis **27**(6): 1204-1213.

Pritchard, J. K., M. Stephens and P. Donnelly (2000). "Inference of population structure using multilocus genotype data." Genetics **155**(2): 945-959.

Puccio, H., D. Simon, M. Cossee, P. Criqui-Filipe, F. Tiziano, J. Melki, C. Hindelang, R. Matyas, P. Rustin and M. Koenig (2001). "Mouse models for Friedreich ataxia exhibit cardiomyopathy, sensory nerve defect and Fe-S enzyme deficiency followed by intramitochondrial iron deposits." Nat Genet **27**(2): 181-186.

Remington, D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt, J. Doebley, S. Kresovich, M. M. Goodman and E. S. t. Buckler (2001). "Structure of linkage disequilibrium and phenotypic associations in the maize genome." Proc Natl Acad Sci U S A **98**(20): 11479-11484.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky and M. W. Feldman (2002). "Genetic structure of human populations." Science **298**(5602): 2381-2385.

Rotig, A., P. de Lonlay, D. Chretien, F. Foury, M. Koenig, D. Sidi, A. Munnich and P. Rustin (1997). "Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich ataxia." Nat Genet **17**(2): 215-217.

Rufini, A., S. Fortuni, G. Arcuri, I. Condo, D. Serio, O. Incani, F. Malisan, N. Ventura and R. Testi (2011). "Preventing the ubiquitin-proteasome-dependent degradation of frataxin, the protein defective in Friedreich's ataxia." Hum Mol Genet **20**(7): 1253-1261.

Ruggiero, B. L. and M. D. Topal (2004). "Triplet repeat expansion generated by DNA slippage is suppressed by human flap endonuclease 1." J Biol Chem **279**(22): 23088-23097.

Ruprecht, R. M., L. D. Rossoni, W. A. Haseltine and S. Broder (1985). "Suppression of retroviral propagation and disease by suramin in murine systems." Proc Natl Acad Sci U S A **82**(22): 7733-7737.

Rusmini, P., F. Simonini, V. Crippa, E. Bolzoni, E. Onesto, M. Cagnin, D. Sau, N. Ferri and A. Poletti (2011). "17-AAG increases autophagic removal of mutant androgen receptor in spinal and bulbar muscular atrophy." Neurobiol Dis **41**(1): 83-95.

Sakamoto, N., P. D. Chastain, P. Parniewski, K. Ohshima, M. Pandolfo, J. D. Griffith and R. D. Wells (1999). "Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia." Mol Cell **3**(4): 465-475.

Sandi, C., S. Al-Mahdawi and M. A. Pook (2013). "Epigenetics in Friedreich's Ataxia: Challenges and Opportunities for Therapy." Genet Res Int **2013**: 852080.

Sansaloni, C. P., C. D. Petroli, J. Carling, C. J. Hudson, D. A. Steane, A. A. Myburg, D. Grattapaglia, R. E. Vaillancourt and A. Kilian (2010). "A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus." Plant Methods **6**: 16.

Scannell, D. A. W. M. J. P. (1983). Flora of Connemara and The Burren, Royal Dublin Society & Cambridge University Press.

Schauer, S. E., S. E. Jacobsen, D. W. Meinke and A. Ray (2002). "DICER-LIKE1: blind men and elephants in Arabidopsis development." Trends Plant Sci **7**(11): 487-491.

Schmitt, J. K. and S. B. Johns (1995). "Altering therapy of type II diabetes mellitus from insulin to tolazamide increases blood pressure in spite of weight loss." Am J Hypertens **8**(5 Pt 1): 520-523.

Schwab, R., S. Ossowski, M. Riester, N. Warthmann and D. Weigel (2006). "Highly specific gene silencing by artificial microRNAs in Arabidopsis." Plant Cell **18**(5): 1121-1133.

Schweitzer, J. K. and D. M. Livingston (1997). "Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast." Hum Mol Genet **6**(3): 349-355.

Schweitzer, J. K. and D. M. Livingston (1999). "The effect of DNA replication mutations on CAG tract stability in yeast." Genetics **152**(3): 953-963.

Shan, G., Y. Li, J. Zhang, W. Li, K. E. Szulwach, R. Duan, M. A. Faghihi, A. M. Khalil, L. Lu, Z. Paroo, A. W. Chan, Z. Shi, Q. Liu, C. Wahlestedt, C. He and P. Jin (2008). "A small molecule enhances RNA interference and promotes microRNA processing." Nat Biotechnol **26**(8): 933-940.

Shindo, C., C. Lister, P. Crevillen, M. Nordborg and C. Dean (2006). "Variation in the epigenetic silencing of FLC contributes to natural variation in Arabidopsis vernalization response." Genes Dev **20**(22): 3079-3083.

Shinozaki, K., K. Yamaguchi-Shinozaki and M. Seki (2003). "Regulatory network of gene expression in the drought and cold stress responses." Curr Opin Plant Biol **6**(5): 410-417.

Sinden, R. R. and R. D. Wells (1992). "DNA structure, mutations, and human genetic disease." Curr Opin Biotechnol **3**(6): 612-622.

Sinnott, R. W. (1984). "Virtues of the Haversine." Sky and Telescope **68**(2): 159.

Skirycz, A., H. Claeys, S. De Bodt, A. Oikawa, S. Shinoda, M. Andriankaja, K. Maleux, N. B. Eloy, F. Coppens, S. D. Yoo, K. Saito and D. Inze (2011). "Pause-and-Stop: The Effects of Osmotic Stress on Cell Proliferation during Early Leaf Development in Arabidopsis and a Role for Ethylene Signaling in Cell Cycle Arrest." Plant Cell **23**(5): 1876-1888.

Soriano, S., J. V. Llorens, L. Blanco-Sobero, L. Gutierrez, P. Calap-Quintana, M. P. Morales, M. D. Molto and M. J. Martinez-Sebastian (2013). "Deferiprone and idebenone rescue frataxin depletion phenotypes in a Drosophila model of Friedreich's ataxia." Gene **521**(2): 274-281.

Spiro, C., R. Pelletier, M. L. Rolfsmeier, M. J. Dixon, R. S. Lahue, G. Gupta, M. S. Park, X. Chen, S. V. Mariappan and C. T. McMurray (1999). "Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats." Mol Cell **4**(6): 1079-1085.

Stern, D. L. and V. Orgogozo (2008). "The loci of evolution: how predictable is genetic evolution?" Evolution **62**(9): 2155-2177.

Stram, Y. and L. Kuzntzova (2006). "Inhibition of viruses by RNA interference." Virus Genes **32**(3): 299-306.

Sturm, B., U. Bistrich, M. Schranzhofer, J. P. Sarsero, U. Rauen, B. Scheiber-Mojdehkar, H. de Groot, P. Ioannou and F. Petrat (2005). "Friedreich's ataxia, no changes in mitochondrial labile iron in human lymphoblasts and fibroblasts: a decrease in antioxidative capacity?" J Biol Chem **280**(8): 6701-6708.

Sugui, J. A., Y. C. Chang and K. J. Kwon-Chung (2005). "Agrobacterium tumefaciens-mediated transformation of Aspergillus fumigatus: an efficient tool for insertional mutagenesis and targeted gene disruption." Appl Environ Microbiol **71**(4): 1798-1802.

Suh, N., L. Baehner, F. Moltzahn, C. Melton, A. Shenoy, J. Chen and R. Blelloch (2010). "MicroRNA function is globally suppressed in mouse oocytes and early embryos." Curr Biol **20**(3): 271-277.

Sureshkumar, S., M. Todesco, K. Schneeberger, R. Harilal, S. Balasubramanian and D. Weigel (2009). "A genetic defect caused by a triplet repeat expansion in Arabidopsis thaliana." Science **323**(5917): 1060-1063.

Tam, O. H., A. A. Aravin, P. Stein, A. Girard, E. P. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R. M. Schultz and G. J. Hannon (2008). "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes." Nature **453**(7194): 534-538.

Tan, G. S., C. H. Chiu, B. G. Garchow, D. Metzler, S. L. Diamond and M. Kiriakidou (2012). "Small molecule inhibition of RISC loading." ACS Chem Biol **7**(2): 403-410.

Tiurenkov, I. N., E. V. Volotova, D. V. Kurkin, D. A. Bakulin, I. O. Logvinov and T. A. Antipova (2014). "[Neuroprotective effect of neuroglutam under conditions of activated free radical oxidation]." Eksp Klin Farmakol **77**(8): 16-19.

Undurraga, S. F., M. O. Press, M. Legendre, N. Bujdoso, J. Bale, H. Wang, S. J. Davis, K. J. Verstrepen and C. Queitsch (2012). "Background-dependent effects of polyglutamine variation in the Arabidopsis thaliana gene ELF3." Proc Natl Acad Sci U S A **109**(47): 19363-19367.

Ura, K. and J. J. Hayes (2002). "Nucleotide excision repair and chromatin remodeling." Eur J Biochem **269**(9): 2288-2293.

Vachon, C. M., D. J. Schaid, J. N. Ingle, D. L. Wickerham, M. Kubo, T. Mushiroda, M. P. Goetz, E. E. Carlson, S. Paik, N. Wolmark, Y. Nakamura, L. Wang, R. Weinshilboum and F. J. Couch (2015). "A polygenic risk score for breast cancer in women receiving tamoxifen or raloxifene on NSABP P-1 and P-2." Breast Cancer Res Treat.

