Insight into the Disease Processes of Fragile X Premutation Carriers Associated Pathologies

Expression-profile characterization and identification of a novel pathogenic mechanisms

Elisabet Mateu Huertas

TESIS DOCTORAL UPF / 2013

DIRECTORS DE LA TESIS

Dra. Eulàlia Martí Puig

Dr. Xavier Estivill Pallejà

BIOINFROMATICS AND GENOMICS DEPARTMENT,

CENTER FOR GENOMIC REGULATION (CRG)

upf. Universitat Pompeu Fabra Barcelona

Als meus pares

“Hope lies in dreams, in imagination and in the courage of those who dare to make dreams into reality.”

Jonas Salk

ACKNOLEDGEMENTS

Ara si, sembla que ha arribat el moment d'escriure els agraïments. Han passat mes de cinc anys des de aquell primer dia que vaig arribar al laboratori. I, sense cap dubte, moltes són les persones a qui he d'agraïr el recolzament que m'ha donat durant tot aquest temps tant a nivell professional com personal.

Primer de tot vull agrair a en Xavier i l'Eulàlia l'oportunitat que he tingut de formar-me en un laboratori i en un centre tan excepcional com el CRG, tot un privilegi. Gracies per l'esforç i la dedicació que heu fet perquè aquest projecte sigui possible. Eulàlia, merci per donar-me sempre una opinió sincera de la feina feta. Són infites les coses que he parés durant aquets cinc anys. Merci

Agraïr als membres del comitè de tesis que durant aquests cinc anys han seguit d'aprop aquest projecte, Juan, Mara i Montse. Gràcies per les vostres aportacions al projecte, per les critiques constructives i les ganes que hi heu posat tots perquè aquest treball sigui el mes fructífer possible. En especial, vull agraïr-li a en Juan que m'hagi escoltat i recolzat en els moments en que mes ho he necessitat. Gràcies de tot cor. I per suposat, també agraïr als membres del tribunal de tesis per ser tan propers i mostrar el seu entusiasme per participar en un dia tan important per a mi. Moltíssimes gràcies.

Evidentment, he de reconèixer que he tingut especialment sort de poder compartir el laboratori amb tanta gent excepcional: Marta, Anna, Laia,

Georgia, Elisa, Esther, Kelly, Dani, Johanna, Tere, Anna H. Aparna, Marc, Joan, Laura, Hyun, Charlotte, i altres que ja han marxat com la l'Elena, la Lorena, la Monica, la Birgit, la Susana, l'Ester, la Silvia i en Sergi. Espero no deixar-me a ningú! Entre tots heu aconseguit que la muntanya russa que suposa fer una tesis estigui plena de moments inoblidables i anècdotes divertides. Mònica, per acompanyarme durant els meus primers passos dins del laboratori. Birgit, sincerament ha estat tot un plaer poder treballar amb tu. Però del cert, el que mes he apreciat ha estat la teva honestedat i l'amistat que hem compartit durant aquest temps que, de segur durarà molts anys. Marta, saps que sense tu aquets laboratori seria una trinxera! Elisa, Laia i Johanna, finalment hi hem arribat noies. Sou genials, i m'heu ajudat un munt en aquest últim tram de la tesis. Puig, ha estat un plaer poder discutir els protocols i treballar amb tu aquesta última etapa de la tesis. I en Joan i la Laura, les dues gotes d'aigua fresca que han arribat al laboratori. Sou genials! Ens aquets mesos d'escriptura us he trobat molt a faltar. Merci per els ànims, la comprensió i el recolzament que m'heu donat. Són d'aquelles coses que no s'obliden així com així! Segur que la vostra tesis us anirà molt bé. Gaudiu molt de la oportunitat que teniu, i molta sort.

També hi ha hagut un munt de persones d'altres laboratoris cal que esmenti, persones amb qui ens hem ajudat a treballar o a fer passar les hores d'una manera mes agradable. Em refereixo sobretot a persones de l'antic programa de Gens i Malalties: Chiara, Kriszti, Balducci, Dani, Xevi, Anabel, Susana, Monica, Davide, Meri, Maria, Ignasi, Esteban, Vicky, Ana,... Companys la beca de la Caixa: Camilla, Giovanni, Johanna, Chang, Cristina, Gabriel i Judith i moltíssimes altres persones com la Aina, Caro, Amy, Iris, Martin, Isma, Aitor,... Entre tos heu fet que la tesis sigui una experiencia inoblidable. I com oblidar a les persones del meu casi segon laboratori: Anne, Peter, Erik, Kiana,

Judith, Besray, Marie, Tony, Raik, Almer, Tobias, Sira, Maria i Berni. I també els companys del equip de volei platja, Salva, Maria, Maik, Jordi, Oliver Leszek, Silvia and Romain.

I also have to thank my second family. Something happened during my first year PhD, the perfect conditions where placed for the formation of a great friendship that without doubts will persist in time. Almer, Raik, LNA, JOJO, Meri, Kiki, Empanado, Esther and Marc. You have been my allies for the last 5 years. It's countless the time we spend together having a coffee, going out, traveling, or just supporting each other. Thanks for everything.

A en Jordi i la Judit. Ja fa dotze anys que ens vam conèixer a la UAB! Anys plens de moments i llocs inoblidables. Merci per escoltar-me i recolzar-me, m'heu ajudat moltíssim tan els moments bons com els que no eren tan bons. Sou els millors amb diferència!

I evidentment a la meva família que sempre m'ha demostrat el seu suport incondicional. A tu Alfred, perquè m'has fet veure que tot es possible amb esforç i treball. I el meus pares, merci per fer possible l'educació privilegiada que he tingut, tan dins com fora de casa, per ensenyar-me a ser independent i per ser la veu de la raó. Són infinites les coses que us podria agraïr, gràcies!

And last but not least, you Tobias. Thanks for being by my side during big part of this journey. Thanks for your moral support, helping me manage difficult situations, for the trips together, the fun, … I could just keep on for long time. This end is a new beginning that I can't wait to start! So let's GO!!

Summary

Male premutation (PM) carriers presenting between 55 and 200 CGG repeats in the fragile X mental retardation *(FMR1)* gene are at risk to develop fragile X tremor/ataxia syndrome (FXTAS), and females to undergo premature ovarian failure (POF). These pathologies are likely caused by the toxic gain of function of the premutated *FMR1* mRNA. In this thesis, we have characterized the transcriptome alterations associated to *FMR1* premutation and further evaluated the relevance of the biogenesis and activity of small RNAs formed by the repeated CGG (sCGG) in neuronal dysfunction linked to the *FMR1*-PM. In blood of *FMR1* premutation carriers (fXPCs) we have detected a strong deregulation of genes enriched in FXTAS-relevant biological pathways. We have also identified a deregulated gene *(EAP1)* that may underlie POF1 in female fXPCs. In addition, we found increased levels of sCGG in different models of *FMR1*-PM and further demonstrated the neurotoxic activity of sCGG through a mechanism dependent on RNA induced silencing machinery. We propose that the activity of sCGG may contribute to transcriptome perturbations with downstream pathogenic consequences. Overall, we provide mechanistic insight into the disease process and further suggest targets for FXTAS diagnosis to the myriad of phenotypes associated with fXPCs.

Resum

Els homes portadors d'una premutació (PM) en el gen del retard mental del cromosoma X fràgil (*FMR1*) tenen una expansió de trinucleòtids CGG d'entre 55 i 200 repeticions, presenten un alt risc de desenvolupar el síndrome de tremolor/atàxia associat al cromosoma X fràgil (FXTAS), i les dones fallida ovàrica precoç (POF). Aquestes malalties estan causades probablement per la funció tòxica de l'ARN missatger del gen *FMR1* premutat. En aquesta tesi, s'han caracteritzat les alteracions del transcriptoma associades a la premutació en el gen *FMR1* (*FMR1*-PM) i s'ha analitzat la rellevància de la biogènesi i activitat d'un ARN de mida petita, format per CGG repetits (sCGG) en la disfunció neuronal relacionada amb al *FMR1*-PM. En sang de portadors de la premutació de *FMR1* (fXPCs) s'ha detectat una pronunciada desregulació de gens enriquits en vies biològiques rellevants en FXTAS. També hem identificat un gen desregulat (*EAP1*) que podria contribuir a POF1 en dones portadores de la permutació. A més, hem trobat un augment dels nivells de sCGG en diferents models de *FMR1*-PM i hem demostrat la seva activitat neurotòxica a través d'un mecanisme dependent de la maquinària de silenciament gènic. Proposem que l'activitat dels sCGG podria contribuir a causar alteracions en el transcriptoma i desencadenar mecanismes patogènics. Aquest treball ofereix un nou enfoc en procés de la malaltia i proporciona un conjunt de dianes amb

possible utilitat diagnòstica per a la gran varietat de fenotips associats amb fXPCs.

PREFACE

Aging population is more sensitive to suffer progressive lost of neuronal function, leading to motor and cognitive behavioral deficits. Such disorders, called neurodegenerative disorders, can be hereditary or sporadic and are often associated with atrophy at the central or peripheral structures of the nervous system. As life expectancy increases, the number of aged population and thus the necessity to prevent those age related disorders also expands.

In the last 20 years, DNA expansions of trinucleotide repeats (TNR) have been related to several neurodegenerative disorders, the so-called trinucleotide repeat disorders (TREDs). There are around 30 different TREDs and they are all characterized by an expansion of a triplet DNA sequence beyond a specific threshold. Some of the most studied are the Huntington's disease, a number of spinocerebellar ataxias and myotonic dystrophy among others.

This thesis is focused in the study of the TNR associated with the *FMR1* locus and the multiple number of phenotypes related to it. In particular, this thesis aims to provide a better understanding of the biological consequences of the CGG TNR expansion in the rage 50 and 200 repeats occurring at the 5'UTR of the *FMR1* gene. The thesis provides a comprehensive approach to understand the overall transcriptome changes linked to this TRED. In addition, it presents a novel pathogenic mechanism of action for the expanded TNR, which has been already reported in another TRED.

Barcelona, August 2013

Table of Contents

ACKNOLEDGEMENTS V
SUMMARY IX
RESUM XI
PREFACE XIII
TABLE OF CONTENTS XV
FIGURE LIST XIX
TABLE LIST XXIII
SUPPLEMENTARY TABLE LIST XXV
1. INTRODUCTION 1
1.1. Trinucleotide repeat disorders (TREDs) 1
1.2. The Fragile X Site 7
a. *FMR1* alleles 7
b. *FMR1* instability 8
1.3. Clinical features of the fragile X families 10
a. Full mutation (FM) alleles 10
b. Premutation (PM) alleles 12
1.4. Fragile X tremor/ataxia syndrome (FXTAS) 14
a. Clinical features 14
* FXTAS symptomatology 14
* Radiological and neuroanatomical features of FXTAS 15
b. Diagnostic criteria 16
c. Neuropathology of FXTAS 17
d. Molecular signature of FXTAS 19
e. Models for RNA gain of function toxicity in FXTAS 21
* Titration of RNA binding proteins by CGG repeats 21
* RNA-mediated protein aggregation 23

* Antisense-RNA effects on expression..... 23
* RAN-mediated expression of alternative peptides..... 24
* siRNA formation form TNR structures..... 25
f. Cellular pathology in FXTAS 26
g. Modeling of FXTAS 27
* Fly model 27
* Mouse model..... 29
h. Treatment.....32
1.5. Fragile X associated premature ovarian insufficiency (FXPOI) .33
a. Terminology 33
b. Clinical features.....34
c. Modeling FXPOI.....35
1.6. Non-coding RNA 36
a. Small ncRNA 38
* Biogenesis of regulatory small RNAs..... 39
* Silencing mechanism..... 42
b. Long non-coding RNAs (lncRNA) 44
2. HYPOTHESIS AND OBJECTIVES 47
3. RESULTS..... 51
3.1. Evaluation of transcriptome changes in peripheral blood samples from *FMR1* premutation carriers 51
a. Identification of the biological pathways altered in peripheral blood of premutation carriers 51
b. Validation of candidate genes altered in peripheral blood samples of *FMR1* carriers in brains of the CGG-KI mouse model ...60
c. Assessment of the contribution of the levels of expression of the premutated allele in modulating the transcriptome. 64
3.2. Evaluation of the involvement of the gene silencing machinery in PM-FMR1 pathogenesis.....71

a. Identification of small CGG molecules 72
* Determination of sCGG in human carriers of the *FMR1* PM 72
* Determination of sCGG species in brain samples of the KI model 74
* Determination of sCGG species in an in vitro cell model expressing the *FMR1* premutation 75
* Characterization of sCGG molecules 77
b. Evaluation of the toxic effects of small CGG 78
c. Evaluation of the involvement of the siRNA biogenesis machinery in the formation of small CGG molecules 83
d. Evaluation of whether the toxic effects of small CGG molecules are dependent on the RNA induced silencing complex (RISC) 86
e. Identification of putative silencing targets of the small CGG molecules 87
3.3. Supplementary information 91
4. DISCUSSION 113
4.1. Evaluation of transcriptome changes in peripheral blood samples from *FMR1* PM carriers 113
4.2. Evaluation of the involvement of the gene silencing machinery in fXPCs pathologies 119
5. CONCLUSIONS 123
6. MATERIAL AND METHODS 125
6.1. Human Samples 125
6.2. Mouse samples 126
6.3. Total and small RNA isolation 126
6.4. Microarray Hybridization and Analysis 127
6.5. Pathway enrichment analysis 128
6.6. qPCR validations 128

6.7. Generation of *FMR1* 5'-UTR vectors 129
6.8. Cell culture 130
6.9. Transfections 131
6.10. sCGG amplification 132
6.11. sCGG sequencing 133
6.12. Western blotting 134
6.13. Cell toxicity assays 134
6.14. Probes for ROS detection and mitochondrial labelling 135
6.15. In silico identification of sCGG targets 135

7. ANNEX 137
7.1. Publications 137
7.2. Communication to scientific meetings 138
a. Poster presentations 138

8. BIBLIOGRAPHY 141

Figure list

Figure 1. Base excision repair (BER) and nucleotide excision repair (NER) mechanisms in long expansions of trinucleotide repeat (TNR)..... 3
Figure 2. Slippage model for short expansions in trinucleotide repeats..... 4
Figure 3. *FMR1* allele..... 8
Figure 4. Clinical manifestations in *FMR1* premutation carriers 13
Figure 5. Radiological images form FXTAS brains..... 16
Figure 6. Histopathological changes in FXTAS brains..... 19
Figure 7 *FMR1* transcriptional start sites..... 20
Figure 8. Representation of DGCR8 role in FXTAS pathogenesis..... 22
Figure 9. Model for CGG RAN Translation in FXTAS..... 25
Figure 10. Age of menopause among FMR1 alleles..... 35
Figure 11. Transcriptional patterns of ncRNAs..... 37
Figure 12. Function of long-non coding RNAs..... 45
Figure 13. Gene expression analysis in peripheral blood samples from fXPC..... 53
Figure 14. EAP1 expression in fXPCs females..... 59
Figure 15. Cluster analysis of differential expressed genes in brains from the CGG-KI mouse model compared to control mouse..... 63
Figure 16. *FMR1* 5'-UTR expression vectors..... 65

Figure 17. Expression quantification for the *FMR1* 5'-UTR vectors..... 66
Figure 18. Cluster analysis of differential expressed genes upon transfection of the *FMR1* 5'-UTR vectors..... 68
Figure 19. Expression profile of 5 different gens upon transfection of the *FMR1* 5'-UTR vectors..... 69
Figure 20. Expression profile of 5 different gens upon transfection of 21*CGG at two different concentrations..... 70
Figure 21. Identification and quantification of sCGG in lymphocytes cell lines of FXTAS patients..... 73
Figure 22. Identification and quantification of sCGG in peripheral blood from fXPCs 74
Figure 23. Cetection of sCGG molecules in brains of CGG-KI mouse model..... 75
Figure 24. Identification and quantification of sCGG in SH-SY5Y neuroblastoma cell line..... 76
Figure 25. sCGG sequence in lymphocyte cell lines of FXTAS 77
Figure 26. CGG-siRNA and 79*CGG vector induce toxicity in neuroblastoma human cells 79
Figure 27. Increase of superoxide anion concentration in cells expressing 79*CGG vector..... 80
Figure 28. Small RNA derived from cells expressing 79*CGG vector produce neuronal toxicity 81
Figure 29. Anti-sCGG prevent cell toxicity mediated by small RNA fraction isolated from cells expressing 79*CGG vector 82

Figure 30. Knock down (KD) of Dicer reduces cellular toxicity mediated by 79*CGG 84
Figure 31. Expression levels of *Fmr1* and *Dicer* in brains of control and CGG-KI mouse model..... 85
Figure 32. Knock down (KD) of Ago2 prevents cellular toxicity mediated by 79*CGG and CGG-siRNA 86
Figure 33. Venn diagram showing overlap between down-regulated genes in fXPCs and putative sCGG targets 88

Table List

Table 1. Clinical and molecular characteristics of some neurological expansion repeat disorders 6
Table 2. Clinical signs associated to FM alleles 11
Table 3. Effects of directed CGG repeat expression in different *Drosophila M.* tissues 28
Table 4. Comparison of FXTAS with the CGG KI mouse model 31
Table 5. Classification of small ncRNAs 38
Table 6. Classification of lncRNA 44
Table 7. Clinical and molecular characteristics from male fXPC 52
Table 8. Pathway enrichment analysis of differential expressed genes in blood of FXTAS patients vs. control individuals 55
Table 9. Cluster distribution of differentially expressed lncRNA, in patients with FXTAS vs. control individuals 56
Table 10. Clinical and molecular characteristics of female fXPC 57
Table 11. qPCR relative quantification of specific FXTAS-blood-DEG in different brain areas of the CGG-KI FXTAS mouse model 62
Table 12. sCGG seed regions in putative targets 89

Supplementary table list

Supplementary Table 1 Predicted upstream regulators and its activation state in FXTAS patients, in IPA of blood-DEG 91

Supplementary Table 2. Pathway enrichment analysis of differential expressed genes in the hippocampus, prefrontal cortex and striatum of the CGG-KI mouse model vs. control mouse 93

Supplementary Table 3. Predicted upstream regulators and its activation state in hippocampus samples of the CGG-Ki mouse modelsamples, according to IPA 94

Supplementary Table 4. Predicted upstream regulators and its activation state in prefrontal cortex samples of the CGG-Ki mouse modelsamples, according to IPA..... 99

Supplementary Table 5. Genes commonly deregulated in blood samples of FXTAS patients vs. control individuals and in neuronal cells overexpressing the permutated *FMR1* 5'-UTR (79*CGG8x) vs. cells expressing the wild-type *FMR1* 5'-UTR (21*CGG) 104

Supplementary Table 6. Genes commonly deregulated in blood samples of AP vs. SP and in 79*CGG vs. 79*CGG8x expressing cells..... 105

Supplementary Table 7. Genes commonly deregulated in 21*CGG vs. 79*CGG8x expressing cells and 79*CGG vs. 79*CGG8x expressing cells 106

Supplementary Table 8. Genes commonly deregulated in 21*CGG vs. 79*CGG expressing cells and 21*CGG vs. 79*CGG8x expressing cells. 111

1. INTRODUCTION

1.1. Trinucleotide repeat disorders (TREDs)

In genetics, mutations are described as change in the sequence of DNA. Insertions, deletions, translocations and substitutions from a single nucleotide to a large fragment of DNA can lead to changes in the phenotype of a living organism. Most of the changes in the DNA are consequence of unrepaired DNA damage or errors in DNA replication. Mutations have an important role for normal and abnormal biological processes, such as evolution, cancer, and developmental disorders.

In 1981, sequence analysis upstream of the globin gene revealed a varying number of short repetitive sequence motifs (1). Those repetitive sequence and other of mono-, di-, tri-, tetra-, penta-, and hexanucleotides are named microsatellites – also refereed as short tandem repeats (STRs) or simple sequence repeats (SSRs) –. Since then an enormous amount of microsatellites have been identified scattered in the human genome. They are characterized as one of the most variable types of DNA sequences in the genome (2), and their variability relies in length rather than in the primary sequence. Although they present a polymorphic nature, microsatellites are only stable within a specific

length beyond which they become unstable. Such characteristic has named microsatellite expansions as dynamic mutation, in which presentation of a mutant phenotype depends on the number of expansions. Dynamic mutations are hereditability unstable upon transmission and present anticipation phenomenon in subsequent generations. Anticipation happens when the symptomatology of a disease manifests at early ages and with increase severity in subsequent generations. In the last 20 years, nearly 30 hereditary disorders had been identified to be consequence of an increase or expansion of the number of repeats in microsatellites (3). Specifically, trinucleotide repeats (TNRs) are causative of around 20 different neuromuscular and neurodegenerative disorders called trinucleotide repeat disorders (TREDs) (3). Most models for TNRs length variability agree that expansions occur through formation of a intermediate loop or hairpin, which is incorporated into the DNA (3). Large expansions seam to occur in non-dividing cells during repair of single strand breaks (SSBs) either by base excision repair (BER) or by nucleotide excision repair (NER) facilitating the formation of the loop (Figure 1) (4, 5). Short expansions are the result of simple polymerase slippage during DNA replication, which also facilitate the formation of a loop (4) (Figure 2).

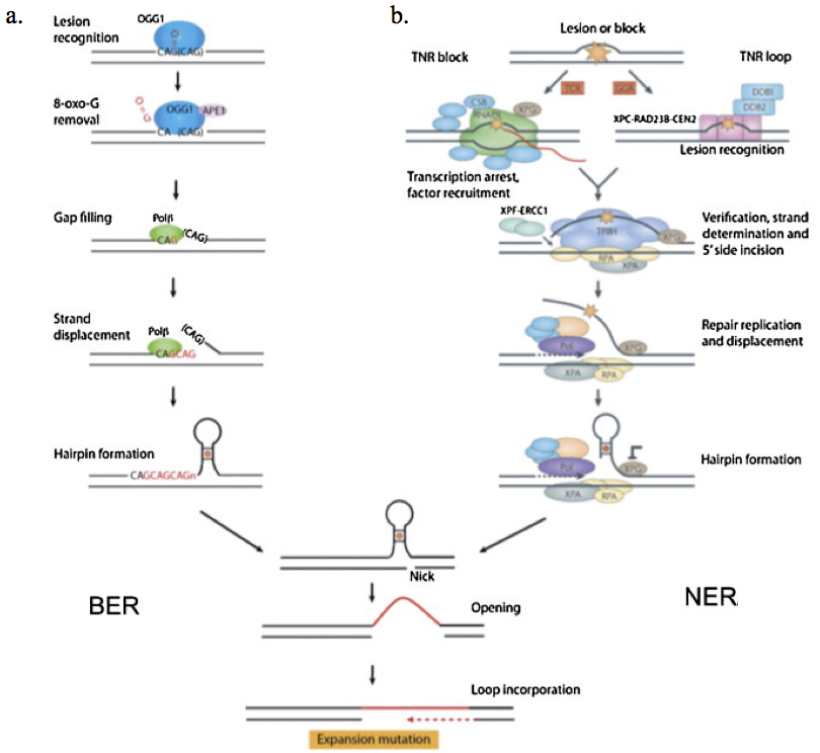


Figure 1. Base excision repair (BER) and nucleotide excision repair (NER) mechanisms in long expansions of trinucleotide repeat (TNR)

Expansion scheme in a BER. 7,8-dihydro-8-oxoguanine DNA glycosylase (OGG1) (blue oval) recognizes and removes oxidized guanines (O = G) (red) in the DNA template. After the oxidized guanidine has been removed an apurinic/apyrimidinic endonuclease 1 (APE1, also known as APEX1) (pink oval) produces a suitable 3'site for gap-filing by the DNA polymerase (Pol) (green oval). During that process the TNR expansion is displaced forming a loop/hairpin. Expansion occurs when the hairpin is ligated into the DNA. b | Expansion scheme in NER. There are two possibilities: transcription-coupled repair (TCR) (left) and global genome repair (GGR) (right). In both cases RNA polymerase II (RNAPII) transcription is blocked (orange star) by the presence of the CG-rich regions. Rescue of transcription progression depends on the recruitment of several proteins. Action of the XPF–ERCC1 complex (composed of XPF and excision cross complementing repair 1 (ERCC1)) or the endonuclease XPG, two nucleases, a nick is produce in the DNA. This action permits a displacement of the DNA and the generation of a TNR hairpin. As previously, the loop is inserted into the DNA leading to a TNR expansion. Taken from Budhworth & McMurray et al. 2013, (5).

In any case, the loop is incorporated to the DNA resulting in an expansion of the microsatellite repeats. The microsatellites with higher amount of related disorders are the TNR. It is believed that the loop structures generated by TNR during replication or DNA repair mechanisms are structures with a relatively high amount of base pair complementarities, thus stabilizing the loop and permitting the expansion (3).

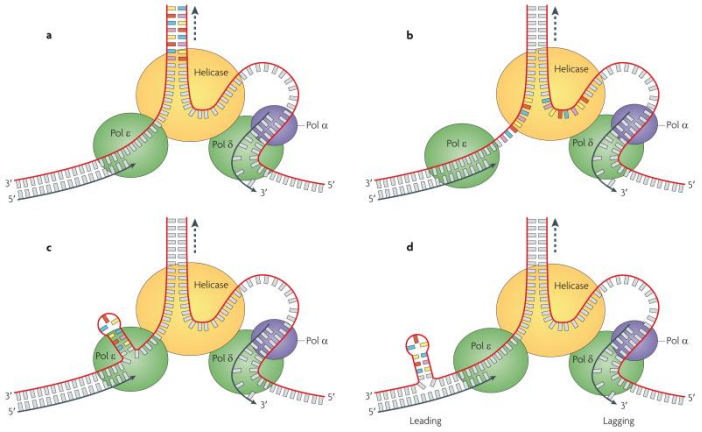


Figure 2. Slippage model for short expansions in trinucleotide repeats

Illustrations show the error in slippage that leads to TNR expansions. a | DNA polymerase on the leading strand (Pol ε) and helicase perfectly couple before the TNR. b | Unlike the helicase, Pol ε velocity is slowed due to the TNR presence. c | To avoid uncoupling of the two enzymes, Pol ε bypasses a segment of unreplicated TNR segment. d | After bypassing the TNR region, DNA segment will reassemble resulting I instability of the TNR region.

Each expansion disease is categorized upon three different criteria. One is the sequence composition of the expansions. Although highly heterogeneous, most of the repeat sequences are composed by CAG, CTG and GCG trinucleotide repeats. Another criterion is the location of the expansion within the gene, such as the promoter regions, 5' untranslated regions (UTRs), introns, exons and 3'-UTRs. Finally, it is also important to consider the mechanism of pathogenesis such as

polyglutamine gain of function, polyalanine gain of function, RNA gain of function or RNA loss of function. Polyglutamine gain of function mechanisms are the consequence of CAG expansions repeats coding for polyglutamine tracks in coding regions and are responsible for Huntington's Disease (HD) and a number of Spinocerebellar Ataxias (SCAs). Equally, GCN repeat expansions in the coding region of a gene are responsible for polyalanine gain of function disorders as in Oculopharyngeal Muscular Dystrophy (OPMD). In both mechanisms, the mutant protein becomes toxic and deleterious for the cell homeostasis. Trinucleotide expansions located outside the coding region, cause RNA gain of function and RNA loss of function. RNA gain of function is associated to Fragile X Tremor/Ataxia Syndrome (FXTAS) and Myotonic Dystrophy type 1 (DM1). Extensive repeat expansion outside the coding region leads to the loss of activity of the harbouring gene as occurs in Fragile X Syndrome (FXS or FRAXA) and Friedreich's Ataxia (FRDA). Table 1 shows a resume of the most common expansion disorders and their characteristics. Clinical consequences of repeat expansion related disorders range from congenital syndromes to late-onset neurodegenerative disorders.

Disease	Main clinical features	Causal repeat (gene)	Repeat location	Mechanism or category	Comments
DM1	Muscle weakness, myotonia, cardiac-endocrine-GI disease, MR	CTG (*DM1*)	3′ UTR	RNA GOF	A very common form of muscular dystrophy
DM2	Muscle weakness, myotonia, cardiac-endocrine-GI disease	CTG (*ZNF9P*)	Intron	RNA GOF	A striking phenocopy of DM1
DRPLA	Seizures, choreoathetosis, ataxia, cognitive decline	CAG (*ATN1*)	Coding region	Polyglutamine GOF	Very rare, most patients are in Japan
FRAXA	MR, facial dysmorphism, autism	CGG (*FMR1*)	5′ UTR	LOF	Most common inherited MR
FRAXE	MR, hyperactivity	GCC (FMR2)	5′ UTR	LOF	Needs to be ruled out in X-linked MR
FRDA	Ataxia, sensory loss, weakness, diabetes mellitus, cardiomyopathy	GAA (FXN)	Intron	LOF	Most common inherited ataxia in Caucasian ethnicity
FXTAS	Ataxia, intention tremor, parkinsonism	CGG (FMR1)	5′ UTR	RNA GOF	Premutation carriers only
HD	Chorea, dystonia, cognitive decline, psychiatric disease	CAG (HTT)	Coding region	Polyglutamine GOF	One of the most common inherited diseases in humans
HDL2	Chorea, dystonia, cognitive decline	CTG (JPH3)	3′ UTR	Not Determined	A striking phenocopy of HD
OPMD	Eyelid weakness, dysphagia, proximal limb weakness	GCG (PABPN1)	Coding region	Polyglutamine GOF	Modest expansion causes disease
SBMA	Proximal limb weakness, lower motor neuron disease	CAG (AR)	Coding region	Polyglutamine GOF	Phenotype includes LOF androgen insensitivity
SCA1	Ataxia, dysarthria, spasticity, ophthalmoplegia	CAG (ATXN1)	Coding region	Polyglutamine GOF	Accounts for 6% of all dominant ataxia
SCA2	Ataxia, slow eye movement, hyporeflexia, motor disease	CAG (ATXN2)	Coding region	Polyglutamine GOF	ATXN2 protein may not reside in the nucleus
SCA3	Ataxia, dystonia, lower motor neuron disease	CAG (ATXN3)	Coding region	Polyglutamine GOF	Most common dominant ataxia
SCA6	Ataxia, dysarthria, sensory loss, occasionally episodic	CAG (CACNA1A)	Coding region	Polyglutamine GOF	Causal gene encodes a subunit of a P/Q-type Ca2+ channel
SCA7	Ataxia, dysarthria, cone-rod dystrophy retinal disease	CAG (ATXN7)	Coding region	Polyglutamine GOF	Clinically distinct as patients have retinal disease
SCA8	Ataxia, dysarthria, nystagmus, spasticity	CTG/CAG (ATXN8)	Untranslated RNA	RNA GOF& polyglutamine GOF	Many cases of reduced penetrance
SCA10	Ataxia, dysarthria, seizures, dysphagia	ATTCT (ATXN10)	Intron	RNA GOF?	Huge repeats; only Mexican ancestry?
SCA12	Tremor, ataxia, spasticity, dementia	CAG (PPP2R2B)	5′ UTR	Unknown	Causal gene encodes a phosphatase
SCA17	Ataxia, dementia, chorea, seizures, dystonia	CAG (TBP)	Coding region	Polyglutamine GOF	Causal gene encodes a common transcription factor (TBP)

Table 1. Clinical and molecular characteristics of some neurological expansion repeat disorders

AR, androgen receptor; *ARX*, aristaless-related homeobox; *ATN1*, atrophin 1; *ATXN*, ataxin; *CACNA1A*, voltage-dependent P/Q-type calcium channel subunit α-1A; *CSTB*, cystatin B; DM, myotonic dystrophy; *DMPK*, DRPLA, dentatorubral-pallidoluysian atrophy; FRAXA, fragile X mental retardation syndrome; FRAXE, fragile X E mental retardation; FRDA, Friedreich's ataxia; *FXN*, frataxin; FXTAS, fragile X tremor ataxia syndrome; GI, gastrointestinal; GOF, gain of function; HD, Huntington's disease; HDL2, Huntington's disease-like 2; *HTT*, huntingtin; *JPH3*, junctophilin 3; LOF, loss of function; MR, mental retardation; OPMD, oculopharyngeal muscular dystrophy; *PABPN1*, poly(A)-binding protein, nuclear 1; *PPP2R2B*, protein phosphatase 2 regulatory subunit B, β isoform; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; *TBP*, TATA box-binding protein; *ZNF9*, zinc finger 9. Adapted from LaSpada A, 2010 (6)

1.2. The Fragile X Site

a. *FMR1* alleles

The first identified gene related to expansion disorders was *FMR1* (OMIN; 309550) (7). *FMR1* is a highly conserved gene located on the X chromosome, Xq27.3, that consists of 17 exons spanning ~38 kb (8). *FMR1* encodes the fragile X mental retardation protein (FMRP) (9). FMRP is an RNA binding protein that is highly expressed in neurons and glial cells (10, 11). FMRP has selective affinity for mRNA, including its own transcript and other targets essential for neuronal development and plasticity (12). FMRP is involved in the trafficking from the nucleus to the cytoplasm of the mRNA for which it has affinity (13). However, in neurons, the vast majority of FMRP is localized to the cytoplasm where it binds mRNA and polyribosomes and functions primarily as a regulator of translation (14). Mutations in the *FMR1* gene can lead to a wide variety of disorders. The majority of deleterious mutations in *FMR1* account for a CGG trinucleotide repeat expansion in the 5'-UTR of the gene. Depending on the size of the expansion, *FRM1* alleles can be grouped in 4 different types: normal allele, intermediate alleles or grey zone (GZ), premutation (PM) allele and full mutation (FM) allele. In the general population the trinucleotide repeat spreads between 6 and 44 CGG repeats (15), where most of the alleles have 29-30 CGG repeats and are usually stable upon transmission (16). GZ alleles account for an intermediate size between normal alleles and PM (from 45 to 54 repeats) and have shown to be slightly unstable upon transmission. Those alleles could potentially expand and lead to a PM in future generations (17). Alleles larger than

GZ but <200 repeats, called PM, exhibit somatic and germ line instability. PM alleles may also expand to a FM allele only upon maternal transmission to the next generation. Transition to FM depends on the size of the expansion; the smallest CGG expansion repeat known to expand to a FM is 59 CGG repeat (18, 19). Finally, FM alleles are those over 200 repeats (Figure 4).

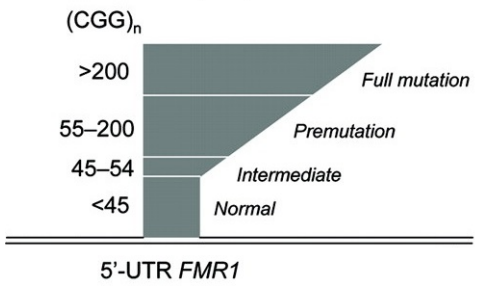


Figure 3. *FMR1* allele

Representation of the different *FMR1* alleles. Schematic visualization of the number of CGG repeat present in the different *FMR1* alleles. Taken form Filipovic-Sadic et al, 2010 (20).

b. *FMR1* instability

The classification and characterization of the different *FMR1* alleles is based on the association of the expansion size with different clinical features. A good genetic counseling strategy relays on the analysis of the *FMR1* allelic distribution in the world population and the risk factors for the alleles to undergo expansion. Several reports, that have used protocols by Fu et al. (21) to quantify CGG repeats, made possible some cross-population comparisons. In a recent review (22), comparisons of populations from different continents has been performed. The study shows that for each population, a large number

of allelic variants for the CGG repeat exist, being 29 or 30 CGG repeat the most abundant allele in most of the populations. Though, Mexican population has a significant number of long alleles (34-40 repeats). Specifically, in Mestizos and Tarahumaras populations the most common variant is of 32 repeats.

Concerning the risk for CGG expansion there are two important variables: position and number of AGG interruptions together with the CGG repeat size. Characteristically, *FMR1* normal alleles have two AGG interruption spaced every 9-10 CGG repeats: [(CGG)9AGG(CGG)9AGG(CGG)n]. Like CGG repeat length, AGG presents a level of instability, such as the loss of an interruption during transmission (23, 24). Recently, it has been also reported that such interruption along the CGG repeats reduced the risk of transition of PM below 100 repeats to a FM. More in detail, expansions between 70-80 repeats have a 60% decreased probability to expand to a FM when interrupted with AGG triplet (25). Moreover, evidences show that variations in the length of the CGG repeat are important to determine the degree of instability. While no direct transition from a normal allele to a FM has been detected, the risk of expansion from a PM to a FM is almost 100% in PM >99 CGG repeats (26). It has been reported that variations in the repeat length appear to be polar (23). Increments of CGG repeats happens distally from the most 3' AGG interruption (27). A recent study demonstrates that both AGG interruptions and the length of uninterrupted CGG repeat at the 3'-end are correlated with repeat instability on transmission. In addition, a greater risk of instability is linked to maternally alleles without AGG interruptions (28).

1.3. Clinical features of the fragile X families

In accordance with his high variety of alleles, *FMR1* gene is also related to a plethora of phenotypes.

a. Full mutation (FM) alleles

Full mutation alleles are characterized by the presence of >200 CGG repeats in the 5'-UTR of the *FMR1* (21). Through a mechanism that is still not well understood, these expansions lead to the hypermethylation of the CGG repeat and a 1 kb upstream CpG island, in the promoter region (29). Methylation of the DNA is a common mechanism for gene silencing that results in the lack of gene product. In addition to the altered methylation status, the *FMR1* in the FM alleles also presents other epigenetics changes. Remarkably, FM alleles present deacetilation of histones H3 and H4, reduced methylation of lysine 4 (K4) and increased methylation of lysine 9 (K9) in histone H3 (30). The combination of such epigenetic changes induces a heterochromatic configuration that prevents the binding of transcription factors and the machinery for transcription (31), thus turning off gene expression. Although most of the FM alleles are usually methylated, some unmethylated FM alleles have also been described with milder clinical features than methylated full mutation alleles (29).

The absence of functional FMRP is the cause for FXS, which has an incidence of 1/4000 males and 1/6000 females (24, 32). The lost of FMRP is usually caused by the trinucleotide repeat expansion in the 5'-UTR of the gene, but there are other deletions and missense mutations

liked to FXS (33, 34). FXS is the most common form of inherited metal retardation and the major genetic cause for autism spectrum disorders (21, 35). FXS has also other co-morbid features like depression, anxiety and other behavioral problems. Though FXS occurs in both genders, females are often less severely affected because of the protective effect of the second X chromosome (36). Individuals with mosaicism also develop milder symptomatology. There are two types of FM mosaicism: repeat size and methylation mosaicism. Repeat size mosaicism happens when PM alleles expand to FM in early embryonic states, leading to the coexistence of both PM and FM alleles in the same individual. Methylation mosaicism happens the FM allele is methylated in some cells but not in others. As a consequence, individuals with mosaicism can synthesize some FMRP, therefore having a variable phenotype between FM and PM (37).