Walley, J. W., S. Coughlan, M. E. Hudson, M. F. Covington, R. Kaspi, G. Banu, S. L. Harmer and K. Dehesh (2007). "Mechanical stress induces biotic and abiotic stress responses via a novel cis-element." PLoS Genet **3**(10): 1800-1812.

Wang, G. Y. and L. R. Sun (2007). "[Effect of Mycobacterium phlei F.U.36 suspended liquor on culture and proliferation of dendritic cells derived from human umbilical cord blood in vitro]." Zhongguo Shi Yan Xue Ye Xue Za Zhi **15**(6): 1257-1260.

Wang, X. H., R. Aliyari, W. X. Li, H. W. Li, K. Kim, R. Carthew, P. Atkinson and S. W. Ding (2006). "RNA interference directs innate immunity against viruses in adult Drosophila." Science **312**(5772): 452-454.

Watanabe, H., F. Tanaka, M. Doyu, S. Riku, M. Yoshida, Y. Hashizume and G. Sobue (2000). "Differential somatic CAG repeat instability in variable brain cell lineage in dentatorubral pallidoluysian atrophy (DRPLA): a laser-captured microdissection (LCM)-based analysis." Hum Genet **107**(5): 452-457.

Watanabe, T., Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki and H. Sasaki (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-543.

Watashi, K., M. L. Yeung, M. F. Starost, R. S. Hosmane and K. T. Jeang (2010). "Identification of small molecules that suppress microRNA function and reverse tumorigenesis." J Biol Chem **285**(32): 24707-24716.

Weatherall, D. J. and J. B. Clegg (2001). "Inherited haemoglobin disorders: an increasing global health problem." Bull World Health Organ **79**(8): 704-712.

Webb, D. A. (1961). Noteworthy Plants of the Burren. G. Section B: Biological, and Chemical Science. Proceedings of the Royal Irish Academy.

Weir, B. S. (1996). Genetic Data Analysis II : Methods for Discrete Population Genetic Data by Bruce S. Weir

Wells, R. D. (2009). "Discovery of the role of non-B DNA structures in mutagenesis and human genomic disorders." J Biol Chem **284**(14): 8997-9009.

Wells, R. D. and T. Ashiuzawa (2006). Genetic instabilities and neurological diseases, Academic Press.

Wells, R. D., R. Dere, M. L. Hebert, M. Napierala and L. S. Son (2005). "Advances in mechanisms of genetic instability related to hereditary neurological diseases." Nucleic Acids Res **33**(12): 3785-3798.

Werner, J. D., J. O. Borevitz, N. H. Uhlenhaut, J. R. Ecker, J. Chory and D. Weigel (2005). "FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions." Genetics **170**(3): 1197-1207.

Willadsen, K., M. D. Cao, J. Wiles, S. Balasubramanian and M. Boden (2013). "Repeat-encoded poly-Q tracts show statistical commonalities across species." BMC Genomics **14**: 76.

Wu, M. J., L. W. Chow and M. Hsieh (1998). "Amplification of GAA/TTC triplet repeat in vitro: preferential expansion of (TTC)n strand." Biochim Biophys Acta **1407**(2): 155-162.

Wyman, C. and R. Kanaar (2006). "DNA double-strand break repair: all's well that ends well." Annu Rev Genet **40**: 363-383.

Xia, H., Y. Cao, X. Dai, Z. Marelja, D. Zhou, R. Mo, S. Al-Mahdawi, M. A. Pook, S. Leimkuhler, T. A. Rouault and K. Li (2012). "Novel frataxin isoforms may contribute to the pathological mechanism of Friedreich ataxia." PLoS One **7**(10): e47847.

Xiong, L. and J. K. Zhu (2002). "Molecular and genetic aspects of plant responses to osmotic stress." Plant Cell Environ **25**(2): 131-139.

Yamaguchi, K., Y. Takahashi, T. Berberich, A. Imai, A. Miyazaki, T. Takahashi, A. Michael and T. Kusano (2006). "The polyamine spermine protects against high salt stress in Arabidopsis thaliana." FEBS Lett **580**(30): 6783-6788.

Yandim, C., T. Natisvili and R. Festenstein (2013). "Gene regulation and epigenetics in Friedreich's ataxia." J Neurochem **126 Suppl 1**: 21-42.

Yang, L. L., Y. C. Liang, C. W. Chang, W. S. Lee, C. T. Kuo, C. C. Wang, H. M. Lee and C. H. Lin (2002). "Effects of sphondin, isolated from Heracleum laciniatum, on IL-1beta-induced cyclooxygenase-2 expression in human pulmonary epithelial cells." Life Sci **72**(2): 199-213.

Yin, P., J. Kang, F. He, L. J. Qu and H. Gu (2010). "The origin of populations of Arabidopsis thaliana in China, based on the chloroplast DNA sequences." BMC Plant Biol **10**: 22.

Yokoi, S., F. J. Quintero, B. Cubero, M. T. Ruiz, R. A. Bressan, P. M. Hasegawa and J. M. Pardo (2002). "Differential expression and function of Arabidopsis thaliana NHX Na+/H+ antiporters in the salt stress response." Plant J **30**(5): 529-539.

Young, D. D., C. M. Connelly, C. Grohmann and A. Deiters (2010). "Small molecule modifiers of microRNA miR-122 function for the treatment of hepatitis C virus infection and hepatocellular carcinoma." J Am Chem Soc **132**(23): 7976-7981.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich and E. S. Buckler (2006). "A unified mixed-model method for association mapping that accounts for multiple levels of relatedness." Nat Genet **38**(2): 203-208.

Yu, Z., X. Teng and N. M. Bonini (2011). "Triplet repeat-derived siRNAs enhance RNA-mediated toxicity in a Drosophila model for myotonic dystrophy." PLoS Genet **7**(3): e1001340.

Zambon, R. A., V. N. Vakharia and L. P. Wu (2006). "RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster." Cell Microbiol **8**(5): 880-889.

Zhu, J. K. (2001). "Plant salt tolerance." Trends Plant Sci **6**(2): 66-71.

Zoghbi, H. Y. and H. T. Orr (2000). "Glutamine repeats and neurodegeneration." Annu Rev Neurosci **23**: 217-247.

# CHAPTER 5. Thesis Summary & Conclusions

## 5.1 Irish *A. thaliana* collection

Collectively in this this thesis, I have contributed to *A. thaliana* research by collecting and proving extensive phenotypic data for >500 wild *A. thaliana* samples from Ireland representing >130 distinct populations. This work presents the 1st population genetic study of *A. thaliana* from Ireland. This Irish collection has vast potential for many different applications. This work has revealed substantial genetic and phenotypic variation among the Irish accessions of *A. thaliana,* which can now be used for natural variation studies for several traits. Since it is a local population with clearly defined collection sites, they would be useful for investigating mechanisms through which plants respond to environmental cues. Furthermore, it will also help understand post-glacial colonization of Ireland. Through DArT-SEQ, these accessions have been genotyped across the genome providing a genotyped local collection for further analyses. Seeds from all wild accessions presented in this thesis will be made available to the stock center. Therefore this study has produced wild accessions that could be useful for widespread applications.

## 5.2 Recovery of the repeat expansion at the *IIL1* locus in wild Irish accessions

I have shown that even mutations with potentially detrimental effects can be maintained in nature if there phenotypic effects are conditional and masked. This thesis reports the recovery of at least 8 accessions displaying extreme variability in repeat length, maintaining dramatically expanded alleles. These 8 accessions display *iil* phenotypes at 27 °C SD (of variable severity) and reduced *IIL1* gene expression upon qPCR analysis. These accessions are in a similar, yet distinctly different genetic background to Bur-0 as determined by SSLP analysis

and DArT sequence data. This frequency is too high to be detrimental allowing us to suggest that this repeat expansion is possibly neutral under local Irish environments and thus presents a cryptic genetic variation maintained in nature.

## 5.3 Similarities between the effects of repeat expansion observed in plants and humans

Our results indicate striking similarities between the Bur-0 TNR expansion defect and observations found in non-coding human Friedreich's ataxia. This thesis further provides evidence for similarities at the fundamental molecular level between the distinct systems. Given these similarities in the repeat dynamics when inherited across generations and the observation of genetic anticipation suggested that Bur-0 could well be used as a model for in depth investigation into the underlying mechanisms governing genetic anticipation at all levels, including the epigenetic influences and repeat dynamics.

Our results also indicate that the DNA MMR pathway protein MSH2 is contributing to the increased severity of the Bur-0 TNR expansion defect at 27 °C. The phenotypic consequences and affects on gene expression in *35S::amiRMSH2* plants are consistent with findings in other FRDA model systems. This provides a platform for study into the molecular mechanisms through which DNA repair pathways influence TNR expansion pathology.

## 5.4 Investigating the role of smallRNAs in FRDA

This thesis demonstrates for the first-time that smallRNAs are dysregulated in FRDA. I have shown a 3.5 fold increase in siRNAs that map to *FXN* in presence of the expansion suggesting a potential mechanism, which may underlie the transcriptional downregulation seen in FRDA. This appears to be clinically significant as blocking smallRNA pathways using molecules such as ATA, Suramin and Oxidopamine HCl appears to increase *FXN* expression in FRDA cells.