Mutation Type	**nº of CGG Repeats**	**Methylation Status of *FMR1***	**Clinical Status**	
			Male	**Female**
Full mutation	>200	Methylated	100% MR	50% MR
Repeat Size Mosaicism	PM or FM in ≠ cells	Methylated+ Unmethylated	~100% ID	Variable
Methylation Mosaicism	>200	Mixture	~100% ID	Variable
Unmethylated FM	>200	Unmethylated	~100% ID	Variable

Table 2. Clinical signs associated to FM alleles

Abbreviations: ID, intellectual disability; MR, mental retardation. Adapted from Saul and Tarleton (38).

b. Premutation (PM) alleles

Premutation alleles are those with a CGG repeat length between 54 and 200 repeats. In contrast to FM allele, PM alleles are not methylated and the *FMR1* gene expression is increased (39). The clinical phenotypes associated with PM alleles are at least two: fragile X tremor/ataxia syndrome (FXTAS) (OMIM: #311360) and the fragile X premature ovarian insufficiency (FXPOI) (OMIM: #300623) (40). Both pathologies are restricted to PM alleles, suggesting a distinct mechanism of pathogenesis from FM alleles. Such observations coupled with an overexpression of *FMR1* mRNA in PM carriers, lead to the postulation of a RNA toxic gain of function in PM carriers (41, 42).
FXTAS is a neurodegenerative disorder that develops mainly in men over 50 years of age (40). It is characterized by a progressive action tremor, gait ataxia and other frequent variable features of cognitive decline; especially executive dysfunction, parkinsonism, neuropathy, and autonomic dysfunction (43). FXTAS patients may also present psychiatric symptoms, such as anxiety, mood liability and depression (44, 45). FXTAS prevalence is 1 in ~400 males and ~250 females. Penetrance in females is around 8-16% and in males near 40%, but it increases with age to a 60-70% in males of around 80 years (24, 46). These data suggest that FXTAS is one of the most common monogenic forms of gait ataxia and tremor in older males.

Another usual clinical manifestation in female fragile X premutation carrier (fXPCs) is premature ovarian failure (POF1). POF1 is characterized by the cessation of menstruation prior to the age of 40 (47). fXPCs has a 20-fold increase incidence of POF1, now known as

FXPOI (48). The prevalence of FXPOI is estimated to be between 20% and 28% of the female fXPCs (49).

Moreover, recent studies in premutation alleles carriers provide new evidences that PM alleles can cause other clinical involvement beyond FXTAS and FXPOI. Children fXPCs frequently present neurodevelopmental phenotypes like attention deficit and hyperactivity disorders and autism spectrum disorders (50, 51). Also seizures are a significant co-morbid clinical aspect of the PM (51). Psychiatric symptoms like depression and anxiety are also co-morbid aspects of the premutation in adults (52). Nonetheless, psychiatrics symptoms are more intense in individuals presenting FXTAS (45).

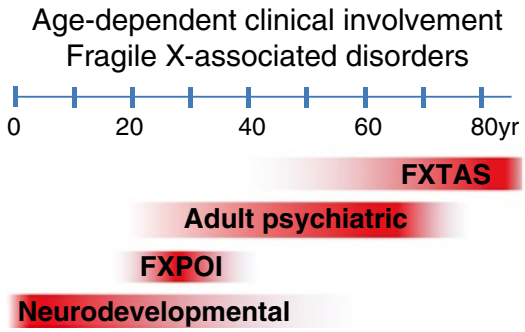


Figure 4. Clinical manifestations in *FMR1* premutation carriers

Schematic visualization from clinical signs associated with *FMR1* PM alleles. Correlation between age of onset and clinical signs. Taken from Hagerman, 2013 (53)

1.4. Fragile X tremor/ataxia syndrome (FXTAS)

a. Clinical features

* FXTAS symptomatology

Observations in Fragile X families indicate a high occurrence of a neurodegenerative phenotype in grandfathers of children who suffered FXS, now named FXTAS (41, 54). 30% of FXTAS patients started presenting the symptomatology at the age of 50, whereas 70% start at the age of 75 (46, 55). The most common clinical feature of FXTAS patients is tremor. At least 50% present mild tremor and another 17% moderate tremor, which resembles in appearance to essential tremor. Parkinsonism resting tremor is uncommon in FXTAS patients, but 57% have mild bradykinesia and 71% present rigidity (55). FXTAS patients commonly also present ataxia (balance problems) and autonomic dysfunction, such as erectile dysfunction (80%), hypertension and loss of bowel and bladder function (30–55%) (54). Other symptoms like peripheral neuropathy (60%), lower limb proximal muscle weakness, hearing loss, dysphagia, sleep apnea, hypertension and immune-mediated disorders are also present in FXTAS patients (41, 54–61).

Cognitive decline is also current in FXTAS patients with short-term memory loss and difficulty in performing other cognitive tasks like taking decisions (43). FXTAS patients exhibit executive dysfunction, which prevents them from organizing, planning, anticipating, and carrying out everyday tasks (62). Furthermore, neuropsychological problems have also been diagnosed in FXTAS patients. Patients present

anxiety, depression, apathy, reclusive behavior or social phobias and mood changes which might include increased irritability, outbursts of anger, and inappropriate or impulsive behavior (45, 63, 64).

Ultimately, significant dementia may also occur in FXTAS patients (54, 65). It has been suggested that dementia in FXTAS patients is defined as frontal-subcortical dementia, which develops after movement disorders (45) .

* Radiological and neuroanatomical features of FXTAS

A diagnostic criterion for FXTAS patients is based on clinical features as well as in radiological findings. Magnetic resonance imaging (MRI) examination in FXTAS brains revealed an abnormal brain pattern. The most important radiological signs are white matter alterations manifested as increased T2 signal intensity in the middle cerebellar peduncle (MCP) (41, 66), which are present in 60% of the FXTAS patients (67). White matter disease is not restricted to MCP but present in the overall of the brain. T2-weighted MRI and diffusion tensor imaging techniques have identified white matter disease in the frontal and parietal cortex and in the superior cerebellar peduncle (SCP), MCP, the fornix, and stria terminalis, respectively (57, 68). Images also reveal generalized brain atrophy and volume loss in brainstem, cerebellum, cerebral cortex, amygdalo-hippocampal complex and thalamus (69, 70).

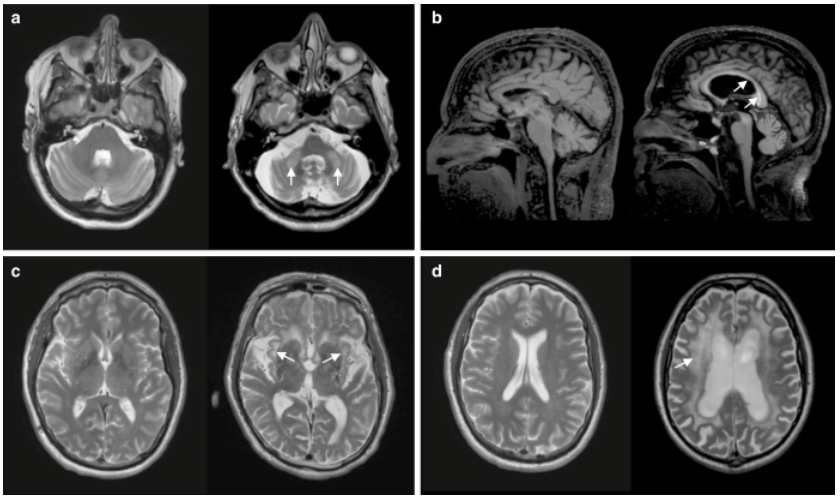


Figure 5. Radiological images form FXTAS brains

MRI images from FXTAS brains with white matter disease and structural abnormalities. Each pair of image represents a control (left) and a FXTAS (right) case. a | Arrows indicate increased signal intensity on T2 turbo spin-echo sequences in the middle cerebellar peduncle. b | Thinning and increased signal of the trunk and splenium of the corpus callosum.c | subinsular white matter. d | Cerebral white matter disease. Taken form Hagerman, 2012 (71).

b. Diagnostic criteria

Ten years ago, Jacquemont et al. (41) proposed a diagnostic criteria based on the study of 26 male fXPCs. The study compiled the clinical and radiological data of the patients and classifies them into three different groups: definite, probable and possible FXTAS. The classification is based on three different criteria

Molecular *FMR1* CGG Repeat Size 55–200

Clinical

Major signs	Intention tremor
	Gait ataxia
Minor signs	Parkinsonism

	Short term memory deficits Executive function deficits

Radiological

Major signs	MRI white matter lesions in the middle cerebellar peduncle (MCP sign)
Minor signs	Moderate to severe generalized atrophy MRI white matter lesions in cerebral white matter

Clinical and radiological criteria are divided in to two main groups, major and minor signs. The division is based on statistical measurements of clinical observation, being the major signs those more commonly presented by FXTAS patients. Upon these criteria the FXTAS diagnosis categories are:

Diagnostic Categories

Definite	Presence of one major radiological sign plus one major clinical symptom
Probable	Presence of either one major radiological sign plus one minor clinical symptom or has two major clinical symptoms
Possible	Presence of one minor radiological sign plus one major clinical symptom

c. Neuropathology of FXTAS

Autopsies carried out in brains of symptomatic fXPCs correlate with MRI findings and show moderate to severe cortical, brainstem, and pontine atrophy; ventriculomegaly and loss of deep white matter.

Cerebrum, cerebellum and brain steam present significant white matter disease (65, 72). White matter histological changes are distinct from those seen in other pathologies or vascular lesions. Precisely, T2 hyperintensities indicated by MRI correlate with spongiosis present in subcortical and deep white matter, MCP and deep cerebellar white matter (41, 66). Other histological changes associated with the pathology in gray and white matter are the presence of reactive astrocytes, axonal loss and glial cell lost, axonal swelling, cerebellar Purkinje cell lost and Bergmann gliosis (Figure 6) (65, 72).

Postmortem criteria for a definite FXTAS is the presence of eosinophilic, ubiquitin-positive intranuclar inclusions in neurons and astrocytes (Figure 6) (65, 72). FXTAS inclusions are negative for α-synuclein and tau, characteristics for Parkinson's disease and Huntington's disease (HD), respectively. Inclusions are abundant in cerebral cortex, amygdala, hippocampus, brainstem nuclei, cerebellum and choroid plexus (65, 72, 73). Interestingly, a strong correlation has been found between premutation length and presence of inclusions (72). In a recent study, intranuclear inclusion where also found throughout the peripheral nervous system and systemic organs (74). Among other, inclusions where found in the pancreas, thyroid, adrenal gland, pituitary gland, pineal gland, heart, mitral valve, testes, epididymis, and kidneys. In resume, the three major neuropathological features of FXTAS are white matter disease, astrocyte pathology and the presence of intranuclear inclusions.

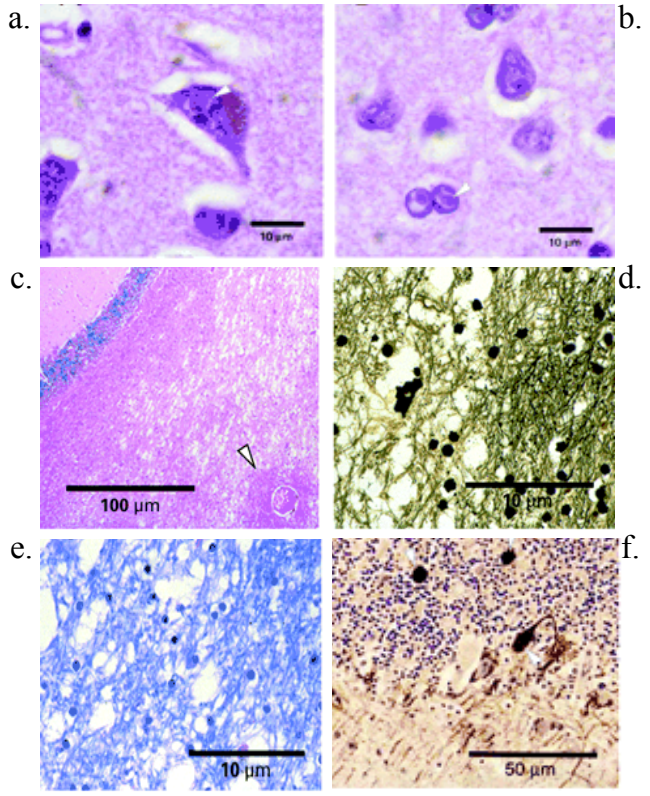


Figure 6. Histopathological changes in FXTAS brains

Intranuclear inclusions in a | cortical neurons. b | astroglia. c | Spongiform changes in the white matter of the cerebellum. d | Cerebellar white matter with axonal loss. e | Cerebellar white matter with myelin loss. f | Purkinje cell dropout and the presence of axonal swelling in the cerebellum Adapted form Greco et al., 2002 (65).

d. Molecular signature of FXTAS

Differences between normal alleles, FM alleles and PM alleles are also detectable at a molecular level. The most characteristic trait for PM alleles is the increased expression of the *FMR1* mRNA. Expression of the transcript is between 2-8 times higher than in normal alleles, despite normal or slightly reduced FMRP protein levels (75, 76). The

mechanism by which overexpression occurs, are still uncertain. A hypothesis is that high number of CGG repeats at the 5'-UTR of the *FMR1* gene induced chromatin opening permitting the accessibility to transcription machinery. Chromatin changes induced by CGG repeat expansions induce the transcription of the *FMR1* gene and of its antisense *FMR1* transcript (*ASFMR1*) (77). Interestingly, *ASFMR1* has different transcription start sites; all of them before the CGG repeat region (-208 to -99) and another before it (+10243). In PM alleles, A*SFMR1* is predominantly transcribed from the transcription start site that expands the premutation (77).

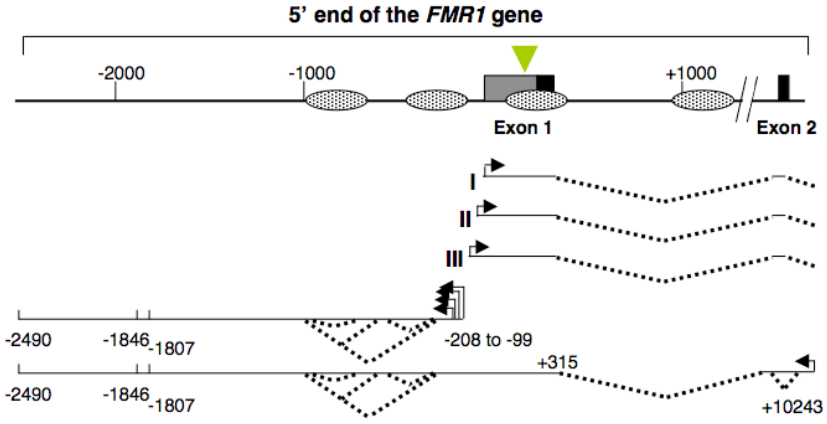


Figure 7 *FMR1* transcriptional start sites.

Schematic representation of the *FMR1* genomic region. Each arrow represents the initiation position and direction of a transcript

FXTAS has been postulated as a toxic RNA gain of function. There are several findings that support this argument. The first is that FXTAS clinical symptomatology is not present in expanded alleles that are transcriptionally inactive. Thus, the expression of the RNA molecule is necessary for the development of the symptomatology. As described above, *FMR1* mRNA expression is higher than in normal alleles. Moreover, expanded CGG repeats, as RNA, results in reduced cell

viability in primary neural progenitor and established neural cell lines (78). In accordance with these results, it has also been reported that expression of a 90 CGG repeats mRNA is sufficient to cause degeneration in a flies (42). Being this phenotype neuron specific and CGG repeat dosage sensitive (79). Finally, *FMR1* mRNA is present in the intranuclear inclusions characteristic of FXTAS patients (80).

Further investigations for the molecular pathogenesis of FXTAS were performed to determine the composition of the intranuclear inclusions. Several laboratories have identified more than 30 different proteins by mass spectroscopy and immnunohistochemistry assays (IHC) (65, 72, 80–84). From those we can remark: αB-crystallin, lamine A/C, α-tubulin, Glial fibrillary acidic protein (GFAP), Phosphorilated histone 2A member X (γH2AX), Src-associated in mitosis 68 kD (SMA68), DiGeorge syndrome critical region gene 8 (DGCR8), Double-Stranded RNA-Specific Endoribonuclease (DROSHA), muscleblind-like splicing regulator 1 (MBNL1), Heterogeneous Nuclear Ribonucleoprotein A2/B1 (hnRNP A2/B1), heat shock protein 70 (HSP70) and HSP27.

e. Models for RNA gain of function toxicity in FXTAS

Several mechanisms have been proposed to be the consequences of the clinical features associated with the PM.

* Titration of RNA binding proteins by CGG repeats

The most established mechanism proposed is protein sequestration model by the RNA molecules with CGG expansions. The model

proposes that *FMR1* mRNA is able to sequestrate RNA binding proteins (RNA-BP). Specifically, that some RNA-BP have a high affinity, and therefore bind, to RNA CGG expansions. The association of the RNA with the proteins led to the formation of the intranuclear inclusions present in FXTAS patients. But more importantly, the consequence of the sequestration is a partial or total loss of function of the sequestrated proteins. Some of the proteins present in the inclusions are important factors for splicing, mRNA trafficking, transcription and stress response (65, 72, 80–84), which can explain the clinical features of FXTAS. The latest example for this mechanism was published early this year by Sellier et al, (83). They demonstrate that CGG repeats can bind DGCR8, an enzyme in the biogenesis of microRNA (miRNA). The result is a decreased production of miRNA molecules, key modulators of gene expression, which results in lower neuronal viability and lower dendritic complexity.

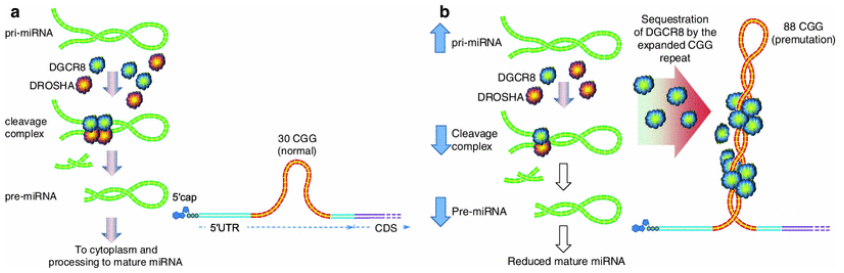


Figure 8. Representation of DGCR8 role in FXTAS pathogenesis

a | Normal conditions. DGCR8 binds to a specific helix region of the pri-miRNA and recruits DROSHA. The complex is called microprocessor, which cleaves pri-miRNA to pre-miRNA (miRNA precursors). b | PM conditions. Expanded CGG repeat forma a helix structure that recruits DGCR8. The microprocessor is not assembled and the levels of pri-miRNA in the nucleus increase while the levels of mature miRNA decrease in the cytoplasm. Taken form Hagerman, 2012 (71).

* RNA-mediated protein aggregation

Another proposed toxic mechanism is the prionic-like mechanism. The CGG expanded RNA would start a conformation changes is proteins with prion-like domains, developing a cascade of aggregations (85, 86). In agreement with this argument, a protein found within the inclusions, hnRNP A2/B1, is among the highly "prionogenic" proteins (85), likely contributing FXTAS pathogenesis.

* Antisense-RNA effects on expression

As previously described, *FMR1* loci also codifies for an antisense transcript. The transcript is transcribed by RNA polymerase II, is spliced and polyadenylated (77, 87). It's expression is coupled with the FMR1 expression, being silenced in FM alleles and overexpressed in PM alleles (77, 87). A study points that some of the spliced variants of the *ASFMR1* are specific to PM alleles (77). In agreement, some splicing factors and modulators like SAM68 and MBNL1 are present within the FXTAS inclusions. These results suggest that some of the spliced forms of the *ASFMR1* could present an associated pathogenic mechanism. In contrast, another study shows that knockdown of *ASFMR1* in cultured cell lines altered cell cycle progression and induced cell death, whereas up-regulation of *ASFMR1* resulted in enhanced cell proliferation (87). While, *ASFMR1* seems to have an important role in cell survival and proliferation, more studies are need to understand its relevance in FXTAS.

* RAN-mediated expression of alternative peptides

Translation initiation independent form the AUG codon has been described to happen in transcripts with hairpin-forming repeats (88, 89). This repeat-associated non-AUG-initiated (RAN) translation occurs when the ribosome assembles and stalls due to the presence of a hairpin structure, thus prompting transcription activation. RAN translation has been reported in several TREDs involving CAG repeats and CGG repeats (88, 89). This mechanism has been proposed to be a related toxic mechanism in FXTAS (89). Translation initiation upstream form the AUG codon and the CGG expansions produce short peptides rich in polyglycine (ploy**Gly**) and polyalanine (polyAla), GGC and GCG codify for glycine and Alanine, from two different reading frames respectively (89). In agreement, this kind of peptides have been found to accumulate in FXTAS inclusions (89). Polyarginine (PolyArg) peptides, arising from the third reading frame, have not been detected. Nonetheless, further studies are needed to evaluate the function and the consequences of such short peptides.

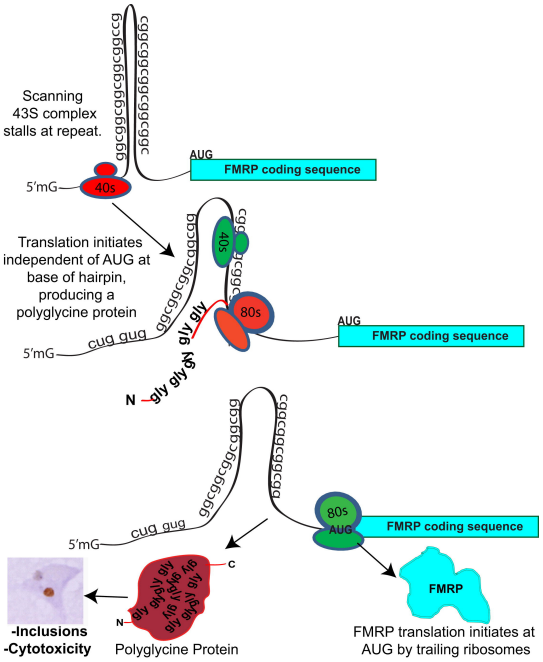


Figure 9. Model for CGG RAN Translation in FXTAS

Schematic representation of RNA translation a | Ribosomes assemble on the 5'-UTR *FMR1* message and search for the AUG codon. The expanded CGG repeat is folded into a hairpin structure that stalls the 43S (RED) triggering RAN translation initiation close to the AUG codon. b | CGG is translated as a glycine. The result is a long polyglycine that accumulates in the nucleus. Trailing ribosomes (green) may not stall at the hairpin, thus initiating translation normally at the AUG for FMRP.

* siRNA formation form TNR structures

TNR structures based on CNG repeats form A-helices stabilized by C-G and G-C pairs acting as strong support (90, 91). Transcripts containing those long hairpin structures composed of CNG repeats can be cleaved by the ribonuclease III DICER (92). Indeed, a study shows that mutant transcripts involved in TREDs, such as myotonic dystrophy type 1 (DM1), Huntington's disease (HD) and spinocerebellar ataxia 1 (SCA1) are DICER substrates in fibroblasts of patients with these

diseases. DICER activity results in the formation of a double-stranded short-repeated CNG molecules similar to siRNAs/miRNAs that use the RNA interference pathway to trigger downstream silencing effects (92). The detrimental effect of small repeated CAG (sCAG) biogenesis was recently demonstrated by our group, in HD (93). This recent study demonstrates that short repeated CAG RNAs of around 21 nt in length (sCAG) are produced in a DICER dependent manner from the mutant transcript involved in HD. sACG detrimental mechanism of action involves the RISC silencing machinery.

Structural characteristics from the mutant transcripts involved in DM1, HD and SCA1 are also present in FXTAS PM transcript. The formation of hairpin structures in the case of long CGG repeat structures has been supported by several studies (90, 91). Moreover, CGG repeat structures are also a substrate of dicer (92, 94). But yet, no studies have further investigated whether CGG molecules are involved in the pathologies associated with fXPCs.

f. Cellular pathology in FXTAS

Although the toxic mechanism for FXTAS RNA gain of function are still unclear, some advances have been done in neuronal cells. RNA molecules with a CGG repeat within the permutation range can induce intranuclear inclusions and disrupt the structure of *lamin A/C* in the nucleus (78). Also, reduced cell viability, development of dendritic arbors, and synaptic function have been observed in FXTAS tissue and FXTAS mouse model, respectively (95, 96). These observations agree with the mitochondrial dysfunction observed in fibroblast of FXTAS patients, in hippocampal cultured cells from the FXTAS mouse model

and in post-mortem central nervous system tissue from FXTAS patients (97–99). There is also a lower import of nuclear encoded mitochondrial proteins due to a reduced loading of zinc into metalloproteases that process proteins for import (97). As a result, some of the nuclear encoded proteins from the electronic chain transport are deficient in the mitochondria, thus leading to a lower oxidative phosphorylation capacity and a increase in oxidative stress response observed in FXTAS brain tissue (97, 99). Finally, induction of mTOR, involved in regulation of organismal growth and homeostasis (100), can ameliorate neurodegeneration in the FXTAS fly model. Alterations in the mTOR pathway are related to axon regeneration, dendrite arboritzation and spine morphology (101).

g. Modeling of FXTAS

Genetic models for human disease have greatly contributed to the understanding of the molecular causes underlying a disorder. For FXTAS the introduction of expanded CGG repeats alone or within the *FMR1* gene into the models organisms has provided a great inside in in the pathogenesis of FXTAS. Two different models have been used to study FXTAS, an invertebrate model (fly) and a mammalian vertebrate model (mouse)

* Fly model

The fly model for FXTAS was generated inserting a fragment of the human FMR1 DNA containing 60 or 90 CGG repeats followed by the green fluorescent protein (*GFP*) reporter gene (42). Different tissue

specific promoter drove the expression of the vector in the eyes, in all neurons of the peripheral and central nervous system, in the embryo and in the epithelial cells. Strong expression of 90 CGG repeats in the embryo and in the cells of the central and peripheral nervous system is lethal for the fly. Expression in the eyes induced a rough eye phenotype, loss of pigmentation and severe cell death. In contrast, no effect was inflicted when the expression was directed to epithelial cells (42). The same exploration was performed to flies expressing moderate levels of the 90CGG vector and high to moderate levels of the 60CGG vector. The results showed that moderate expression of the 90CGG vector has similar consequences as strong expression of the 60CGG vector, and there are both milder to the ones performed by strong expression of 90CGG vector. In addition, moderate expression of 60CGG vector did not produce any phenotypic change. These results manifest the importance of the CGG repeat length and its levels of expression in FXTAS phenotype.

GAL4 line	Expression Pattern	(CGG)60-EGFP Moderate*	(CGG)60-EGFP Strong*	(CGG)90-EGFP Moderate*	(CGG)90-EGFP Strong*
gmr-GAL4	All eye cells posterior to the furrow including photoreceptor neurons and pigment cells	No effect	Rough eye, loss of pigmentation, and holes in the tangential sections	Mild rough eye and holes in the tangential sections	Rough eye, loss of pigmentation, and severe cell death
elav-GAL4	All neurons of the peripheral and central nervous system	No effect	Reduced viability	Reduced viability	Lethal
Act5C-GAL	Ubiquitous expression in embryo	No effect	Late larval lethal	Male lethality and reduced viability in females	Late larval lethal
dpp-GAL4	Along the anterior-posterior boundary of imaginal discs, epithelial cells	No effect	No effect	No effect	No effect

Table 3. Effects of directed CGG repeat expression in different *Drosophila M.* tissues

*Referred to the levels of vector expression. Adaptation form Jin et al, (42).

Interestingly, flies expressing high levels of the 90CGG vector had a worsening of their phenotypic characteristics with age, agreeing with the late onset neurodegenerative disorder in humans. The fly model also presents inclusions positive for ubiquitin and *HSP70*. Unlike human inclusions that are only present in the nucleus, the fly model present inclusions in the nucleus and the cytoplasm (42).

The presence of inclusions in this model permitted the investigation of the RNA-BP sequestration model for the RNA toxic gain of function. A genetic screening identified CUG binding protein (*CUGBP1*) as a modifier of the rough eye phenotype in 90CGG expressing flies (102). The screen revealed that overexpression of CUGBP1 suppresses the rough eye phenotype (102). The binding it is mediated by *hnRNP A2/B1*, which its presence has been detected in human nuclear inclusions (81, 102). In agreement, overexpression of *hnRNP A2/B1* and its two fly homologues can also suppress the eye phenotype in flies (102). All the studies in the fly model have been crucial to support the RNA-BP sequestration model for the RNA toxic gain of function in FXTAS.

* Mouse model

A mouse model for FXTAS was generated by Bontekoe et al. (103) when generating a model to study the repeat instability in humans. The model is a knock-in (KI) mouse model for the *FMR1* gene, which the 8 CGG repeat expansion of the mouse *Fmr1* gene has been replaced by 98 CGG repeats from human origin. After the discovery that PM alleles are associated with late onset neurodegenerative disorders, this model

has undergone exhaustive studies of the molecular aspects, the cellular pathology and the clinical signs displayed for FXTAS patients.

The KI mouse model exhibits a big proportion of the pathophysiology and the behavioral deficits encountered in FXTAS patients. Both, human FXTAS patients and KI mouse model, display elevated levels *Fmr1* gene and slight decrease of FMRP expression in brain is found in the KI mouse model (39, 103, 104). In addition, they both also show ubiquitin-positive intranuclear inclusions in neurons and astrocytes, which are also negative for α-synuclein and tau (72, 105, 106). So far, three proteins have been identified within the mouse inclusions; Hsp40, 20S proteasome and the DNA repair ubiquitin-associated (HR23B) (106, 107). In parallel with human inclusions, mouse inclusions are also fund through out the brain, including the cerebral cortex, olfactory nucleus, parafascicular thalamic nucleus, medial mammillary nucleus and colliculus inferior, cerebellum, amygdala, pontine nucleus and dentate gyrus at 72 weeks of age (104, 106). In agreement with the late onset of FXTAS neurodegenerative disorder, the presence of inclusions in the KI mouse increases in number and size during the course life (106). Importantly, brain regions with intranuclear inclusions correlate with clinical features in FXTAS patients (106). However, some of the histopathological changes present in humans are not recapitulated by the KI mouse including, neuronal cell loss, gliosis and Purkinje cell death (106).

Motor deficits such as tremor and ataxia are a major clinical signature of FXTAS patients. Studies in the KI mouse model have also demonstrated motor deficits on the rota-rod test at 70 weeks (108). The mouse model also showed deficits in the Morris water maze test, indicating deficits in spatial learning and memory (108). The KI mouse

also shows deficiencies in cognitive function, which exacerbate with aging (109). In concrete, the mice show insufficiencies in spatial pattern separation and spatial object recognition.

Finally, the KI mouse model also displays some of the psychological changes that FXTAS patients suffer. The mice show increased anxiety in open field behavior tests and high levels of corticosterone levels after the exposure to a mild stressor (108, 110). These behaviors suggest that an abnormal functioning of the hypothalamic-pituitary-adrenal axis (HPA) may contribute to the pathophysiology of FXTAS (110). In agreement, intranuclear inclusions have been found in the pituitary, the adrenal gland and the amygdala of the KI mouse model (110).

Core pathology	**Human FXTAS**	**CGG KI mouse**
CGG repeat expansion length on FMR1	55–200 CGG repeat length, repeat instability	70–300 CGG repeat length, modest repeat instability
Elevated *FMR1* mRNA expression	2–8-fold increase	1.5–3.0-fold increase
FMRP levels	Reduced in several brain regions	Reduced in several brain regions
Motor impairment	Tremor/ataxia, postural sway, parkinsonism	Impaired on Rotorod, ladder rung task
Cognitive impairment	Poor working memory, anxiety, depression, social phobia	Spatial memory deficits, anxiety in elevated plus maze
Intranuclear inclusions	Neurons and astrocytes, highly correlated with CGG repeat length, frequency increases with aging	Neurons and astrocytes, related to length of CGG repeat, frequency increases with aging

Table 4. Comparison of FXTAS with the CGG KI mouse model.

Adapted from Berman et al, 2012 (96).

Another two FXTAS mouse model have been also developed since the discovery of FXTAS. The first was generated by Entezam et al. (2007)

where the mouse Fmr1 CGG repeat was expanded by serial ligations of CGG·CCG-repeat tracts until 130 CGG·CCG (111). The mouse shows repeat instability, increase levels of *Fmr1*, decreased levels of FMRP, Purkinje cell dropout, hyperactivity and decreased dendritic arborization (111, 112). Yet, the mouse has no motor deficiencies one of the most important traits of FXTAS patients.

The other mouse model was generated by Hashem et al. (2009), in which they express the 90 CGG repeats from human origin followed by Fmr1 or GFP under the control of a Purkinje promoter (113). Their results show that neurodegeneration is dependent on the expression of a 90 CGG repeat. The mouse present ubiquitin-positive intranuclear, axonal swellings and progressive age-dependent decline in neuromotor learning ability.

h. Treatment

Patients diagnosed with FXTAS can only be treated to alleviate the symptoms associated to FXTAS. Mainly, FXTAS patients are treated to ameliorate tremor, ataxia, mood changes, anxiety, cognitive decline, dementia, neuropathic pain and fibromyalgia.

Research in FXTAS is mostly focused in a better understanding of the pathology to develop efficient drugs to slow-down the progression of the disorder and cure it. Studies in flies have shown that histone acetyltransferase inhibitors can change the chromatin structure of the *FMR1* loci reducing the expression levels of the gene to normal levels. The consequence is the reversion of the rough eye phenotype, thus suppressing the toxic RNA gain of function (114). The results indicate

that drugs that modulate the activity of histone acetylases could be a good therapeutic strategy.

Another chemical screen performed in the fly model has also found a molecule that recues the CGG RNA repeat toxicity by inhibiting the phospholipases A2 (*Pla2*) activity (115). *PLA2* has also previously been liked to other neurodegenerative disorders and brain trauma (116). Nonetheless, more efforts are needed to study the application of these results in treating patients and in the discovery of alternative therapeutic targets for the disorder.

1.5. Fragile X associated premature ovarian insufficiency (FXPOI)

a. Terminology

In 1998, a study conducted over 147 females presenting idiopathic POF resulted in the association females fXPC with premature ovarian failure (POF1 or POF). POF1 is defined as onset of menopause before the age of 40 years (117). One year later, an international collaborative study of 760 women from fragile X families demonstrated that 16% of the females fXPCs presented menopause before the age of 40, none of the full mutation presented POF and 0.7% of the controls presented menopause before the age of 40 (118). Now we know that one of the most common known causes of POF is the PM allele of the *FMR1* gene. In particular, POF cases related to fXPCs are called FXPOI.

POF is a disorder that causes impaired fertility and hormonal deficiencies in females (119). POF is a heterogeneous disorder with different levels of severity. For that reason, it was proposed to use primary ovarian insufficiency (POI) as a more accurate terminology to describe the spectrum of decreased ovarian function by using specific modifiers (120). Based on this last terminology, *occult* POI describes normal menstrual cycles and normal follicle stimulating hormone (FSH) but reduced fertility. *Biochemical* POI occurs in females with normal menstrual cycles but elevated FSH levels and reduced fertility. And *over* POI or POF happens when menstrual cycles are abnormal or absent, levels of FSH are elevated and fertility is severely reduced (119). Nonetheless, the lack of specification for ovarian dysfunction age of onset is an important limitation for the use of POI terminology. In this chapter POF will still be used in reference to *overt* POI before the age of 40.

b. Clinical features

The prevalence of POF is between a 15% and 24% within fXPCs. Yet, with the most severe form of the disorder, when menstrual cycles ceases, ovarian follicles are present and can function spontaneously and unpredictably (119). Moreover, the reduced ovarian function characteristic of the patients causes estrogen deficiencies manifested as hot flashes, vaginal dryness, insomnia and decreased bone mineral density (121, 122).

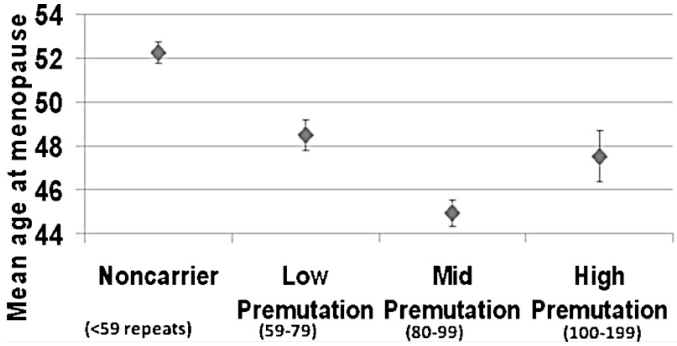


Figure 10. Age of menopause among FMR1 alleles

The graph shows the relation between the number of CGG repeats at the *FMR1* 5'-UTR and age of menopause onset. Taken from Sullivan et al. (123)

On average, the length of reproductive years in female fXPCs is shortened by 5 (124, 125). There are several signs for premature ovarian ageing, such as short menstrual cycles and elevated levels of FSH in fXPCs (120). Increased levels of FSH suggest decreased follicle number. In agreement, anti-mullerian hormone (AMH) levels, which indicate the number of growing small follicles, are reduced in fXPCs (126). As foreseen, fXPCs show reduced fertility when compared to control females.

c. Modeling FXPOI

The availability of mouse models with PM alleles used to study FXTAS has been also used recently to study FXPOI. The mouse model generated by Entezam et al. (2007), previously described, displays fast lost of all follicles classes, suggesting that the problem is intrinsic to the ovary. The size of the follicles is small with less granulosa cells (GCs) (127).

Another mouse model generated in 2010 by Peier et al. also show some ovarian abnormalities. The model was generated using the latest techniques in homologous recombination of yeast artificial chromosome (YAC). A PM allele was used in for homologous recombination with a YAC that encoded for a normal allele of the *FMR1* gene. Female of the mouse model show reduction in the number of growing follicles, which is sufficient to impair female fertility (128). They also show alteration FSH levels, like the alteration seen in human female carriers.

1.6. Non-coding RNA

During the last years, the ENCODE (ENCyclopedia of DNA Elements) consortium has shown that at least 90 % of the genome is transcribed in the different cells and that most of the transcribed gens does not code for proteins. These non-coding RNA (ncRNA) are structurally very diverse and a significant part is functional. Such ncRNA characteristics are the bases of the so-called pervasive transcription, in which transcriptome and proteome are exposed to a complex layer of regulation and modification (129, 130).

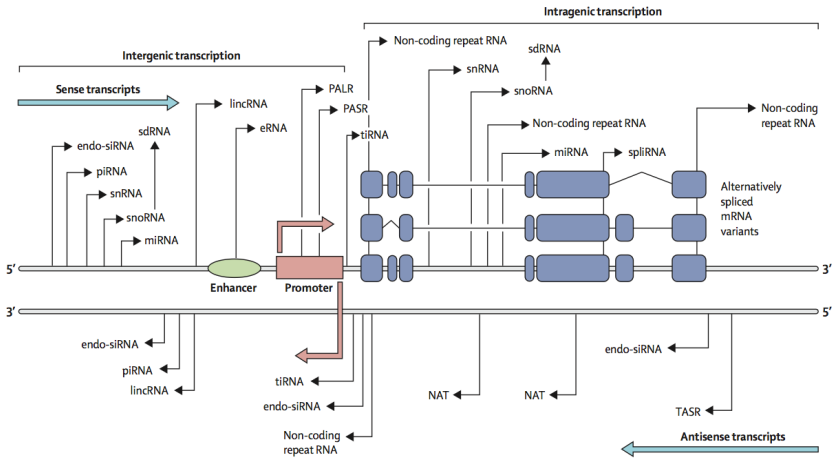


Figure 11. Transcriptional patterns of ncRNAs

Schematic representation of the transcriptional complexity of ncRNA. The figure represents the two strands of DNA with basic building blocks of a gene. Black arrows indicate the transcription site location for a given ncRNA. ncRNA; non-coding RNA. endo-siRNA; endogenous small interfering RNA. piRNA; piwi-interacting RNA. snRNA; small nuclear RNA. snoRNA; small nucleolar RNA. miRNA; microRNA. sdRNA; sno-derived RNA. lincRNA; long intergenic non-coding RNA. eRNA; enhancer RNA. PALR; promoter-associated long RNA. PASR; promoter-associated short RNA. tiRNA; transcription initiation RNA. spliRNA; splice junction-associated RNA. mRNA; messenger RNA. NAT; natural antisense transcript. TASR; termini-associated short RNA. Taken from Salta et al. 2012 (131).

NcRNAs are expressed in all cell types, and especially in the central nervous system. Their importance relay on many aspects of the normal function and development of the CNS, such as neurogenesis, neuronal patterning, neurotransmission and synaptic plasticity (132–134). Increasing evidences of ncRNA deregulation or mutation in the CNS have been related to several neurological disorders, including some neurodegenerative disorders (93, 135, 136).

a. Small ncRNA

Characteristically, ncRNA shorter than 200bp are called small ncRNAs. The first identified short RNAs were the transfer RNA (tRNA) in 1868 followed by the ribosomar RNA (rRNA) more than 100 years after. Although, in the last years the number of short ncRNA has increased vastly.

	Origen	Mechanisms and functions
Small		
miRNAs	Sense, intergenic or intronic	Incorporate into RISC (miRISC), base pair to 3′-UTR of mRNA targets, and mainly induce translational repression or deadenylation and degradation
endo-siRNAs	Sense or antisense, intergenic or exonic	Incorporate into RISC (siRISC), base pair to mRNA target, and induce degradation or heterochromatin formation
piRNAs	Sense or antisense, intergenic	Epigenetic and possible translational control via complementarity with DNA or RNA sequences
PASRs1	Sense, intergenic (promoter region)	Unknown
TASRs1	Antisense, intergenic (3′-UTR end of genes)	Unknown
snoRNAs	Sense, intergenic or intronic	Pre-RNA processing or nucleoside modification (2′-O-ribose methylation and pseudouridylation) of other RNA molecules
snRNAs	Sense, intergenic or intronic	pre-mRNA splicing
tiRNAs	Sense or antisense, intergenic (5′-UTR transcription initiation sites)	Possibly promotes transcription via epigenetic regulation
spliRNAs	Sense, exonic (splice donor site)	Possible epigenetic regulation

Table 5. Classification of small ncRNAs

miRNAs; micro RNAs. endo-si RNAs; endogenous small interfering RNAs. piRNAs; piwi-interacting RNAs. PASRs; promoter-associated short RNAs. TASRs; termini-associated short RNAs. snoRNAs; small nucleolar RNAs. snRNAs; small nuclear RNAs. tiRNAs; transcription initiation RNAs. spliRNAs; splice junction-associated RNAs. Adapted from Salta et al. 2012 (131)

They are involved in most of essential mechanisms of the cell, such as biosynthesis of proteins, removal of introns, site-specific RNA modification, telomere synthesis, transcription, modulation of protein function and regulation of gene expression.

An interesting subgroup of the small ncRNA is the regulatory small ncRNAs. This group is represent by micro RNAs (miRNAs), small interfering RNA (siRNA) and PIWI-interacting RNAs (piRNAs). But recently, many other have been identified product of the processing of long RNAs, such as microRNA-offset RNAs (moRs), mirtrons, miRNA-like sRNAs derived from snoRNAs (H/ACA sRNAs), sRNAs derived from tRNAs (tsRNAs), short hairpin RNAs (shRNAs), and vault RNA (svRNAs), QDE-2 interacting small RNAs (qiRNAs) and C/D box snoRNAs derived small RNAs (C/D sRNAs).

This subset of small ncRNA is 18-35nt long, and associates with Argonaute (AGO) protein family. The Ago family is divided into two subfamilies, PIWI subfamily and the Argonaute subfamily. Classification of this subset of small ncRNA is according to the biogenesis mechanism and the type of AGO family protein they interact.

* Biogenesis of regulatory small RNAs

The most extensively characterized regulator small ncRNA are the miRNAs. They are transcribed by RNA polymerase II from miRNA genes, introns or coding protein genes (137). The transcription product is a pri-miRNA, an 80nt transcript with a 5'cap, a polyadenylated tail and a stem loop structure. Then the microprocessor is formed by the union of RNase III enzyme (DROSHA), the Di George Syndrome Critical region gene 8 (DGCR8); a double stranded RNA binding protein and the newly synthetized pri-miRNA. The microprocessor recognizes the stem loop of the pri-miRNA and cleaves it, resulting in a

70bp pre-miRNA (138, 139). Exportine 5 will transport the pre-miRNA to the cytoplasm (140). Once in the cytoplasm, another RNase III, Dicer, further processes the pre-miRNA. Dicer generates a 21-25 RNA duplex with a 5' phosphate and 2-3nt overhangs. Finally, one of the strands from the RNA duplex, usually the thermodynamically less stable at 5' base paring, is loaded into the RNA-induced silencing complex (RISC). The loading is performed by direct binding one of the RNA duplex strands, which becomes the mature miRNA, with one of the AGO proteins, main component of RISC (139, 141) (Figure 12-A).

Endogenous siRNAs (endo-siRNA) are about 21nt length. They originate from transposon transcripts, from sense-antisense transcripts pairs and long stem loop structures (Figure 12)(142–144). Endo-siRNA are synthesized in the nucleus and exported to the cytoplasm where the action of Dicer trims them to a 21nt RNA duplex (145). Once the duplex is from they interact with RISC promote mRNA cleavage or repression of translation (146).

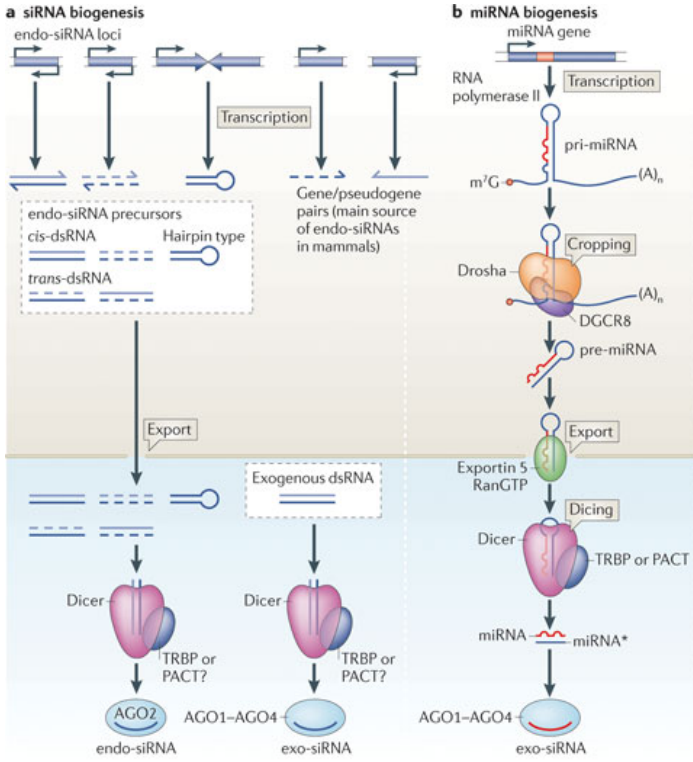


Figure 12. Biogenesis of small RNAs

Schmeatic representation of the biogenesis of a | Small interfering RNA (siRNA). Different types of transcripts that can generate endo-siRNAs. Once in the cytoplasm, Dicer processes them. Finaly, they are loaded to AGO2. Bottom right; model for the generation of exogenous siRNA (exo-siRNA). The double strand RNA (dsRNA) is also processed by dicer and loaded to any of the AGO proteins. b | miRNAs. Is transcribed by RNa polymerase II and processed by DROSHA and DICER. The pre-miRNA resulting is transported to the cytoplasm and cleaved by Dicer. The mature miRNA is loaded into one of the AGO proteins. Taken from Siomi et al, 2011 (147)

In addition to the action of DROSHA and DICER, there are also other ribonucleases involved in the biogenesis of other types of short ncRNAs (Figure 13).

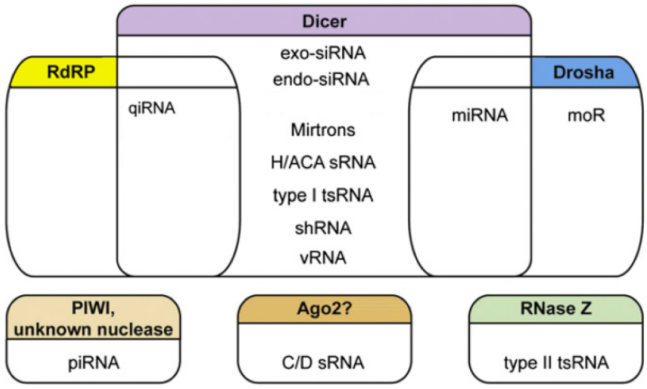


Figure 13. Classification of small ncRNAs depending on the ribonuclease required during their biogenesis.

Schematic representation of enzymes involved in the biogenesis of small ncRNAs. Each square represents the action of a specific ribonuclease. In the square, headings indicate the ribonuclease; and the body represent the types of small ncRNA processed by a specific ribonuclease. Overlapping squares indicate the requirement of the overlapping ribonucleases for the biogenesis of the specific short ncRNAs. Taken from Röther et al. 2011 (148)

* Silencing mechanism

MiRNA regulation of gene expression occurs when the mature miRNA within the RISC complex binds to an mRNA with a complementary sequence. The miRNA seed region specifies the sequence complementarity region, which is determined from the nucleotide 2 to 8 of the mature miRNA. The outcome of the interaction is translation inhibition or mRNA destabilization/cleavage. Translational repression by miRNA is not yet well understood and several studies show contradictory results. Same laboratories have shown that miRNA translational repression can be caused by the impediment of circularization of the mRNA molecule, which is important for translation initiation (149, 150). The circularization is a consequence of the union of the eukaryotic translation initiation factor (eIF4G), a

scaffolding protein with the mRNA cap structure and the polyA-binding protein (PABP), which binds to the poly A tale at the 3'-UTR of the mRNA. Yet, RNA lacking polyA tales can still be targeted by miRNA. Other models propose that miRNA promotes the unloading of the ribosome. It has also been proposed that the miRNA loaded in the RISC complex can compete with the 60S subunit of the ribosome or with eIF4E to binding the mRNA or the cap structure respectively, thus blocking translation initiation (149).

Destabilization of mRNAs can also be performed by miRNA-RISC with partially complementary mRNA targets (151). Trinucleotide repeat-containing gene 6A (TNRC6A), B or C interact with the Ago proteins and the poly A binding protein (PABP). These proteins can also recruit deadenylases CCR4 and CAF1, which will remove the polyA tale from the target mRNA. Thus, the mRNA is destabilized promoting its degradation (151, 152). In many organisms, perfect complementarity of the miRNA/siRNA with the target mRNA can lead to the cleavage of the mRNA by the AGO protein. In humans, AGO2 still preserves the endonucleolitic activity to perform such degradation (153, 154).

The other classes of small ncRNA except piRNA, also interact with AGO proteins, leading to target specific gene silencing. piRNAs are a special case of small ncRNA that interact with PIWI (P-element induced wimpy testis) proteins (155–158). Studies in mammals have shown that piRNA can also induce gene silencing and epigenetic changes.

b. Long non-coding RNAs (lncRNA)

ncRNA longer than 300nt are called long non-coding RNAs (lncRNAs) and most are transcribed by transcribed by RNA polymerase II (159). Taking in account the transcriptional pattern, lncRNA can be classified as it follows:

	Origen	Mechanisms and functions
Long		
lincRNAs31	Sense or antisense, intergenic	Epigenetic regulation
NATs32	Antisense, exonic or intronic	mRNA transcription, splicing, stability, and translation, epigenetic modifications, and precursors of endo-siRNAs
RNA expansion repeats33	Sense or antisense, exonic or intronic	Epigenetic regulation and RNA toxic effects
ENORs34	Sense or antisense, intergenic	Transcriptional regulation, genomic imprinting, precursors of other short and long ncRNAs
eRNAs35	Sense, intergenic (enhancer region)	Transcriptional regulation
PALRs1	Sense, overlap with promoter and first exon of genes	Unknown

Table 6. Classification of lncRNA

Adapted from Salta et al. 2012 (131)

Increasing evidences suggest that an important role of lncRNA is the regulation of protein expression. lncRNA can modify protein expression by recruiting chromatin remodeling complexes to specific genomic loci. Many lncRNA have a specific spatiotemporal pattern of expression, which allows determinate changes in the chromatin structure in a cell specific manner at a specific moment (160, 161). Transcription can be also regulated in trans by lncRNA recruiting repressors or activators of transcription at the promoter site or by directly binding by complementarity to the DNA. This last mechanism leads to the formation of a triple helix that obstructs the binding of transcriptional machinery (162, 163). This last same principal, the recognition and binding of complementary sequences, is adapted for the

post-transcriptional modifications regulated by lncRNA. Expression of complementary RNA sequences, like antisense transcripts, allows the binding of two RNA molecules blocking the function of proteins implicated in splicing, editing, transport, translation and degradation processes (164).

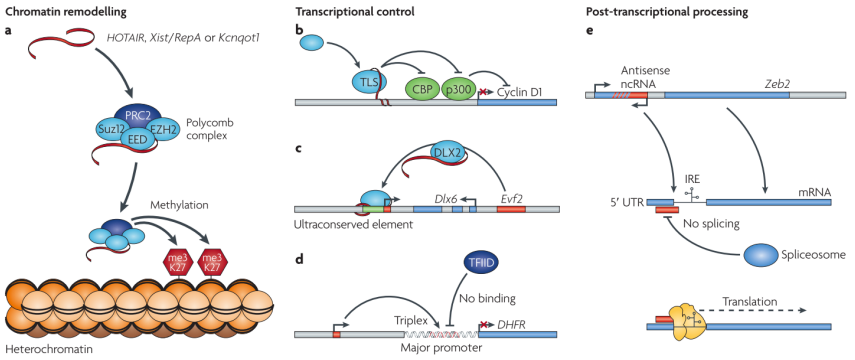


Figure 12. Function of long-non coding RNAs

lncRNA functional examples in a | Chromatin remodelling. b | Transcriptional regulation. c | post-transcriptional regulation.

Taken form Mercer et al. (2009)(165).

In light of the molecular pathways in which lncRNAs are implicated, is not surprising that increasing evidences show their involvement in disease, including neurodegenerative diseases. In particular lncRNA alterations have been described in Huntington's disease (HD), Parkinson's disease (PD) and Alzheimer's disease (AD) (136, 166). Indeed, it was been recently demonstrated that pseudogenes, lncRNAs can act as a sponges of miRNAs competing with other transcripts for the binding of miRNAs (166–168). Therefore, changes in the quantity of lncRNA within the cell could lead to variations in the miRNA pool of the cell, thus resulting in harmful changes for the cell.

2. HYPOTHESIS AND OBJECTIVES

Carriers of CGG trinucleotide repeat expansions, between 55 and 200 repeats, in the 5'-UTR of the *FMR1* gene are at a risk of suffering fragile X tremor/ataxia syndrome (FXTAS) and females premature ovarian failure (POF1 or POF). Carriers of these alleles, also called premutation (PM) alleles, present eosinophilic positive intranuclear inclusions throughout neurons and astrocytes and abnormal high levels of *FMR1* transcript (73). Despite the high levels of *FMR1* transcript, no increase in *FMR1* protein (FMRP) has been detected (39, 76). Those findings prompted the suggestion that the pathogenic mechanism in the *FMR1* PM alleles is related to a toxic mRNA gain of function. Evidences that corroborate this hypothesis where found in mouse and fly models, reproducing features of FXTAS neuropathology by expressing PM alleles outside the *Fmr1* genomic context (42, 113). Additionally, in vitro assays show that *FMR1* expanded CGG repeats affect cell viability in primary neural progenitor cells and in established neural cell lines (78).

Histopathological studies in inclusions found in FXTAS brain samples revealed the presence of *FMR1* mRNA within the inclusions (7). Additional studies demonstrate that the premutated transcript binds to some RNA binding proteins and sequesters them, thus forming the

inclusions (79, 83, 102). Some of the proteins present in the inclusions have important roles in transcription splicing, mRNA trafficking, translation and miRNA biogenesis (79, 81, 83, 84, 102). Sequestration of these proteins prevents their normal function, likely causing an impact on the transcriptome.

Moreover, several studies suggest that CGG repeats, such as the one present in the 5'-UTR of the *FMR1* gene, can form a secondary structure similar to that of miRNA precursors, and that this structure can be cleaved by DICER, giving rise to a CGG duplex of about 21nt (sCGG) (92, 94, 169). Thus, premutated *FMR1* transcript could generate sCGG and trigger downstream silencing effects.

Based on these data, our hypothesis is that changes in the transcriptome induced by the expanded *FMR1* 5'-UTR and the activity of sCGG biogenesis contribute to downstream pathogenic effects. The aim of this thesis is to study changes in gene expression in FXTAS and FXPOI, and to evaluate the involvement of the gene silencing machinery in these pathologies. For this purpose the following objectives have been established:

- Evaluation of transcriptome changes in peripheral blood samples from *FMR1* premutation carriers.
    - Identification of the biological pathways altered in PM carriers.
    - Validation of genes altered in blood samples of *FMR1* carriers in brains of the FXTAS KI mouse model.
    - Assessment of the contribution of the levels of expression of the premutated allele to the observed transcriptional changes.

- Evaluation of the involvement of the gene silencing machinery in expanded-*FMR1* pathologies
    - Identification of small CGG (sCGG) molecules in human and mouse carriers of the *FMR1* premutation.
    - Evaluation of the toxic effects related to the sCGG molecules
    - Evaluation of the involvement of the siRNA biogenesis machinery in the formation of sCGG molecules
    - Evaluation of whether the toxic effects of sCGG molecules are dependent on the RNA induced silencing complex (RISC).
    - Identification of putative silencing targets of the small CGG molecules.

3. RESULTS

3.1. Evaluation of transcriptome changes in peripheral blood samples from *FMR1* premutation carriers

In this section we focused on the description and identification of gene expression perturbations and biological pathways altered in fragile X premutation carriers (fXPCs). Subsequently, we validated gene expression alterations in a mouse model for the *FMR1* PM. In addition, we studied the effects of expanded *FMR1* expression levels in perturbing the transcriptome.

a. Identification of the biological pathways altered in peripheral blood of premutation carriers

To identify mechanisms with possible relevance in the nature and progression of the FXTAS phenotype, we have examined changes in the global gene expression profile in blood samples of male fXPCs. We included nine PM carriers: five FXTAS-symptomatic patients (SP),

four FXTAS-asymptomatic patients (AP) and five age-matched controls (Table 7). During the course of the study, two of the AP fXPC initiated FXTAS symptomatology and therefore became SP. *FMR1* mRNA expression levels in the two other AP, were similar to control individuals, while SP subjects presented a clear up-regulation of *FMR1* mRNA when compared to controls (~2-fold change, $p<0.01$) (Table 7). Accordingly, clinical characteristics of individuals of another study reported AP with no increase in the expression of the *FMR1* gene (97).

Subject	Age	(CGG)n	mRNA *FMR1*[1]	% FMRP[2]	Clinical Signs	Radiological Signs	Diagnostic[3]	FXTAS Onset
fXPC 1	63	106	1,56	67	2 major	1major + 1minor	DEF	62
fXPC 2	65	113	3,2	88	1 major + 1minor	1 major	DEF	64
fXPC 3	62	60	1,72	100	2 major	1major + 1minor	DEF	60
fXPC 4	43	60	1,22	94	1minor	1 major	PROB	64
fXPC 5	69	75	2,06	83	2minor + 1major	1 minor	POS	62
fXPC 6	65	83	1,93	78	1 major	2 major	DEF	65
fXPC 7	55	73	3,82	95	1 major + 2 minor	1 minor	POS	57
fXPC 8	65	80	1,04	78	None	1minor	NO	-
fXPC 9	42	73	0,95	86	None	None	NO	-
Contol 1	72	28	1,09	-	None	None	-	-
Contol 2	77	29	1,17	-	None	None	-	-
Contol 3	69	23	1	-	None	None	-	-
Contol 4	71	29	1,45	-	None	None	-	-
Contol 5	77	23	0,92	-	None	None	-	-

Table 7. Clinical and molecular characteristics from male fXPC

[1]RNA levels are reported as fold change over those in normal sex-matched control individuals. [2]Percentage of hair root positive for FMRP by immune-hystochemical staining. [3]Diagnostic criteria described by Jaquemont et al. (74). DEF: definite, PROB: Probable, POS: Possible.

We performed gene expression profiling using Agilent-based microarrays (SurePrint G3 Human GE 8x60K Microarray) (GSE48873). Transcripts that passed stringent filtering criteria (Fold Change $\geq |1.5|$ and FDR rate $\leq 1\%$) were considered as differentially expressed genes (DEG). According to these criteria, a total of 1,660 RNAs were differentially expressed when comparing SP with control individuals, being 1003 genes down-regulated, and 657 up-regulated. Interestingly, more than 30% of the DEG were long non-coding RNAs

(lncRNAs), being most of them (93%) down-regulated. Heatmap analysis of DEG across different samples showed three main clusters consisting in controls, AP and SP (Figure 13). Because *FMR1* mRNA levels are different between SP and AP, both the CGG expansion per se and the expression levels of the PM *FMR1* gene may account for the gene deregulation observed in PM carriers.

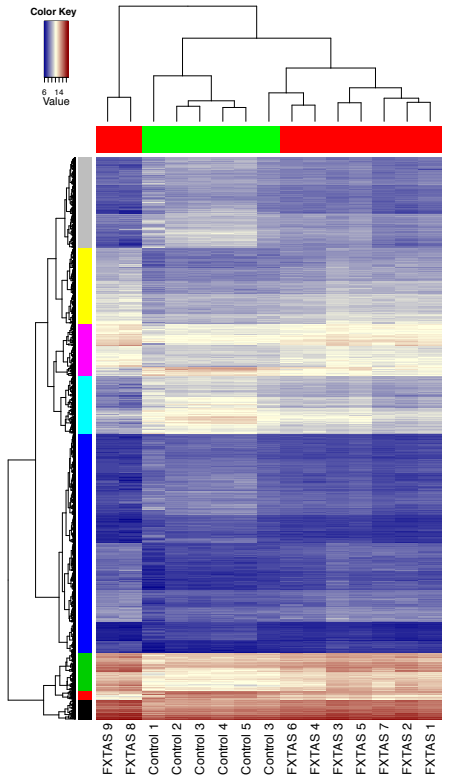


Figure 13. Gene expression analysis in peripheral blood samples from fXPC

Heatmap plot from 1,660 differential expressed RNAs (Fold Change ≥ | 1.5 | and false discovery rate ≤1%). Genes with similar expression profiles are grouped in eight different clusters, labeled with different colors. Columns represent samples and rows show the genes. For each gene, red indicates up-regulation and blue down-regulation of expression relative to the mean.

In the heatmap analysis, we identified eight clusters of DEG, containing between 1,260 (blue) and 51 (red) RNAs (Figure 13). Ingenuity Pathway Analysis (IPA) in each cluster revealed enrichment in several biological pathways that may underlie FXTAS-pathogenic aspects (Table 8), including mitochondrial dysfunction, cell death and survival, inflammatory response, reproductive system disease, and neurological disorders. Thus, DEG in blood identifies pathways related to different neuropathological aspects of the FXTAS phenotypes. Heatmap analysis showed that 45% of the deregulated lncRNA clustered with protein-coding genes involved in inflammatory response, and another 28% with genes involved in nervous system development and function (Table 9), suggesting a participation in analogous functions.

Cluster	Physiological System Development & Function	Disease&Disorders	Molecular & Cellular Function	Canonical Pathways
	Behavior **Nervous System** Organismal Development Tissue Development Cardiovascular System	Hereditary Disorder **Neurological Disease** **Psychological Disorders** Cardiovascular Disease Connective Tissue Disorders	Cell-to-Cell Signalling & Interacion Amino Acid Metabolism Molecular Transport Small Molecule Biochemistry Protein Synthesis	Glutamate Receptor
	-	-	Post-Transletional Modifications **Cell Death & Survival**	**Assembly of RNA polymerase II Complex**
	Hematological System Organismal Development Cardiovascular System Connective Tissue Embryonic Development	Renal & Urological disease **Neurological Disease** Cancer Gastrointestinal Disease Hepatic System Disease	**Cell Death & Survival** Gene Expression Cell Cycle Cellular Movement Protein Synthesis	**EIF2 Signaling** **Regulation of eIF4 & p70S6K Signaling** **mTOR Signaling** **Mitochondrial Dysfunction**
	Skeletal & Muscular System Connective Tissue Cardiovascular System Organ morphology Tissue morphology	Cardiovascular Disease Developmental Disorders Organismal Injury & Abnormalities **Reproductive System Disease** Skeletal & Muscular Disorders	Cell-to-Cell Signalling & Interacion Amino Acid Metabolism Carbohydrate Metabolism Cell Cycle **Cell Death & Survival**	
	Connective Tissue Tissue development Organismal Function Hematological System Embryonic Development	Cancer Dermatological Disease & Conditions **Inflammatory Response** Ophthalmic Disease Respiratory Disease	Cellular Assembly & Organization Cellular Function & Maintanance Cell-to-Cell Signalling & Interacion Cellular Growth & Proliferation Cellular Development	**Role of BRCA1 in DNA Damage Response** RAN Signaling
	Hematological System Immune Cell Trafficking Embryonic Development Heatopoiesis Lymphoid Tissue Structure	Cancer Hematological Disease Renal and Urological disease Gastrointestinal disease Hepatic System Disease	RNA Post-Transciptional Modification Protein Synthesis Gene Expression Cellular Movement Cellular Development	**EIF2 Signaling** **Regulation of eIF4 & p70S6K Signaling** **mTOR Signaling** **Mitochondrial Dysfunction**
	Connective Tissue Embryonic Development **Nervous System** Organ Morphology Organismal Development	Cardiovascular Disease Connective Tissue Disorders Dermatological Disease & Conditions Developmental Disorders Gastrointestinal Disease	Carbohydrate Metabolism **Cell Death & Survival** Cellular Assembly & Organization DNA Replication, Recombination & Repair Lipid Metabolism	-
	Cell-mediated Immune Response Hematological System Organismal Development Tissue Development	Hereditary Disorder Developmental Disorder **Infectious Disease**	**RNA Post-Transciptional Modification** **Protein Synthesis** **Gene Expression**	**EIF2 Signaling** **Regulation of eIF4 & p70S6K Signaling** **mTOR Signaling**

Table 8. Pathway enrichment analysis of differential expressed genes in blood of FXTAS patients vs. control individuals

List of the significantly enriched canonical pathways, molecular and cellular functions, diseases and physiological system development and functions altered for each of the eight gene clusters according to IPA.

Cluster	nº CDS	nº CPG	nº lncRNA	lncNRA %	lncRNA % within cluster
	450	187	226	28.39	50.22
	250	229	15	1.88	6
	191	142	31	3.89	16.23
	292	127	136	17.09	46.58
	1004	554	356	44.72	35.46
	141	82	10	1.26	7.09
	44	22	19	2.39	43.18
	69	54	3	0.38	4.35

Table 9. Cluster distribution of differentially expressed lncRNA, in patients with FXTAS vs. control individuals

In each cluster the table shows the number of coding sequences (CDS), number of coding protein genes (CPG) among the CDS, and number of lncRNA among the CDS. The percentage of lncRNA is refereed to the total number of deregulated lncRNA and the percentage of lncRNA within the cluster is refereed to the total number of deregulated CDS in the cluster.

To identify possible modulators explaining DEG, IPA uses two statistical tools to determine the activation state of a given molecule: the Z-score, calculated within our data set, and the overlap p-value, which also takes into account already known regulators for a given gene. With this approach, IPA identified several cytokines, transcription factors and a number of miRNAs that could contribute to deregulation of gene expression in FXTAS (Table S1).

Microarray expression data were validated by TaqMan-based qPCR for 15 genes selected for their probable functional link to FXTAS (Table 11). These genes are involved in nervous system function, inflammatory response, cell death, mitochondrial function, and oxidative stress. We validated 14 out of 15 genes, using two different reference genes: *ACTB* and *TBP*. The only gene that we could not replicate was the potassium voltage-gated channel member 3 (*KCNC3*), which was not consistently expressed in our samples (Ct values: 38-40). One of the validated genes is the early at menopause 1 (*EAP1*) gene. *EAP1* is an important component in the hypothalamic control for the

initiation of puberty and maintenance of female reproductive cycles in rodents and in non-human primates (170). Since POF is characteristic of female fXPC, we determined the expression *of EAP1* in peripheral blood samples of 25 female fXPC, 12 of them presenting POF1. As controls we included eight females without the permutation, four presenting premature menopause (onset before the age of 40 years) (Table 10).

Subject	(CGG)n	Clinical Signs	Menopause Onset
Patient 1	94	POF1	< 40 years
Patient 2	105	POF1	< 40 years
Patient 3	119	POF1	< 40 years
Patient 4	154	POF1	< 40 years
Patient 5	82	POF1	< 40 years
Patient 6	74	POF1	< 40 years
Patient 7	126	POF1	< 40 years
Patient 8	71	POF1	< 40 years
Patient 9	62	POF1	< 40 years
Patient 10	73	POF1	< 40 years
Patient 11	168	non	> 40 years
Patient 12	80	non	> 40 years
Patient 13	60	non	> 40 years
Patient 14	96	non	> 40 years
Patient 15	116	non	> 40 years
Patient 16	69	non	> 40 years
Patient 17	60	non	> 40 years
Patient 18	70	non	> 40 years
Patient 19	76	non	> 40 years
Patient 20	190	non	> 40 years
Patient 21	73	POF1	< 40 years
Patient 22	60	POF1	< 40 years
Patient 23	80	non	> 40 years
Patient 24	61	non	> 40 years
Patient 25	81	non	> 40 years
Patient 26	<45	POF1	< 40 years
Patient 27	<45	POF1	< 40 years
Patient 28	<45	POF1	< 40 years
Patient 29	<45	POF1	< 40 years
Control 1	<45	non	> 40 years
Control 2	<45	non	> 40 years
Control 3	<45	non	> 40 years
Control 4	<45	non	> 40 years

Table 10. Clinical and molecular characteristics of female fXPC

Gene Symbol	Gene Name	Microarray results		qRT-PCR results		Biological Processes	Disease Processes
		FC	adj. P value	FC	adj. P value		
CASP3	caspase 3, apoptosis-related cysteine peptidase	1.86	3.06E-02	1.39	1.60E-03	apoptosis, cell death, cell cycle progression, DNA damage response, commitment, response, necrosis, development	neurodegeneration: AD, PD, ALS, SCA7; purkinje cell degeneration
DFFA	DNA fragmentation factor, 45kDa, alpha polypeptide	-1.72	3.46E-02	-1.40	6.20E-03	apoptosis, fragmentation in cell death, laddering in condensation, signal transduction in excitotoxicity	crystalline corneal dystrophy
BCL2L1	BCL2-like 1 or Apoptosis regulator Bcl-X	-2.23	4.51E-02	-3.07	2.00E-04	apoptosis, cell death, survival, proliferation, transmembrane potential, growth, cell cycle progression, quantity, cell viability, autophagy	neurodegeneration: AD, PD, ALS, ; swelling
BCL2L11	BCL2-like 11 (apoptosis facilitator) or BIM	-2.49	8.25E-03	-1.64	4.00E-04	apoptosis, number, cell death, homeostasis, quantity, survival, function, abnormal morphology, differentiation, proliferation	neurodegeneration: AD ; K/BxN serum transfer arthritis
APP	amyloid beta (A4) precursor protein	1.87	1.84E-02	1.31	2.00E-03	cell death, apoptosis, cell viability	neurodegeneration: AD; cognitive impairment, dystrophy, gliosis, dementia
AKT1	v-akt murine thymoma viral oncogene homolog 1	-3.24	7.65E-03	-1.51	2.00E-04	apoptosis, survival, proliferation, cell death, migration, growth, size, phosphorylation, transformation, differentiation	cancer, schizophrenia, neurological disorder
COX6C	cytochrome c oxidase subunit VIc	2.92	1.04E-02	1.17	2.00E-04	Respiratory electron transport chain	Mitochondrial Dysfunction, Bipolar Disorder
COX7B	cytochrome c oxidase subunit VIIb	4.36	1.05E-02	2.09	2.00E-04	Respiratory electron transport chain	Mitochondrial Dysfunction, Huntington's Disease Signaling
SOD1	superoxide dismutase 1, soluble	1.80	4.24E-03	1.48	2.00E-04	cell death, apoptosis, survival, degeneration	neurodegeneration: ALS, AD; motor dysfunction; gliosis
RNF10	ring finger protein 10	-2.32	2.72E-02	-2.89	2.00E-04	Regulation of Schwann cell proliferation; Regulation of myelination	Parkinson's disease
HDAC5	histone deacetylase 5	-3.14	9.70E-03	-1.52	1.00E-03	differentiation, organization, modification, silencing, remodeling, expression in, development, activation, growth, acetylation	Huntington's Disease Signaling
ATXN7	ataxin 7	-3.95	4.91E-03	-1.51	2.00E-04	cell death, quantity, inhibition in, degeneration, expression in dysfunction	SCA7, ataxia, cone-rod dystrophy, neurological disorder, neurodegeneration
ATXN3	ataxin 3	1.51	2.58E-03	1.31	2.00E-04	cell death, synaptic transmission, endoplasmic reticulum stress response, axonal transport	SCA3, ataxia, weight loss, psoriasis, neurodegeneration, neurological disorder
IRF2BPL/ EAP1	interferon regulatory factor 2 binding protein-like enhanced at puberty 1	-2.67	6.91E-03	-1.60	4.00E-04	transcription factor that play a role in regulating female reproductive function	Premature Ovarian Failure (POF)
KCNC3	potassium voltage-gated channel member 3	-2.84	5.35E-02	--	--	cell death, ion transport, regulation of neurotransmitter secretion synaptic transmission	SCA3, ataxia, tremor, weight gain, weight loss, hyperactive behavior, myoclonic seizure

Table 11.Validation of expression changes by qPCR. Comparison between qPCR and microarray results for selected genes is shown

RT-PCR analysis showed a significant down-regulation of *EAP1* in fXPC patients when compared with any of the two control groups, thus confirming *EAP1* deregulation both in male and female fXPCs. Down-regulation was stronger (24%) in female fXPCs with POF1 compared with female fXPCs without POF1 (Figure 14). However, a significant *EAP1* decrease was also detected in controls with POF, thus suggesting that *EAP1* low levels may be associated both with the *FMR1* PM and with POF. These findings are in agreement with a common involvement of *EAP1* decline in loss/disruption of the menstrual cycle, in different genetic conditions. Thus, a possibility exists that decreased *EAP1* levels in female fXPCs, contribute to the high prevalence of POF1 in fXPCs.

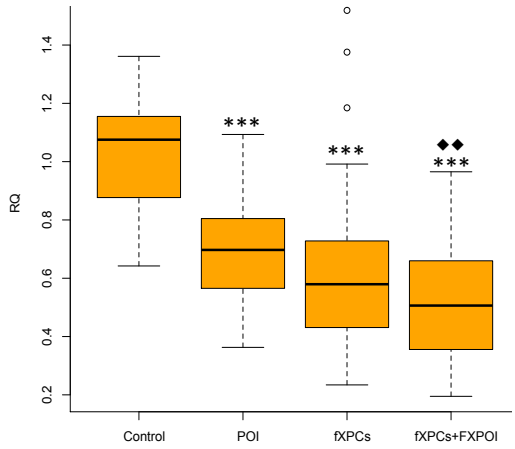


Figure 14. EAP1 expression in fXPCs females

EAP1 mRNA expression levels in control females presenting or not POF and female fXPCs presenting or not POF1. Expression levels are referred to a control sample for RQ. Determinations were performed using two independent reference genes. Plots show distribution and mean RQ ± SE (**P<0.01; ***P<0.001; ◆◆P<0.01; ◆◆◆P<0.001; using a linear mixed effects model). Bonferroni correction was applied for multiple comparisons. Significant differences with respect to control individuals is indicated with "*" and significant differences between fXPCs and fXPC+POF1 indicated with "◆".

b. Validation of candidate genes altered in peripheral blood samples of *FMR1* carriers in brains of the CGG-KI mouse model

The altered expression of a large amount of transcripts in peripheral blood samples of FXPC raises the question on whether similar changes occur in other tissues affected by the CGG-repeat expansion, especially in brain samples for the FXTAS phenotype. We therefore analyzed the expression of selected genes in brain samples of the CGG-KI mouse model, in which the endogenous mouse CGG repeats in the murine *Fmr1* gene have been replaced by the human *FMR1* gene carrying 150 CGG repeats (103). We analyzed the expression of 12 blood validated DEG with relevant functions in neuronal homeostasis (Table 11), in brain samples of three CGG-KI and four control mice. We performed TaqMan-based qPCR using two different reference genes in six brain areas: prefrontal cortex (PFCx), motor cortex (MCx), striatum (STR), hippocampus (HC), cerebellum (CRBL) and brainstem (BS). These areas where selected according to their association with the symptomatology in FXTAS patients and their affectation/involvement in human tissue and mouse samples. MCx, PFCx and HC show high content of intranuclear inclusions in the human brain as well as in the CGG-KI mouse model (105, 106). In addition, working memory is impaired in the CGG-KI mouse model, suggesting that hippocampal-dependent impairments in spatial processing may occur (109). Intranuclear inclusions are also present in specific BS nuclei and CRBL layers of FXTAS CGG-KI mice (105, 106). Moreover, neuroimaging studies in human fXPCs also revealed volume loss in cerebral cortex, HC, CRBL and BS and white matter disease in BS and CRBL (69, 70). Though, intranuclear inclusions are rare in the STR of CGG-KI mice

(105, 106), other types of alterations in these areas may explain some of the motor deficits in FXTAS patients, including resting tremor and rigidity (171, 172).