Furthermore, we discovered several new and novel miRNAs dysregulated in FRDA, which can be additionally explored for therapeutic potential. This is again consistent with the results obtained in the plant system further demonstrating the potential for translational approaches.

## 5.5 Identification natural compounds that increase *FXN* expression in FRDA

We have identified compounds that significantly increased *FXN* expression.
In addition to the three compounds identified as increasing *FXN* expression by blocking potentially toxic smallRNA pathways in FRDA, this thesis also identifies two additional compounds, Visnagin and Sphondin, also found to increase *FXN* expression in FRDA cells. Overall, this thesis not only reveals novel mechanisms that underlie repeat expansion associated genetic defects, but also exploits the same for therapeutics and makes a case for translational biology using the plant model.

# Appendix

| ecotypes with DArT-Seq data | latitude | longitude | altitude (masl) | soil | annual rainfall (mm/an) | leaf serration scale 1-9 low-high | *IIL1* GAA/TTC repeat number | 23sd root length (mm) | 23sd hypocotyl length (mm) | 23ld flowering time (dtf) | 23sd flowering time (dtf) | Germination 23sd | Germination 23ld |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 53°06'31.65" | -9°08'0.44" | 29 | gravel | 2,000 | 4.5 | 17 | 11.585 55556 | 5.3416 46825 | 36.684 21053 | 110.5 | all | all |
| 15 | 53°05'32.24" | -9°09'0.78" | 23 | dirt | 2,000 | 2 | n/a | 12.340 15789 | 6.23 | 40.5 | 109 | all | all |
| 23 | 53°02'34.17" | -9°03'3.58" | 127 | limestone | 2,000 | 2 | 11 | 10.817 11538 | 6.4007 17037 | 35.083 33333 | 96.333 33333 | all | partial |
| 24 | 53°16'44.12" | -9°03'3.64" | 5 | dirt | 1,400 | 4.75 | 25 | 9.5139 35484 | 5.9318 60328 | 28 | 108.5 | all | all |
| 34 | 53°23'46.67" | -9°55'0.13" | 11 | gravel | 1,600 | 8.25 | n/a | 12.512 14634 | 7.4290 01623 | 27.380 95238 | 110.33 33333 | n/a | n/a |
| 39 | 53°23'46.67" | -9°55'0.13" | 11 | gravel | 1,600 | 5.5 | 317 | 9.3653 125 | 4.0762 10526 | 23.75 | 129 | all | all |
| 40 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | n/a | n/a | 7.1675 11628 | 5.4684 88882 | n/a | n/a | n/a | n/a |
| 41 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | n/a | n/a | 7.1675 11628 | 5.4684 88882 | n/a | n/a | n/a | n/a |
| 50 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | n/a | 11 | 7.1675 11628 | 5.4684 88882 | n/a | n/a | n/a | n/a |
| 60 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | n/a | n/a | 7.1675 11628 | 5.4684 88882 | n/a | n/a | n/a | n/a |
| 64 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | n/a | n/a | 7.1675 11628 | 5.4684 88882 | n/a | n/a | n/a | n/a |
| 66 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | 1 | 17 | 14.156 28302 | 5.6943 53225 | 28.357 14286 | 90 | n/a | n/a |
| 66 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | 1 | 17 | 14.156 28302 | 5.6943 53225 | 28.357 14286 | 90 | n/a | n/a |
| 67 | 53°33'33.90" | -9°53'2.84" | 30 | garden | 2,400 | 1.5 | 26 | 9.3233 125 | 4.9138 88889 | 32.5 | 113.5 | all | all |
| 69 | 53°33'35.00" | -9°53'2.69" | 29 | garden | 2,400 | 4 | 25 | 10.004 54762 | 7.2601 53043 | 27.368 42105 | 109.5 | all | partial |
| 104 | 53°16'34.85" | -9°03'3.61" | 7 | gravel | 1,400 | 4 | 9 | 9.8853 75 | 6.3719 58333 | 30.9 | 115.5 | all | partial |
| 108 | 52°52'00.08" | -8°58'4.03" | 7 | dirt | 1,400 | 2.25 | 9 | 9.0277 5 | 5.9788 25397 | 34.333 33333 | 139 | all | partial |
| 130 | 52°54'40.33" | -9°03'1.97" | 27 | limestone | 1,600 | n/a | 22 | n/a | n/a | n/a | n/a | n/a | n/a |
| 141 | 52°54'39.37" | -9°05'0.78" | 24 | limestone | 1,600 | 2.75 | 25 | 10.432 02778 | 7.7553 33333 | 27.75 | 99 | n/a | n/a |
| 146 | 52°54'39.37" | -9°05'0.78" | 24 | limestone | 1,600 | 4.75 | 22 | n/a | n/a | 30 | 89.5 | all | all |
| 158 | 52°54'39.37" | -9°05'0.78" | 24 | limestone | 1,600 | 2.25 | 25 | 9.5746 90476 | 6.4438 83641 | 30.761 90476 | 142 | all | no |
| 171 | 53°16'28.62" | -9°03'4.11" | 6 | gravel | 1,400 | 4 | 11 | 8.1885 | 5.1887 77778 | 31.846 15385 | 92.5 | all | all |
| 172 | 53°16'28.62" | -9°03'4.11" | 6 | gravel | 1,400 | n/a | 14 break of 5 nucleotides then 9 more | 11.466 69231 | n/a | 37 | n/a | n/a | n/a |
| 193 | 52°39'31.8" | -8°37'4.97" | 15 | garden | 1,200 | n/a | 25 | n/a | n/a | n/a | n/a | n/a | n/a |
| 203 | 52°14'24.95" | -7°03'4.97" | 9 | dirt | 1,200 | 2.5 | 26 | 9.059 | 5.256 | 30 | 107 | all | all |
| 220 | 52°16'31.5" | -7°59'5.4" | 59 | garden | 2,000 | n/a | 11 | n/a | n/a | n/a | n/a | n/a | n/a |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 238 | 52°20' 44.6" | -7°24'2 7.2" | 10 | rock | 1,400 | n/a | 354 | 11.888 78571 | 4 | 68.5 | 97 | n/a | n/a |
| 246 | 52°20' 45.6" | -7°19'2 1.6" | 17 | dirt | 1,200 | n/a | 8° | 10.601 57895 | 3.8253 125 | 43.1 | n/a | n/a | n/a |
| 247 | 52°06' 44.50" | -7°30'4 1.16" | 40 | rock | 1,400 | 7.5 | 17 | n/a | n/a | 100 | 162 | all | all |
| 248 | 52°08' 18.0" | -7°55'5 8.2" | 38 | rock | 1,400 | 4.5 | 10 | n/a | n/a | 34 | 115 | all | all |
| 249 | 52°22' 25.8" | -7°55'3 2.5" | 70 | dirt | 1,000 | n/a | n/a | n/a | n/a | 37.884 61538 | 85 | n/a | n/a |
| 250 | 52°21' 38.9" | -7°41'5 2.9" | 28 | dirt | 1,400 | 3 | 25 | 8.4085 40541 | 7.1506 67857 | 35 | 89 | all | all |
| 262 | 52°20' 18.0" | -6°32'1 9.3" | 113 | dirt | 1,200 | 3 | n/a | 7.1751 53846 | 5.2955 | 30 | 125 | all | all |
| 268 | 53°11' 09.47" | -6°07'4 6.22" | 56 | dirt | 1,000 | 4 | 26 | 6.8077 17949 | | 28 | 90 | partial | all |
| 270 | 52°23' 34.3" | -6°56'3 9.4" | 16 | dirt | 1,200 | 3 | 16 | 8.2764 61538 | 6.4499 | 100 | 162 | all | all |
| 272 | 52°30' 09.3" | -6°34'1 4.9" | 39 | dirt | 1,200 | 5 | n/a | 7.2532 | 5.8195 | 40 | 103 | all | all |