Increased levels of *Fmr1* mRNA (2-3 FC) were detected in all brain areas of the CGG-KI mouse (Table 12), which agrees with previous publications (171, 172). Out of the remaining 11 genes, we validated nine in BS, eight in STR and seven MCx (Table 12) using two independent reference genes (*Hprt1* and *Tbp*). Moreover, half of the genes studied showed the same expression direction in at least three of the six brain regions. Specifically, down-regulation of potassium voltage-gated channel member 3 (*Kcnc3*) and *Bcl-2* interacting mediator of cell death (*Bim*) could be validated in four out of six different brain regions. Also, the ring finger protein 10 (*Rnf10*), histone deacetylase 5 (*Hdac5*), Bcl-2-like protein 1 (*Bcl-X*), spinocerebellar ataxia type 7 protein (*Atxn7*) and *Eap1* showed down-regulation in half of the tested brain regions. Interestingly, we found that *Eap1* was significantly down-regulated in the BS and CB of CGG-KI mouse, suggesting that *Eap1* levels in certain brain areas could contribute to POF in this model (128). All together, these results suggest that gene expression profiling in blood of fXPC partially reflects changes in the brain transcriptome that may underlie neuropathological aspects in FXTAS.

Gene ID	PFCx	MCx	Estriatum	Hippocampus	Cerebellum	Bstem	Array
Rnf10	-	Down -1.14	Down -1.38	-	-	Down -1.87	Down
Kcnc3	-	Down -1.42	Down -1.98	-	Down -1.24	Down -1.33	Down
Eap1	-	-	-	-	Down -1.19	Down -1.85	Down
Hdac5	-	Down -1.12	Down -1.35	Down -1.22	-	-	Down
Bcl2l11	Down -1.36	Down -1.44	Down -1.33	-	-	Down -1.40	Down
Bcl2l1	-	Down -1.22	Down 1.37	-	-	Down -1.32	Down
Dffa	-	-	Down -1.28	-	-	-	Down
Atxn7	-	Down -1.20	Down -1.21	-	Down -1.16	-	Down
Fmr1	Up 2.59	Up 1.99	Up 2.17	Up 2.67	Up 2.28	Up 2.86	Up
App	Up 1.38	-	-	-	-	Up 1.18	Up
Atxn3	Up 1.22	-	-	-	-	Up 1.34	Up
Sod1	Up 1.33	-	-	-	-	Up 1.32	Up

Table 11. qPCR relative quantification of specific FXTAS-blood-DEG in different brain areas of the CGG-KI FXTAS mouse model

Determinations were performed in the prefrontal cortex, motor cortex, striatum, hippocampus, cerebellum and brain stem. The deregulation pattern and the corresponding fold change are indicated for each gene and brain area. Down or up indicate a statistically significant deregulation in CGG-KI vs. control mice, using a linear mixed effects model, and "-" indicates no deregulation of transcript levels. Deregulation pattern observed with 2 independent endogenous reference genes is indicated in black labeling and deregulation detected with one of the two reference genes is indicated in grey. The last column indicates the deregulation pattern observed in blood of FXTAs patients according to the array and qPCR validations.

Given the overlap in gene expression among peripheral blood and brain samples, we further aimed to elucidate whether a broader correlation existed between those two systems. For that reason we performed microarray expression profiling in three of the dissected brain areas: PFCx, STR and HC. The number of samples available for the expression profiling, 3 CGG-KI mouse and 4 controls, was a limiting parameter to get good statistical power. Nevertheless, taking in account

p-value < 0,05 and Fold Change ≥ |1,2|, we found 329, 693 and 198 DEG in PFCx, STR and HC, respectively. Hierarchical cluster analysis of the three DEG list can distinguish CGG-KI mouse model from the control mouse in HC and STR (Figure 15). Although some of the DEG genes in peripheral blood could be validated in some brain areas of the FXTAS mouse model, we did not detect a significant overlap of DEG in the CGG-KI mouse model and peripheral blood human samples. However, functional analysis of the DEG indicates that the altered pathways are the same. IPA functional analysis of the of the DEG in the three brain areas rank nervous system development and function as the top physiological system altered and neurological disease, infectious disease and inflammatory response as the most ranked disorders (Table S2).

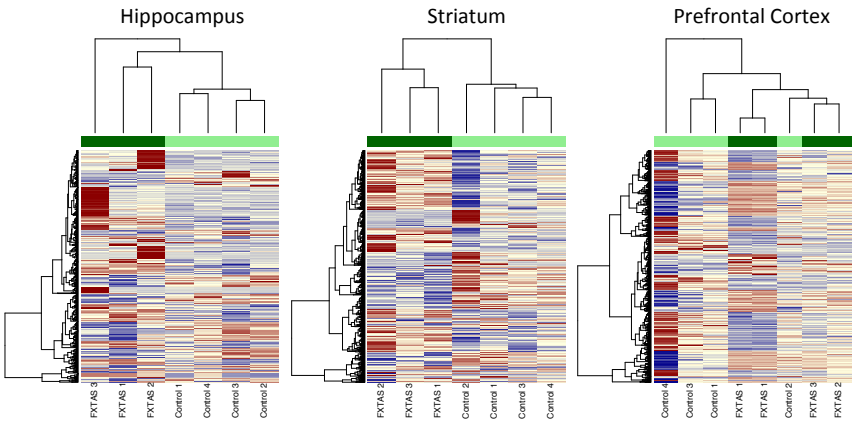


Figure 15. Cluster analysis of differential expressed genes in brains from the CGG-KI mouse model compared to control mouse

Our previous results demonstrate that expression of CGG expansion in the premutation range gives rise to a high deregulation of ncRNA in blood samples of fXPCs. Therefore, we wanted to investigate whether this also the case in brain samples of the mouse model. Data from our

microarray experiments show that, again as seen in peripheral blood samples of FXTAS patients, half of the DEG (Fold Change ≥ |1,2| and p-value ≤ 0.05) were ncRNAs. We identified 149, 346 and 144 ncRNA deregulated in PFCx, STR and HC respectively, where 21 are common in the three brain regions. Prediction of the upstream regulators by IPA revealed that miR-221-3p putative targets showed a tendency for being enriched in HC and PFCx of CGG-KI mouse model, similarly to the analysis of DEG in blood (Table S3 & S4).

c. Assessment of the contribution of the levels of expression of the premutated allele in modulating the transcriptome.

According to patient clustering analysis (Figure 13) our data suggest that SP and AP present differential gene expression profiles which may be explained the levels of expression of *FMR1* 5'-UTR. To determine whether premutated 5'-UTR *FMR1* expression levels could contribute to neuronal gene expression deregulation, we used differentiated post-mitotic SH-SY5Y neuroblastoma cells transfected with expanded and unexpanded 5'-UTR *FMR1* expression vectors. We used genomic DNA from human immortalized lymphocyte cell lines to clone the *FMR1* 5'-UTR of two different CGG lengths. The construct containing 21 CGG repeats (21*CGG) was considered the wild-type (wt) unexpanded allele, whereas a construct containing 79 CGG-repeats (79*CGG) was used as the premutated/expanded allele.

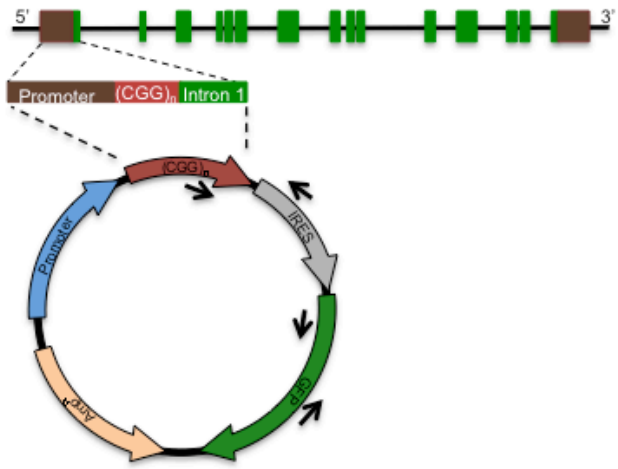
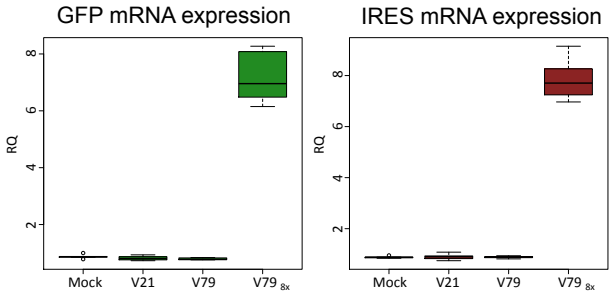


Figure 16. *FMR1* 5'-UTR expression vectors

CGG expanded (79*CGG) and unexpanded (21*CGG) *FMR1* 5'-UTR were cloned into the bicistronic pCAGIG vector and transfected into differentiated SH-SY5Y cells. 79*CGG Two sets of primers (indicated by arrows) were used to quantify the expression of the vectors.

Differentiated SH-SY5Y cells were transiently transfected with the empty vector, 21*CGG-5'FMR1 (V21), 79*CGG-5'FMR1 (V79) or increased amounts of 79*CGG-5'FMR1 (V79_8x) in six independent experiments (Figure 16). The expression levels of the vectors was determined 24 h after transfection using two different custom TaqMan assays designed to amplify GFP or pIRES (Figure 17). 79*CGG-5'FMR1 (V79) was expressed at similar levels as 21*CGG-5'FMR1 (V21) or 8-fold overexpressed (V79_8x). These overexpression levels of premutated-*FMR1* mRNA have been also reported in FXTAS patients (39, 76). Gene expression was then profiled in the six independent biological replicas, using the Agilent SurePrint G3 Human GE 8x60K Microarray (GSE48903).



	GFP		FMR1-IRES	
	FC	p-Value	FC	p-Value
Mock vs. V21	0.91	2.00E-02	0.99	8.90E-01
Mock vs. V79	0.89	8.00E-04	0.98	5.70E-01
Mock vs. V79 8x	8	2.00E-04	8.6	2.00E-04

Figure 17. Expression quantification for the *FMR1* 5'-UTR vectors

Quantification of the expression of the different vectors using two independent sets of primers, covering GFP (right panel) or FMR1-IRES (left panel) fragments. Quantification was normalized using two different reference genes. Expression levels are referred to a control mock-transfected sample for RQ.

Heatmap analysis of the probe sets differently expressed in V79_8x vs. V21 (Fold Change ≥ |1.2| and FDR ≤5%) showed strong clustering of cells overexpressing the 79*CGG vector with respect to the rest of the samples and also discriminated between cells expressing 79*CGG vector, the 21*CGG vector or the empty vector (Figure 18). We found 427 DEG (Fold Change ≥ |1.2| and FDR rate ≤5%) when comparing cells transfected with 21*CGG expressing vector with those overexpressing 79*CGG (8-fold). A total of 16 of these genes presented a similar deregulation pattern in blood when comparing control individuals with SP (Fold Change ≥ |1.2| and FDR ≤5%) (Table S5). We also detected 197 DEG in cells expressing the 79*CGG vector vs. cells overexpressing the same vector. From those, 21 genes presented a similar deregulation pattern in blood when comparing AP with SP (Fold

Change $\geq |1.5|$, Table S6). Although little overlap was observed between DEG in patient's blood and the cellular model, IPA suggests that similar biological pathways are perturbed. Thus, overexpression of expanded 5'-UTR *FMR1* might be sufficient to deregulate a subset of genes involved in inflammation and nervous system development and function. Moreover, 93% (178 out of 192) of DEG in cells expressing the 79*CGG construct vs. cells overexpressing 79*CGG vector (8-fold) were also differentially expressed in cells expressing 21*CGG vectors vs. overexpressing 79*CGG vector (Table S7). This suggests that overexpression of *FMR1* 5'-UTR leads to quantitative rather than qualitative changes in gene expression. In line with this, the levels and pattern of deregulation of 68% of DEG was more substantial in 21*CGG vs. 79*CGG (8-fold) than in 79*CGG vs. 79*CGG (8-fold). A similar phenomenon was observed for 41 genes that showed a tendency to deregulation (not reaching statistical significance) when comparing 21*CGG vs. 79*CGG transfected cells (Table S8). Again, the deregulation levels of those genes was increased in 21*CGG vs. 79*CGG (8-fold) expressing cells.

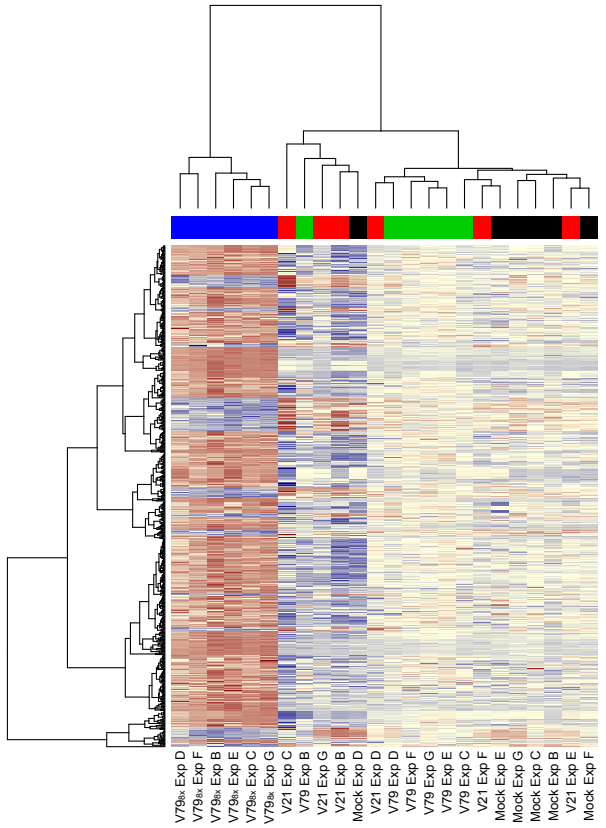


Figure 18. Cluster analysis of differential expressed genes upon transfection of the *FMR1* 5'-UTR vectors

Heatmap plot of 427 DEG comparing 79*CGG8x vs. 21*CGG expressing cells (Fold Change ≥ | 1.2 | and false discovery rate ≤5%) 24 h after transfection of the vectors. Columns represent samples and rows represent genes. For each gene, red indicates up-regulation and blue down-regulation of expression relative to the mean.

To study if the sensitivity of DEG to 79*CGG dosage was specific to the premutation/expansion, we evaluated whether a similar phenomenon was detected upon overexpression of the 21*CGG control vector. Using TaqMan qPCR we validated the 79*CGG dosage sensitivity for four genes out of five examined (Figure 19). For these genes we did not observe significant changes when overexpressing

21*CGG (Figure 20) suggesting selective sensitivity for overexpression of the expanded *FMR1* 5'-UTR.

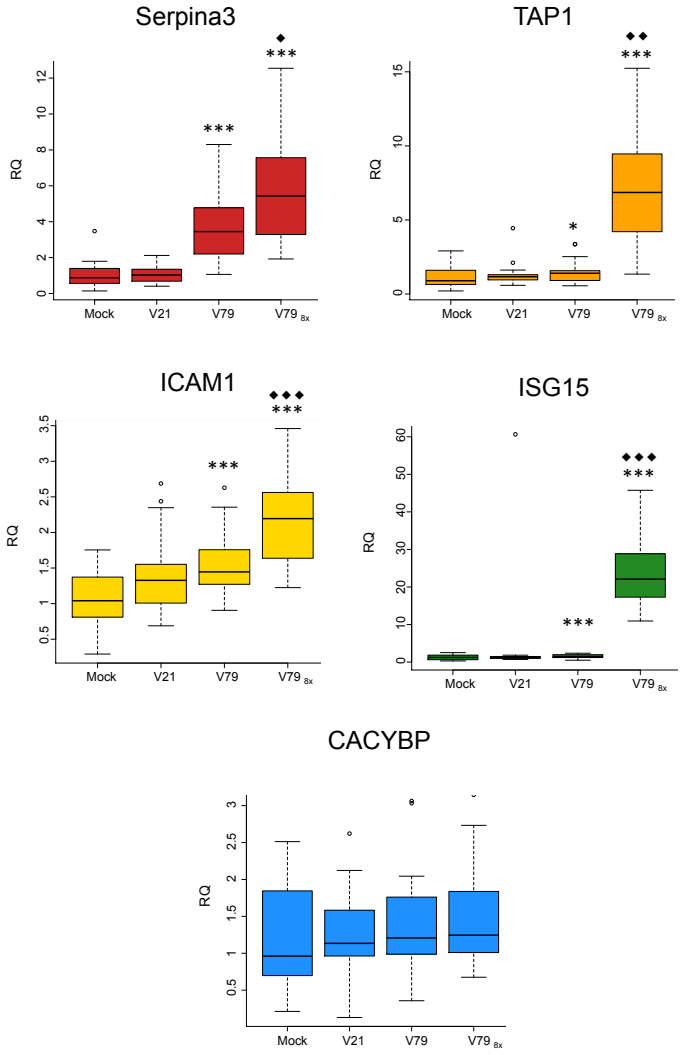
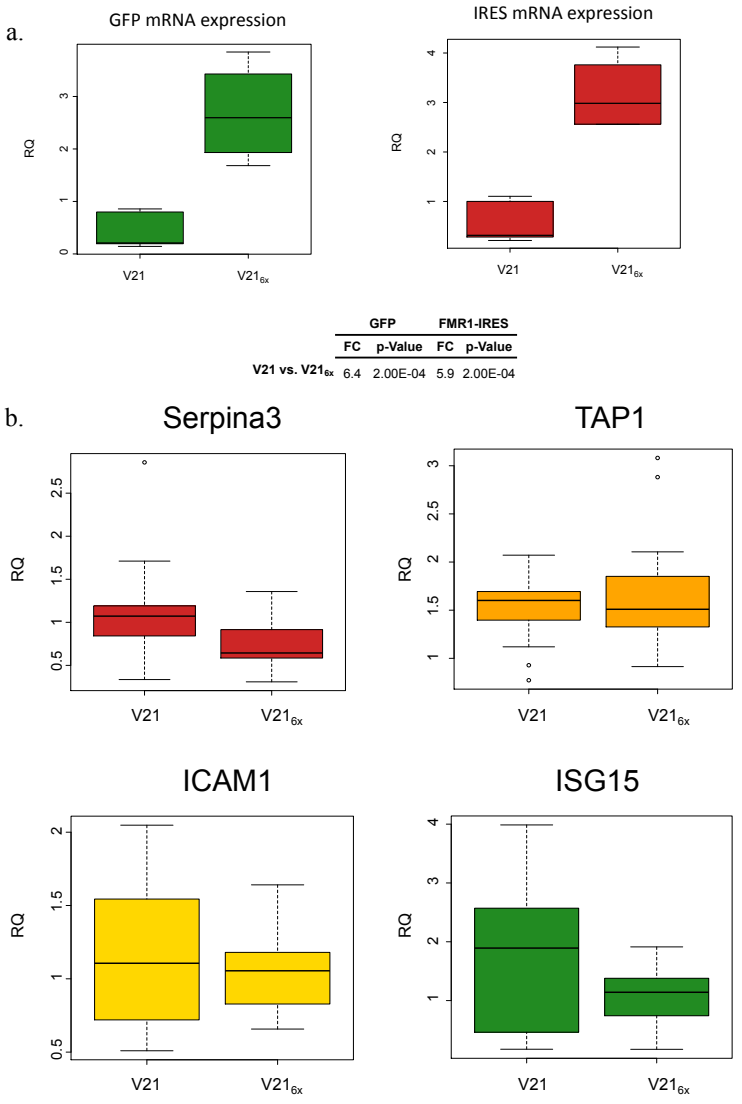


Figure 19. Expression profile of 5 different gens upon transfection of the *FMR1* 5'-UTR vectors

qPCR validation of the expression pattern of SERPINA3, *TAP1, CACYBP, ICAM1* and *ISG15* for which the array showed significant deregulation when comparing 79*CGG or 79*CGG8x vs. 21*CGG expressing cells. Plots show distribution and mean RQ ± SE. Expression levels are referred to a control mock transfected cell for RQ. In all cases quantification was normalized using two different reference genes. (**P<0.01; ***P<0.001; ◆◆P<0.01;

◆◆◆P<0.001;using a linear mixed effects model). Bonferroni correction was applied for multiple comparisons. Significant differences in respect to cells transfected with V21 is indicated with "*" and significant differences between cells transfected with V79 and V79_8x is indicated with "◆".



	GFP		FMR1-IRES	
	FC	p-Value	FC	p-Value
V21 vs. V21$_{6x}$	6.4	2.00E-04	5.9	2.00E-04

Figure 20. Expression profile of 5 different gens upon transfection of 21*CGG at two different concentrations

Relative quantification of 21*CGG overexpression using two independent sets of primers, covering GFP (left panel) or FMR1-IRES (right panel) fragments.

(B) qPCR analysis of SERPINA3, *TAP1, ICAM1* and *ISG15* in SH5Y-SY cells over-expressing the 21*CGG *FMR1* 5'-UTR (V216x). Plots show distribution and mean RQ ± SE. Expression levels are referred to the 21*CGG (V21) transfected cell for RQ and in all cases quantification was normalized using two different reference genes.

We only found 10 differentially expressed lncRNA (Fold Change $\geq |1.2|$ and FDR $\leq$5%) in the *in vitro* cellular system, which contrasts with the large number of lncRNA deregulated in blood of fXPC. Nevertheless, it is worth mentioning that only 15% (919 probes) of the lncRNA probes in the V1 version of the SurePrint G3 Human Gene Expression 8x60K are contained in the newer V2 version. Using the 919 common lncRNA probes for both arrays we found 383 to be differentially expressed in blood samples of patients with FXTAS, with only one being also deregulated in our *in vitro* cell system: zinc finger protein 815 (ZNF815), non-coding RNA.

3.2. Evaluation of the involvement of the gene silencing machinery in PM-FMR1 pathogenesis

In this part of the thesis we aimed to determine whether PM range 5'-UTR of the *FMR1* gene could generate small CGG (sCGG) and cause neuronal dysfunction.

a. Identification of small CGG molecules

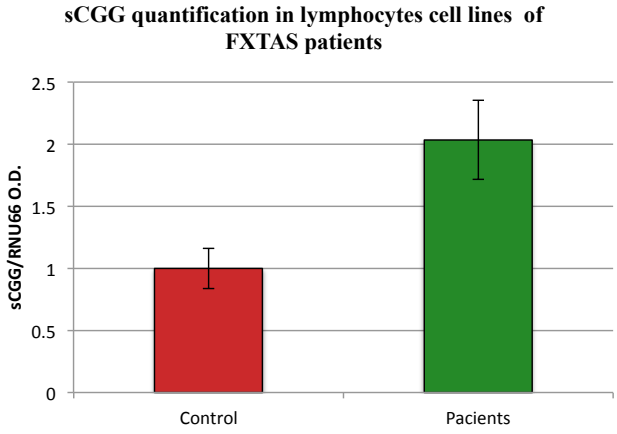
Transcripts with long CGG repeats form hairpin like secondary structures that can be cleaved by the ribonuclease DICER, forming sRNA of 21nt (sCGG) (92, 169, 173). The first objective was to determine weather carriers of PM alleles and a cellular model for the *FMR1* 5'-UTR presented increased levels of sCGG molecules.

* Determination of sCGG in human carriers of the *FMR1* PM

We first intended to identify sCGG molecules in two models that presented the entire genomic background of the *FMR1* gene. We isolated total RNA from two different sources: lymphocytes cell lines (LCL) from FXTAS male patients and peripheral blood samples of female fXPCs. Initially, northern blot was used to identify small RNA species product from *FMR1* 5'-UTR in both models. Unfortunately, this technique was not sensitive enough to detect sCGG. Therefore, polyadenylation of total RNA followed by a RT-PCR and a semi-quantitative PCR was used as an alternative technique to detect RNA products.

We established four LCL lines from male FXTAS patients. Peripheral blood lymphocytes were isolated and subsequently immortalized by infection with Epstein-Barr virus (EBV). Established LCL presented CGG expansion alleles ranging from 60 to 100 CGG repeats. Four age matched controls samples where obtained from the International haplotype map (HapMap) project repository. Detection of sCGG by polyadenilation reaction followed by PCR reaction indicated the presence of sCGG molecules in LCL. sCGG levels tend to increase in

cell lines of FXTAS patients when compared to control individuals, although differences did not reach statistical significance (Figure 21). An increased number samples would have likely resulted in statistically significant increases of sCGG.



Sample	nºCGG
Control 1	<45
Control 2	<45
Control 3	<45
Control 4	<45
FXTAS 1	60
FXTAS 2	70
FXTAS 3	80
FXTAS 4	100

Figure 21. Identification and quantification of sCGG in lymphocytes cell lines of FXTAS patients

sCGG levels are increased in lymphocyte cell line (LCL) of FXTAS carriers when compared to control individuals. sCGG were quantified by semi quantitative RT-PCR using RNU66 as the reference sRNA. Values represent mean fold change from optic density (O.D.) measurements with respect the control samples ± SEM (n=3).

Identification and quantification of sCGG was also performed in peripheral blood samples of female fXPCs, previously described in Table 10. In this case samples where divided in 18 fXPCs and 8 controls. The results indicate a positive correlation between the formation of sCGG molecules and the presence of *FMR1* 5'-UTR premutation (Figure 22).

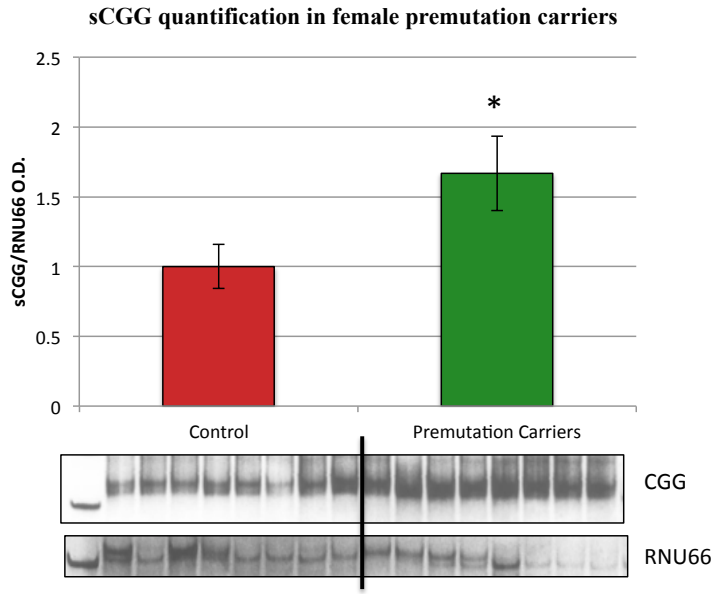


Figure 22. Identification and quantification of sCGG in peripheral blood from fXPCs

sCGG levels are increased in peripheral blood samples of fXPCs when compared to control individuals. sCGG were quantified by semi quantitative RT-PCR using RNU66 as the reference sRNA. Values represent mean fold change from optic density (O.D.) measurements with respect the control samples ± SEM (*p<0.05; using *t*-test).

* Determination of sCGG species in brain samples of the KI model

We also evaluated the presence of sCGG in the brain of the CGG-KI FXTAS mouse model. We isolated total RNA from four different brain regions: HC, PFCx, CRBL and BS, in three KI-CGG and three control mice, and subsequently determined sCGG abundances (Figure 23). We detected increased sCGG levels in two out of the four brain areas analysed: HC and PFCx. Interestingly, sCGG molecules were

significantly increased in the brain areas with higher number of inclusions, PFCx and HC (105).

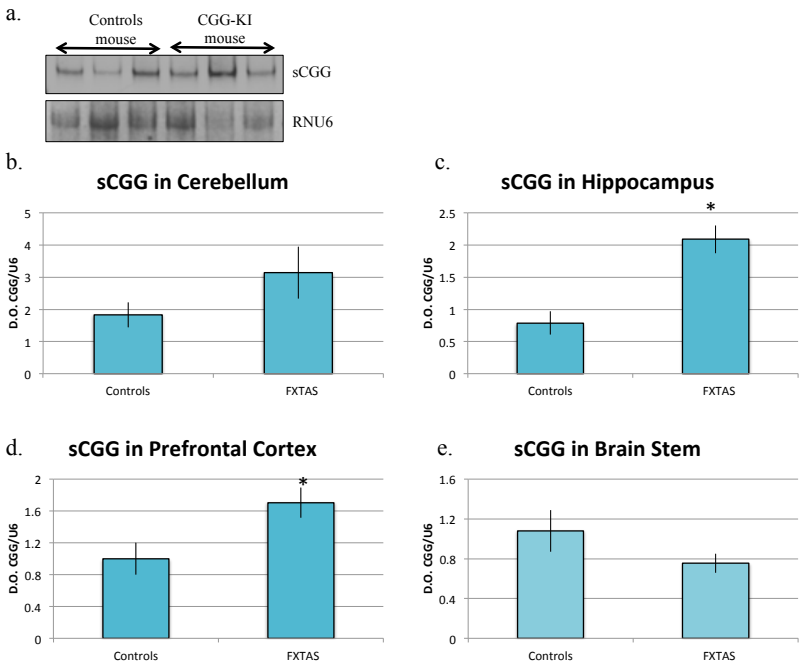


Figure 23. Cetection of sCGG molecules in brains of CGG-KI mouse model

a | Resolution of sCGG acrylamide gel. sCGG and RNU6 were detected by semi quantitative RT-PCR. b | Measurements of sCCG molecules in Cerebellum c | Hippocampus d | Prefrontal cortex e | Brain stem. Quantification is normalized with RNU6 as a reference sRNA. Values represent mean fold change from optic density (O.D) measurements with respect the control samples ± SEM (n=3; *p<0.05; using *t*-test).

* Determination of sCGG species in an in vitro cell model expressing the *FMR1* premutation

We used a neuronal cell line transiently expressing the 5'-UTR of the *FMR1* gene, as previously explained (Figure 16). sCGG levels were determined in cells expressing the normal *FMR1* 5'-UTR with 21 CGG (21*CGG) and cells expressing the mutant *FMR1* 5'-UTR (79*CGG), 48 h after transfection of the corresponding vectors. sCGG molecules were significantly increased in cells expressing the expanded construct (Figure 24).

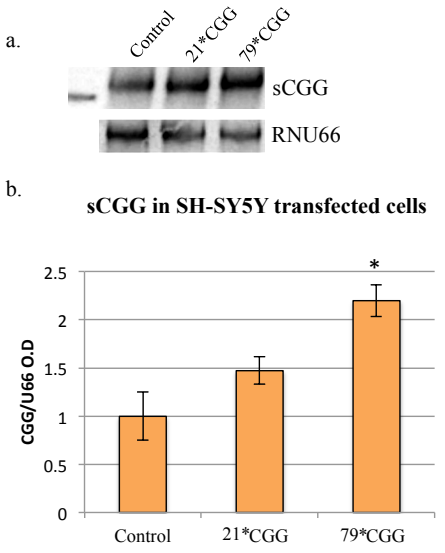


Figure 24. Identification and quantification of sCGG in SH-SY5Y neuroblastoma cell line

sCGG levels are increased in cells transfected with 79*CGG- when compared to those transfected with the 21*CGG-vectors. a | sCGG were quantified by semi quantitative RT-PCR and normalized by transfection efficiency using GFP expression and RNU66 as the reference sRNA. b | Values represent mean fold change from optic density (O.D) measurements with respect the control samples ± SEM (n=3; *p<0.05; using *t*-test).

* Characterization of sCGG molecules

Our results show in different models that the presence of expanded *FMR1* 5'-UTR correlates positively with an increase expression of sCGG. To characterize the identity of sCGG products we cloned and sequenced them. For these experiment we used RNA isolated from the LCL samples because, due to their immortalized condition, they are an unlimited source of RNA in the laboratory, unlike the other samples. We used four control LCL lines and 4 FXTAS LCL lines. The results show the presence of sCGG of 6 repeats length (CGG_{6x}) in all the samples and CGG_{7x} in 3 out of the 4 samples of the FXTAS LCL. Three independent clones have been sequence for each LCL line.

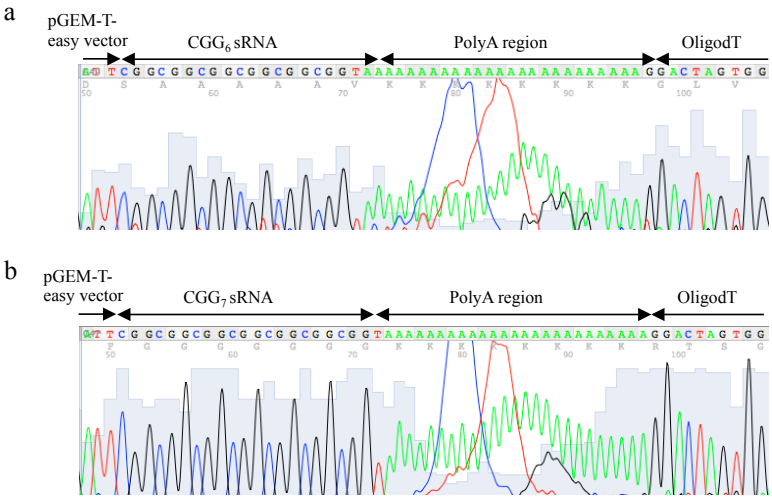


Figure 25. sCGG sequence in lymphocyte cell lines of FXTAS

Histogram with the sequencing results of FXTAS LCL sample expressing a | $(CGG)_{x6}$ sRNA and b | $(CGG)_{x7}$ sRNA

b. Evaluation of the toxic effects of small CGG

According to our data, the presence of sCGG molecules is increased in the presence of PM alleles of the *FMR1* gene. Yet, it remains uncertain whether this event is related with detrimental changes in cellular processes. Therefore, we aimed to assess neuronal viability and oxidative stress responses linked to sCGG up-regulation, two major hallmarks of neurodegenerative processes. Neuroblastoma cells SH-SY5Y were differentiated to a post-mitotic state and transfected either with synthetic CGG-siRNA, similar to the sCGG that are being produced from expanded FMR1, or with the 79*CGG vector, from which sCGG are produced through Dicer endonuclease activity (92, 94). At different time points, cell viability was assed by determinations of lactate dehydrogenase enzymatic activity (LDH), a stable cytosolic enzyme that is released upon permeabilization of the cellular membrane and cell death. We measured cell toxicity 24h and 48h after transfection of the synthetic siRNAs and 14h, 24h, 48h and 72h after transfection of the *FMR1* 5'-UTR expression vectors. Biological effects were compared to scramble siRNA or 21*CGG vector for CGG-siRNA and 79*CGG vector, respectively. LDH activity measurements indicate that cells transfected with either synthetic CGG-siRNA or 79*CGG vector have decreased cell viability compared with cells transfected with scrambled siRNA or 21*CGG vector, respectively (Figure 26). In addition, transfection of CGG-siRNA has faster detrimental effects on cell viability than 79*CGG vector. Nonetheless, the 79*CGG vector induced toxicity at a later time point. The effects of the 79*CGG construct may reflect the need for expanded *FMR1* 5'-UTR processing and/or the contribution of later detrimental effects induced by expanded *FMR1*.

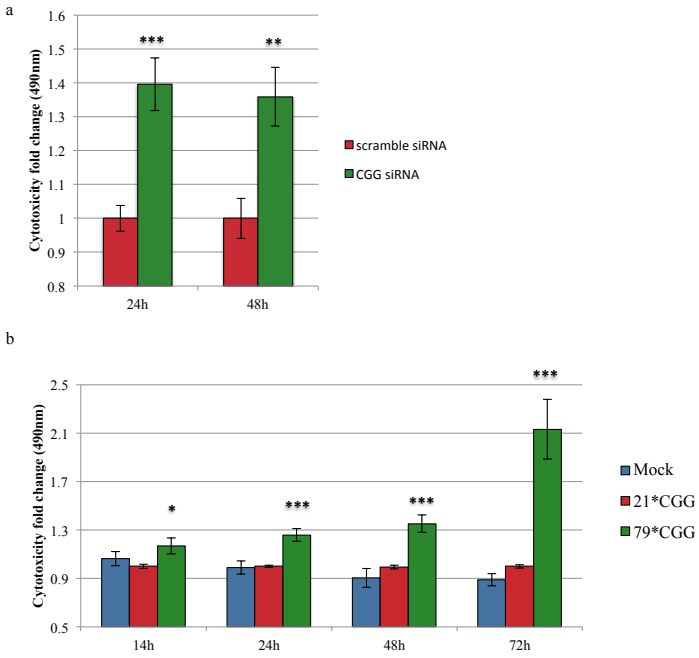


Figure 26. CGG-siRNA and 79*CGG vector induce toxicity in neuroblastoma human cells

a | CGG siRNA induces neuronal toxicity. LDH assay on differentiated neuroblastoma cell line transfected with CGG siRNA (green bars) and scramble siRNA (red bars) at 24 and 48h. b | 79*CGG vector induces neuronal toxicity. LDH assay on differentiated neuroblastoma cell line transfected with empty vector (blue bars), 21*CGG- (red bars) or the 79*CGG- vectors (green bars). Mock condition was used as a negative control. Bars represent fold change induction upon scramble siRNA or 21*CGG vector respectively. Experiments were performed with quintuplicates (n=3; *$p<0.05$; **$p<0.01$; ***$p<0.001$; using *t*-test)

Previous results in the laboratory showed that CGG-siRNA could trigger oxidative stress responses in differentiated neurobalstoma cell line (un published results, Bañez-Coronel et al.). Interestingly, a recent study has shown that MnSOD, an antioxidant enzyme that dismutates superoxide anion to hydrogen peroxide, was decreased in FXTAS fibroblasts (99). Reduced expression of MnSOD would lead to an

increase of the superoxide radical in the cell and an increase in the oxidative stress responses. Therefore, we measured concentration of the superoxide anion upon transfection of 79*CGG vector compared to 21*CGG vector in the same cellular system. We used a red fluorogenic dye that specifically targets mitochondria in living cells. The dye produces red fluorescence upon oxidation by superoxide. We specially used a red dye to be able to distinguish GFP positive cells and oxidative stress response signal. Our results show that cells transfected with 79*CGG vector present a significant increase of superoxide anion when compared to 21*CGG vector (Figure 27).

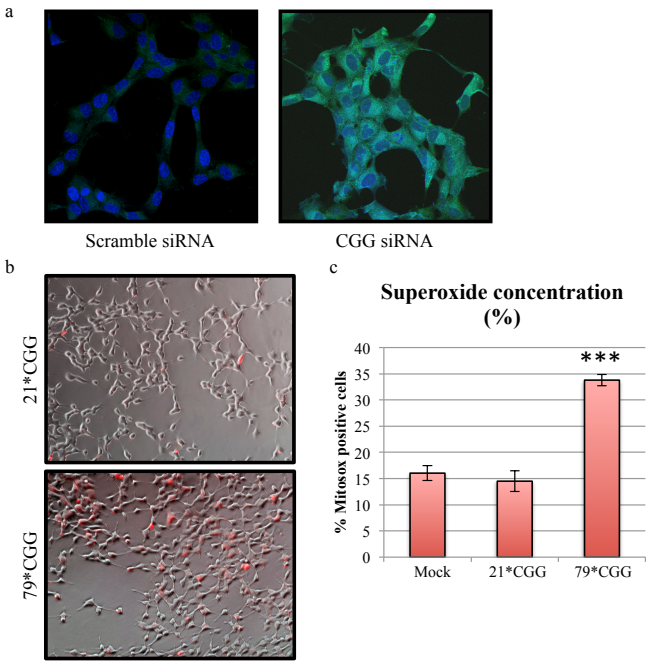


Figure 27. Increase of superoxide anion concentration in cells expressing 79*CGG vector

a | Confocal images of ROS detection with a green fluorescent probe CM-H2-DCFDA and DAPI staining of differentiated SH-SY5Y cells 48h after transfection of either scramble siRNA (left) or CGG-siRNA (right) (Unpublished results, Bañez-Coronel). b | Superoxide detection within the

mitochondria with a red fluorescent probe MitoSOX 72 h after transfection with 21*CGG vector (upper panel) or the 79*CGG vector (lower panel). c | Graph shows the mean percentage of cells positive for superoxide detection ± SEM (n=3; *** p<0.001; using *t*-test).

In addition, to test whether toxicity associated with 5'-UTR *FMR1* PM alleles is related with sRNA mechanisms, we purified the sRNA fraction (<100nt) from neuronal cells expression the mock, 21*CGG and 79*CGG vectors. Isolation of sRNA was performed 48h after transfection of the vectors, when the sCGG rise was detected and the cell viability of the transfected cells was not excessively compromised. The population of sRNA isolated from cells expressing the 79*CGG vector induced a cell death response when compared to those isolated from cells expressing the 21*CGG vector (Figure 28).

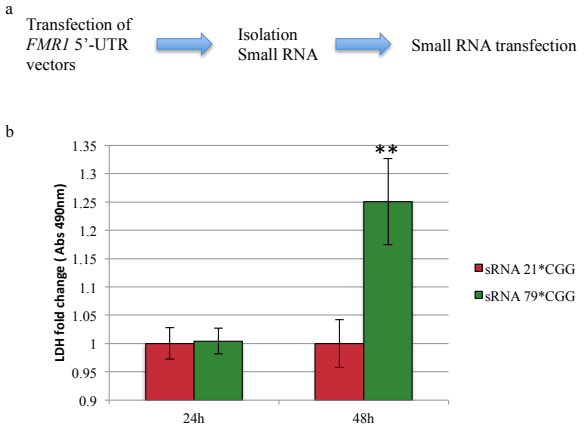


Figure 28. Small RNA derived from cells expressing 79*CGG vector produce neuronal toxicity

Transfection of sRNA obtained from cells expressing the 79*CGG vector induce neuronal toxicity. a | schematic representation of the transfection procedure. b | LDH assay on differentiated neuroblastoma cell line transfected with sRNA (<100nt) obtained from cells expressing 21*CGG (red bars) or 79*CGG (green bars) at 24 and 48h. LDH activity was detected measured by a colorimetric reaction at 490nm. Bars represent fold change induction referred to scramble siRNA or 21*CGG vector respectively. Experiments were performed with quintuplicates (n=3; **p<0.01; using *t*-test)

Further on, we aimed to determine the contribution of sCGG molecules in the toxicity induced by the sRNA fraction form cells transfected with 79*CGG vector. For that reason we designed a synthetic siRNA anti-sCGG, which has the ability to sequester complementary sequences such as the sCGG molecule. Therefore, we transfected again sRNA isolated from cells expressing the 79*CGG vector together with a scramble anti-siRNA or a siRNA anti-sCGG. Preliminary results indicate that toxicity induced by sRNA isolated from 79*CGG expressing cells can be prevented by anti-sCGG but not by a scrambled anti-siRNA (Figure 29). These results suggest that sCGG molecules are an important element in the toxicity mediated by expanded *FMR1* alleles.

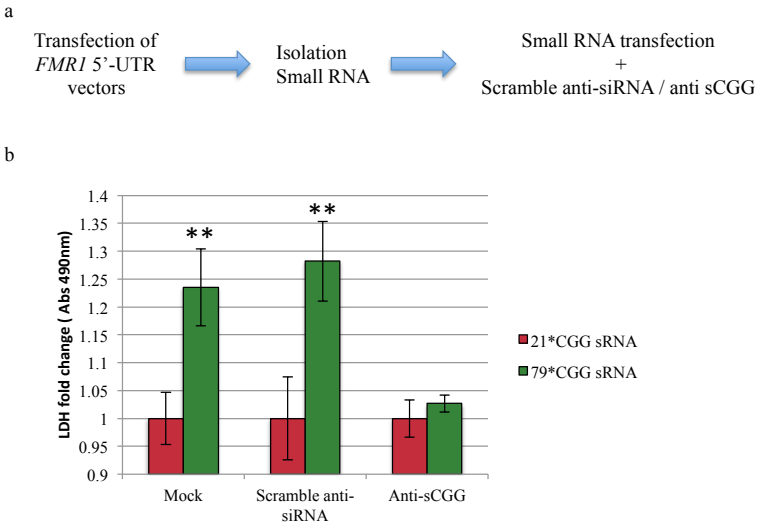


Figure 29. Anti-sCGG prevent cell toxicity mediated by small RNA fraction isolated from cells expressing 79*CGG vector

Transfection of synthetic siRNA anti-sCGG prevents neuronal toxicity induced by sRNA obtained from cells expressing 79*CGG vector. a | schematic representation of the transfection procedure. b | LDH assay on differentiated neuroblastoma cell line transfected with sRNA (<100nt) obtained from cells

expressing 21*CGG (red bars) or 79*CGG (green bars) together with a CGG siRNA or a scramble siRNA. 79*CGG sRNA toxicity is prevented upon co-transfection of anti-sCGG but not scramble anti-siRNA. Bars represent fold change induction upon scramble siRNA or 21*CGG vector respectively. Experiments were performed with quintuplicates (n=1; **p<0.01; using *t*-test)

c. Evaluation of the involvement of the siRNA biogenesis machinery in the formation of small CGG molecules

We have previously described an increase presence of a small CGG molecule in different models for the *FMR1* premutation. Furthermore, we have reported that such sCGG molecules are toxic for neuronal cells and likely account for the toxicity associated with expanded *FMR1* 5'-UTR. Based on the knowledge that long CGG repeat RNA transcripts can be cleaved by Dicer (92, 94), we wanted to evaluate whether depletion of Dicer in cells expressing expanded FMR1 5'-UTR could abolish the formation of sCGG molecules. We used the endogenous RNA interference (RNAi) machinery of the cell to knock down (KD) gene expression of Dicer, the limiting endonuclease in the siRNA biogenesis pathway. A synthetic siRNA targeting Dicer mRNA was transfected into differentiated SHSY-5Y cells. We measured cell viability and expression of sCGG in differentiated neuroblastoma cells KD for Dicer and expressing the 79*CGG vector. Knockdown (KD) of Dicer protein showed to protect the cells against cell death induced by expanded *FMR1* 5'-UTR (Figure 30A). Besides, such protective effect was correlated with a partial disruption of the sCGG biogenesis in Dicer KD cells (Figure 30B).

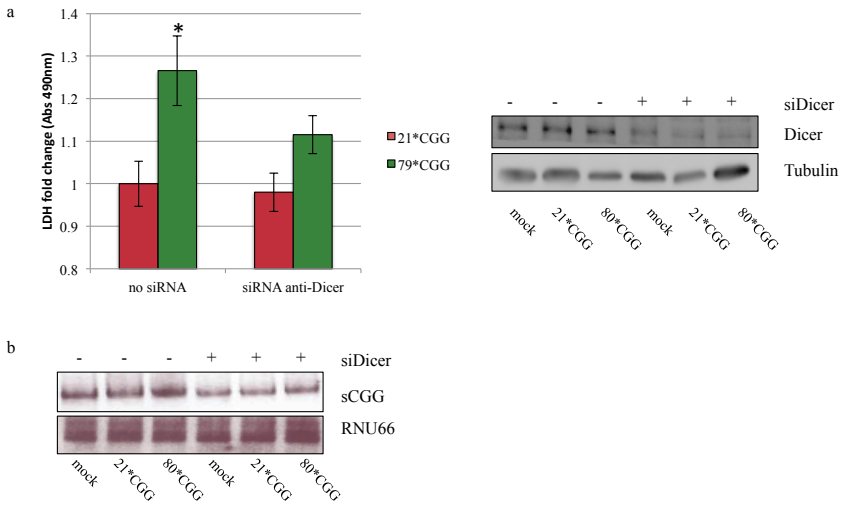


Figure 30. Knock down (KD) of Dicer reduces cellular toxicity mediated by 79*CGG

a | LDH assay on differentiated neuroblastoma cell line KD for Dicer (bars on the right) expressing with 21*CGG vector (red bars) or 79*CGG vector (green bars). Bars represent fold change induction referred to cells transfected with the 21*CGG vector. Experiments were performed with quintuplicates (n=3; **p<0.01; using *t*-test). KD of Dicer protein is represented in the left panel. b | Expression of sCGG molecules is reduced in cells KD for Dicer.

In addition, we have previously shown that sCGG molecules are not equally overexpressed in the different brain areas of the CGG-KI mouse model. In an effort to understand the causes of these differences we have evaluated the expression levels of *Fmr1* and *Dicer* mRNAs in the different brain areas. We used to independent endogenous controls for relative quantification of both mRNAs (*Tbp* and *Hprt1)*. According to previous studies, the expression of *Fmr1* mRNA between 2-3 fold in CGG-KI mouse model when compared to control mouse (106). Nonetheless, within the same group no significant differences in the expression level of *Fmr1* can be observed among the different brain areas (Figure 31a). Moreover, no changes in the expression levels of *Dicer* are appreciated between control and CGG-KI mouse model in

BS, HC and CRBL (Figure 13b). However, *Dicer* expression in control and CGG-KI mouse model is decreased by a 50% in PFCx, a region with increased expression of sCGG molecules. These results suggest that sCGG production might be rated to Dicer activity in the distinct brain areas rather than to its expression.

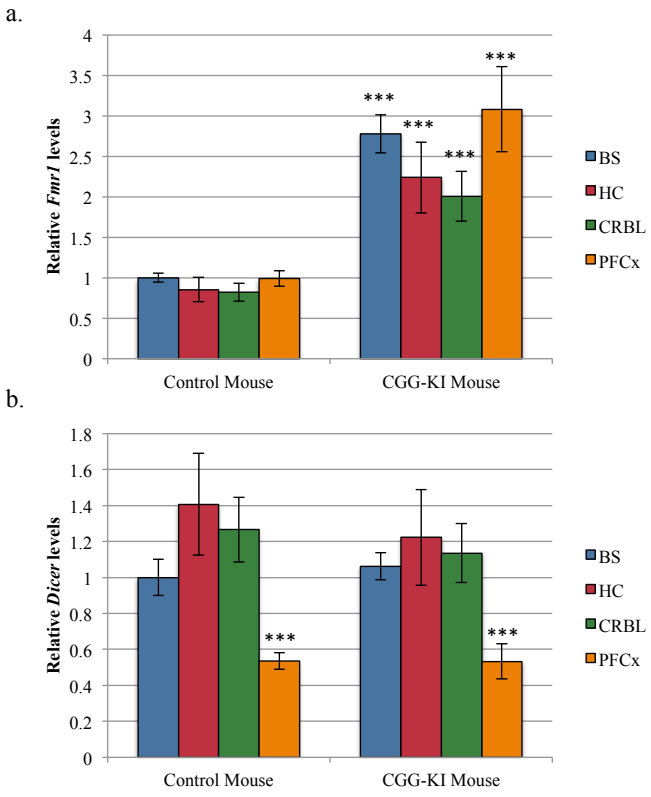


Figure 31. Expression levels of *Fmr1* and *Dicer* in brains of control and CGG-KI mouse model

qPCR validation of the expression pattern of a | *Fmr1* mRNA. b | *Dicer* mRNA. Bars represent fold change induction normalized to brainstem of control mouse ± SE. Experiments were performed with two independent RT-PCR and quadruplicates for each PCR sample. Quantification was normalized using two different reference genes. (***$p<0.001$; using a linear mixed effects model). BS, Brainstem; HC, Hippocampus; CRBL, Cerebellum; PFCx, Prefrontal cortex.

d. Evaluation of whether the toxic effects of small CGG molecules are dependent on the RNA induced silencing complex (RISC)

We have shown that sCGG biogenesis from expanded FMR1 5'-UTR depends on Dicer. Further, we wanted to study whether sCGG molecules related toxicity is associated with pathogenic gene silencing. To evaluate whether sCGG can act through RNA silencing mechanisms, like siRNAs/miRNAs, we KD Argonaure 2 (AGO2), the active component of the RISC silencing machinery. KD of AGO2 was conducted with siRNA against AGO2 mRNA in differentiated neuroblastoma cells. The decrease in cell viability induced by a CGG siRNA and the 79*CGG vector was prevented upon AGO2 KD (Figure 32), thus indicating that AGO2 protein is important factor to mediate sCGG associated toxicity.

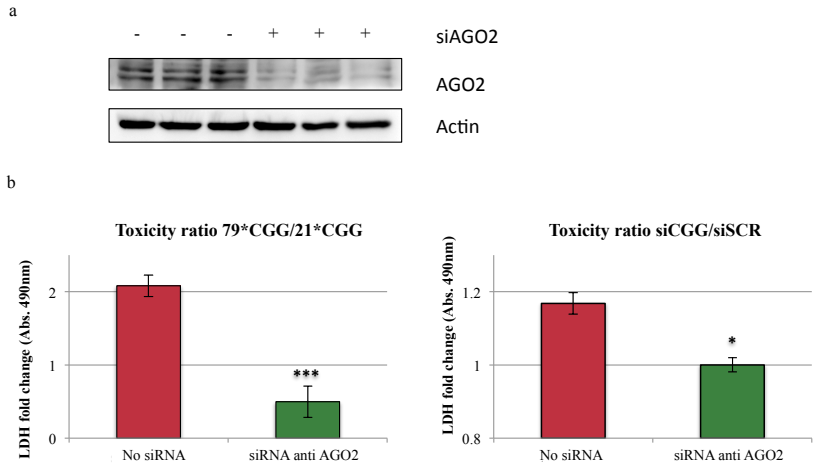


Figure 32. Knock down (KD) of Ago2 prevents cellular toxicity mediated by 79*CGG and CGG-siRNA

a | KD of AGO2 protein . b | LDH assay on differentiated neuroblastoma cell line KD for AGO2. Left graph shows the ratio between 79*CGG/21*CGG toxic effect (red bar) and the same ratio in cells KD for AGO2 (green bar). Right graph shows the ratio between scramble siRNA/CGG siRNA toxic effect (red bar) and the same ratio in cells KD for AGO2 (green bar). Bars represent fold change induction referred to 21*CGG vector. Experiments were performed with quintuplicates (n=3; *p<0.05; ***p<0.001; using *t*-test).

e. Identification of putative silencing targets of the small CGG molecules

We aimed at identifying some putative mRNA silenced by sCGG. We used the transcriptome data available from peripheral blood samples from fXPCs. We first identified mRNA with complementary sequence to the sCGG molecule. Our search was not restricted to a complementarity of the sCGG in the 3'-UTR, but to the entire coding region. Next, we evaluated the possibilities for the sCGG molecule to bind to the mRNA with the complementary sequence. This analysis was performed *in silico* with the RNAhybrid, a tool for finding the minimum free energy hybridisation between a long and a short RNA. For such purpose, the algorithm takes in account the complementary nucleotide sequence, of 6 to 7nt length, to appear three times in the mRNA sequence and its secondary structure. Finally, the list of genes with the 300 mRNA most likely to be target of the sCGG was crossed with the list of probes deregulated in peripheral blood samples of fXPCs. This analysis showed no specific enrichment in genes containing CGG stretches among the down-regulated genes (Figure 33).

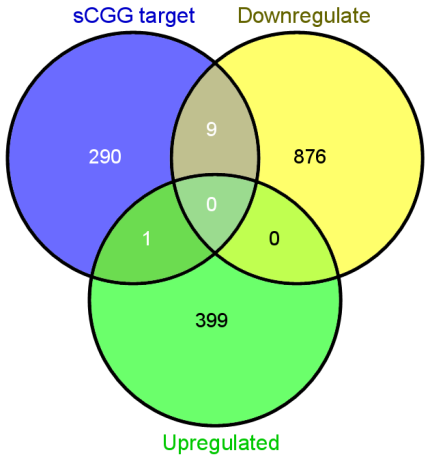


Figure 33. Venn diagram showing overlap between down-regulated genes in fXPCs and putative sCGG targets

sCGG targets predicted by RNAhybrid are represented in blue. Transcript probes down-regulated in microarray expression of fXPCs peripheral blood are represented in yellow, while up-regulated probes are in green.

Yet, we found deregulation of nine genes that overlap with genes down-regulated in the expression arrays of fXPCS. Five of the genes have important roles in apoptosis, neuronal function and gene expression regulation: *BCL2L11, EAP1, HDAC5, KCNC3* and *RNF10* (Figure 21). Interestingly, we have validated by quantitative PCR the down-regulation of those genes in peripheral blood of fXPCs and in brain samples of CGG-KI mouse model (Table 5,6). The location of the GCC seed region, complementary to the CGG repeat, is shown in table 12. Although the demonstration of the direct targeting of these genes by sCGG requires additional experiments, these data suggest that sCGG may affect the expression of specific genes.

Gene	5'-UTR	Exon	Intron	3'-UTR
HDAC5	6	1	-	-
EAP1	-	7	-	-
KCNC3	4	-	1	-
BCL2L11	5	-	-	-
RNF10	4	1	-	-

Table 12. sCGG seed regions in putative targets

Number complementary regions to the sCGG molecules of 6-7nt length. The number indicates the number of seed regions present in each structural part of the gene.

3.3. Supplementary information

Supplementary Table 1 Predicted upstream regulators and its activation state in FXTAS patients, in IPA of blood-DEG

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
BCR (complex)		complex	Activated	2.116	2.82E-01
CD3		complex	Inhibited	-5.183	1.79E-02
PDGF BB		complex	Inhibited	-2.132	3.23E-01
CXCL12		cytokine	Activated	2.201	1.00E+00
C5		cytokine	Inhibited	-2.152	1.00E+00
CRH		cytokine	Inhibited	-2.216	1.00E+00
NPR1		enzyme	Activated	2.204	1.48E-01
HRAS		enzyme	Activated	2.483	1.00E+00
FAAH	-1.869	enzyme	Inhibited	-2.121	1.14E-03
CDC42		enzyme	Inhibited	-2.19	1.01E-01
FN1		enzyme	Inhibited	-2.066	1.00E+00
RAC1		enzyme	Inhibited	-2.219	1.00E+00
Hedgehog		group	Activated	2	5.55E-01
Hdac		group	Inhibited	-2	1.00E+00
Akt	-3.241	group	Inhibited	-2.669	1.00E+00
FGF1	-2.409	growth factor	Inhibited	-2.215	1.00E+00
BMP7		growth factor	Inhibited	-2.646	1.00E+00
ANXA7		ion channel	Activated	2.433	1.00E+00
INSR		kinase	Activated	3.111	3.21E-01
JAK1		kinase	Inhibited	-2	5.38E-01
AKT1	-2.602	kinase	Inhibited	-2.168	1.00E+00
MAPKAPK2		kinase	Inhibited	-2.219	1.00E+00
ESRRA		ligand-dependent nuclear receptor	Activated	2.57	4.88E-04
RORA		ligand-dependent nuclear receptor	Inhibited	-2	1.00E+00
AR		ligand-dependent nuclear receptor	Inhibited	-2.177	1.00E+00
RARA	-3.751	ligand-dependent nuclear receptor	Inhibited	-2.416	1.00E+00
miR-361-3p		mature microRNA	Activated	2.997	4.91E-02
miR-640		mature microRNA	Inhibited	-2.616	1.58E-03
miR-433-3p		mature microRNA	Inhibited	-5.003	2.50E-03
miR-3913-3p		mature microRNA	Inhibited	-2.714	3.38E-03
miR-590-3p		mature microRNA	Inhibited	-8.106	3.76E-03
miR-802		mature microRNA	Inhibited	-2.744	4.24E-03
miR-4432		mature microRNA	Inhibited	-2.209	1.09E-02
miR-139-5p		mature microRNA	Inhibited	-2.091	1.18E-02
miR-16-5p		mature microRNA	Inhibited	-4.866	1.26E-02
miR-181b-1-3p		mature microRNA	Inhibited	-3.872	1.53E-02
miR-503-5p		mature microRNA	Inhibited	-2.159	1.65E-02
miR-622		mature microRNA	Inhibited	-2.81	1.66E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
miR-4633-5p		mature microRNA	Inhibited	-2.34	1.83E-02
miR-628-3p		mature microRNA	Inhibited	-2.201	1.88E-02
miR-4724-5p		mature microRNA	Inhibited	-3.108	1.94E-02
miR-646		mature microRNA	Inhibited	-2.765	2.62E-02
miR-140-3p		mature microRNA	Inhibited	-2.958	2.69E-02
miR-103-3p		mature microRNA	Inhibited	-4.225	2.69E-02
miR-203b-5p		mature microRNA	Inhibited	-2.012	2.84E-02
miR-1295b-5p		mature microRNA	Inhibited	-3.671	3.04E-02
miR-875-5p		mature microRNA	Inhibited	-3.001	3.06E-02
miR-217-5p		mature microRNA	Inhibited	-4.062	3.24E-02
miR-31-5p		mature microRNA	Inhibited	-2.279	3.42E-02
miR-4471		mature microRNA	Inhibited	-2.274	3.50E-02
miR-216b-5p		mature microRNA	Inhibited	-3.043	3.54E-02
miR-298		mature microRNA	Inhibited	-2.078	3.60E-02
let-7i-5p		mature microRNA	Inhibited	-5.229	3.76E-02
miR-499-5p		mature microRNA	Inhibited	-4.076	3.87E-02
miR-548m		mature microRNA	Inhibited	-2.408	3.89E-02
miR-340-5p		mature microRNA	Inhibited	-7.006	4.19E-02
miR-221-3p		mature microRNA	Inhibited	-2.564	4.53E-02
miR-627		mature microRNA	Inhibited	-2.423	4.82E-02
COBRA1		other	Activated	2	4.50E-02
LGALS1	1.491	other	Activated	2.2	3.30E-01
CUL4B		other	Inhibited	-2.646	1.43E-03
MYCN		ranscription regulato	Activated	8.206	1.80E-21
MYC		ranscription regulato	Activated	4.69	9.63E-05
NRF1		ranscription regulato	Activated	3.212	1.00E-04
NFE2L2		ranscription regulato	Activated	4.872	4.14E-02
SKIL		ranscription regulato	Activated	2.433	2.11E-01
PPARGC1A		ranscription regulato	Activated	2.276	4.72E-01
CBFB	2.088	ranscription regulato	Activated	2	1.00E+00
TFAP2A	-2.202	ranscription regulato	Inhibited	-2.343	1.10E-01
IRF4		ranscription regulato	Inhibited	-2.673	1.54E-01
FOXC2		ranscription regulato	Inhibited	-2.236	4.53E-01
SMAD4		ranscription regulato	Inhibited	-2.053	1.00E+00
PAX6		ranscription regulato	Inhibited	-2.168	1.00E+00
NANOG		ranscription regulato	Inhibited	-2.236	1.00E+00
ERG		ranscription regulato	Inhibited	-2.333	1.00E+00
SMAD2		ranscription regulato	Inhibited	-2.429	1.00E+00
EIF4E	1.493	translation regulator	Activated	3.951	7.25E-04
IGF1R		ansmembrane recept	Activated	2.481	1.48E-03
ITGB3		ansmembrane recept	Activated	2	1.00E+00
CD28	1.834	ansmembrane recept	Inhibited	-3.523	3.91E-01

Supplementary Table 2. Pathway enrichment analysis of differential expressed genes in the hippocampus, prefrontal cortex and striatum of the CGG-KI mouse model vs. control mouse

Cluster	Physiological System Development & Function	Disease&Disorders	Molecular & Cellular Function	Canonical Pathways
STR	**Nervous System** Tissue morphology Tissue Development Organ morphology Embryonic Development	**Neurological Disease** Organismal Injury & Abnormalities **Hereditary Disorder** Skeletal &Muscle Disorders Infectious Disease	Cell-to-Cell Signalling & Interacion Cell Morphology Molecular Transport DNA Replication, Recombination & Repair Cellular Assembly and Organization	DNA Double-Strand Break Repair by Non-Homologous End Joining **TR/RXR Activation**
HC	**Nervous System** Tissue Development Organ morphology Tissue morphology **ehavior**	Cardiovascular Disease **Neurological Disease** Infectious Disease Developmental Disorders Organismal Injury & Abnormalities	Cell-to-Cell Signalling & Interacion Cell Morphology Cellular function & Maintenance Cellular Development Cellular assembly & Organization	DNA Double-Strand Break Repair by Non-Homologous End Joining **TR/RXR Activation**
PFCx	**Nervous System** Organ morphology Embryonic Development Organ Development Organismal Development	Cardiovascular Disease Inflmatory Response Infectious Disease **Neurological Disease** **Cancer**	Cell-to-Cell Signalling & Interacion Cellular function & Maintenance Cellular assembly & Organization Cellular Movement **Cell Death & Survival**	Epoxysqualene Biosynthesis Superpathway of Cholesterol Biosynthesis Ketogenesis ALS Pathway

List of the significantly enriched canonical pathways, molecular and cellular functions, diseases and physiological system development and functions altered for each of the eight gene clusters according to IPA.

Supplementary Table 3. Predicted upstream regulators and its activation state in hippocampus samples of the CGG-Ki mouse modelsamples, according to IPA

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
ADAM15		peptidase			3.24E-02
AGPAT9		enzyme			4.81E-02
Alpha catenin		group		0.218	1.12E-02
APOA1		transporter			2.34E-02
AQP7		transporter			4.81E-02
ARNT		transcription regulator			2.28E-02
ATG9B		other			8.19E-03
ATR		kinase			4.81E-02
B4GALT1		enzyme			4.03E-02
BAG6		enzyme			3.24E-02
BDKRB2		G-protein coupled receptor			4.89E-03
BIRC5		other			4.05E-02
C5		cytokine			3.07E-02
CAV1		other			4.31E-02
CCL13		cytokine			6.68E-03
CD2AP		other			2.44E-02
CD40LG		cytokine		0.025	1.00E-02
CDKN3		phosphatase			4.81E-02
CEBPA		transcription regulator		0.647	3.60E-03
CELA1		peptidase			1.63E-02
Cg		complex		-0.97	2.52E-02
CGGBP1		other			1.63E-02
CIITA		transcription regulator			4.00E-04
CMA1		peptidase			3.24E-02
COL14A1		other			8.19E-03
COL4A3		other			1.50E-02
Collagen Alpha1		group			4.81E-02
Collagen type I		complex		-1.715	1.10E-04
COMMD1		transporter			1.15E-02
COMMD3-BMI1		transcription regulator			1.04E-02
CRP		other			2.19E-02
CSF2RA		transmembrane receptor			4.81E-02
CSF3R		transmembrane receptor			2.31E-03
CTSB		peptidase			4.89E-03
CXCL5		cytokine			2.44E-02
CXCR3		G-protein coupled receptor			4.03E-02
CYR61		other			2.65E-02
DDR2		kinase			3.24E-02
DEPDC1		transcription regulator			8.19E-03
DKK1		growth factor			2.19E-02
DVL1		other			4.03E-02
DVL2		other			4.81E-02
EBF1		transcription regulator			1.26E-02
EBF3		other			2.44E-02
EDNRA		transmembrane receptor			8.43E-05
EGFR		kinase		-1.982	9.52E-02
EGLN2		enzyme			4.03E-02
EGLN3		enzyme			4.81E-02
EGR2		transcription regulator		-1	1.52E-02
EN2		transcription regulator			4.81E-02
Eotaxin		group			2.44E-02
EPAS1		transcription regulator		-0.371	1.78E-03
EPM2A		phosphatase			2.44E-02
EPO		cytokine		-1.729	3.70E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
ERAP1		peptidase			6.65E-05
ERBB2		kinase			4.47E-02
ERBB3		kinase		-0.218	2.33E-03
ESR1		igand-dependent nuclear receptc		-1.98	6.25E-02
F2		peptidase		-0.223	3.36E-02
F2R		G-protein coupled receptor			4.43E-02
FCGR1A	1.199	transmembrane receptor			4.03E-02
FN1		enzyme			1.75E-02
FOXC2		transcription regulator			3.50E-02
FOXJ1		transcription regulator			4.81E-02
FOXO1		transcription regulator		0.762	2.26E-02
FOXO3		transcription regulator			1.20E-03
FSH		complex		-1.067	1.78E-02
FZD4		G-protein coupled receptor			8.19E-03
Gata		group			4.03E-02
GATA2		transcription regulator			1.14E-02
GATA6		transcription regulator			4.83E-02
Gcn5l		group			4.81E-02
GDF5		growth factor			2.44E-02
GPR34		G-protein coupled receptor			1.63E-02
GPR98		G-protein coupled receptor			8.19E-03
GRB10		other			2.44E-02
GRM1		G-protein coupled receptor			3.24E-02
GSN		other			2.44E-02
GTF2IRD1		transcription regulator			2.44E-02
HAMP		other			3.24E-02
HDGF	-1.219	growth factor			2.44E-02
HEPH		transporter			4.03E-02
HFE2		other			2.44E-02
HIF1A		transcription regulator		0	1.23E-02
HISTONE		group			7.36E-04
HMOX1		enzyme			5.21E-03
HOXD10		transcription regulator		0	1.38E-03
HOXD3		transcription regulator			4.03E-02
HPGDS		enzyme			3.24E-02
HPX		transporter			1.63E-02
HRAS		enzyme			1.78E-02
HRG		other			1.50E-02
HSD11B2		enzyme			2.44E-02
HTR3A		ion channel			1.63E-02
HTT		transcription regulator			1.49E-03
IFNG		cytokine		-0.603	7.19E-04
IGF2		growth factor			2.87E-02
IKBKB		kinase		-0.323	8.16E-03
IL10R		group			1.63E-02
IL13		cytokine		-0.928	1.13E-02
IL1B		cytokine		0.716	4.92E-03
IL22RA2		transmembrane receptor			4.03E-02
IL27		cytokine	Inhibited	-2	3.83E-03
IL6		cytokine		1.216	1.53E-02
IL8		cytokine			4.63E-02
INHA		growth factor			4.65E-02
INPPL1	-1.323	phosphatase			4.03E-02
INSR		kinase			3.69E-02
Interferon alpha		group		0.246	2.48E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation Stat	Z-score	p-value overlap
ITGB1		transmembrane receptor			1.14E-02
ITGB3		transmembrane receptor			3.50E-02
ITGB5		other			3.24E-02
KCNJ9		ion channel			3.24E-02
KLF2		transcription regulator			1.61E-02
KNG1		other			8.42E-04
KRAS		enzyme		0	7.81E-04
KSR1		kinase			3.24E-02
LEPR		transmembrane receptor			4.59E-03
LHX2		transcription regulator			4.89E-03
LHX6		transcription regulator			2.44E-02
LIPE		enzyme			1.35E-02
LNPEP		peptidase			1.63E-02
MAP4K4		kinase			4.03E-02
Mapk		group			3.16E-02
MAPK14		kinase	Inhibited	-2	2.05E-02
mGLUR Group I		group			4.03E-02
miR-133-3p		mature microRNA		0	3.89E-02
miR-137-3p		mature microRNA		1.767	1.00E+00
miR-138-5p		mature microRNA		1.98	1.00E+00
miR-18-5p		mature microRNA		0.109	3.30E-02
miR-203b-3p		mature microRNA			2.65E-02
mir-205		microRNA			3.24E-02
miR-205-5p		mature microRNA		1.966	3.73E-01
mir-214		microRNA			3.24E-02
miR-214-3p		mature microRNA		1.562	3.29E-02
miR-22-3p		mature microRNA		-1.982	1.00E+00
miR-221-3p		mature microRNA		0.6	9.63E-03
miR-28-3p		mature microRNA			1.04E-02
miR-297-5p		mature microRNA			2.19E-02
miR-3090-3p		mature microRNA		-1.982	1.66E-01
miR-3117-5p		mature microRNA			4.05E-02
miR-3160-5p		mature microRNA			2.90E-03
miR-320b		mature microRNA	Activated	2.007	4.08E-01
miR-338-3p		mature microRNA		0.239	2.50E-02
miR-3473b		mature microRNA	Inhibited	-2	2.09E-01
miR-3589		mature microRNA			1.50E-02
miR-3664-3p		mature microRNA			3.45E-02
miR-3714		mature microRNA			4.31E-02
miR-4276		mature microRNA			1.43E-02
miR-4310		mature microRNA			3.68E-02
miR-4421		mature microRNA			3.56E-02
miR-4504		mature microRNA			6.68E-03
miR-4517		mature microRNA			3.95E-04
miR-4530		mature microRNA		-1.981	7.74E-02
miR-4640-3p		mature microRNA		-0.218	4.58E-02
miR-4655-5p		mature microRNA			4.05E-02
miR-4704-5p		mature microRNA			3.15E-02
miR-4747-3p		mature microRNA			9.81E-03
miR-4758-3p		mature microRNA		1.01	2.47E-02
miR-4761-3p		mature microRNA			2.97E-02
miR-4786-5p		mature microRNA			3.66E-02
miR-591		mature microRNA			1.97E-02
miR-623		mature microRNA			1.83E-02
miR-638		mature microRNA			2.04E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
miR-802		mature microRNA			3.86E-02
miR-95		mature microRNA			2.19E-02
MKX		other			2.44E-02
MLXIP		other			1.63E-02
MMP13		peptidase			4.81E-02
MTF1		transcription regulator			2.44E-02
MTF2		transcription regulator			4.03E-02
MTHFR		enzyme			8.19E-03
MYBBP1A		transcription regulator			3.24E-02
MYC		transcription regulator		-0.268	1.51E-02
MYOF		other			8.19E-03
N-cor		group			3.32E-02
NACA		other			8.19E-03
NFkB (complex)		complex		-0.374	5.71E-03
NKIRAS1		enzyme			1.63E-02
NKX2-1		transcription regulator			8.94E-03
NKX2-3		transcription regulator		-0.514	1.49E-02
NLRC5		other			1.36E-03
NLRP12		other			1.38E-02
NOD2		other			2.04E-02
NONO		other			1.63E-02
NOS2		enzyme		-0.152	2.74E-02
NOTCH1		transcription regulator			2.26E-02
NR3C2		igand-dependent nuclear receptc			2.79E-02
NSF		transporter			1.63E-02
Nuclear factor 1		group			1.38E-02
OLR1		transmembrane receptor			2.04E-02
OSM		cytokine		0.059	3.13E-02
PACS2		other			8.19E-03
PDGFC		growth factor			8.35E-03
PDX1		transcription regulator			3.63E-02
PI3K (complex)		complex		-0.784	1.49E-03
PIK3CG		kinase			2.65E-02
PITX2		transcription regulator			4.63E-02
PLAUR		transmembrane receptor			1.38E-02
POR		enzyme		0.447	1.73E-03
PPARA		igand-dependent nuclear receptc		1.969	1.46E-01
PPARD		igand-dependent nuclear receptc		0.152	2.93E-02
PPP1R14B		phosphatase			4.81E-02
PPRC1		other			4.05E-02
PRKAG3		kinase			8.48E-03
PRKD2		kinase			1.63E-02
PROC		peptidase			4.63E-02
PSEN1		peptidase		1.987	1.70E-02
PSMB9		peptidase			3.24E-02
RBCK1		transcription regulator			3.24E-02
RELA		transcription regulator		0.832	4.95E-02
REST		transcription regulator			3.16E-02
RETN		other			2.81E-02
RFX5		transcription regulator			1.83E-05
RFXANK		transcription regulator			3.24E-02
RFXAP		transcription regulator			4.03E-02
RGS4		other			4.81E-02
RHOA		enzyme			1.03E-02
Rock		group			2.19E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation Stat	Z-score	p-value overlap
RTKN		other			4.03E-02
SDHA		enzyme			2.44E-02
SEMA6A		transmembrane receptor			8.19E-03
SERPINA12		other			4.03E-02
SERPINH1		other			2.44E-02
SH2B3		other			3.24E-02
SIN3A		transcription regulator			2.98E-02
SLC2A4RG		transcription regulator			8.19E-03
SMAD2		transcription regulator			9.32E-03
SNED1		other			4.81E-02
SNORD21		other			8.19E-03
SOX2		transcription regulator			2.28E-02
SP1		transcription regulator			1.89E-02
SP6		transcription regulator			1.63E-02
SPARC		other			5.68E-03
STAT3		transcription regulator		-1.342	5.33E-03
SWAP70		other			4.81E-02
T		transcription regulator			8.35E-03
TAP1		transporter			2.44E-02
TAPBP		transporter			1.63E-02
TGFB1		growth factor		-0.473	3.67E-02
TGFBR2		kinase		-1.067	4.01E-02
THPO		cytokine			2.19E-02
TNF		cytokine		-0.265	3.52E-02
TNFRSF10A		transmembrane receptor			4.03E-02
TNFSF10		cytokine			7.50E-03
TP53		transcription regulator		0.5	1.77E-04
TPH2		enzyme			2.44E-02
TRAM2		other			8.19E-03
Tropomyosin		group			1.63E-02
TSH		complex			5.56E-03
TXNIP		other			2.65E-02
UTS2		other			1.63E-02
UXT		other			3.24E-02
VAMP7		transporter			2.44E-02
VHL		other			2.49E-02
VTN		other			1.50E-02
WNT3A		other		-0.692	7.00E-03
WWOX		enzyme			5.68E-03
XPA		other			4.03E-02
XPC		other			4.03E-02
ZNF224		other			1.63E-02
ZNF384		transcription regulator			4.81E-02