| 273 | 52°30' 39.6" | -6°34'6. 1" | 31 | dirt | 1,200 | n/a | n/a | 8.7404 34783 | 6.2230 37037 | 68.7 | n/a | n/a | n/a |
| 280 | 52°20' 05.2" | -6°28'1 7.6" | 46 | garden | 1,200 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 300 | 52°59' 34.62" | -6°04'3 3.52" | 14 | dirt | 1,000 | 3 | 25 | 9.3827 14286 | n/a | 29 | 85 | partial | all |
| 307 | 52°16' 00.92" | -7°06'0 3.83" | 27 | dirt | 1,200 | 4.5 | 11 | 5.7 | 5.1134 | 29 | 93 | all | all |
| 310 | 52°17' 13.53" | -6°31'2 0.20" | 67 | dirt | 1,200 | 3 | 9 | 7.4491 42857 | 4.8968 | 30 | 98 | partial | partial |
| 317 | 53°07' 43.31" | -8°57'1 8.21" | 15 | dirt | 1,400 | 2 | 26 | 10.102 16981 | 5.5561 21212 | 31.307 69231 | 90.5 | all | all |
| 319 | 53°07' 43.31" | -8°57'1 8.21" | 15 | dirt | 1,400 | 2 | 26 | 10.684 3336 | 6.1703 45238 | 32.066 66667 | 105 | all | all |
| 351 | 53°07' 36.5" | -9°03'1 1.9" | 27 | gravel | 1,400 | 5 | 9 | 11.367 37931 | 8.0308 46154 | 50 | 118 | all | all |
| 352 | 53°07' 36.5" | -9°03'1 1.9" | 27 | gravel | 1,400 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 358 | 53°07' 36.5" | -9°03'1 1.9" | 27 | gravel | 1,400 | 6.5 | 9 | 9.8248 75 | 5.5384 12698 | 34 | 106 | all | all |
| 359 | 53°16' 44.09" | -9°03'4 1.49" | 10 | garden | 1,400 | n/a | 17° | 7.7406 | 5.6775 | 38 | n/a | n/a | n/a |
| 361 | 53°16' 19.7" | -8°58'4 5.2" | 11 | gravel | 1,400 | 5 | 26 | 7.6089 28571 | 5.8203 42105 | 37 | 100 | all | no |
| 362 | 53°16' 25.0" | -8°56'2 5.4" | 7 | gravel | 1,200 | 8 | 9 | 6.4738 | 4.4263 33333 | 32 | 131 | all | all |
| 363 | 53°09' 19.6" | -8°59'2 6.8" | 5 | gravel | 1,200 | 8 | 9 | 10.074 89286 | n/a | 43 | 107 | all | all |
| 364 | 53°06' 54.1" | -9°08'5 9.3" | 5 | gravel | 1,600 | 3 | 9 | 7.4931 03448 | 6.0966 66667 | 39 | 111 | partial | no |
| 365 | 53°02' 26.2" | -9°04'3 2.2" | 127 | gravel | 2,000 | 6.5 | 14 | 11.461 | 4.6681 5 | 44.833 33333 | 126 | all | all |
| 366 | 53°07' 12.8" | -9°04'1 5.4" | 13 | gravel | 1600 | 5 | 9 | 8.2593 33333 | 6.6669 77241 | 29.071 42857 | 114 | all | no |
| 369 | 53°03' 50.33" | -9°04'3 7.41" | 149 | dirt | 1,400 | 4 | 262 | 8.8791 66667 | 5.236 | 37 | 88 | all | all |
| 370 | 53°07' 42.85" | -8°57'1 9.96" | 15 | dirt | 1,400 | 3.5 | 22 | 9.1595 22727 | 6.1381 70213 | 31.5 | 73 | all | all |
| 376 | 53°57' 07.37" | -9°19'0 7.48" | 232 | dirt | 2,000 | 4 | n/a | 8.1224 77273 | 4.6284 77445 | 28.8 | 101 | partial | all |
| 379 | 53°07' 08.13" | -9°04'1 5.44" | 6 | gravel | 1,600 | 3 | n/a | 9.3926 21622 | 3.0822 16146 | 37 | 126 | all | all |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 383 | 53°16' 13.25" | -8°55'1 9.01" | 7 | dirt | 1,200 | 4 | 22 | 7.1767 33333 | 5.5459 85652 | 24.166 66667 | 88.5 | all | all |
| 390 | 53°07' 23.5" | -9°05'5 6.0" | 161 | dirt | 1,400 | 3 | 26 | 6.264 | 4.7771 14833 | 27 | 103.75 | all | all |
| 396 | 53°26' 07.3" | -9°18'2 7.4" | 8 | gra vel | 1,600 | n/a | 28 | 9.4987 27273 | 6.9308 89372 | 28.285 71429 | 112.5 | all | partial |
| 405 | 53°14' 47.1" | -9°16'3 2.24" | 14 | gra vel | 1,400 | 7 | 9 | 9.4299 53488 | 4.9106 66667 | n/a | 109 | partial | no |
| 407 | 53°15' 59.9" | -9°12'2 7.7" | 35 | gra vel | 1,400 | n/a | n/a | 6.025 | 6.7723 33333 | 34.772 72727 | n/a | all | all |
| 408 | 53°15' 59.9" | -9°12'2 7.7" | 35 | gra vel | 1,400 | n/a | 26 | 5.8086 5 | 5.7583 58974 | n/a | n/a | all | all |
| 415 | 55°56' 54.39" | 3° 08'34. 52" | 59 | dirt | 668 | 3 | 9 | 8.3807 | 4.9858 22917 | 35 | 73 | partial | no |
| 434 | 5-6°18'3. 81" | 4°19'4 0.32" | 137 | gra vel | 3063 | 2 | 298 | 9.2274 | 3.6893 40909 | 40 | | all | partial |
| 451 | 53°21' 41.90" | -8°04'5. 80" | 55 | gra vel | 1,200 | 1.5 | 3 | 8.7962 28571 | 5.7543 61111 | 29.904 7619 | 115 | all | partial |
| 454 | 53°16' 30.1" | -8°38'1 1.4" | 40 | dirt | 1,200 | 6 | 9 | 9.8745 625 | | 33 | 162 | all | partial |
| 466 | 53°28' 11.50" | -8°44'1 7.30" | 64 | dirt | 1,200 | 7.5 | 11° | 6.2233 | 5.3597 93189 | 34 | 107.5 | n/a | n/a |
| 467 | 53°41' 25.50" | -7°35'5 9.56" | 87 | gra vel | 1,200 | 2 | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 468 | 53°41' 52.55" | -7°36'5 4.49" | 80 | gra vel | 1,200 | 2 | 11 | n/a | n/a | n/a | n/a | n/a | n/a |
| 469 | 53°17' 23.5" | -8°44'3 6.5" | 37 | dirt | 1,200 | 2 | 11 | n/a | n/a | n/a | n/a | n/a | n/a |
| 490 | 52°53' 40.8" | -8°33'3. 1" | 45 | gra vel | 1,200 | 4 | n/a | 4.5447 69231 | 8.279 | 30 | 118 | all | all |
| 494 | 52°52' 19.07" | -8°36'5 7.70" | 46 | gra vel | 1,200 | 7 | 22 | 10.161 85 | 4.5883 92857 | 28.333 33333 | 97 | all | all |
| 533 | 53° 9'16.2 7" | -9° 4'58.1 9" | 13 | dirt | 1,400 | 5 | 487 | 11.333 6 | 4.6702 03947 | 40.2 | 94.5 | n/a | n/a |
| 100-102 | 53°28' 42.08" | -9°07'5 4.45" | 17 | gra vel | 1,400 | 4.25 | 25 | 11.603 72 | 5.2765 04274 | 40 | 79 | all | partial |
| 109-135 | 52°54' 40.33" | -9°03'1 9.77" | 27 | roc k | 1,400 | 3.25 | 9 | 10.812 90417 | 6.2934 60898 | 38.375 | 95.5 | all | all |
| 112° | 52°54' 40.33" | -9°03'1 9.77" | 27 | roc k | 1,400 | 1.75 | n/a | 17.031 92857 | 5.0276 66667 | 30 | 162 | all | all |
| 136-140 | 52°54' 41.20" | -9°03'5 7.4" | 25 | dirt | 1,400 | 3 | 9 | 7.6318 07692 | 6.3931 25969 | 31.727 27273 | 109 | all | all |
| 161-162 | 53°16' 28.67" | -9°03'2 3.28" | 7 | dirt | 1,400 | 3.75 | 11 | 7.1978 | | 25.312 5 | 63 | all | partial |
| 164-170 | 53°16' 03.19" | -9°03'1 5.68" | 3 | gar den | 1,400 | 3.75 | 26 | 9.0948 5 | 7.1861 09108 | 29 | 83 | all | all |
| 186-201 | 52°39' 31.8" | -8°37'4 0.97" | 15 | gar den | 1,200 | 3 | n/a | n/a | 8.4831 66667 | 23 | n/a | partial | partial |
| 204-207 | 52°17' 43.36" | -7°15'3 7.21" | 30 | gar den | 1,200 | 3.5 | n/a | n/a | n/a | 24 | 77 | all | all |
| 208-217 | 52°21' 37.4" | -7°34'4 0.8" | 15 | dirt | 2,000 | 4 | 11 | 7.8874 33333 | 6.2910 88235 | 30 | 97 | all | all |
| 239-243 | 52°08' 11.9" | -7°28'0 6.2" | 75 | dirt | 1,400 | 4 | 11 | 7.8019 13043 | 6.2024 2402 | 25 | 116 | all | partial |
| 244-245 | 51°57' 20.3" | -7°50'5 4" | 6 | gar den | 1,000 | n/a | 9 | 8.4236 11111 | 4.4448 27381 | 45 | n/a | n/a | n/a |
| 252-254 | 52°20' 45.6" | -7°19'2 1.6" | 14 | gar den | 1,200 | 4 | 9 | 5.8001 66667 | 6.6159 08929 | 31 | 123 | all | all |
| 263-267° | 52°20' 18.9" | -6°27'5 1.7" | 31 | gar den | 1,200 | 4 | n/a | n/a | n/a | 40 | 126 | partial | partial |
| 298-299 | 52°35' 24.06" | -6°29'3 3.94" | 45 | dirt | 1,200 | n/a | 9 | 10.347 75 | 6.0367 82409 | 37.266 66667 | 116 | n/a | n/a |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 301-306 | 52°16'00.92" | -7°06'03.83" | 27 | dirt | 1,200 | 5 | 9 | n/a | n/a | n/a | 123 | partial | no |
| 311-312 | 54°20'59.11" | -7°39'09.53" | 52 | dirt | 1,400 | 3 | 11 | 8.8053 92857 | 7.2115 625 | 33 | 98 | partial | partial |
| 314-316 | 53°07'43.31" | -8°57'18.21" | 15 | dirt | 1,400 | 2.25 | 26 | 7.0293 19444 | 6.1 | 32.05 | 90.5 | all | all |
| 320-321 | 53°07'43.31" | -8°57'18.21" | 15 | dirt | 1,400 | 2 | 26 | 8.5640 69444 | 5.8242 43601 | 34.5 | 76 | all | partial |
| 321-322 | 53°07'43.31" | -8°57'18.21" | 15 | dirt | 1,400 | 1.75 | 350/26 | 7.4212 66667 | 6.0193 01857 | 35.866 66667 | 83 | all | all |
| 339-348 | 53°05'06.65" | -8°59'12.30" | 76 | dirt | 1,600 | n/a | 22° | 11.184 98795 | 3.0884 82143 | 38.384 61538 | n/a | n/a | n/a |
| 370 (2) | 53°07'43.31" | -8°57'18.21" | 15 | dirt | 1,400 | 2.5 | 316 | 9.1595 22727 | n/a | 22 | 105 | all | partial |
| 370 (3) | 53°07'43.31" | -8°57'18.21" | 15 | dirt | 1,400 | 2.5 | 26 | 9.1595 22727 | n/a | n/a | 115 | all | no |
| 370 (4) | 53°07'43.31" | -8°57'18.21" | 15 | dirt | 1,400 | 2 | 350 | 9.1595 22727 | n/a | n/a | 96 | partial | partial |
| 370 (5) | 53°07'43.31" | -8°57'18.21" | 15 | dirt | 1,400 | 1.5 | 26 | 9.1595 22727 | n/a | 46 | 100 | all | partial |
| 394-395 | 53°25'41.19" | -9°19'14.16" | 11 | gravel | 1,400 | 7.25 | n/a | n/a | 6.1055 2 | 32.285 71429 | 120 | n/a | n/a |
| 397-399 | 53°07'08.13" | -9°04'15.44" | 6 | dirt | 1,600 | 2.5 | n/a | 5.6788 4 | 4.9235 88235 | 32.7 | 74 | n/a | n/a |
| 400-404 | 53°19'28.8" | -9°44'9.9" | 24 | gravel | 1,600 | 5 | 8 | 8.1282 85714 | 7.6465 76923 | 45 | 119.5 | all | all |
| 409-412 | 53°16'54.6" | -9°08'19.3" | 58 | gravel | 1,600 | 5 | 25 | 8.9261 85185 | | 32.533 33333 | 92.5 | partial | no |
| 413-414 | 53°25'42.8" | -9°19'12.1" | 18 | gravel | 1,400 | 4 | 11 | 8.8504 65116 | 5.6751 08899 | 31.4 | 94 | all | all |
| 421-424 | 53°4'45.47" | -8°48'6.10" | 7 | dirt | 593 | 3 | n/a | n/a | n/a | 100 | 162 | all | all |
| 425-442 | 5-6°10'42.61" | 4°23'8.28" | 23 | gravel | 3063 | 7 | 25 | 6.369 | 6.7135 | 39 | 88.5 | n/a | n/a |
| 444-448 | 53°17'51.40" | -8°44'34.70" | 35 | dirt | 1,200 | 4 | 18 | 10.364 52632 | 6.5798 67647 | 37.25 | 144 | all | partial |
| 452-453 | 53°19'24.40" | -8°26'26.77" | 107 | dirt | 1,200 | 3.75 | 19 | n/a | n/a | 42 | n/a | n/a | n/a |
| 454-459 | 53°16'30.1" | -8°38'11.4" | 40 | dirt | 1,200 | 3.5 | 18 | 9.8745 625 | 7.8137 83947 | 29.5 | 95.833 33335 | n/a | n/a |
| 460-462 | 53°16'29.21" | -8°38'28.97" | 40 | gravel | 1,200 | 4 | 22 | 8.7024 54545 | 5.8880 83333 | 36 | 84 | no | no |
| 463-465 | 53°19'32.88" | -8°13'17.22" | 39 | gravel | 1,000 | 7 | 9 | 7.5606 22222 | 4.593 | 29.5 | 109.33 33334 | n/a | n/a |
| 467-468 | 53°41'25.50" | -7°35'59.56" | 87 | gravel | 1,200 | 2 | 9 | 7.7224 28571 | 5.4921 48148 | 23.25 | 83.5 | partial | all |
| 469-472 | 53°17'23.5" | -8°44'36.5" | 37 | gravel | 1,200 | 3 | n/a | 6.5324 78261 | 5.105 | 34.75 | 89.5 | all | partial |
| 472-475 | 53°17'23.5" | -8°44'36.5" | 37 | gravel | 1,200 | 6 | 9 | 7.0311 5 | 4.2655 96154 | 31.736 84211 | 106 | n/a | n/a |
| 472-478 | 52°54'21.54" | -8°32'34.68" | 37 | gravel | 1,200 | 8 | 22 | 8.4348 5 | 5.4921 48148 | 31.625 | 112 | n/a | n/a |
| 50-59 | 53°33'33.90" | -9°53'26.84" | 30 | garden | 2,400 | 1.75 | 26 | 6.3765 | 6.9236 84524 | 24.357 14286 | 117 | all | all |
| 506-510 | 53°24'4.26" | -9°0'27.29" | 20 | dirt | 1,200 | 2 | n/a | 9.4213 47826 | 5.0997 22222 | 26 | 70 | partial | all |
| 529-530 | 53°22'47.0" | -8°51'23.5" | 31 | dirt | 1,200 | 4 | 9 | n/a | n/a | 34 | 92 | partial | no |
| 566-575 | unknown | unknown | n/a | unknown | n/a | 5 | 17 | 8.2102 28571 | 8.9907 3539 | 30.692 30769 | 77.25 | n/a | n/a |
| 75-76 | 53°33'10.66" | -9°56'58.76" | 29 | dirt | 2,400 | 2 | 26 | 8.9821 53846 | 6.7399 75 | 26.25 | 86.5 | all | no |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 77-79 | 53°30' 54.41" | -9°26'5 0.81" | 9 | gravel | 2,800 | 2.25 | 25 | 10.101 85185 | | 34.857 14286 | 145 | all | partial |
| 80-86 | 53°30' 54.53" | -9°26'5 1.73" | 10 | dirt | 2,800 | 2.75 | 25 | 10.272 72 | 6.5053 54167 | 38.2 | 93.5 | all | all |
| 91-99 | 53°28' 42.08" | -9°07'5 4.45" | 17 | dirt | 1,400 | 3.5 | n/a | 9.092 | 7.4337 97619 | 40 | 159 | partial | all |
| bur-0 | n/a | n/a | n/a | limest one | 2,000 | 3.5 | 400+ | 9.8163 61111 | 4.3763 53333 | 90 | 163 | all | all |
| col-0 | n/a | n/a | n/a | n/a | n/a | 6.5 | 22 | 7.4983 66667 | 2.9675 25253 | 22 | 112.2 | all | all |
| *ler* | n/a | n/a | n/a | n/a | n/a | 1 | 16 | n/a | 5.6409 09091 | n/a | n/a | n/a | n/a |
| ns_10 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | 13.294 09756 | 4.8660 03521 | 50 | 158 | n/a | n/a |
| ns_15 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | 13.484 14583 | n/a | 33 | 98 | n/a | n/a |
| wang 35 | n/a | n/a | n/a | n/a | n/a | n/a | 19 | n/a | n/a | n/a | n/a | n/a | n/a |
| wang 25 | n/a | n/a | n/a | n/a | n/a | n/a | 10 | n/a | n/a | n/a | n/a | n/a | n/a |
| 32 | n/a | n/a | n/a | n/a | n/a | 5.25 | 9 | 8.8827 77778 | 5.6969 33114 | 40 | 142 | all | all |
| 313** | n/a | n/a | n/a | n/a | n/a | n/a | 11 | n/a | n/a | n/a | n/a | n/a | n/a |