Supplementary Table 4. Predicted upstream regulators and its activation state in prefrontal cortex samples of the CGG-Ki mouse modelsamples, according to IPA

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
ACOX1		enzyme		-0.447	1.98E-02
ACTN4		other			2.33E-02
ADCY2		enzyme			1.17E-02
ADRA2B		G-protein coupled receptor			4.61E-02
AGT		growth factor	Activated	2.56	1.95E-01
ALAS2		enzyme			2.33E-02
ANK2		other			3.48E-02
APH-1		group			4.61E-02
ARAF		kinase			2.33E-02
ARHGEF2		other			1.17E-02
ATE1		enzyme			3.48E-02
ATF2		transcription regulator			4.58E-02
ATP7B		transporter			9.24E-03
AXIN1		other			1.66E-02
BAG1		other			2.77E-03
BARX1		transcription regulator			1.17E-02
BCL11A		transcription regulator			2.33E-02
BDNF		growth factor			2.38E-02
BNIP3L		other			2.31E-02
BRCA1		transcription regulator			3.42E-02
BTRC		enzyme			1.48E-02
CALCOCO1		transcription regulator			1.17E-02
calpain		complex			1.48E-02
CAPNS1		peptidase			1.99E-03
caspase		group			2.71E-02
CCL17		cytokine			1.17E-02
CCND1	2.032	other			3.71E-02
CCND2		other			4.61E-02
CCNDBP1		other			1.17E-02
CDC25B		phosphatase			4.61E-02
CDH11		other			4.61E-02
CDH16		enzyme			3.48E-02
CDK20		kinase			1.17E-02
CEL		enzyme			4.61E-02
CELA1		peptidase			2.33E-02
CFTR		ion channel			4.92E-02
CGGBP1		other			2.33E-02
CHD4		enzyme			8.37E-03
CINP		other			4.61E-02
CITED2	-1.868	transcription regulator			2.06E-02
Collagen type VI		complex			4.61E-02
COMMD3-BMI1	-1.262	transcription regulator			2.71E-02
Creb		group		0.75	2.43E-02
CREB1		transcription regulator			4.99E-02
CSNK2B		kinase			4.61E-02
CST5		other			1.17E-02
CTAG1B (includes othe		other			3.48E-02
Ctbp		group			1.85E-02
CTNNAL1		other			1.17E-02
CTNNB1		transcription regulator		0.394	1.08E-02
CTNNBIP1		other			4.61E-02
CYP2C9		enzyme			2.33E-02
DIO3		enzyme			4.81E-02
DLX4		transcription regulator			2.33E-02
DNAJB4		other			4.61E-02
DNMT3B		enzyme			3.94E-02
DTNA		other			2.33E-02
E2F4		transcription regulator			1.28E-02
ECT2		other			2.33E-02
EGF		growth factor		0.638	4.86E-02
EGFR		kinase		-0.334	6.71E-03
EIF2B1		translation regulator			3.48E-02
EIF4H		translation regulator			1.17E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
ELAVL3		other			3.48E-02
EMD		other			3.48E-02
EPAS1		transcription regulator		0	1.28E-02
ERAP1		peptidase			4.61E-02
ERBB4		kinase		0.816	3.81E-04
ERK1/2		group		1.96	2.21E-01
FAM3B		cytokine			2.49E-02
FAM57A		other			4.61E-02
FBXO31		other			1.17E-02
FBXO4		enzyme			2.33E-02
Fc gamma receptor		group			4.81E-02
Fgf		group			4.81E-02
FGF2		growth factor		1.937	4.69E-01
FGF4		growth factor			1.66E-02
FHL2		transcription regulator			2.71E-02
FOXO3		transcription regulator		0.246	7.96E-03
FOXO4		transcription regulator		0.109	5.76E-06
FZD4		G-protein coupled receptor			1.17E-02
GPC3		other			2.33E-02
GRM1		G-protein coupled receptor			4.61E-02
GTPBP4		enzyme			1.17E-02
Hedgehog		group			2.09E-02
HISTONE		group			3.19E-02
HLTF		transcription regulator			2.33E-02
HMGA1		transcription regulator			3.97E-02
HOXC6		transcription regulator			9.94E-03
HTT		transcription regulator			5.79E-03
ID2		transcription regulator			3.97E-02
IFNB1		cytokine		-1.98	8.85E-02
IFNG		cytokine		0.625	2.96E-02
IGF1		growth factor		0.511	4.15E-02
IL13		cytokine		0	4.93E-02
IL24		cytokine			3.97E-02
ILKAP		phosphatase			2.33E-02
INSIG1		other		-1.982	3.02E-02
Integrin alpha 5 beta 1		complex			1.17E-02
Interferon alpha		group		-1.331	3.48E-02
IQGAP2		other			4.61E-02
IRF2		transcription regulator			1.39E-02
JARID2		transcription regulator			4.61E-02
KCNJ12		ion channel			1.17E-02
KCNJ6		ion channel			2.33E-02
KHDRBS1		transcription regulator			4.61E-02
KLF7		transcription regulator			4.61E-02
KRAS		enzyme		0.816	3.07E-02
LHX2		transcription regulator			1.48E-02
MAFK		transcription regulator			1.30E-02
MAPK12		kinase			1.48E-02
MAPT		other			4.82E-02
MCF2		other			4.61E-02
MCF2L		other			2.33E-02
MED15		transcription regulator			1.17E-02
MEFV		other			2.33E-02
miR-105-5p		mature microRNA			3.97E-02
miR-1207-5p		mature microRNA	Inhibited	-2.561	7.63E-02
miR-1237-5p		mature microRNA		-0.417	2.40E-02
miR-124-3p		mature microRNA		0.693	4.57E-03
miR-1291		mature microRNA		-1.634	2.10E-02
miR-1307-3p		mature microRNA			2.95E-02
miR-130-3p		mature microRNA	Activated	2.022	1.53E-03
miR-132-3p		mature microRNA		0.685	3.83E-02
miR-133-3p		mature microRNA		-0.309	4.05E-02
miR-135-5p		mature microRNA		-0.701	7.64E-03

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
ITGB1		transmembrane receptor			1.14E-02
ITGB3		transmembrane receptor			3.50E-02
ITGB5		other			3.24E-02
KCNJ9		ion channel			3.24E-02
KLF2		transcription regulator			1.61E-02
KNG1		other			8.42E-04
KRAS		enzyme		0	7.81E-04
KSR1		kinase			3.24E-02
LEPR		transmembrane receptor			4.59E-03
LHX2		transcription regulator			4.89E-03
LHX6		transcription regulator			2.44E-02
LIPE		enzyme			1.35E-02
LNPEP		peptidase			1.63E-02
MAP4K4		kinase			4.03E-02
Mapk		group			3.16E-02
MAPK14		kinase	Inhibited	-2	2.05E-02
mGLUR Group I		group			4.03E-02
miR-133-3p		mature microRNA		0	3.89E-02
miR-137-3p		mature microRNA		1.767	1.00E+00
miR-138-5p		mature microRNA		1.98	1.00E+00
miR-18-5p		mature microRNA		0.109	3.30E-02
miR-203b-3p		mature microRNA			2.65E-02
mir-205		microRNA			3.24E-02
miR-205-5p		mature microRNA		1.966	3.73E-01
mir-214		microRNA			3.24E-02
miR-214-3p		mature microRNA		1.562	3.29E-02
miR-22-3p		mature microRNA		-1.982	1.00E+00
miR-221-3p		mature microRNA		0.6	9.63E-03
miR-28-3p		mature microRNA			1.04E-02
miR-297-5p		mature microRNA			2.19E-02
miR-3090-3p		mature microRNA		-1.982	1.66E-01
miR-3117-5p		mature microRNA			4.05E-02
miR-3160-5p		mature microRNA			2.90E-03
miR-320b		mature microRNA	Activated	2.007	4.08E-01
miR-338-3p		mature microRNA		0.239	2.50E-02
miR-3473b		mature microRNA	Inhibited	-2	2.09E-01
miR-3589		mature microRNA			1.50E-02
miR-3664-3p		mature microRNA			3.45E-02
miR-3714		mature microRNA			4.31E-02
miR-4276		mature microRNA			1.43E-02
miR-4310		mature microRNA			3.68E-02
miR-4421		mature microRNA			3.56E-02
miR-4504		mature microRNA			6.68E-03
miR-4517		mature microRNA			3.95E-04
miR-4530		mature microRNA		-1.981	7.74E-02
miR-4640-3p		mature microRNA		-0.218	4.58E-02
miR-4655-5p		mature microRNA			4.05E-02
miR-4704-5p		mature microRNA			3.15E-02
miR-4747-3p		mature microRNA			9.81E-03
miR-4758-3p		mature microRNA		1.01	2.47E-02
miR-4761-3p		mature microRNA			2.97E-02
miR-4786-5p		mature microRNA			3.66E-02
miR-591		mature microRNA			1.97E-02
miR-623		mature microRNA			1.83E-02
miR-638		mature microRNA			2.04E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
miR-137-3p		mature microRNA	Activated	2.917	1.34E-01
miR-143-3p		mature microRNA		0.881	1.61E-02
miR-149-3p		mature microRNA	Inhibited	-2.332	2.75E-01
miR-153		mature microRNA	Activated	2.067	1.02E-02
miR-1827		mature microRNA	Inhibited	-2.173	4.55E-01
miR-183-5p		mature microRNA		1.97	1.47E-01
miR-1909-3p		mature microRNA	Inhibited	-2.151	3.47E-01
miR-1913		mature microRNA	Inhibited	-2.099	6.48E-02
miR-1915-3p		mature microRNA	Inhibited	-2.393	8.70E-02
miR-194-5p		mature microRNA		0.582	1.69E-02
miR-19b-3p		mature microRNA		1.723	1.50E-04
miR-203-3p		mature microRNA	Activated	2.473	3.26E-01
miR-208-3p		mature microRNA		-1.982	1.87E-01
miR-21-5p		mature microRNA		0.702	2.00E-02
miR-218-5p		mature microRNA		0.056	2.27E-03
miR-219-5p		mature microRNA		1.275	2.41E-02
miR-221-3p		mature microRNA		1.683	1.67E-02
miR-23-3p		mature microRNA		-0.133	3.56E-02
miR-296-5p		mature microRNA	Inhibited	-2.661	3.09E-03
miR-3120-5p		mature microRNA		-1.067	3.45E-03
miR-3128		mature microRNA		-0.9	4.79E-02
miR-3162-3p		mature microRNA		1.98	1.38E-01
miR-3173-5p		mature microRNA	Inhibited	-2.2	5.50E-01
miR-3180-3p		mature microRNA	Inhibited	-2.586	1.48E-01
miR-3194-5p		mature microRNA		-1.414	3.84E-02
miR-320b		mature microRNA		1.588	1.46E-02
miR-330-3p		mature microRNA			1.77E-02
miR-3473b		mature microRNA	Inhibited	-2.449	2.15E-01
miR-34-5p		mature microRNA		-0.886	8.09E-03
miR-3605-5p		mature microRNA		-0.13	3.43E-03
miR-3619-3p		mature microRNA		-0.331	4.58E-03
miR-362-5p		mature microRNA			2.95E-02
miR-3620-3p		mature microRNA		-1.982	6.42E-02
miR-3648		mature microRNA		-1.402	8.60E-04
miR-365-3p		mature microRNA		-1.953	5.46E-01
miR-3656		mature microRNA		-1.498	5.48E-03
miR-3676-3p		mature microRNA			4.52E-02
miR-3679-3p		mature microRNA		1.777	1.23E-01
miR-3688-5p		mature microRNA		0.371	4.42E-02
miR-376-3p		mature microRNA		0.696	1.27E-04
miR-381-3p		mature microRNA		0.589	1.75E-02
miR-411-5p		mature microRNA		0.915	3.06E-03
miR-421-3p		mature microRNA	Activated	2.396	4.84E-01
miR-4283		mature microRNA		-1.358	2.95E-03
miR-4292		mature microRNA	Inhibited	-2.804	3.72E-01
miR-433-3p		mature microRNA		-0.351	1.87E-02
miR-448-3p		mature microRNA		0.845	2.46E-03
miR-4530		mature microRNA		-1.977	2.56E-01
miR-455-5p		mature microRNA		1.195	3.48E-02
miR-4639-3p		mature microRNA		-1.177	7.15E-03
miR-4640-3p		mature microRNA		-1.715	2.27E-02
miR-4642		mature microRNA		-0.179	6.78E-03
miR-4667-3p		mature microRNA		-1.762	3.06E-02
miR-4691-5p		mature microRNA		-1.588	7.57E-03
miR-4701-5p		mature microRNA		-0.371	4.16E-02
miR-4715-3p		mature microRNA	Inhibited	-2.213	9.88E-02
miR-4732-3p		mature microRNA	Inhibited	-2.219	2.35E-02
miR-4790-5p		mature microRNA			1.48E-02
miR-485-5p		mature microRNA	Inhibited	-2.178	3.61E-01
miR-495-3p		mature microRNA		0.813	4.14E-02
miR-504		mature microRNA	Inhibited	-2	1.00E+00
miR-539-5p		mature microRNA	Activated	2.473	4.48E-01
miR-554		mature microRNA		-1.969	1.65E-02

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation State	Z-score	p-value overlap
miR-137-3p		mature microRNA	Activated	2.917	1.34E-01
miR-143-3p		mature microRNA		0.881	1.61E-02
miR-149-3p		mature microRNA	Inhibited	-2.332	2.75E-01
miR-153		mature microRNA	Activated	2.067	1.02E-02
miR-1827		mature microRNA	Inhibited	-2.173	4.55E-01
miR-183-5p		mature microRNA		1.97	1.47E-01
miR-1909-3p		mature microRNA	Inhibited	-2.151	3.47E-01
miR-1913		mature microRNA	Inhibited	-2.099	6.48E-02
miR-1915-3p		mature microRNA	Inhibited	-2.393	8.70E-02
miR-194-5p		mature microRNA		0.582	1.69E-02
miR-19b-3p		mature microRNA		1.723	1.50E-04
miR-203-3p		mature microRNA	Activated	2.473	3.26E-01
miR-208-3p		mature microRNA		-1.982	1.87E-01
miR-21-5p		mature microRNA		0.702	2.00E-02
miR-218-5p		mature microRNA		0.056	2.27E-03
miR-219-5p		mature microRNA		1.275	2.41E-02
miR-221-3p		mature microRNA		1.683	1.67E-02
miR-23-3p		mature microRNA		-0.133	3.56E-02
miR-296-5p		mature microRNA	Inhibited	-2.661	3.09E-03
miR-3120-5p		mature microRNA		-1.067	3.45E-03
miR-3128		mature microRNA		-0.9	4.79E-02
miR-3162-3p		mature microRNA		1.98	1.38E-01
miR-3173-5p		mature microRNA	Inhibited	-2.2	5.50E-01
miR-3180-3p		mature microRNA	Inhibited	-2.586	1.48E-01
miR-3194-5p		mature microRNA		-1.414	3.84E-02
miR-320b		mature microRNA		1.588	1.46E-02
miR-330-3p		mature microRNA			1.77E-02
miR-3473b		mature microRNA	Inhibited	-2.449	2.15E-01
miR-34-5p		mature microRNA		-0.886	8.09E-03
miR-3605-5p		mature microRNA		-0.13	3.43E-03
miR-3619-3p		mature microRNA		-0.331	4.58E-03
miR-362-5p		mature microRNA			2.95E-02
miR-3620-3p		mature microRNA		-1.982	6.42E-02
miR-3648		mature microRNA		-1.402	8.60E-04
miR-365-3p		mature microRNA		-1.953	5.46E-01
miR-3656		mature microRNA		-1.498	5.48E-03
miR-3676-3p		mature microRNA			4.52E-02
miR-3679-3p		mature microRNA		1.777	1.23E-01
miR-3688-5p		mature microRNA		0.371	4.42E-02
miR-376-3p		mature microRNA		0.696	1.27E-04
miR-381-3p		mature microRNA		0.589	1.75E-02
miR-411-5p		mature microRNA		0.915	3.06E-03
miR-421-3p		mature microRNA	Activated	2.396	4.84E-01
miR-4283		mature microRNA		-1.358	2.95E-03
miR-4292		mature microRNA	Inhibited	-2.804	3.72E-01
miR-433-3p		mature microRNA		-0.351	1.87E-02
miR-448-3p		mature microRNA		0.845	2.46E-03
miR-4530		mature microRNA		-1.977	2.56E-01
miR-455-5p		mature microRNA		1.195	3.48E-02
miR-4639-3p		mature microRNA		-1.177	7.15E-03
miR-4640-3p		mature microRNA		-1.715	2.27E-02
miR-4642		mature microRNA		-0.179	6.78E-03
miR-4667-3p		mature microRNA		-1.762	3.06E-02
miR-4691-5p		mature microRNA		-1.588	7.57E-03
miR-4701-5p		mature microRNA		-0.371	4.16E-02
miR-4715-3p		mature microRNA	Inhibited	-2.213	9.88E-02
miR-4732-3p		mature microRNA	Inhibited	-2.219	2.35E-02
miR-4790-5p		mature microRNA			1.48E-02
miR-485-5p		mature microRNA	Inhibited	-2.178	3.61E-01
miR-495-3p		mature microRNA		0.813	4.14E-02
miR-504		mature microRNA	Inhibited	-2	1.00E+00
miR-539-5p		mature microRNA	Activated	2.473	4.48E-01
miR-554		mature microRNA		-1.969	1.65E-02

Supplementary Table 5. Genes commonly deregulated in blood samples of FXTAS patients vs. control individuals and in neuronal cells overexpressing the permutated *FMR1* 5'-UTR (79*CGG8x) vs. cells expressing the wild-type *FMR1* 5'-UTR (21*CGG)

GeneName	Systematic Name	Blood Arrays		In vitro cellular system Array		Description
		FC FXTAS vs. Controls	adjPVal FXTAS vs. Controls	FC V21 vs. V79 8x	adjPVal V21 vs. V79 8x	
ANKRD54	NM_138797	**-1.84**	**1.40E-02**	**-1.24**	**4.57E-02**	ankyrin repeat domain 54 (ANKRD54), mRNA
ARMCX1	NM_016608	**1.52**	**3.07E-02**	**1.33**	**7.54E-04**	armadillo repeat containing, X-linked 1 (ARMCX1), mRNA
BEX2	NM_032621	**1.58**	**1.65E-02**	**1.29**	**3.29E-03**	brain expressed X-linked 2 (BEX2), mRNA
BIRC3	NM_001165	**1.82**	**3.11E-02**	**1.43**	**6.05E-05**	baculoviral IAP repeat-containing 3 (BIRC3), transcript variant 1, mRNA
CTSO	NM_001334	**1.51**	**7.65E-03**	**1.44**	**5.06E-04**	cathepsin O (CTSO), mRNA
IL15	NM_172174	**1.61**	**2.24E-02**	**1.54**	**2.39E-04**	interleukin 15 (IL15), transcript variant 1, mRNA
ISG15	NM_005101	**1.57**	**4.73E-02**	**14.31**	**1.09E-15**	ISG15 ubiquitin-like modifier (ISG15), mRNA
LMO2	NM_005574	**1.56**	**2.08E-02**	**1.72**	**1.49E-03**	LIM domain only 2 (rhombotin-like 1) (LMO2), transcript variant 1, mRNA
MKI67	NM_002417	**-1.45**	**3.64E-02**	**-1.34**	**2.55E-02**	antigen identified by monoclonal antibody Ki-67 (MKI67), transcript variant 1, mRNA
PLIN3	NM_005817	**-1.36**	**4.62E-02**	**-1.28**	**4.43E-02**	perilipin 3 (PLIN3), transcript variant 1, mRNA
PNPT1	NM_033109	**1.58**	**1.60E-02**	**1.97**	**8.15E-06**	polyribonucleotide nucleotidyltransferase 1 (PNPT1), mRNA
SSX2IP	NM_001166417	**-1.21**	**1.75E-02**	**-1.30**	**2.66E-02**	synovial sarcoma, X breakpoint 2 interacting protein (SSX2IP), transcript variant 1, mRNA
STAT2	NM_005419	**1.33**	**3.16E-02**	**1.77**	**6.27E-08**	signal transducer and activator of transcription 2, 113kDa (STAT2), mRNA
TNFRSF10B	NM_003842	**1.55**	**3.80E-02**	**1.29**	**1.38E-02**	tumor necrosis factor receptor superfamily, member 10b (TNFRSF10B), transcript variant 1, mRNA
TNFRSF14	NM_003820	**1.37**	**4.48E-02**	**1.81**	**8.00E-05**	tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator) (TNFRSF14), mRNA
ZNF815	NR_023382	**1.40**	**3.54E-02**	**1.23**	**2.50E-02**	zinc finger protein 815 (ZNF815), non-coding RNA

Supplementary Table 6. Genes commonly deregulated in blood samples of AP vs. SP and in 79*CGG vs. 79*CGG8x expressing cells

		Blood Arrays	In vitro cellular system Array		
		FC	FC	adjPVal	
GeneName	Systematic Name	AP vs. SP	V79 vs. V79 8x	V79 vs. V79 8x	Description
APOL6	NM_030641	**-1.81**	**1.56**	**2.74E-06**	apolipoprotein L, 6 (APOL6), mRNA
B2M	NM_004048	**2.05**	**1.70**	**4.04E-02**	beta-2-microglobulin (B2M), mRNA
CASP1	NM_033292	**1.91**	**1.68**	**9.16E-10**	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase) (CASP1), transcript variant alpha, mRNA
CMPK2	NM_207315	**2.00**	**7.91**	**5.14E-13**	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial (CMPK2), nuclear gene encoding mitochondrial protein, mRNA
CXCL10	NM_001565	**2.28**	**2.29**	**9.44E-10**	chemokine (C-X-C motif) ligand 10 (CXCL10), mRNA
DDX60	NM_017631	**1.64**	**4.03**	**8.54E-10**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 (DDX60), mRNA
GBP3	NM_018284	**1.69**	**1.58**	**1.86E-03**	guanylate binding protein 3 (GBP3), mRNA
GIMAP2	NM_015660	**1.71**	**1.49**	**1.45E-02**	GTPase, IMAP family member 2 (GIMAP2), mRNA
IF144	NM_006417	**1.64**	**9.85**	**3.61E-15**	interferon-induced protein 44 (IFI44), mRNA
IL15	NM_172175	**1.91**	**1.37**	**1.65E-02**	interleukin 15 (IL15), transcript variant 2, mRNA
INHBA	NM_002192	**1.54**	**1.35**	**2.93E-03**	inhibin, beta A (INHBA), mRNA
LMO2	NM_005574	**1.93**	**1.51**	**4.58E-02**	LIM domain only 2 (rhombotin-like 1) (LMO2), transcript variant 1, mRNA
NMI	NM_004688	**1.54**	**1.79**	**6.36E-06**	N-myc (and STAT) interactor (NMI), mRNA
OAS1	NM_002534	**1.52**	**6.96**	**3.61E-15**	2'-5'-oligoadenylate synthetase 1, 40/46kDa (OAS1), transcript variant 2, mRNA
OAS2	NM_001032731	**1.60**	**1.97**	**1.46E-06**	2'-5'-oligoadenylate synthetase 2, 69/71kDa (OAS2), transcript variant 3, mRNA
PARP10	NM_032789	**1.51**	**1.70**	**6.33E-07**	poly (ADP-ribose) polymerase family, member 10 (PARP10), mRNA
PLSCR1	NM_021105	**1.83**	**4.03**	**9.28E-10**	phospholipid scramblase 1 (PLSCR1), mRNA
RGAG4	NM_001024455	**1.80**	**1.22**	**9.49E-03**	retrotransposon gag domain containing 4 (RGAG4), mRNA
TAP2	NM_000544	**1.77**	**2.58**	**4.39E-06**	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP) (TAP2), transcript variant 1, mRNA
UBA7	NM_003335	**1.67**	**2.43**	**6.61E-06**	ubiquitin-like modifier activating enzyme 7 (UBA7), mRNA
USP18	NM_017414	**1.83**	**5.76**	**4.35E-10**	ubiquitin specific peptidase 18 (USP18), mRNA

Supplementary Table 7. Genes commonly deregulated in 21*CGG vs. 79*CGG8x expressing cells and 79*CGG vs. 79*CGG8x expressing cells

GeneName	Systematic Name	FC	adjPVal	FC	adjPVal	Description
		V21 vs. V79 8x	V21 vs. V79 8x	V79 vs. V79 8x	V79 vs. V79 8x	
ACO1	NM_002197	**-1.29**	**0.08**	**-1.35**	**0.04**	aconitase 1, soluble (ACO1)
ACP5	NM_001611	**1.40**	**0.00**	**1.40**	**0.00**	acid phosphatase 5, tartrate resistant (ACP5), transcript variant 4
ADAR	NM_001111	**1.81**	**0.00**	**1.64**	**0.00**	adenosine deaminase, RNA-specific (ADAR), transcript variant 1
AK074144	AK074144	**1.50**	**0.00**	**1.42**	**0.00**	gb| mRNA for FLJ00217 protein.
ANGPTL4	NM_139314	**1.40**	**0.00**	**1.41**	**0.01**	angiopoietin-like 4 (ANGPTL4), transcript variant 1
APOBEC3F	NM_001006666	**1.39**	**0.00**	**1.36**	**0.00**	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F (APOBEC3F), transcript variant 2
APOL2	NM_145637	**1.77**	**0.00**	**1.77**	**0.00**	apolipoprotein L, 2 (APOL2), transcript variant beta
APOL6	NM_030641	**1.57**	**0.00**	**1.56**	**0.00**	apolipoprotein L, 6 (APOL6)
ARMCX1	NM_016608	**1.33**	**0.00**	**1.23**	**0.04**	armadillo repeat containing, X-linked 1 (ARMCX1)
ATP10A	NM_024490	**2.12**	**0.00**	**2.04**	**0.00**	ATPase, class V, type 10A (ATP10A)
B2M	NM_004048	**1.98**	**0.00**	**1.74**	**0.02**	beta-2-microglobulin (B2M)
BATF2	NM_138456	**6.05**	**0.00**	**5.90**	**0.00**	basic leucine zipper transcription factor, ATF-like 2 (BATF2)
BIRC3	NM_001165	**1.43**	**0.00**	**1.35**	**0.00**	baculoviral IAP repeat containing 3 (BIRC3), transcript variant 1
BST2	ENST00000252593	**10.26**	**0.00**	**8.22**	**0.00**	ens|bone marrow stromal cell antigen 2 [Source:HGNC Symbol;Acc:1119]
C12orf70	NM_001145010	**1.28**	**0.00**	**1.23**	**0.04**	chromosome 12 open reading frame 70 (C12orf70)
C1orf38	NM_001039477	**2.47**	**0.00**	**2.13**	**0.00**	chromosome 1 open reading frame 38 (C1orf38), transcript variant 2
C1S	NM_001734	**2.51**	**0.00**	**1.78**	**0.00**	complement component 1, s subcomponent (C1S), transcript variant 2
C4orf33	NM_173487	**1.26**	**0.11**	**1.36**	**0.02**	chromosome 4 open reading frame 33 (C4orf33), transcript variant 1
C5orf56	NR_045116	**2.57**	**0.00**	**2.23**	**0.00**	chromosome 5 open reading frame 56 (C5orf56), non-coding RNA
C7orf41	NM_152793	**-1.35**	**0.00**	**-1.31**	**0.02**	chromosome 7 open reading frame 41 (C7orf41)
C8orf31	NM_173687	**1.48**	**0.00**	**1.30**	**0.05**	chromosome 8 open reading frame 31 (C8orf31)
CASP1	NM_033292	**1.71**	**0.00**	**1.68**	**0.00**	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase) (CASP1), transcript variant alpha
CASP4	NM_033306	**1.43**	**0.00**	**1.24**	**0.05**	caspase 4, apoptosis-related cysteine peptidase (CASP4), transcript variant gamma
CCL5	NM_002985	**1.84**	**0.00**	**1.68**	**0.00**	chemokine (C-C motif) ligand 5 (CCL5)
CERK	NM_022766	**-1.64**	**0.00**	**-1.67**	**0.00**	ceramide kinase (CERK)
CFB	NM_001710	**1.70**	**0.00**	**1.65**	**0.00**	complement factor B (CFB), mRNA [NM_001710]
CGNL1	NM_032866	**1.45**	**0.00**	**1.26**	**0.01**	cingulin-like 1 (CGNL1), transcript variant 2, mRNA [NM_032866]
CMPK2	NM_207315	**8.23**	**0.00**	**7.91**	**0.00**	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial (CMPK2), nuclear gene encoding mitochondrial protein
CNP	NM_033133	**1.50**	**0.00**	**1.28**	**0.00**	2',3'-cyclic nucleotide 3' phosphodiesterase (CNP)
CTSO	NM_001334	**1.44**	**0.00**	**1.38**	**0.00**	cathepsin O (CTSO)
CXCL10	NM_001565	**2.35**	**0.00**	**2.29**	**0.00**	chemokine (C-X-C motif) ligand 10 (CXCL10)
CXCL11	NM_005409	**1.90**	**0.00**	**2.04**	**0.00**	chemokine (C-X-C motif) ligand 11 (CXCL11)
CXCL14	NM_004887	**1.49**	**0.00**	**1.35**	**0.00**	chemokine (C-X-C motif) ligand 14 (CXCL14)
CYP2J2	NM_000775	**2.00**	**0.00**	**1.72**	**0.00**	cytochrome P450, family 2, subfamily J, polypeptide 2 (CYP2J2)
DDX58	NM_014314	**6.67**	**0.00**	**5.85**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (DDX58)
DDX60	NM_017631	**4.85**	**0.00**	**4.03**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 (DDX60)
DDX60L	NM_001012967	**3.15**	**0.00**	**3.05**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60-like (DDX60L)
DHX58	NM_024119	**3.80**	**0.00**	**3.43**	**0.00**	DEXH (Asp-Glu-X-His) box polypeptide 58 (DHX58)
DTX3L	NM_138287	**2.03**	**0.00**	**2.01**	**0.00**	deltex 3-like (Drosophila) (DTX3L)
EBI3	NM_005755	**1.52**	**0.00**	**1.25**	**0.02**	Epstein-Barr virus induced 3 (EBI3)
EIF2AK2	NM_002759	**2.53**	**0.00**	**2.55**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 1
EIF2AK2	NM_001135652	**2.24**	**0.00**	**2.43**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 3

GeneName	Systematic Name	FC	adjPVal	FC	adjPVal	Description
		V21 vs. V79 8x	V21 vs. V79 8x	V79 vs. V79 8x	V79 vs. V79 8x	
ACO1	NM_002197	**-1.29**	**0.08**	**-1.35**	**0.04**	aconitase 1, soluble (ACO1)
ACP5	NM_001611	**1.40**	**0.00**	**1.40**	**0.00**	acid phosphatase 5, tartrate resistant (ACP5), transcript variant 4
ADAR	NM_001111	**1.81**	**0.00**	**1.64**	**0.00**	adenosine deaminase, RNA-specific (ADAR), transcript variant 1
AK074144	AK074144	**1.50**	**0.00**	**1.42**	**0.00**	gb| mRNA for FLJ00217 protein.
ANGPTL4	NM_139314	**1.40**	**0.00**	**1.41**	**0.01**	angiopoietin-like 4 (ANGPTL4), transcript variant 1
APOBEC3F	NM_001006666	**1.39**	**0.00**	**1.36**	**0.00**	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F (APOBEC3F), transcript variant 2
APOL2	NM_145637	**1.77**	**0.00**	**1.77**	**0.00**	apolipoprotein L, 2 (APOL2), transcript variant beta
APOL6	NM_030641	**1.57**	**0.00**	**1.56**	**0.00**	apolipoprotein L, 6 (APOL6)
ARMCX1	NM_016608	**1.33**	**0.00**	**1.23**	**0.04**	armadillo repeat containing, X-linked 1 (ARMCX1)
ATP10A	NM_024490	**2.12**	**0.00**	**2.04**	**0.00**	ATPase, class V, type 10A (ATP10A)
B2M	NM_004048	**1.98**	**0.00**	**1.74**	**0.02**	beta-2-microglobulin (B2M)
BATF2	NM_138456	**6.05**	**0.00**	**5.90**	**0.00**	basic leucine zipper transcription factor, ATF-like 2 (BATF2)
BIRC3	NM_001165	**1.43**	**0.00**	**1.35**	**0.00**	baculoviral IAP repeat containing 3 (BIRC3), transcript variant 1
BST2	ENST00000252593	**10.26**	**0.00**	**8.22**	**0.00**	ens|bone marrow stromal cell antigen 2 [Source:HGNC Symbol;Acc:1119]
C12orf70	NM_001145010	**1.28**	**0.00**	**1.23**	**0.04**	chromosome 12 open reading frame 70 (C12orf70)
C1orf38	NM_001039477	**2.47**	**0.00**	**2.13**	**0.00**	chromosome 1 open reading frame 38 (C1orf38), transcript variant 2
C1S	NM_001734	**2.51**	**0.00**	**1.78**	**0.00**	complement component 1, s subcomponent (C1S), transcript variant 2
C4orf33	NM_173487	**1.26**	**0.11**	**1.36**	**0.02**	chromosome 4 open reading frame 33 (C4orf33), transcript variant 1
C5orf56	NR_045116	**2.57**	**0.00**	**2.23**	**0.00**	chromosome 5 open reading frame 56 (C5orf56), non-coding RNA
C7orf41	NM_152793	**-1.35**	**0.00**	**-1.31**	**0.02**	chromosome 7 open reading frame 41 (C7orf41)
C8orf31	NM_173687	**1.48**	**0.00**	**1.30**	**0.05**	chromosome 8 open reading frame 31 (C8orf31)
CASP1	NM_033292	**1.71**	**0.00**	**1.68**	**0.00**	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase) (CASP1), transcript variant alpha
CASP4	NM_033306	**1.43**	**0.00**	**1.24**	**0.05**	caspase 4, apoptosis-related cysteine peptidase (CASP4), transcript variant gamma
CCL5	NM_002985	**1.84**	**0.00**	**1.68**	**0.00**	chemokine (C-C motif) ligand 5 (CCL5)
CERK	NM_022766	**-1.64**	**0.00**	**-1.67**	**0.00**	ceramide kinase (CERK)
CFB	NM_001710	**1.70**	**0.00**	**1.65**	**0.00**	complement factor B (CFB), mRNA [NM_001710]
CGNL1	NM_032866	**1.45**	**0.00**	**1.26**	**0.01**	cingulin-like 1 (CGNL1), transcript variant 2, mRNA [NM_032866]
CMPK2	NM_207315	**8.23**	**0.00**	**7.91**	**0.00**	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial (CMPK2), nuclear gene encoding mitochondrial protein
CNP	NM_033133	**1.50**	**0.00**	**1.28**	**0.00**	2',3'-cyclic nucleotide 3' phosphodiesterase (CNP)
CTSO	NM_001334	**1.44**	**0.00**	**1.38**	**0.00**	cathepsin O (CTSO)
CXCL10	NM_001565	**2.35**	**0.00**	**2.29**	**0.00**	chemokine (C-X-C motif) ligand 10 (CXCL10)
CXCL11	NM_005409	**1.90**	**0.00**	**2.04**	**0.00**	chemokine (C-X-C motif) ligand 11 (CXCL11)
CXCL14	NM_004887	**1.49**	**0.00**	**1.35**	**0.00**	chemokine (C-X-C motif) ligand 14 (CXCL14)
CYP2J2	NM_000775	**2.00**	**0.00**	**1.72**	**0.00**	cytochrome P450, family 2, subfamily J, polypeptide 2 (CYP2J2)
DDX58	NM_014314	**6.67**	**0.00**	**5.85**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (DDX58)
DDX60	NM_017631	**4.85**	**0.00**	**4.03**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 (DDX60)
DDX60L	NM_001012967	**3.15**	**0.00**	**3.05**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60-like (DDX60L)
DHX58	NM_024119	**3.80**	**0.00**	**3.43**	**0.00**	DEXH (Asp-Glu-X-His) box polypeptide 58 (DHX58)
DTX3L	NM_138287	**2.03**	**0.00**	**2.01**	**0.00**	deltex 3-like (Drosophila) (DTX3L)
EBI3	NM_005755	**1.52**	**0.00**	**1.25**	**0.02**	Epstein-Barr virus induced 3 (EBI3)
EIF2AK2	NM_002759	**2.53**	**0.00**	**2.55**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 1
EIF2AK2	NM_001135652	**2.24**	**0.00**	**2.43**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 3