**Table S1. Irish *A. thaliana* samples genotyped through DArT-Seq.** Where location, altitude and soil were recorded from the collection site. Table includes various phenotypic and genotypic data obtained from the collected accessions. Yellow highlights point out earlier flowering times observed in the NS plants relative to full length TTC/GAA expansion Bur-0 plants at 23 °C. n/a indicates missing data. * Asterisks indicate wild accessions with dramatic TTC/GAA expansions.

| Gene | Forward primer | Reverse primer | PCR fragment (in bp) |
|---|---|---|---|
| FXN (SKB_2407& 2408) | AAGGAAGTGGTAGAGGGTGTTTCACGAGG A | TTTGGATCCAACTCTGCTGACAACCCAT GCTGTCCACA | 748 + GAA/TTC |
| FXN* (SKB_2409& 2410) | CTGGGTGGAGATCTAGGAACC | TTTCCCAGTCCAGTCATA | 69 |
| GAPDH* (2024 &2025) | GTTTACATGTTCCAATATGA | TCTCCATGGTGGTGAAGACG | 152 |
| SLC39A10*( SKB_2081& 2080) | ACTTGGGCCTTGGAGAGAGA | GTGATGACGTAGGCGGTGAT | 281 |
| MARK4*(SK B_2082&208 3) | GTGGTTCCACCATCCGCAG | TGGTGAAGAGGTTGGTGGTG | 171 |
| GDI1*(SKB_ 2084&2085) | TTGCCAGCACTACTGTGGAG | CACCTGGCTCTCACAACCAT | 140 |
| FOXG1*(SK B_2086&208 7) | GAGGTGCAATGTGGGGAGAA | TTCTCAAGGTCTGCGTCCAC | 196 |
| DAPK1* (SKB_2088& 2089) | CTGCTTCGGAGTGTGAGGAG | CCCCCTCATGCATCAGTCTC | 268 |
| ATF6B*(SK B_2091&209 2) | TGATGGCTCCTCAGGCAAAG | CTGGATGGCATTGGGACAGT | 90 |
| TARDBP*(S KB_2093&2 094) | GGTTGGGGATCAGCATCCAA | CAACCACAACCCCACTGTCT | 114 |
| SLC39A10*( SKB_2097& 2098) | GGAGAACAGAACCGGGGAGT | TTTCCTCCACAGGAAGGTGAG | 234 |
| IIL1 (SKB_608,6 09 &561) | GCCTGCAATGACCTTCTTGT | GGGATGACAATGACGGAGAA or GGGATGACAATGACGGAGAA | ~648 + TTC/GAA or ~604+ TTC/GAA |
| IIL1*(SKB_2 423&2434) | TCGCTTGAATGAACCTCAAGTGTG | TGCCCTTCTTTGTGTCCCATCC | 78 |
| TUB2* (1016&1017) | CCAGCTTTGGTGATTTGAAC | CAAGCTTTCGGAGGTCAGAG | 100 |

**Table S2. List of oligonucleotides used to determine *IIL1* and *FXN* genomic expansion as well as various gene expressions.** Gene targeted and PCR fragment size are indicated. * indicates oligos are specific for gene CDS and used in q-RT-PCR analysis. *FXN* & *IIL1* PCR fragment sizes are dependent on length of GAA/TTC tract. For q-RT-PCR analysis amplification products for each gene were run on 1% agarose gels (+Ethidium bromide) to confirm a single product of expected size. *FXN* gDNA oligos and PCR cycling parameters were taken from (Evans-Galea, Carrodus et al. 2012). All other oligonucleotides were designed using (NCBI, Primer-BLAST). [NB: Expected PCR product size for 608 & 609 primers in absence of repeat = 1093 bp. Expected PCR product size for 608 & 561 in absence of repeat = 604 bp. Thus product size varies between the Bur-0 and various natural suppressors]. All sequences are given in the 5'-3'.

| Chr | Marker | Position (bps) | Forward primer (5'-3') | Reverse primer (5'-3') | PCR product size in Bur-0 (bps) | PCR product size in Col-0 (bps) | PCR product size in At322 +/- *iil* (bps) |
|---|---|---|---|---|---|---|---|
| I | nga59 (SKB_2011&2012) | 8642 | GCATCTGTGTTCACTCGCC | TTAATACATTAGCCCAGACCCG | 120 | 100 | 110 |
| I | F5114 (SKB_1031&1032) | 24370345 | CTGCCTGAAATTGTCGAAAC | GGCATCACAGTTCTGATTCC | 305 | 220 | 300 |
| II | MSat2.5 (SKB_1033&1034) | 208179 | TGAGAGGGACAGATAGGAA | ATCAAAGGGATACTGACAA | 50 | 100 | 75 |
| II | F504-3 (SKB_983&984) | 512733 | GAATGTTTTGAAGGATATCTCAG | GAAAAATGGAGCCACGAAATAAGC | 200 | 300 | 200 |
| II | T10J7-T7 (SKB_1308&1309) | 4851327 | TCTCTGTGCTTTCTCTTTCCTGAC | GCAATGCTACCGCTCTGATAG | 500 | 300 | 500 |
| II | LUGSSLP531 (SKB_1352&1353) | 5172823 | CTGGTTTTGGACATTAACAGCAA | CAAGATGATGATGACGATGAGGAT | 100 | 100 | 75 |
| II | LUGSSLP765 (SKB_1350&1351) | 7228683 | GGATGTGAGATTACTGCACAAGGAT | ACGAAGAGTGACAATGGTTGGA | 150 | 150 | 150 |
| II | SKB (SKB_677&678) | 8090763 | ACGGATCGAAGTGGGTATTC | AGCTGCTCAAGGAAAAGCTG | 1,000 | 800 | 1,000 |
| II | SKB (SKB_600&601) | 8111803 | TTTGAAAGGAGGGTTTGCGTTAGA | AGCTTGCCAAAGGAAAGTGA | 300 | 300 | 300 |
| II | Msat2.11 (SKB_1392&1393) | 8220744 | GATTTAAAAGTCCGACCTA | CCAAAGAGTTGTGCAA | 275 | 300 | 275 |
| II | Msat2.36 (SKB_1398&1399) | 8678440 | GATCTGCCTCTTGATCAGC | CCAAGAACTCAAAACCGTT | 200 | 175 | 200 |
| II | PLS7 (SKB_1394&1395) | 9806670 | GATGAATCTTCTCGTCCAAAAT | GACAAACTAAACAACATCCTTCTT | 175 | 150 | 175 |
| II | Msat2.17 (SKB_1396&1397) | 10726715 | GATTCCACCATATGTGGAT | CTTCGTCACTGCCAATAC | 150 | 110 | 110 |
| II | Msat2.9 (SKB_1335&1336) | 18152579 | TAAAAGAGTCCCTCGTAAAG | GTTGTTGTTGTGGCATT | 175 | 150 | 170 |
| III | Msat3.2 (SKB_1225&1226) | 905503 | AAGGTACGGCGGTGGATATTG | CGGGGATTTCTTCTTCCTGTG | 150 | 175 | 150 |
| III | Msat3.13 (SKB_1300&1301) | 20386139 | TTGTGTGTTTGCGATC | CATATCCGTTTTTATGTTTT | 175 | 200 | 200 |
| IV | nga8 (SKB_1302&1303) | 5628813 | GAGGGCAAATCTTTATTTCGG | TGGCTTTCGTTTATAAACATCC | 250 | 300 | 250 |
| IV | MSat4.35 (SKB_1325&1326) | 7549254 | CCCATGTCTCCGATGA | GGCGTTTAATTTAATTTGCATTCT | 280 | 270 | 280 |
| IV | JAERI-4 (SKB_1288&1289) | 7768715 | AGGAGAAACTTGCAACAGAA | CAACAAAGCTCTGCACAATA | 150 | 200 | 150 |
| IV | G-1311 (SKB_1201&1202) | 9963075 | GGTAATTTTCTTGGAGACCC | CTAACTAGATCGTCCCTCGT | 140 | 150 | 140 |
| IV | N-1200 (SKB_1151&1152) | 18096262 | GCGAAAAAACAAAAAAATCA | CGACGAATCGACAGAATTAGAGG | 110 | 150 | 130 |
| V | NGA151 (SKB_1011&1012) | 4670058 | GTTTTGGGAAGTTTTGCTGG | CAGTCTAAAAGCGAGAGTATGATG | 750 | 750 | 750 |
| V | Nga139 (SKB_1013&1014) | 8326953 | GGTTTCGTTTCACTATCCAGG | AGAGCTACCAGATCCGATGG | 300 | 320 | 300 |

**Table S3.** Genotype results from 22 SSLP markers distributed across *A. thaliana's* 5 chromosomes. Showing marker sizes in Bur-0, Col-0 and At322- a wild accession segregating for the *iil* trait.

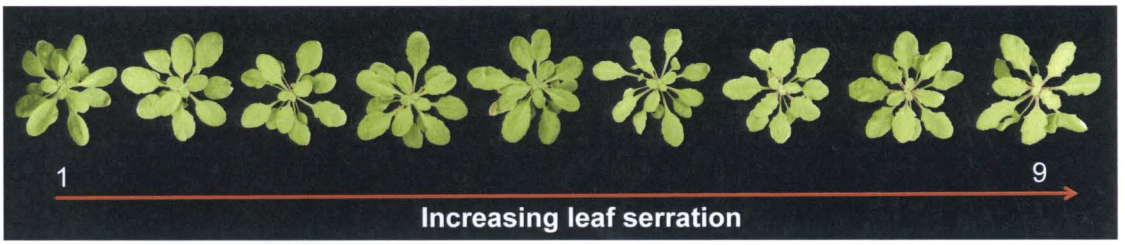



**Figure S1. Leaf Serration internal scale generated from wild Irish *A. thaliana* accessions used in this study.** Scale was given to 3 independent volunteers and asked to assign a number from 1-9 characterizing each accessions degree of leaf serration. The average was taken for use in this study.

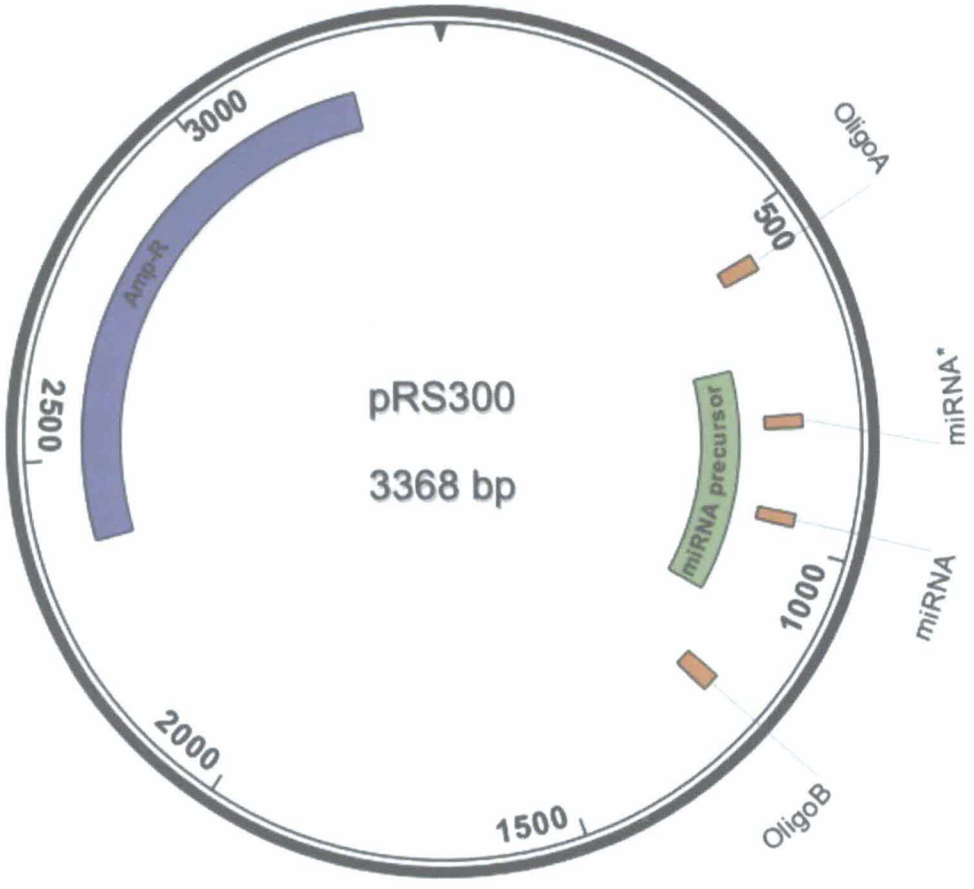



**Figure S2. *pRS300* vector.** Showing the *Ampicillin resistance* gene (*Amp*$^{R}$), the location of the amiRNA backbone precursor (green) and the binding sites for primers 'A' and 'B' (primers SKB_792 & SKB_793 respectively, Table S2). The vector was designed using the SeqBuilder softeare (DNASTAR, USA) and supplied by Mr Aleksej Stevanovic.

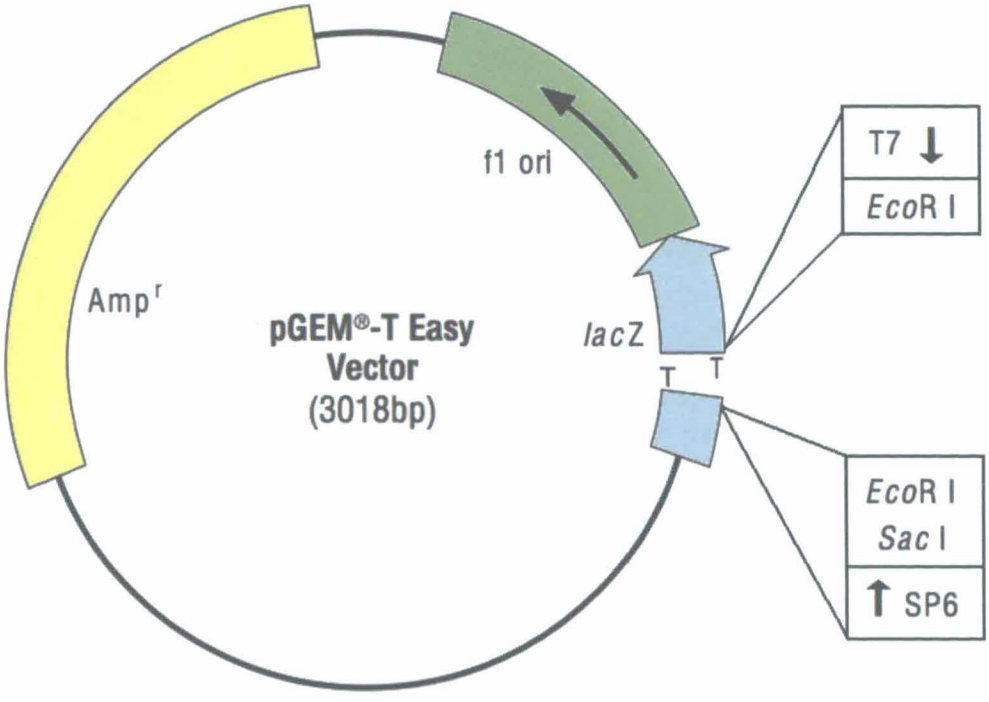



**Figure S3. *pGEM-T easy* vector** containing the *Ampicillin resistance* gene (*Amp*$^{R}$), the origin of replication (f1 ori) and the *β–Lactamase* gene (*lacZ*) for blue/white colony selection. The vector contains the T7 and SP6 regions, which were used to design primers in order to sequence the inserts. (Modified from the '*pGEM-T* and *pGEM-T easy* Vector Systems Manual', Promega)

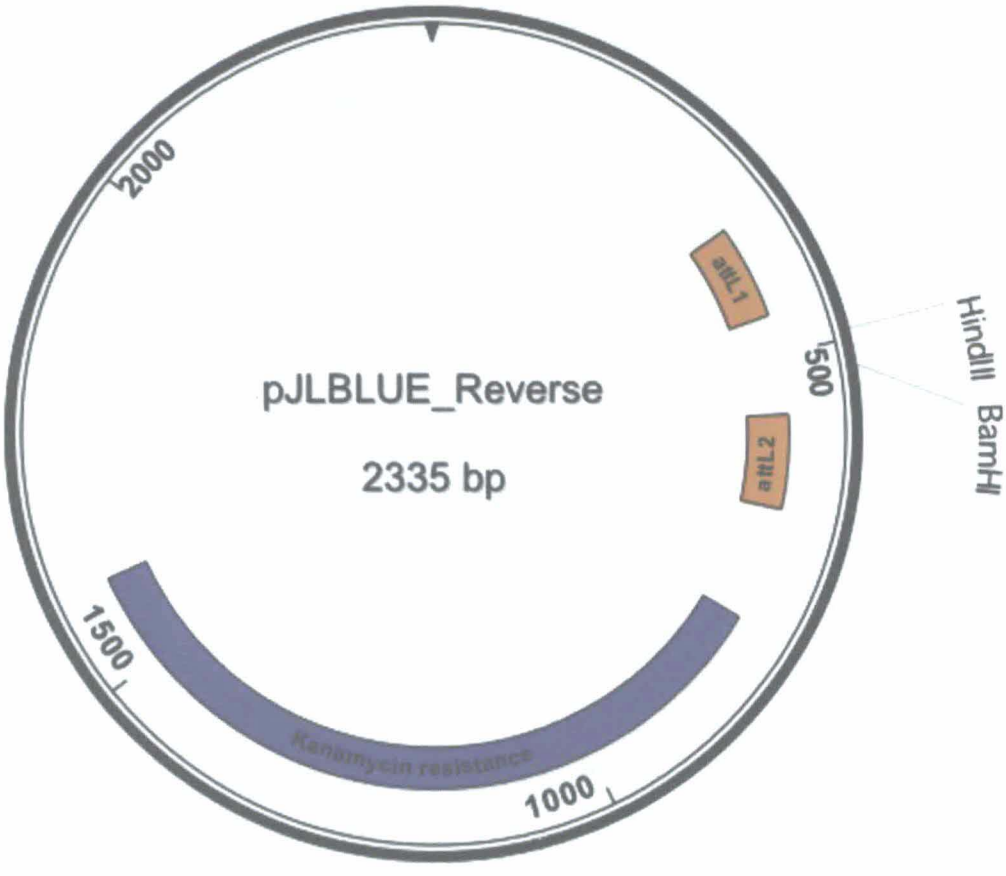



**Figure S4. *pJLBlue reverse* vector** showing the attL1 and attL2 sites, the Kanamycin resistance gene (in blue), and the restriction sites used to verify the presence of the insert. The vector was designed using SeqBuilder software (DNASTAR, USA) and supplied by Mr Aleksej Stevanovic.