GeneName	Systematic Name	FC	adjPVal	FC	adjPVal	Description
		V21 vs. V79 8x	V21 vs. V79 8x	V79 vs. V79 8x	V79 vs. V79 8x	
ACO1	NM_002197	**-1.29**	**0.08**	**-1.35**	**0.04**	aconitase 1, soluble (ACO1)
ACP5	NM_001611	**1.40**	**0.00**	**1.40**	**0.00**	acid phosphatase 5, tartrate resistant (ACP5), transcript variant 4
ADAR	NM_001111	**1.81**	**0.00**	**1.64**	**0.00**	adenosine deaminase, RNA-specific (ADAR), transcript variant 1
AK074144	AK074144	**1.50**	**0.00**	**1.42**	**0.00**	gb| mRNA for FLJ00217 protein.
ANGPTL4	NM_139314	**1.40**	**0.00**	**1.41**	**0.01**	angiopoietin-like 4 (ANGPTL4), transcript variant 1
APOBEC3F	NM_001006666	**1.39**	**0.00**	**1.36**	**0.00**	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F (APOBEC3F), transcript variant 2
APOL2	NM_145637	**1.77**	**0.00**	**1.77**	**0.00**	apolipoprotein L, 2 (APOL2), transcript variant beta
APOL6	NM_030641	**1.57**	**0.00**	**1.56**	**0.00**	apolipoprotein L, 6 (APOL6)
ARMCX1	NM_016608	**1.33**	**0.00**	**1.23**	**0.04**	armadillo repeat containing, X-linked 1 (ARMCX1)
ATP10A	NM_024490	**2.12**	**0.00**	**2.04**	**0.00**	ATPase, class V, type 10A (ATP10A)
B2M	NM_004048	**1.98**	**0.00**	**1.74**	**0.02**	beta-2-microglobulin (B2M)
BATF2	NM_138456	**6.05**	**0.00**	**5.90**	**0.00**	basic leucine zipper transcription factor, ATF-like 2 (BATF2)
BIRC3	NM_001165	**1.43**	**0.00**	**1.35**	**0.00**	baculoviral IAP repeat containing 3 (BIRC3), transcript variant 1
BST2	ENST00000252593	**10.26**	**0.00**	**8.22**	**0.00**	ens|bone marrow stromal cell antigen 2 [Source:HGNC Symbol;Acc:1119]
C12orf70	NM_001145010	**1.28**	**0.00**	**1.23**	**0.04**	chromosome 12 open reading frame 70 (C12orf70)
C1orf38	NM_001039477	**2.47**	**0.00**	**2.13**	**0.00**	chromosome 1 open reading frame 38 (C1orf38), transcript variant 2
C1S	NM_001734	**2.51**	**0.00**	**1.78**	**0.00**	complement component 1, s subcomponent (C1S), transcript variant 2
C4orf33	NM_173487	**1.26**	**0.11**	**1.36**	**0.02**	chromosome 4 open reading frame 33 (C4orf33), transcript variant 1
C5orf56	NR_045116	**2.57**	**0.00**	**2.23**	**0.00**	chromosome 5 open reading frame 56 (C5orf56), non-coding RNA
C7orf41	NM_152793	**-1.35**	**0.00**	**-1.31**	**0.02**	chromosome 7 open reading frame 41 (C7orf41)
C8orf31	NM_173687	**1.48**	**0.00**	**1.30**	**0.05**	chromosome 8 open reading frame 31 (C8orf31)
CASP1	NM_033292	**1.71**	**0.00**	**1.68**	**0.00**	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase) (CASP1), transcript variant alpha
CASP4	NM_033306	**1.43**	**0.00**	**1.24**	**0.05**	caspase 4, apoptosis-related cysteine peptidase (CASP4), transcript variant gamma
CCL5	NM_002985	**1.84**	**0.00**	**1.68**	**0.00**	chemokine (C-C motif) ligand 5 (CCL5)
CERK	NM_022766	**-1.64**	**0.00**	**-1.67**	**0.00**	ceramide kinase (CERK)
CFB	NM_001710	**1.70**	**0.00**	**1.65**	**0.00**	complement factor B (CFB), mRNA [NM_001710]
CGNL1	NM_032866	**1.45**	**0.00**	**1.26**	**0.01**	cingulin-like 1 (CGNL1), transcript variant 2, mRNA [NM_032866]
CMPK2	NM_207315	**8.23**	**0.00**	**7.91**	**0.00**	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial (CMPK2), nuclear gene encoding mitochondrial protein
CNP	NM_033133	**1.50**	**0.00**	**1.28**	**0.00**	2',3'-cyclic nucleotide 3' phosphodiesterase (CNP)
CTSO	NM_001334	**1.44**	**0.00**	**1.38**	**0.00**	cathepsin O (CTSO)
CXCL10	NM_001565	**2.35**	**0.00**	**2.29**	**0.00**	chemokine (C-X-C motif) ligand 10 (CXCL10)
CXCL11	NM_005409	**1.90**	**0.00**	**2.04**	**0.00**	chemokine (C-X-C motif) ligand 11 (CXCL11)
CXCL14	NM_004887	**1.49**	**0.00**	**1.35**	**0.00**	chemokine (C-X-C motif) ligand 14 (CXCL14)
CYP2J2	NM_000775	**2.00**	**0.00**	**1.72**	**0.00**	cytochrome P450, family 2, subfamily J, polypeptide 2 (CYP2J2)
DDX58	NM_014314	**6.67**	**0.00**	**5.85**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (DDX58)
DDX60	NM_017631	**4.85**	**0.00**	**4.03**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 (DDX60)
DDX60L	NM_001012967	**3.15**	**0.00**	**3.05**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60-like (DDX60L)
DHX58	NM_024119	**3.80**	**0.00**	**3.43**	**0.00**	DEXH (Asp-Glu-X-His) box polypeptide 58 (DHX58)
DTX3L	NM_138287	**2.03**	**0.00**	**2.01**	**0.00**	deltex 3-like (Drosophila) (DTX3L)
EBI3	NM_005755	**1.52**	**0.00**	**1.25**	**0.02**	Epstein-Barr virus induced 3 (EBI3)
EIF2AK2	NM_002759	**2.53**	**0.00**	**2.55**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 1
EIF2AK2	NM_001135652	**2.24**	**0.00**	**2.43**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 3

GeneName	Systematic Name	FC	adjPVal	FC	adjPVal	Description
		V21 vs. V79 8x	V21 vs. V79 8x	V79 vs. V79 8x	V79 vs. V79 8x	
ACO1	NM_002197	**-1.29**	**0.08**	**-1.35**	**0.04**	aconitase 1, soluble (ACO1)
ACP5	NM_001611	**1.40**	**0.00**	**1.40**	**0.00**	acid phosphatase 5, tartrate resistant (ACP5), transcript variant 4
ADAR	NM_001111	**1.81**	**0.00**	**1.64**	**0.00**	adenosine deaminase, RNA-specific (ADAR), transcript variant 1
AK074144	AK074144	**1.50**	**0.00**	**1.42**	**0.00**	gb| mRNA for FLJ00217 protein.
ANGPTL4	NM_139314	**1.40**	**0.00**	**1.41**	**0.01**	angiopoietin-like 4 (ANGPTL4), transcript variant 1
APOBEC3F	NM_001006666	**1.39**	**0.00**	**1.36**	**0.00**	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F (APOBEC3F), transcript variant 2
APOL2	NM_145637	**1.77**	**0.00**	**1.77**	**0.00**	apolipoprotein L, 2 (APOL2), transcript variant beta
APOL6	NM_030641	**1.57**	**0.00**	**1.56**	**0.00**	apolipoprotein L, 6 (APOL6)
ARMCX1	NM_016608	**1.33**	**0.00**	**1.23**	**0.04**	armadillo repeat containing, X-linked 1 (ARMCX1)
ATP10A	NM_024490	**2.12**	**0.00**	**2.04**	**0.00**	ATPase, class V, type 10A (ATP10A)
B2M	NM_004048	**1.98**	**0.00**	**1.74**	**0.02**	beta-2-microglobulin (B2M)
BATF2	NM_138456	**6.05**	**0.00**	**5.90**	**0.00**	basic leucine zipper transcription factor, ATF-like 2 (BATF2)
BIRC3	NM_001165	**1.43**	**0.00**	**1.35**	**0.00**	baculoviral IAP repeat containing 3 (BIRC3), transcript variant 1
BST2	ENST00000252593	**10.26**	**0.00**	**8.22**	**0.00**	ens|bone marrow stromal cell antigen 2 [Source:HGNC Symbol;Acc:1119]
C12orf70	NM_001145010	**1.28**	**0.00**	**1.23**	**0.04**	chromosome 12 open reading frame 70 (C12orf70)
C1orf38	NM_001039477	**2.47**	**0.00**	**2.13**	**0.00**	chromosome 1 open reading frame 38 (C1orf38), transcript variant 2
C1S	NM_001734	**2.51**	**0.00**	**1.78**	**0.00**	complement component 1, s subcomponent (C1S), transcript variant 2
C4orf33	NM_173487	**1.26**	**0.11**	**1.36**	**0.02**	chromosome 4 open reading frame 33 (C4orf33), transcript variant 1
C5orf56	NR_045116	**2.57**	**0.00**	**2.23**	**0.00**	chromosome 5 open reading frame 56 (C5orf56), non-coding RNA
C7orf41	NM_152793	**-1.35**	**0.00**	**-1.31**	**0.02**	chromosome 7 open reading frame 41 (C7orf41)
C8orf31	NM_173687	**1.48**	**0.00**	**1.30**	**0.05**	chromosome 8 open reading frame 31 (C8orf31)
CASP1	NM_033292	**1.71**	**0.00**	**1.68**	**0.00**	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase) (CASP1), transcript variant alpha
CASP4	NM_033306	**1.43**	**0.00**	**1.24**	**0.05**	caspase 4, apoptosis-related cysteine peptidase (CASP4), transcript variant gamma
CCL5	NM_002985	**1.84**	**0.00**	**1.68**	**0.00**	chemokine (C-C motif) ligand 5 (CCL5)
CERK	NM_022766	**-1.64**	**0.00**	**-1.67**	**0.00**	ceramide kinase (CERK)
CFB	NM_001710	**1.70**	**0.00**	**1.65**	**0.00**	complement factor B (CFB), mRNA [NM_001710]
CGNL1	NM_032866	**1.45**	**0.00**	**1.26**	**0.01**	cingulin-like 1 (CGNL1), transcript variant 2, mRNA [NM_032866]
CMPK2	NM_207315	**8.23**	**0.00**	**7.91**	**0.00**	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial (CMPK2), nuclear gene encoding mitochondrial protein
CNP	NM_033133	**1.50**	**0.00**	**1.28**	**0.00**	2',3'-cyclic nucleotide 3' phosphodiesterase (CNP)
CTSO	NM_001334	**1.44**	**0.00**	**1.38**	**0.00**	cathepsin O (CTSO)
CXCL10	NM_001565	**2.35**	**0.00**	**2.29**	**0.00**	chemokine (C-X-C motif) ligand 10 (CXCL10)
CXCL11	NM_005409	**1.90**	**0.00**	**2.04**	**0.00**	chemokine (C-X-C motif) ligand 11 (CXCL11)
CXCL14	NM_004887	**1.49**	**0.00**	**1.35**	**0.00**	chemokine (C-X-C motif) ligand 14 (CXCL14)
CYP2J2	NM_000775	**2.00**	**0.00**	**1.72**	**0.00**	cytochrome P450, family 2, subfamily J, polypeptide 2 (CYP2J2)
DDX58	NM_014314	**6.67**	**0.00**	**5.85**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (DDX58)
DDX60	NM_017631	**4.85**	**0.00**	**4.03**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 (DDX60)
DDX60L	NM_001012967	**3.15**	**0.00**	**3.05**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60-like (DDX60L)
DHX58	NM_024119	**3.80**	**0.00**	**3.43**	**0.00**	DEXH (Asp-Glu-X-His) box polypeptide 58 (DHX58)
DTX3L	NM_138287	**2.03**	**0.00**	**2.01**	**0.00**	deltex 3-like (Drosophila) (DTX3L)
EBI3	NM_005755	**1.52**	**0.00**	**1.25**	**0.02**	Epstein-Barr virus induced 3 (EBI3)
EIF2AK2	NM_002759	**2.53**	**0.00**	**2.55**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 1
EIF2AK2	NM_001135652	**2.24**	**0.00**	**2.43**	**0.00**	eukaryotic translation initiation factor 2-alpha kinase 2 (EIF2AK2), transcript variant 3

GeneName	Systematic Name	FC	adjPVal	FC	adjPVal	Description
		V21 vs. V79 8x	V21 vs. V79 8x	V79 vs. V79 8x	V79 vs. V79 8x	
TRIM14	NM_014788	**2.57**	**0.00**	**2.41**	**0.00**	tripartite motif containing 14 (TRIM14), transcript variant 1
TRIM21	NM_003141	**1.74**	**0.00**	**1.70**	**0.00**	tripartite motif containing 21 (TRIM21)
TRIM22	NM_006074	**1.60**	**0.00**	**1.64**	**0.00**	tripartite motif containing 22 (TRIM22), transcript variant 1
TRIM25	NM_005082	**2.76**	**0.00**	**2.60**	**0.00**	tripartite motif containing 25 (TRIM25)
TRIM56	NM_030961	**1.54**	**0.00**	**1.61**	**0.00**	tripartite motif containing 56 (TRIM56)
UBA7	NM_003335	**2.78**	**0.00**	**2.43**	**0.00**	ubiquitin-like modifier activating enzyme 7 (UBA7)
UBE2L6	NM_198183	**2.79**	**0.00**	**2.25**	**0.00**	ubiquitin-conjugating enzyme E2L 6 (UBE2L6), transcript variant 2
UNC93B1	NM_030930	**1.63**	**0.00**	**1.48**	**0.04**	unc-93 homolog B1 (C. elegans) (UNC93B1)
USP18	NM_017414	**7.29**	**0.00**	**5.76**	**0.00**	ubiquitin specific peptidase 18 (USP18)
USP41	ENST00000454608	**3.86**	**0.00**	**3.63**	**0.00**	ens|ubiquitin specific peptidase 41 [Source:HGNC Symbol;Acc:20070]
VSIG10L	NM_001163922	**1.48**	**0.00**	**1.36**	**0.00**	V-set and immunoglobulin domain containing 10 like (VSIG10L)
XAF1	NM_017523	**5.26**	**0.00**	**5.51**	**0.00**	XIAP associated factor 1 (XAF1), transcript variant 1, mRNA [NM_017523]
XLOC_00075	ENST00000440492	**3.69**	**0.00**	**3.53**	**0.00**	linc|BROAD Institute lincRNA (XLOC_000755), lincRNA [TCONS_00000492]
XLOC_00196	ENST00000414795	**1.21**	**0.07**	**1.26**	**0.03**	linc|BROAD Institute lincRNA (XLOC_001966), lincRNA [TCONS_00004118]
XLOC_00247	TCONS_00004564	**1.38**	**0.00**	**1.26**	**0.03**	linc|BROAD Institute lincRNA (XLOC_002473), lincRNA [TCONS_00004564]
XLOC_00705	ENST00000521989	**-1.26**	**0.00**	**-1.25**	**0.01**	linc|BROAD Institute lincRNA (XLOC_007054), lincRNA [TCONS_00014976]
XLOC_l2_00	THC2520127	**1.59**	**0.00**	**1.56**	**0.00**	tc|Q5D1D6_CERAE (Q5D1D6) Guanylate binding protein 1, partial (31%) [THC2520127]
XLOC_l2_01	AL049161	**1.32**	**0.00**	**1.29**	**0.00**	gb|DKFZp434G1019_r1 434 (synonym: htes3) cDNA clone DKFZp434G1019, mRNA sequence [AL049161]
XLOC_l2_01	TCONS_l2_000247	**2.77**	**0.00**	**2.41**	**0.00**	linc|BROAD Institute lincRNA (XLOC_l2_012953), lincRNA [TCONS_l2_00024711]
ZC3HAV1	NM_024625	**2.15**	**0.00**	**1.98**	**0.00**	zinc finger CCCH-type, antiviral 1 (ZC3HAV1), transcript variant 2
ZNFX1	NM_021035	**1.53**	**0.00**	**1.37**	**0.02**	zinc finger, NFX1-type containing 1 (ZNFX1)

Supplementary Table 8. Genes commonly deregulated in 21*CGG vs. 79*CGG expressing cells and 21*CGG vs. 79*CGG8x expressing cells.

		FC	P-Value	FC	adjPVal		
GeneName	Systematic Name	V21 vs. V79	V21 vs. V79	V21 vs. V79 8x	V21 vs. V79 8x	Description
DDX60	NM_017631	**1.20**	**0.08**	**4.85**	**0.00**	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 (DDX60)
IL4I1	NM_172374	**1.20**	**0.03**	**2.44**	**0.00**	interleukin 4 induced 1 (IL4I1), transcript variant 2
TNFRSF10B	NM_003842	**1.20**	**0.00**	**1.29**	**0.01**	tumor necrosis factor receptor superfamily, member 10b (TNFRSF10B), transcript variant 1
NGFR	NM_002507	**1.21**	**0.01**	**1.38**	**0.02**	nerve growth factor receptor (NGFR)
OAF	NM_178507	**1.21**	**0.02**	**1.45**	**0.01**	OAF homolog (Drosophila) (OAF)
OASL	NM_003733	**1.21**	**0.14**	**17.70**	**0.00**	2'-5'-oligoadenylate synthetase-like (OASL), transcript variant 1
GIMAP2	NM_015660	**1.22**	**0.02**	**1.83**	**0.00**	GTPase, IMAP family member 2 (GIMAP2)
HERC5	NM_016323	**1.22**	**0.03**	**3.63**	**0.00**	hect domain and RLD 5 (HERC5)
IFIH1	NM_022168	**1.22**	**0.02**	**6.03**	**0.00**	interferon induced with helicase C domain 1 (IFIH1)
INPP5D	NM_001017915	**1.22**	**0.01**	**1.39**	**0.01**	inositol polyphosphate-5-phosphatase, 145kDa (INPP5D), transcript variant 1
LY96	NM_015364	**1.22**	**0.01**	**1.41**	**0.01**	lymphocyte antigen 96 (LY96), transcript variant 1
MOV10	NM_020963	**1.23**	**0.02**	**1.41**	**0.04**	Mov10, Moloney leukemia virus 10, homolog (mouse) (MOV10), transcript variant 1
TP53INP1	NM_033285	**1.23**	**0.05**	**1.53**	**0.03**	tumor protein p53 inducible nuclear protein 1 (TP53INP1), transcript variant 1
BTN3A1	NM_007048	**1.24**	**0.02**	**1.43**	**0.02**	butyrophilin, subfamily 3, member A1 (BTN3A1), transcript variant 1
PROCR	NM_006404	**1.24**	**0.03**	**1.48**	**0.02**	protein C receptor, endothelial (PROCR)
UBE2L6	NM_198183	**1.24**	**0.06**	**2.79**	**0.00**	ubiquitin-conjugating enzyme E2L 6 (UBE2L6), transcript variant 2
BST2	ENST00000252593	**1.25**	**0.00**	**10.26**	**0.00**	ensjbone marrow stromal cell antigen 2 [Source:HGNC Symbol;Acc:1119]
SCN4B	NM_174934	**1.25**	**0.02**	**1.68**	**0.00**	sodium channel, voltage-gated, type IV, beta (SCN4B), transcript variant 1
HLA-J	NR_024240	**1.26**	**0.10**	**2.85**	**0.00**	major histocompatibility complex, class I, J (pseudogene) (HLA-J), non-coding RNA
ISG15	NM_005101	**1.26**	**0.01**	**14.31**	**0.00**	ISG15 ubiquitin-like modifier (ISG15), mRNA
PARP9	NM_031458	**1.26**	**0.12**	**3.27**	**0.00**	poly (ADP-ribose) polymerase family, member 9 (PARP9), transcript variant 1, mRNA
PDGFRL	NM_006207	**1.26**	**0.01**	**1.67**	**0.00**	platelet-derived growth factor receptor-like (PDGFRL)
C4orf7	NM_152997	**1.27**	**0.00**	**1.56**	**0.00**	chromosome 4 open reading frame 7 (C4orf7)
EBI3	NM_005755	**1.27**	**0.00**	**1.62**	**0.00**	Epstein-Barr virus induced 3 (EBI3
FAM65C	NM_080829	**1.27**	**0.02**	**1.63**	**0.01**	family with sequence similarity 65, member C (FAM65C)
IFIT3	NM_001549	**1.27**	**0.08**	**24.71**	**0.00**	interferon-induced protein with tetratricopeptide repeats 3 (IFIT3), transcript variant 1
IFITM3	NM_021034	**1.27**	**0.00**	**4.80**	**0.00**	interferon induced transmembrane protein 3 (IFITM3)
USP18	NM_017414	**1.27**	**0.06**	**7.29**	**0.00**	ubiquitin specific peptidase 18 (USP18)
IFITM4P	NR_001590	**1.28**	**0.00**	**4.41**	**0.00**	interferon induced transmembrane protein 4 pseudogene (IFITM4P), non-coding RNA
SEMA5B	NM_001031702	**1.28**	**0.00**	**1.36**	**0.03**	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short
TAP1	NM_000593	**1.28**	**0.00**	**5.86**	**0.00**	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP) (TAP1)
VIPR2	NM_003382	**1.30**	**0.01**	**1.58**	**0.01**	vasoactive intestinal peptide receptor 2 (VIPR2)
ICAM1	NM_000201	**1.36**	**0.00**	**1.75**	**0.00**	intercellular adhesion molecule 1 (ICAM1)
IFI44L	NM_006820	**1.36**	**0.08**	**47.23**	**0.00**	interferon-induced protein 44-like (IFI44L)
C1S	NM_001734	**1.41**	**0.00**	**2.51**	**0.00**	complement component 1, s subcomponent (C1S), transcript variant 2
XAF1	NM_017523	**1.41**	**0.09**	**3.20**	**0.00**	XIAP associated factor 1 (XAF1), transcript variant 1
IFIT1	NM_001548	**1.44**	**0.05**	**45.66**	**0.00**	interferon-induced protein with tetratricopeptide repeats 1 (IFIT1), transcript variant 2
IFI6	NM_022873	**1.45**	**0.01**	**12.61**	**0.00**	interferon, alpha-inducible protein 6 (IFI6), transcript variant 3
SERPINA3	NM_001085	**1.48**	**0.00**	**1.82**	**0.00**	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3 (SERPINA3)
IFI27	NM_005532	**1.76**	**0.00**	**37.04**	**0.00**	interferon, alpha-inducible protein 27 (IFI27), transcript variant 2
IFITM1	NM_003641	**2.16**	**0.00**	**56.47**	**0.00**	interferon induced transmembrane protein 1 (9-27) (IFITM1)

4. DISCUSSION

4.1. Evaluation of transcriptome changes in peripheral blood samples from *FMR1* PM carriers.

Population-based studies in newborn for the *FMR1* PM allele have estimated prevalences ranging from 1/209 in females to 1/430 in males (174). Fragile X PM carriers (fXPC) are at risk to develop a myriad of different disorders, including tremor and ataxia (FXTAS), early menopause (POF1), thyroid dysfunction, hypertension, fibromyalgia, and chronic muscle pain (7-10).

To evaluate the constellation of gene expression changes linked to *FMR1* CGG-repeat premutation we have performed a comprehensive gene expression profiling in peripheral blood samples of male FXPC, including FXTAS patients. The pattern of DEG identified eight premutation-associated clusters, each containing genes enriched in key physiological processes, whose disruption is likely relevant in several fXPC phenotypes.

A large number of DEG encode for respiratory chain subunits, which are essential for ATP production. We validated by qPCR the up-regulation of cytochrome c oxidase subunit VIc and subunit VIIb (*COX6C* and *COX7B*). The changes we observed in blood of FXTAS patients was also detected in fibroblasts and in brain samples from FXTAS patients, suggesting a shared disease response in different tissues (97, 99). It has been proposed that the increase in the expression of mitochondrial function related transcripts in FXTAS patients is a compensatory response to overcome the diminished oxidative phosphorylation capacity (97, 99). A compromised oxidative phosphorylation affects ATP production and increases cellular stress by generating mitochondrial reactive oxygen species, which can damage cell components and lead to cell death. In line with this, we detected an up-regulation of the superoxide dismutase 1 (*SOD1)*, involved in neutralizing free superoxide radicals. Importantly, our data revealed increased expression of genes implicated in other stress responses, including spinocerebellar ataxia type 3 protein (*ATXN3*), involved in degradation of misfolded chaperone substrates (48), and the amyloid beta (A4) precursor protein (*APP),* participating in endoplasmic reticulum stress-induced apoptosis (175, 176). Among the DEG we also detected an enrichment in oxidative stress response and in oxidative phosphorylation pathways, which are both largely associated with neurodegenerative processes (177).

Other gene clusters showed enrichment in mTOR signaling pathway and its downstream effectors: eIF2- eIF4-, and p70S6K-signaling regulation. mTOR integrates cellular signals to regulate organismal growth and homeostasis. In the nervous system, mTOR is critical for long-term memory formation, axon regeneration, dendrite arboritzation and spine morphology (101). The relevance of mTOR in cellular and

organismal homeostasis is further supported by its deregulation in aging and neurodegenerative disorders (178). An upstream regulator of mTOR is v-akt murine thymoma viral oncogene homolog 1 (*AKT1*), for which we detected decreased expression levels in blood samples of SP. It has been shown in the knock-down of *AKT1* in the fly model of FXTAS enhances the neurodegenerative phenotype in the eye (100), indicating its importance FXTAS neuropathology.

Blood profiling showed DEG involved in gene expression regulation that are important for normal neuronal homeostasis. *ATXN7* is involved chromatin remodeling and showed decreased expression in SP. Mutations in *ATXN7* leads to spinocerebellar ataxia type 7, a neurodegenerative disorder closely related to FXTAS (179). We also detected decreased expression of *RNF10*, a transcription factor involved in expression regulation of myelin-associated glycoprotein (MAG) (180). Studies with cerebral and cerebellar tissue from FXTAS patients exhibit white matter disease, characterized among other factors by neuron demyelination (65).

We also found decreased expression of *HDAC5* in SP. Interestingly, histone deacetylases (HDACs) activity is critical to suppress the expanded CGG-repeat dependent phenotype in a fly model of FXTAS (114). HDACs activation has been shown to rescue the rough eye phenotype in a FXTAS drosophila model. Therefore decreased expression of *HDAC5* may be involved in FXTAS phenotype. In addition, abnormal levels of genes involved in cell death and survival pathways were also enriched in SP. Decreased expression of *Bcl-X*, a potent inhibitor of cell death, and DNA fragmentation factor subunit alpha (*DFFA*), an inhibitor of DNA fragmentation, together with the upregulation of apoptosis-related cysteine peptidase 3 (*CASP3*),

responsible for apoptosis execution, indicate a higher predisposition in SP towards apoptosis activation.

Finally, we identified *EAP1* as a gene deregulated in male and female blood samples of fXPC. EAP1 is an important component in the hypothalamic control for the initiation of puberty and maintenance of female reproductive cycles in rodents and in non-human primates (181, 182). Female FXPC are at a risk for fragile X-associated premature ovarian failure (POF1) (118), which is characterized by menopause symptomatology. Studies show that 20% of female carriers will present FXPOI before 40 years of age and 30% before 45 years of age (121). Thus, decreased expression of *EAP1* in female fXPC could contribute to the appearance of POF1. In summary, the present study elucidates gene expression profiles in blood samples of SP that, at least partially, overlap with DEG in brain and fibroblast samples of FXTAS patients.

The study of DEG in human blood samples showed a high overlapping deregulation pattern in the BS, MCx and STR of a FXTAS mouse model. The PFCx and the BS displayed an up-regulation of genes involved in the response to cellular stress, including *App*, *Atxn3* and *Sod1*. Moreover, four different brain regions showed decreased expression of the histone modifiers *Atxn7* and *Hdac5*, coinciding both in MCx and STR and *Rnf10* in the MCx, STR and BS. We also determined a decreased expression of *Kcnc3* in four out of six brain areas. Knock out mice for *Kcnc3* display motor deficits such as tremor and severe ataxia (183), similar to what is observed in FXTAS patients. All together, these results suggest that gene expression profiling in blood partially reflects changes in the brain transcriptome that underlie neuropathological aspects in FXTAS.

In peripheral blood of fXPC we identified gene expression profiles that could distinguish between AP and SP. A previous report (28) and our own data suggest that *FMR1* expression levels contribute to the clinical manifestation of FXTAS. The expression profiles in a neuronal cell model indicated that mutant *FMR1* levels are relevant for transcriptome deregulation involving cell death and survival pathways, inflammatory response, and nervous system development and function. We therefore propose that transcriptional changes driven by different levels of *FMR1* PM may contribute to clinical/phenotypic differences between FXTAS patients.

Changes in gene expression can also be the consequence of variations in miRNA profiles, activity or concentration. Alteration in miRNA processing machinery has been reported recently in FXTAS (83), leading to a decreased expression of many miRNAs. In agreement with this, upstream regulators IPA predicts a decreased activity of 31 out of 32 miRNAs. Of these, four are down-regulated in the blood of FXTAS patients (184). One of the interesting candidates is miR-221-3p/miR-221 (Z score = -2.56 and overlap p-value = 4.53e-2), which has an important role in neuronal differentiation and participates in neuronal apoptosis and inflammation processes (185, 186). In our data set we found 49 up-regulated putative miR-221-3p targets, according to Targetscan predictions and increased levels of two validated miR-221-3p targets: cyclin-dependent kinase inhibitor 1B (CDKN1B or $p27^{Kip1}$) and Bcl-2-modifying factor (BMF). Interestingly, miR-221-3p targets the 3'-UTR of *FMR1* mRNA modulating its expression (187). Thus, decreased levels of miR-221-3p might contribute to *FMR1* up-regulation in FXTAS patients.

Increasing evidences suggest that lncRNAs, by perturbing gene expression, play an important role in human disease, including neurodegenerative diseases. We report a significant FXTAS-associated deregulation of lncRNAs in blood samples, which may trigger downstream deregulation of protein coding genes in FXPC. We found that maternally expressed 3 *(MEG3)* and tuarine up-regulated gene 1 (*TUG1*) are deregulated in our data set. Both lncRNA are known to associate with polycomb repressive complex 2 *(PRC2)*, acting as epigenetic enhancers or repressors in cis- or in trans-, respectively (188, 189). Another lncRNA deregulated in our data set is Nuclear Enriched Abundant Transcript 1 *(NEAT1)*, which is essential for the formation and maintenance of paraspeckles (190), an irregular-shaped compartment found in the nucleus's interchromatin space (191) involved in regulation of gene expression (192). Interestingly, *TUG1, MEG3* and *NEAT1* are similarly deregulated in HD (136, 189), which suggests a common involvement CNS dysfunction.

In summary, we have shown here a large number of genes with abnormal expression deregulation in fXPCs. In blood, the main affected functional pathways overlap with previous findings in the brain and primary fibroblasts of FXPCs, suggesting that blood RNA profiling reflects FXTAS pathogenic alterations. Specifically, we have identified several genes whose expression deregulation likely contributes to FXTAS and/or POF1, including *EAP1*, *HDAC5* and *KCNC3*. Moreover, we provide new evidences for a role of expanded 5'-UTR *FMR1* expression levels in transcriptome modulation. Longitudinal studies in fXPC addressing the relationship between the dynamic changes in the expression of selective genes (specially *FMR1*) and the clinical evolution may shed light onto the pathogenic mechanisms and further target molecules with therapeutic potential.

4.2. Evaluation of the involvement of the gene silencing machinery in fXPCs pathologies

In the last ten years, a number of different studies have suggested a toxic gain of function mechanism for the PM alleles of the *FMR1* gene. Up to date, several RNA toxic mechanisms have been proposed to contribute in fXPCs (65, 72, 77, 81–87, 89, 193), but none has explored the generation and toxic activity of small RNA molecules containing repeated CGG (sCGG). According to previous in vitro studies, transcripts with large triplet expansions form stable hairpin structures that can be a substrate for Dicer, a ribonuclease III enzyme involved in the processing of siRNAs and miRNAs, generating 21nt length RNA molecules (92, 173). This event has been reported for several transcripts involved in expanded trinucleotide repeat neurodegenerative disorders, such as HD, DM1 and SCA1 (92, 173). We have recently described in vivo sCAG biogenesis from expanded HTT gene and further demonstrated their neurotoxic activity in HD context (93).

In the case of long-CGG RNA transcripts, a study has described that they form hairpins structures thermodynamically more stable than other CNG repeat transcripts (N = A, U, or G) (91). In deed, several evidences indicate that FMR1 PM alleles form a hairpin that can be cleaved by Dicer (94, 194). In agreement with these findings, we found that increases in sCGG molecules abundance correlated with the presence of FMR1-PM alleles in peripheral blood samples blood, LCL from FXTAS patients and a human neuronal cell line overexpressing the PM 5'-UTR. Moreover, we demonstrate in a human neuronal cell system that sCGG biogenesis from PM-FMR1 5'UTR could be

prevented upon Dicer KD, thus indicating that sCGG formation was strongly dependent on Dicer activity. We cannot exclude the possibility that sCGG molecules could be also originated from bidirectional transcription form the *FMR1* loci. fXPCs present increased expression of the *FMR1* antisense gene (ASFMR1) spanning the CGG repeat region (77); and thus FMR1/ASFMR1 duplexes may contribute to Dicer-dependent sCGG biogenesis. Therefore, the origin of sCGG detected in human samples may be related to DICER activity on hairpin repeated CNG like structures or on double-stranded RNAs resulting from bidirectional transcription (144)

Studies performed in the CGG-KI mouse model revealed increased levels of sCGG in two of the four brain areas evaluated, HC and PFCx. Study of the *Fmr1* expression indicates that there are no differences in its the levels of expression in the different brain areas, thus not accounting for the differences in the quantity of sCGG molecules. *Dicer* expression levels and/or activity could explain why sCGG molecules are more enriched in certain brain areas. *Dicer* mRNA levels are comparable in HC, CRBL and BS of control and CGG-KI mouse model. Besides, *Dicer* mRNA levels in PFCx, area with increased quantity of sCGG molecules, are 50% reduced when compared to the other brain regions. Therefore, increase abundant of sCGG molecules in PFCx and HC is likely related to DICER protein activity rather than to its expression. In agreement, recent studies indicate that an array of proteins can modulate DICER activity by direct interaction to DICER or the miRNA/siRNA precursor (194, 195). The combination of cell-type specific transcriptome together with the specific regulation of DICER activity in each brain region provides a model to understand how the sCGG molecules affect differently distinct cell types. In agreement, it has been previously shown in our laboratory that sCGG

molecules do not affect breast, pancreas or bladder cell lines, but only affect neuronal cell lines (Bánez-Coronel M., unpublished).

Our results indicate that cells expressing a PM-*FMR1* 5'UTR not only overexpress sCGG molecules but trigger a reduced cell viability and the induction of cellular stress responses, which agrees with previous results in neuronal cell lines (78). The deleterious effect of transfection of si-like sCGG molecules and sRNAs isolated from cells expressing the expanded PM-FMR1-5'UTR further suggests that sCGG molecules are, at least partially, responsible for neuronal toxicity induced by PM-*FMR1*. Dicer and AGO2 KD in cells expressing the PM-*FMR1* 5'UTR demonstrated that this toxic effect was largely dependent on these key members of the RISC machinery. Thus, the detrimental changes might be caused by pathogenic down-regulation of targets harboring complementary CGG sites, similarly to what we have recently described in HD (92). However, analysis of the down-regulated transcripts in peripheral blood samples of fXPCs with complementary CGG regions did not show any significant enrichment. Nonetheless, we identified five down-regulated genes with CGG complementary regions with functions related to fXPCs associated pathologies: *EAP1, HDAC5, KCNC3, BCL2L11* and *RNF10*. Although additional experiments should be performed to establish a direct action of sCGG molecules in the down-regulation of such targets, a possibility exists that sCGG silencing activity onto specific genes triggers downstream secondary pathogenic changes in gene expression.

Alternative biological roles for sCGG molecules cannot be yet discarded. Studies have reported that small RNAs, like miRNAs, can bind and modulate the activity of proteins that regulate gene expression, thus triggering histone modification and DNA methylation of promoter

sites (196). In addition, a possibility exists that sCGG directly bind RNA-biding proteins known to have affinity for the long CGG repeat transcripts (81–84). These might provide additional mechanisms for sCGG-induced perturbations in gene expression. In summary, our results support the generation of sCGG RNA molecules from expanded *FMR1* 5'-UTR with toxic activity. The detrimental effects of sCGG molecules trough RISC machinery provides a dynamic model in which the physiological output depends on the quantity of sCGG molecules and the transcriptome in a specific cell type. These aspects could provide an explanation for the different levels of affection in different cell types. Altogether, this study provides a complementary model to study pathogenesis associated with fXPCs and develop new therapeutic approaches.

5. CONCLUSIONS

- Expression profiles from peripheral blood samples and from a neuronal cellular model indicated a strong gene expression deregulation associated to *FMR1* PM alleles.

- Transcriptome alterations in peripheral blood samples and in a neuronal cellular model show enrichment in pathways involved in biological processes relevant in FXTAS, such as mitochondrial dysfunction, cell death and survival, stress responses and inflammatory response.

- Genes deregulated in peripheral blood samples of fXPCs and involved in relevant biological processes in FXTAS are also deregulated in brain samples of FXTAS mouse model, such as *Hdac5*, *Rnf10*, *Eap1*, *Kcnc3*, *Bcl2l11*, *Bcl2l1*, *Dffa*, *Atxn3*, *Atxn7*, *Sod1* and *App*. Thus, FMR1 PM-5'UTR in different systems similarly deregulate biological pathways relevant for cellular homeostasis and cell survival.

- Changes in gene expression profiles from peripheral blood samples and from a neuronal cellular model are influenced by *FMR1* mRNA

expression levels and the number of CGG repeats in the *FMR1* 5'-UTR.

- We have identified *EAP1*, a gene down-regulated in blood samples of female fXPCs, the deregulation of which could underlie POF1 disorders.

- Increased formation of sCGG molecules is associated with PM alleles for the *FMR1* in a DICER dependent manner.

- sCGG molecules are, at least partially, responsible for the neurotoxic effect of PM *FMR1* 5'-UTR, inducing oxidative stress and mitochondrial dysfunction.

- AGO2 protein is responsible to mediate the sCGG neurotoxicity, thus indicating that gene silencing mechanisms may trigger a toxic effect.

6. MATERIAL AND METHODS

6.1. Human Samples

A total of 15 unrelated male individuals were recruited for the expression profile study: 9 fragile X premutation carriers (FXPC) and 5 control males (average age of 73 years) of age with normal *FMR1* alleles (Supplementary Table S1). All FXPC were recruited from FXS families. At the time the samples where collected, a total of 5 fXPC where diagnosed with FXTAS. In the course of this study, two additional fXPC developed FXTAS. A total of 25 female fXPC were recruited, with 12 presenting POF1 (Supplementary Table S4). In addition, 8 control females were included, with 4 presenting POF. The CGG repeat number and clinical/neurological findings of these patients are summarized in tables S1 and S4. All patients provided written informed consent for testing and for the use of their phenotypic and genetic data. Determination of sCGG molecules was performed in 4 LCL from FXTAS patients from the expression profiling samples and 4 control LCL from the HapMap project repository (age between 65-75 years).

6.2. Mouse samples

We used three male CGG-KI mice for the fragile X premutation (range 150–178 CGG repeats) (103) at 70 week of age and four male wild-type (*wt*) mice of the same age range were included in this study. Mice were sacrificed, and brain tissue was removed from the skull and cut sagittal at the midline into two equal pieces before being frozen and preserved in liquid nitrogen. Subsequent dissection of the different brain areas was performed in transgenic and control mouse: pre-frontal cortex, striatum, brainstem, motor cortex and hippocampus.