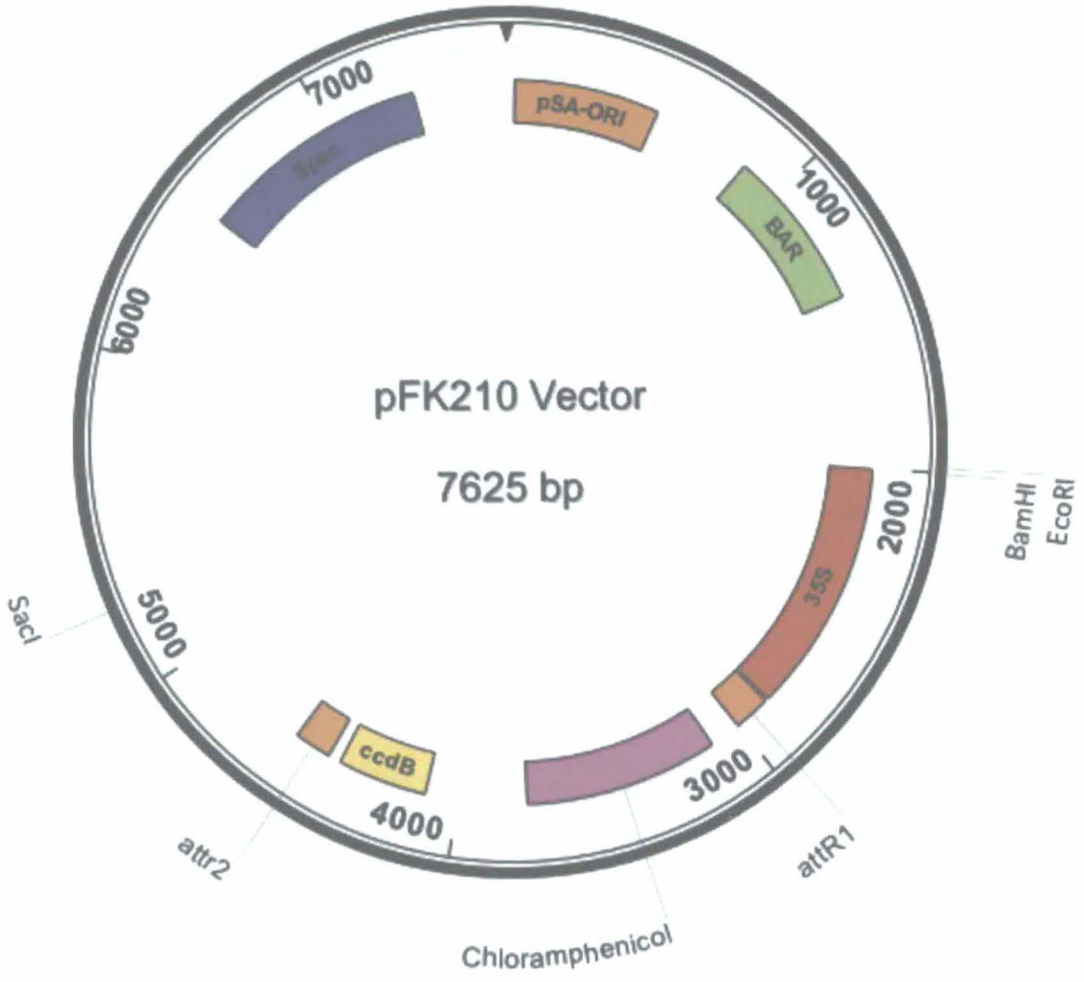



**Figure S5. *pFK210* vector** containing replication origins for both *E. coli* and *A. tumefaciens* cells. The figure shows the attR1 and attR2 sites, the 35S promoter (red), the *Chloramphenicol* gene (purple), the *Spectinomycin resistance* gene (blue), the *BAR* gene (green) and the restriction sites used in cloning. The vector was designed using the SeqBuilder software (DNASTAR, USA) and supplied by Mr Aleksej Stevanovic.

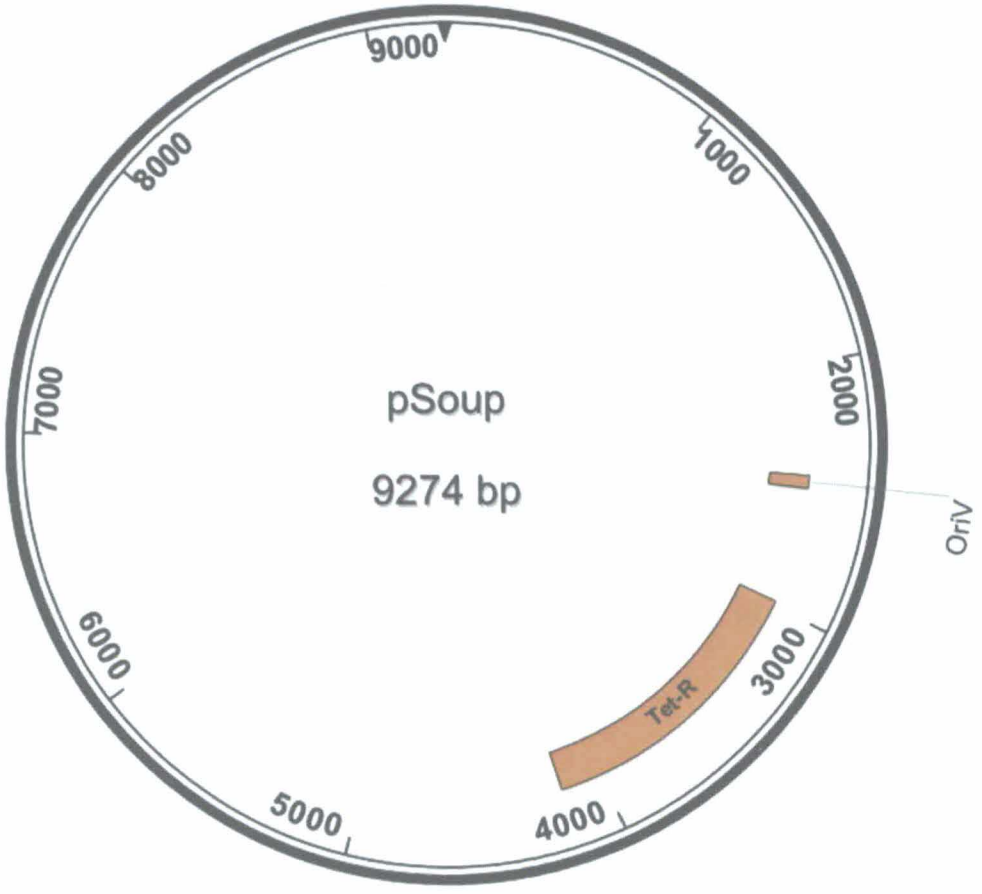



**Figure S6.** ***pSoup*** used as a helper plasmid. The vector contains a *Tetracycline resistance* gene and the origin of replication (OriV). The vector was designed using the SeqBuilder software (DNASTAR, USA) and supplied by Mr Aleksej Stevanovic.

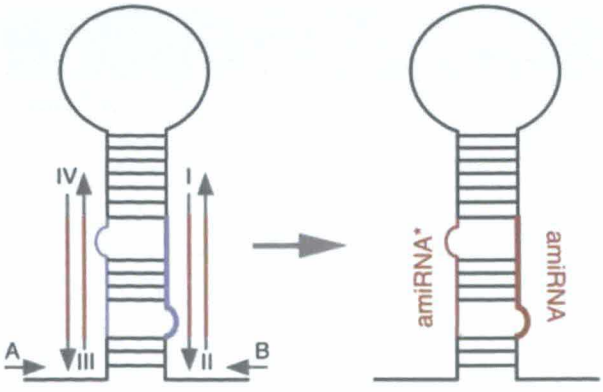



**Figure S7. *pRS300* vector used as template for amiRNA cloning.** The left figure shows construct specific primers I, II, III and IV (designed specifically for each amiRNA). The right figure shows the amiRNA* and amiRNA sequences (red) which are replaced by the amiRNA specific sequences by site-directed mutagenesis (Schwab, Ossowski et al. 2006). Primers 'A' and 'B' depicted by small black arrows on the left are the SKB_792 & SKB_793 primers respectively, (Appendix - Table S3 & S4).

| Gene | Primer (5' to 3') |
|---|---|
| atMSH2 (1) I (SKB_745) | TTACAGTATGGAGTAGCGCAA |
| atMSH2 (1) II (SKB_746) | TTGCGCTACTCCATACTGTAA |
| atMSH2 (1) III (SKB_747) | TTACGCTACTCCAAACTGTAT |
| atMSH2 (1) IV (SKB_748) | ATACAGTTTGGAGTAGCGTAA |
| atMSH2 (2) I (SKB_749) | TAATCATACAGGCTTTTGCAT |
| atMSH2 (2) II (SKB_750) | ATGCAAAAGCCTGTATGATTA |
| atMSH2 (2) III (SKB_751) | ATACAAAAGCCTGAATGATTT |
| atMSH2 (2) IV (SKB_752) | AAATCATTCAGGCTTTTGTAT |

**Table S4. Primers used for construction of the MSH2 amiRNA backbones, MSH2-(1) and MSH2-(2) via site directed mutagenesis.** [NB: 2 different amiRNA constructs targeting the same gene were generated to increase the chances of a successful knock-out].

| Primer | Sequence (5' to 3') |
|---|---|
| SKB_480 attR1 | ATTTAGGTGACACTATAG |
| SKB_482 attR2 | TAATACGACTCACTATAGGGG |
| SKB_478 T7 | GTAAAACGACGGCCAGT |
| SKB_479 SP6 | CAGGAAACAGCTATGAC |
| SKB_792 (A) | CTGCAAGGCGATTAAGTTGGGTAAC |
| SKB_793 (B) | GCGGATAACAATTTCACACAGGAAACAG |

**Table S5. Primers used in MSH2-amiRNA cloning.** Where primers SKB_480 & 482 were used for the attR1 & attR2 gateway cloning from *pJLBlue* into *pFK210* vector. SKB_478 & 479 were used for the T7 & SP6 cloning into *pGEMT* easy vector. SKB_792 & 793 were used for the A & B oligo cloning into *pRS300* easy vector and construction of the *amiRNA-MSH2* backbones.