6.3. Total and small RNA isolation

For each individual, 2.5 mL of peripheral venous blood was collected in 5 mL PAXgene tubes. Whole blood RNA was isolated and purified with the PAXgene Blood RNA Kit according to the manufacturer's instructions. The mouse dissected brain areas were placed immediately in Qiazol solution from Qiagen, followed by RNAextraction with miRNeasy kit (Qiagen) as indicated by the manufacturer. RNA extraction from our neuronal cell model was also performed with miRNeasy (Qiagen) following manufacturer's instructions. Isolation of the small RNA fraction (<100bp) was performed using Nucleospin extract II Kit following manufacturer instructions (Macherey Nagel). The RNA quality and quantity measures were done with a 2100 Bioanalyzer (Agilent Technologies) (Applied Biosystems) and an ND-1000 spectrophotometer (Thermo Scientific), respectively. All RNA samples showed an RNA integration number of seven or more.

6.4. Microarray Hybridization and Analysis

100 ng of total RNA was labelled using LowInputQuick Amp Labeling kit (Agilent) following manufacturer instructions. Briefly: mRNA was reverse transcribed in the presence of T7-oligo-dT primer to produce cDNA. cDNA was then in vitro transcribed with T7 RNA polymerase in the presence of Cy3-CTP to produce labelled cRNA. The labelled cRNA was hybridized to the Agilent SurePrint G3 Human gene expression 8x60K microarray according to the manufacturer's protocol. The arrays were washed, and scanned on an Agilent G2565CA microarray scanner at 100% PMT and 3 μm resolution. Intensity data was extracted using the Feature Extraction software.

Raw data were taken from the Feature Extraction output files and was corrected for background noise using the normexp method (197). To assure comparability across samples we used quantile normalization (198). Differential expression analysis was carried out on non control probes with an empirical Bayes approach on linear models (limma) (199). Results were corrected for multiple testing according to the False Discovery Rate (FDR) method (200). All statistical analyses were performed with the Bioconductor project (http://www.bioconductor.org/) in the R statistical environment (http://cran.r-project.org/) (201). We used SurePrint G3 Human GE 8x60K Microarray for peripheral blood samples of FXPC expression profile, SurePrint G3 Human Gene Expression 8x60K v2 Microarray for our in vitro cellular system and SurePrint G3 Mouse GE 8x60K Microarray for the mouse brain regions analyzed (Agilent).

6.5. Pathway enrichment analysis

We used Ingenuity Pathway Analysis (IPA) online tool (Ingenuity System Inc, www.ingenuity.com) to interpret data in the context of biological processes, pathways, and networks.

6.6. qPCR validations

Gene expression analysis was performed using 0.15 µg total RNA from human peripheral blood and 0.3 µg total RNA from CGG-KI mice model or SH-SY5Y cell line. cDNA synthesis was performed using SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies) following manufacturer's instructions. The cDNA product was diluted to 1/5 with sterile water. Real time PCR reaction (rRT-PCR) was performed using TaqMan gene expression assays, following manufacturer's instructions in an AB 7900HT Fast Real-Time PCR System.

For each experiment, all cases and controls were analysed in the same rRT–PCR experiment, each sample was run in quadruplicates and the cDNA synthesis repeated at least twice. Relative quantification (RQ) as shown in graphs was calculated with the $2^{-\Delta\Delta Ct}$ method (202) using ACTB and MRIP as a reference genes. RQ was calculated to compare all expression values normalized to the reference genes among FXPC and controls samples. These RQ and their statistical significance were obtained from a linear mixed effects model (203) that accounted for the different sources of variation derived from the experimental design (135). TaqMan Expression Assays are found in Supplementary Table 9.

6.7. Generation of *FMR1* 5'-UTR vectors

Lymphocyte cell lines (LCL) from two different FXPC had been established. One LCL is from a female FXPC, which permits the cloning of a normal allele. B-lymphocytes were isolated from peripheral blood samples by gradient centrifugation with Ficoll using Leucoceps tubes (Greiner Bio-One). Peripheral Blood Mononuclear Cells (PBNCs) interphase was collected and washed twice with phosphate buffered solution (PBS), and finally resuspended in 5 mL PBS. PBNCs were incubated with Epstein-Barr virus supernatant at room temperature. After 3 hours incubation 7 mL of incubation medium was added (RPMI 1640 + 20% FBS + 200 ng/mL cyclosporine A). After 5-7 days, 5 mL of medium were replaced for fresh medium. Polymerase chain reaction (PCR) amplification was carried out on genomic DNA (gDNA) isolated from the established LCL. Amplification conditions were as follows: an initial denaturation at 96ºC for 30 s was followed by 25 cycles of 96ºC for 30 s, 64ºC for 30 s and 68ºC for 5 min and a final elongation of 30 min at 68ºC. Each 12,5 μl reaction contained: 3 μl GC-RICH PCR reaction buffer 5x, 3 μl GC-RICH resolution solution 5 M, 0.08 μM each oligonucleotide (FMR1 F and FMR1 R), 150 μM of each dNTP, 200 ng of gDNA and 1 μl GC-RICH Enzyme Mix (Roche). Products were purified through 2% agarose gel and gel extracted using a gel extraction kit (QIAGEN). Primers used were FMR1 F (5′-AGCCCCGCACTTCCACCACCAGCTCCTCCA) and FMR1 R (5′-TTCACTTCC GGTGGAGGGCCGCCTCTGAGC). Gel-purified PCR products were cloned using the pGEM-T Easy Vector Systems (Promega). The recombinant plasmids were transformed into SURE2 supercompetent cells (Stratagene) according to manufacturer

instructions. The recombinant pGEM-T Easy Vector cloned in 5' to 3' orientation was digested with *Sac*II (New England BioLabs) followed by a 3' Polishing with X-Pfu DNA Polymerase (Kyratec) for 30 min at 37ºC to achieve a blunt end. The insert was then realized using *Not*I (New England BioLabs) and gel-purified. The DNA fragment of interest was subcloned into the pCAGIG (Addgene plasmid 11159) (204) and transformed into SURE2 supercompetent cells. Positive plasmids were sequenced using the Big Dye 3.1 Termination Cycle Sequencing Kit and DNA Sequencer (ABI3100) from Applied Biosystems. PCR conditions where optimized with 7% DMSO in the final volume. Sequencing conditions were as follow: an initial denaturation at 94ºC for 1 min was followed by 30 cycles of 94ºC for 30 s, 58ºC for 30 s and 60ºC for 4 min. Each 10 µl reaction contained: 1 µl Big dye, 2 µl Big dye 5x buffer, 0.5 µl of oligonucleotide at a 100nM concentration, 0.7µl DMSO and 300 to 400ng od DNA template; the final volume was adjusted with water. Sequencing was performed using the cloning primers FMR1 F and FMR1 R alternatively.

6.8. Cell culture

SH-SY5Y neuroblastoma cells were grown in the Dulbecco's modified essential medium (DMEM, Invitrogen) supplemented with 10% of inactivated foetal bovine serum (FBS) and 100 units/mL penicillin and 100 mg/mL Streptomycin (GIBCO, Invitrogen). Differentiation protocol for SH-SY5Y consisted of growing media with 10 mM retinoic acid (RA) (SIGMA). After four days exposure, media was removed and replaced by normal growing media plus 80 nM of 12-O-

tetradecanoylphorbol-13-acetate (TPA) (SIGMA) during five additional days (205).

6.9. Transfections

All the transfection experiments were performed using Lipofectamine 2000 (Invitrogen), according to the manufacturer's instructions and at a cell confluence of 60%. For the arrays experiments vectors were transfected at a concentration that permits an equivalent expression of the plasmids (between 0.6 ng/µl and 0.3 ng/µl). Cells were processed 24 hours after transfeccion and total RNA isolated as previously indicated. Subsequently, rRT-PCR determinations of the expression levels of the 5'-UTR transgene were performed, using custom TaqMan assays: EGFP (AGTTCGAGGGCGACACCCTGGTGAA) and FMR1-IRES (GAAGCAGTTCCTCTGGAAGCTTCTT).

Transfection of siRNA were performed with 50 nM per well in a 96 wells multiwall. sCGG-siRNA (5'-CGGCGGCGGCGGCGGCGGCGG-3') and scramble siRNA (5'-CTGTAACACGTGTATACGCCCA-3') where ordered at Dharmacon. LNA modified anti-sRNA were transfected to a final concentration of 60 nM. Anti-sRNAs were purchased from Exiqon: anti sCGG-sRNA (5'-GCCGCCGCCGCCGCCGCC-3') and anti scramble sRNA (5'-GTGTAACACGTCTATACGCCCA-3')

Transfections with sRNA fractions (<100nt) were performed using 35 ng of each sRNA pool per well in 96 wells multiwell plate. Transfections were performed in quintuplicate.

Dicer and Ago2 knockdown experiments were performed by a double transfection procedure consisting in the transfection of Ago2 or Dicer siRNA in the first assay (50 nM), and the co-transfection of the siRNA (75 nM) and 79*CGG or 21*CGG (400 ng), in 6 multiwell plates. Dicer siRNA (5'- GCUCGAAAUCUUACGCAAAUA-3') and Ago2 siRNA (5'-GCACGGAAGUCCAUCUGAA-3') were purchased from Dharmacon. Transfection efficiency in experiments using siRNA or sRNA pools was determined at each experimental condition using siGLO transfection indicator (Dharmacon). Transfection conditions were optimized in order to obtain similar transfection efficiencies (≈90%) in all the cell transfections.

6.10. sCGG amplification

Total RNA was treated with TURBO DNA-free kit following manufacture's instructions (Ambion). In vitro polyadenylation reactions were carried out using 1mg of total RNA using poly (A) polymerase (Ambion) for 1 h at 37ºC in the presence of ATP (1 mM). Samples were then annealed with a polyT-adapter primer (5'-CGAATTCTAGAGCTCGAGGCAGGCGACATGGCTfGGCTAGTT AAGCTTGGTACCGAGCTCGGATCCACTAGTCCTTTTTTTTTTT TTTTTTTTTTTTTAC-3') prior to RT reaction. Anneling was performed at 65ºC for 10 min followed by 10 min at 25ºC and 10 more minutes at 4ºC. Specific primers recognizing the adapter and sCGG allowed the amplification of specific products by PCR. Equal amounts of cDNA were used for the PCR amplification. Five pmol of the forward primer and reverse primer were used in each reaction. Amplification for sCGG molecules was done under the conditions of 15

sec at 95ºC and followed by 55 cycles consisting in 1 min at 70ºC and 2 min at 72ºC. For reference genes amplification was performe under the conditions of 15 sec at 95ºC and followed by 55 cycles consisting in 1 min at 60ºC and 2 min at 72ºC. RNU66 or RNU6B expression was used as a reference small RNA. PCR products were run in a 6% polyacrylamide gel. sCGG expression levels were analyzed by densitometry of the PCR amplified products. Data are presented as the ratio between the normalized expression of sCGG (sCGG/RNU66) or (sCGG/RNU6B). Primers for PCR amplification of the reference genes were: RNU66 forward: 5'-GTAACTGTGGTGATGGAAATGTG-3'; RNU66 reverse: 5'- GACTGTACTAGGATAGAAAGAACC-3'; RNU6B forward: 5'-CGCTTCGGCAGCACATATAC-3'; RNU6B reverse: 5'-TTCACGAATTTGCGTGTCAT-3'.

6.11. sCGG sequencing

CGG PCR products were run on a 15% polyacrylamide gel and visualized by SybrSafe staining (Invitrogene). PCR products were purified and ligated into pGEMT-easy vector. The sequencing reactions of the vectors were carried out using the Big Dye 3.1 Termination Cycle Sequencing Kit and DNA Sequencer (ABI3100) from Applied Biosystems. Sequencing conditions were as follow: an initial denaturation at 94ºC for 1 min was followed by 30 cycles of 94ºC for 30 s, 58ºC for 30 s and 60ºC for 4 min. Each 10 µl reaction contained: 1 µl Big dye, 2 µl Big dye 5x buffer, 0.5 µl of oligonucleotide at a 100nM concentration, 0.7µl DMSO and 300 to 400ng od DNA template; the final volume was adjusted with water. Sequencing was

performed with the oligo corresponding to the polyT adapter after polyadenylation (59-CGAATTCTAGAGCTCGAGGCAGG-39).

6.12. Western blotting

Cells were rinsed with PBS1X, lysed with 0.05% SDS, boiled for 2 min and centrifuged at 10,000g and 4ºC for 10 min. The supernatant was diluted with 6X Laemmli loading buffer and boiled for 2 min prior to loading. Proteins were resolved by 10% SDS–PAGE, and electroblotted onto nitrocellulose membranes using the i-Blot dry transfer system (Invitrogen). Membranes were blocked with blocking solution (either 3% BSA or 10% dry milk, and 0.1% Tween-20, TBS1X) for at least 1 hour at RT. Primary antibodies were incubated overnight at 4ºC then membranes were washed with TBS-0.1% Tween, and incubated for 1 h at RT with secondary antibodies (Abcam) at a dilution of 1:3,000. After washing, membranes were developed with the enhanced chemiluminiscence system (ECL, Amersham Life Sciences). Western blot signal was quantified using ImageJ software. Expression of α-tubulin was used as a loading normalization control. The primary antibodies used anti-GFP (1:2000, Molecular Probes, rabbit), anti-Dicer (1:500, Abcam, mouse), anti-Ago2, (1:500, Abnova, clone 2E12-1C9). Anti-a-Tubulin (1:50000, Sigma, mouse) was used as loading controls.

6.13. Cell toxicity assays

Lactose dehydrogenase (LDH) released from dying cells was determined using the LDH assay (Cytotox 96, Promega) according to the manufacturer's protocol, at different time-points following

transfection. Absorbance was recorded at 490 nm. LDH determinations were performed in quintuplicate.

6.14. Molecular probes for ROS detection and mitochondrial labelling

Cells were plated for transfection in 24 MW plates over 13 mm coverslips, then at 24 and 48 h post-transfection they were incubated with MitoSOX™ Red Mitochondrial Superoxide Indicator at 1 µM final concentration for 20 minutes at 37 ºC. Then cells were rinsed with PBS1X and coverslips were mounted in PBS1x. Images were directly taken with a Zeiss Observer Z1 microscope. For quantification three independent experiments were carried out and five fields were randomly captured for each condition and experiment. Around 400 cells were counted in each experiment. Statistical significance was calculated with two-tailed Student's t-test.

6.15. In silico identification of sCGG targets

The 21-nt long CGG repeat sequence was in silico hybridized to each human RefSeq mRNA to identify candidate high-complementary targets. Specifically, human RefSeq mRNA sequences were downloaded from the UCSC table browser (PMID 14681465). The repeat sequence was hybridized to each of these using RNAhybrid (PMID 16845047) run with default parameters. This analysis was performed three times, reporting the single, two or three best target sites. For the analyses reporting the two or three best target sites,

estimated hybridization energies were summed to a single value for each 3' UTR.

7. ANNEX

7.1. Publications

Work from this thesis has given rise to the following publication in process of submission:

Blood expression profiles of fragile X premutation carriers identify candidate genes involved in neurodegenerative and infertility phenotypes; Mateu-Huertas, E., Rodriguez-Revenga, L., Alvarez-Mora, M.I., Madrigal, I., Willemsen, R., Milà, M., Martí, E. & Estivill, X.

Publication in preparation:

A novel pathogenic mechanism involved in fXPCS associated pathologies.

Other publications:

A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. Bañez-Coronel, M., Porta, S., Kagerbauer, B., Mateu-Huertas, E., Pantano, L., Ferrer, I.,

Guzmán, M., Estivill, X., Martí, E. PLoS Genetics 2012 Feb;8(2):e1002481. Epub 2012 Feb 23.

MicroRNA expression profiling in blood from fragile X-associated tremor/ataxia syndrome patients; Alvarez-Mora, M.I, Rodriguez-Revenga, L., Madrigal, I., Torres-Silva, F., Mateu-Huertas, E., Lizano, E., Friedlander, M., Martí, E., Estivill, X., & Milà, M. Genes Brain Behav. 2013 Aug;12(6):595-603. doi: 10.1111/gbb.12061. Epub 2013 Jul 24.

Evidence of the biogenesis of more than thousand novel human microRNAs; Friedlander, M., Lizano, E., Houben, A.J.S., Bañez-Coronel M., Mateu-Huertas E., Kagerbauer B., Gonzales J., J, Chen K., Martí, E., Estivill, X. (in preparation)

7.2. Communication to scientific meetings

a. Poster presentations

"CAG-repeated small RNAs selectively affects neuronal cell survival". Bañez-Coronel, M., Miñones-Moyano, E., Porta, S., Mateu-Huertas, E., Estivill,X., Martí, E.; at the 6th International Conference on Unstable Microsatellites and Human Disease; January 17-22, 2009, Costa Rica

"A novel mechanism involved in FXTAS neuropathology". Mateu-Huertas, E., Bañez-Coronel, M., Milà, M., Estivill, X., Martí, E.; at the SIROCCO Annual Meeting ;October 11-12, 2010, Heidelberg, Germany

"RNA toxicity in the pathogenesis of Huntington Disease". Bañez-Coronel, M., Porta, S., Kagerbauer, B., Mateu-Huertas, E., Ferrer, I., Guzmán, M., Estivill, X., Martí, E.; at EMBO|EMBL Symposium Non-Coding Genome October 13-16, 2010, Heidelberg, Germany

"mRNA profiling in FMR1 premutation carriers". Mateu-Huertas, E., Bañez-Coronel, M., Friedlander, M., Willemsen, R., Milà, M., Estivill, X., Martí, E.; at Keystone Symposia Protein-RNA Interactions in Biology and Disease (C1); 4-9 March 2012, Santa Fe, NM, USA

"mRNA profiling in FMR1 premutation carriers". Mateu-Huertas, E., Bañez-Coronel, M., Friedlander, M., Willemsen, R., Milà, M., Estivill, X., Martí, E.; at 7th International Conference on Unstable Microsatellites and Human Diseases; 9-14 June 2012, Strasburg, France

8. BIBLIOGRAPHY

1. Spritz RA (1981) Duplication/deletion polymorphism 5' - to the human beta globin gene. *Nucleic acids research* 9:5037–47.

2. Lander E, Linton L, Birren B, Nusbaum C (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

3. Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447:932–40.

4. McMurray CT (2010) Mechanisms of trinucleotide repeat instability during human development. *Nature reviews Genetics* 11:786–99.

5. Budworth H, McMurray CT (2013) Bidirectional transcription of trinucleotide repeats: Roles for excision repair. *DNA repair* 1.

6. La Spada AR, Taylor JP, Spada A La (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature reviews Genetics* 11:247–58.

7. Verkerk AJ et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65:905–14.

8. Eichler EE, Richards S, Gibbs RA, Nelson DL (1993) Fine structure of the human FMR1 gene. *Human molecular genetics* 2:1147–53.

9. Verheij C et al. (1993) Characterization and localization of the FMR-1 gene product associated with fragile X syndrome. *Nature* 363:722–4.

10. Devys D, Lutz Y, Rouyer N, Bellocq JP, Mandel JL (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nature genetics* 4:335–40.

11. Pacey LKK, Doering LC (2007) Developmental expression of FMRP in the astrocyte lineage: implications for fragile X syndrome. *Glia* 55:1601–9.

12. Penagarikano O, Mulle JG, Warren ST (2007) The pathophysiology of fragile x syndrome. *Annual review of genomics and human genetics* 8:109–29.

13. Eberhart DE, Malter HE, Feng Y, Warren ST (1996) The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Human molecular genetics* 5:1083–91.

14. Feng Y et al. (1997) FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. *Molecular cell* 1:109–18.

15. Yu S et al. (1991) Fragile X genotype characterized by an unstable region of DNA. *Science (New York, NY)* 252:1179–81.

16. Kunst CB, Warren ST (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* 77:853–61.

17. Nolin SL et al. (1996) Familial transmission of the FMR1 CGG repeat. *American journal of human genetics* 59:1252–61.

18. Mornet E (1996) Analysis of germline variation at the FMR1 CGG repeat shows variation in the normal-premutated borderline range. *Human Molecular Genetics* 5:821–825.

19. Nolin SL, Glicksman A, Houck GE, Brown WT, Dobkin CS (1994) Mosaicism in fragile X affected males. *American journal of medical genetics* 51:509–12.

20. Filipovic-Sadic S et al. (2010) A novel FMR1 PCR method for the routine detection of low abundance expanded alleles and full mutations in fragile X syndrome. *Clinical chemistry* 56:399–408.

21. Fu YH et al. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 67:1047–58.

22. Peprah E (2012) Fragile X syndrome: the FMR1 CGG repeat distribution among world populations. *Annals of human genetics* 76:178–91.

23. Eichler EE et al. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature genetics* 8:88–94.

24. Dombrowski C et al. (2002) Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Human molecular genetics* 11:371–8.

25. Yrigollen CM et al. (2012) AGG interruptions within the maternal FMR1 gene reduce the risk of offspring with fragile X syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics* 14:729–36.

26. Heitz D, Devys D, Imbert G, Kretz C, Mandel JL (1992) Inheritance of the fragile X syndrome: size of the fragile X premutation is a major determinant of the transition to full mutation. *Journal of medical genetics* 29:794–801.

27. Eichler EE et al. (1996) Haplotype and interspersion analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Human molecular genetics* 5:319–30.

28. Nolin SL et al. (2013) Fragile X AGG analysis provides new risk predictions for 45-69 repeat alleles. *American journal of medical genetics Part A*.

29. Pieretti M et al. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66:817–22.

30. Godler DE, Slater HR, Amor D, Loesch DZ (2010) Methylation analysis of fragile X-related epigenetic elements may provide a suitable newborn screening test for fragile X syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics* 12:595.

31. Kumari D, Usdin K (2001) Interaction of the transcription factors USF1, USF2, and alpha -Pal/Nrf-1 with the FMR1 promoter. Implications for Fragile X mental retardation syndrome. *The Journal of biological chemistry* 276:4357–64.

32. Hagerman PJ (2008) The fragile X prevalence paradox. *Journal of medical genetics* 45:498–9.

33. De Boulle K et al. (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nature genetics* 3:31–5.

34. Collins SC et al. (2010) Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *American journal of medical genetics Part A* 152A:2512–20.

35. Glass IA (1991) X linked mental retardation. *Journal of medical genetics* 28:361–71.

36. De Vries BB et al. (1996) Mental status of females with an FMR1 gene full mutation. *American journal of human genetics* 58:1025–32.

37. Kaufmann WE, Reiss AL (1999) Molecular and cellular genetics of fragile X syndrome. *American journal of medical genetics* 88:11–24.

38. Saul RA *FMR1-Related Disorders*.

39. Tassone F et al. (2000) Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *American journal of human genetics* 66:6–15.

40. Hagerman PJ, Hagerman RJ (2004) Fragile X-associated tremor/ataxia syndrome (FXTAS). *Mental retardation and developmental disabilities research reviews* 10:25–30.

41. Jacquemont S et al. (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *American journal of human genetics* 72:869–78.

42. Jin P et al. (2003) RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron* 39:739–47.

43. Grigsby J et al. (2006) Impairment in the cognitive functioning of men with fragile X-associated tremor/ataxia syndrome (FXTAS). *Journal of the neurological sciences* 248:227–33.

44. Hessl D et al. (2005) Abnormal elevation of FMR1 mRNA is associated with psychological symptoms in individuals with the fragile X premutation. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 139B:115–21.

45. Bacalman S et al. (2006) Psychiatric phenotype of the fragile X-associated tremor/ataxia syndrome (FXTAS) in males: newly described fronto-subcortical dementia. *The Journal of clinical psychiatry* 67:87–94.

46. Jacquemont S et al. (2004) Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA : the journal of the American Medical Association* 291:460–9.

47. Kenneson A, Warren ST (2001) The female and the fragile X reviewed. *Seminars in reproductive medicine* 19:159–65.

48. Cronister a et al. (1991) Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. *American journal of medical genetics* 38:269–74.

49. Oostra BA, Willemsen R (2003) A fragile balance: FMR1 expression levels. *Human molecular genetics* 12 Spec No:R249–57.

50. Farzin F et al. (2006) Autism spectrum disorders and attention-deficit/hyperactivity disorder in boys with the fragile X premutation. *Journal of developmental and behavioral pediatrics : JDBP* 27:S137–44.

51. Bailey DB, Raspa M, Olmsted M, Holiday DB (2008) Co-occurring conditions associated with FMR1 gene variations: findings from a national parent survey. *American journal of medical genetics Part A* 146A:2060–9.

52. Franke P et al. (1996) Fragile-X carrier females: evidence for a distinct psychopathological phenotype? *American journal of medical genetics* 64:334–9.

53. Hagerman P (2013) Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms. *Acta neuropathologica* 126:1–19.

54. Hagerman RJ et al. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* 57:127–30.

55. Leehey MA et al. (2007) Progression of tremor and ataxia in male carriers of the FMR1 premutation. *Movement disorders : official journal of the Movement Disorder Society* 22:203–6.

56. Juncos JL et al. (2012) Olfactory dysfunction in fragile X tremor ataxia syndrome. *Movement disorders : official journal of the Movement Disorder Society* 27:1556–9.

57. Berry-Kravis E et al. (2007) Fragile X-associated tremor/ataxia syndrome: clinical features, genetics, and testing guidelines. *Movement disorders : official journal of the Movement Disorder Society* 22:2018–30, quiz 2140.

58. Hagerman RJ et al. (2007) Neuropathy as a presenting feature in fragile X-associated tremor/ataxia syndrome. *American journal of medical genetics Part A* 143A:2256–60.

59. Hamlin A et al. (2011) Sleep apnea in fragile X premutation carriers with and without FXTAS. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official*

publication of the International Society of Psychiatric Genetics 156B:923–8.

60. Hamlin A a et al. (2012) Hypertension in FMR1 premutation males with and without fragile X-associated tremor/ataxia syndrome (FXTAS). *American journal of medical genetics Part A* 158A:1304–9.

61. Winarni T, Chonchaiya W (2012) Immune-mediated disorders among women carriers of fragile X premutation alleles. *American Journal of ...* 158A:2473–81.

62. Grigsby J et al. (2008) Cognitive profile of fragile X premutation carriers with and without fragile X-associated tremor/ataxia syndrome. *Neuropsychology* 22:48–60.

63. Bourgeois JA et al. (2011) Lifetime prevalence of mood and anxiety disorders in fragile X premutation carriers. *The Journal of clinical psychiatry* 72:175–82.

64. Hagerman RJ, Hagerman PJ (2002) The fragile X premutation: into the phenotypic fold. *Current opinion in genetics & development* 12:278–83.

65. Greco CM et al. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain : a journal of neurology* 125:1760–71.

66. Brunberg J a et al. (2002) Fragile X premutation carriers: characteristic MR imaging findings of adult male patients with progressive cerebellar and cognitive dysfunction. *AJNR American journal of neuroradiology* 23:1757–66.

67. Jacquemont S et al. (2004) Aging in individuals with the FMR1 mutation. *American journal of mental retardation : AJMR* 109:154–64.

68. Hashimoto R, Srivastava S, Tassone F, Hagerman RJ, Rivera SM (2011) Diffusion tensor imaging in male premutation carriers of the fragile X mental retardation gene. *Movement disorders : official journal of the Movement Disorder Society* 26:1329–36.

69. Cohen S et al. (2006) Molecular and imaging correlates of the fragile X-associated tremor/ataxia syndrome. *Neurology* 67:1426–31.

70. Moore CJ et al. (2004) The effect of pre-mutation of X chromosome CGG trinucleotide repeats on brain anatomy. *Brain : a journal of neurology* 127:2672–81.

71. Hagerman PJ (2012) Current Gaps in Understanding the Molecular Basis of FXTAS. *Tremor and other hyperkinetic movements (New York, NY)* 2.

72. Greco CM et al. (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain : a journal of neurology* 129:243–55.

73. Tassone F et al. (2012) Neuropathological, clinical and molecular pathology in female fragile X premutation carriers with and without FXTAS. *Genes, brain, and behavior* 11:577–85.

74. Hunsaker MR et al. (2011) Widespread non-central nervous system organ pathology in fragile X premutation carriers with fragile X-associated tremor/ataxia syndrome and CGG knock-in mice. *Acta neuropathologica*.

75. Tassone F, Hagerman RJ, Chamberlain WD, Hagerman PJ (2000) Transcription of the FMR1 gene in individuals with fragile X syndrome. *American journal of medical genetics* 97:195–203.

76. Kenneson a, Zhang F, Hagedorn CH, Warren ST (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Human molecular genetics* 10:1449–54.

77. Ladd PD et al. (2007) An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Human molecular genetics* 16:3174–87.

78. Arocena DG et al. (2005) Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. *Human molecular genetics* 14:3661–71.

79. Jin P et al. (2007) Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. *Neuron* 55:556–64.

80. Tassone F (2004) FMR1 RNA within the Intranuclear Inclusions of Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS). *Rna Biology*:103–105.

81. Iwahashi CK et al. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain : a journal of neurology* 129:256–71.

82. Li Y, Jin P (2012) RNA-mediated neurodegeneration in fragile X-associated tremor/ataxia syndrome. *Brain research* 1462:112–7.

83. Sellier C et al. (2013) Sequestration of DROSHA and DGCR8 by Expanded CGG RNA Repeats Alters MicroRNA Processing in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell reports* 3:1–12.

84. Sellier C et al. (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *The EMBO journal* 29:1248–61.

85. King OD, Gitler AD, Shorter J (2012) The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain research* 1462:61–80.

86. Renoux AJ, Todd PK (2012) Neurodegeneration the RNA way. *Progress in neurobiology* 97:173–89.

87. Khalil AM, Faghihi MA, Modarresi F, Brothers SP, Wahlestedt C (2008) A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PloS one* 3:e1486.

88. Zu T et al. (2010) Non-ATG – initiated translation directed by microsatellite expansions. 1–6.

89. Todd PKK et al. (2013) CGG Repeat-Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome. *Neuron* 78:1–16.

90. Sobczak K (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Research* 31:5469–5482.

91. Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic acids research* 39:7308–15.

92. Krol J et al. (2007) Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Molecular cell* 25:575–86.

93. Bañez-Coronel M et al. (2012) A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. *PLoS genetics* 8:e1002481.

94. Handa V, Saha T, Usdin K (2003) The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. *Nucleic acids research* 31:6243–8.

95. Chen Y et al. (2010) Murine hippocampal neurons expressing Fmr1 gene premutations show early developmental deficits and late degeneration. *Human molecular genetics* 19:196–208.

96. Berman RF, Murray KD, Arque G, Hunsaker MR, Wenzel HJ (2012) Abnormal dendrite and spine morphology in primary visual cortex in the CGG knock-in mouse model of the fragile X premutation. *Epilepsia* 53 Suppl 1:150–60.

97. Napoli E et al. (2011) Altered zinc transport disrupts mitochondrial protein processing/import in fragile X-associated tremor/ataxia syndrome. *Human molecular genetics* 20:3079–92.

98. Kaplan ES et al. (2012) Early mitochondrial abnormalities in hippocampal neurons cultured from Fmr1 pre-mutation mouse model. *Journal of neurochemistry* 123:613–21.

99. Ross-Inta C, Omanska-Klusek A, Wong S (2010) Evidence of mitochondrial dysfunction in fragile X-associated tremor/ataxia syndrome. *Biochem J* 552:545–552.

100. Lin Y, Tang C, He H, Duan R (2013) Activation of mTOR Ameliorates Fragile X Premutation rCGG Repeat-Mediated Neurodegeneration. *PloS one* 8:e62572.

101. Hoeffer CAC, Klann E (2010) mTOR signaling: at the crossroads of plasticity, memory and disease. *Trends in neurosciences* 33:67–75.

102. Sofola OA et al. (2007) RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron* 55:565–71.

103. Bontekoe CJ et al. (2001) Instability of a (CGG)98 repeat in the Fmr1 promoter. *Human molecular genetics* 10:1693–9.

104. Brouwer JR et al. (2008) CGG-repeat length and neuropathological and molecular correlates in a mouse model for fragile X-associated tremor/ataxia syndrome. *Journal of neurochemistry* 107:1671–82.

105. Wenzel HJ, Hunsaker MR, Greco CM, Willemsen R, Berman RF (2010) Ubiquitin-positive intranuclear inclusions in neuronal and glial cells in a mouse model of the fragile X premutation. *Brain research* 1318:155–66.

106. Willemsen R (2003) The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. *Human Molecular Genetics* 12:949–959.

107. Bergink S et al. (2006) The DNA repair-ubiquitin-associated HR23 proteins are constituents of neuronal inclusions in specific neurodegenerative disorders without hampering DNA repair. *Neurobiology of disease* 23:708–16.

108. Van Dam D et al. (2005) Cognitive decline, neuromotor and behavioural disturbances in a mouse model for fragile-X-

associated tremor/ataxia syndrome (FXTAS). *Behavioural brain research* 162:233–9.

109. Michael R. Hunsaker, H. Jürgen Wenzel, Rob Willemsen and RFB, Hunsaker MR, Wenzel HJ, Willemsen R, Berman RF (2012) Progressive Spatial Processing Deficits in a Mouse Model of the Fragile X Premutation. *Behavioural Neuroscience* 123:1315–1324.

110. Brouwer JR et al. (2008) Altered hypothalamus-pituitary-adrenal gland axis regulation in the expanded CGG-repeat mouse model for fragile X-associated tremor/ataxia syndrome. *Psychoneuroendocrinology* 33:863–73.

111. Entezam A et al. (2007) Regional FMRP deficits and large repeat expansions into the full mutation range in a new Fragile X premutation mouse model. *Gene* 395:125–34.

112. Qin M et al. (2011) A mouse model of the fragile X premutation: effects on behavior, dendrite morphology, and regional rates of cerebral protein synthesis. *Neurobiology of disease* 42:85–98.

113. Hashem V et al. (2009) Ectopic expression of CGG containing mRNA is neurotoxic in mammals. *Human molecular genetics* 18:2443–51.

114. Todd PK et al. (2010) Histone deacetylases suppress CGG repeat-induced neurodegeneration via transcriptional silencing in models of fragile X tremor ataxia syndrome. *PLoS genetics* 6:e1001240.

115. Qurashi A et al. (2012) Chemical screen reveals small molecules suppressing fragile X premutation rCGG repeat-mediated neurodegeneration in Drosophila. *Human molecular genetics* 21:2068–75.

116. Sun GY, Xu J, Jensen MD, Simonyi A (2004) Phospholipase A2 in the central nervous system: implications for neurodegenerative diseases. *Journal of lipid research* 45:205–13.

117. Murray A, Webb J, Grimley S, Conway G, Jacobs P (1998) Studies of FRAXA and FRAXE in women with premature ovarian failure. *Journal of medical genetics* 35:637–40.

118. Allingham-Hawkins DJ et al. (1999) Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study--preliminary data. *American journal of medical genetics* 83:322–5.

119. Nelson LM (2009) Clinical practice. Primary ovarian insufficiency. *The New England journal of medicine* 360:606–14.

120. Welt CK, Smith PC, Taylor a E (2004) Evidence of early ovarian aging in fragile X premutation carriers. *The Journal of clinical endocrinology and metabolism* 89:4569–74.

121. Jacquemont S, Hagerman RJ, Hagerman PJ, Leehey M a (2007) Fragile-X syndrome and fragile X-associated tremor/ataxia syndrome: two faces of FMR1. *Lancet neurology* 6:45–55.

122. Gallagher JC Effect of early menopause on bone mineral density and fractures. *Menopause (New York, NY)* 14:567–71.

123. Sullivan SD, Welt C, Sherman S (2011) FMR1 and the continuum of primary ovarian insufficiency. *Seminars in reproductive medicine* 29:299–307.

124. Murray A, Ennis S, MacSwiney F, Webb J, Morton NE (2000) Reproductive and menstrual history of females with fragile X expansions. *European journal of human genetics : EJHG* 8:247–52.

125. Sullivan AK et al. (2005) Association of FMR1 repeat size with ovarian dysfunction. *Human reproduction (Oxford, England)* 20:402–12.

126. Rohr J et al. (2008) Anti-Mullerian hormone indicates early ovarian decline in fragile X mental retardation (FMR1) premutation carriers: a preliminary study. *Human reproduction (Oxford, England)* 23:1220–5.

127. Hoffman GE et al. (2012) Ovarian abnormalities in a mouse model of fragile X primary ovarian insufficiency. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 60:439–56.

128. Lu C et al. (2012) Fragile X premutation RNA is sufficient to cause primary ovarian insufficiency in mice. *Human molecular genetics* 21:5039–5047.

129. Kapranov P et al. (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, NY)* 316:1484–8.

130. Berretta J, Morillon A (2009) Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO reports* 10:973–82.

131. Salta E, De Strooper B (2012) Non-coding RNAs with essential roles in neurodegenerative disorders. *Lancet neurology* 11:189–200.

132. Cao X, Yeo G, Muotri AR, Kuwabara T, Gage FH (2006) Noncoding RNAs in the mammalian central nervous system. *Annual review of neuroscience* 29:77–103.

133. Mehler MF, Mattick JS (2007) Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiological reviews* 87:799–823.

134. Johnson R, Noble W, Tartaglia GG, Buckley NJ (2012) Neurodegeneration as an RNA disorder. *Progress in neurobiology* 99:293–315.

135. Miñones-Moyano E et al. (2011) MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function. *Human molecular genetics* 20:3067–78.

136. Johnson R (2012) Long non-coding RNAs in Huntington's disease neurodegeneration. *Neurobiology of disease* 46:245–54.

137. Lee Y et al. (2004) MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* 23:4051–60.

138. Lee Y et al. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425:415–9.

139. Gregory RI et al. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432:235–40.

140. Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. *Science (New York, NY)* 303:95–8.

141. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–97.

142. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes & development* 22:2773–85.

143. Ghildiyal M et al. (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science (New York, NY)* 320:1077–81.

144. Okamura K, Balla S, Martin R, Liu N, Lai EC (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. *Nature structural & molecular biology* 15:581–90.

145. Watanabe T et al. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453:539–43.

146. Tam OH et al. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453:534–8.

147. Siomi MC, Sato K, Pezic D, Aravin A a (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nature reviews Molecular cell biology* 12:246–58.

148. Röther S, Meister G (2011) Small RNAs derived from longer non-coding RNAs. *Biochimie* 93:1905–1915.

149. Eulalio A, Huntzinger E, Izaurralde E (2008) Getting to the root of miRNA-mediated gene silencing. *Cell* 132:9–14.

150. Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature reviews Genetics* 12:99–110.

151. Wu L, Fan J, Belasco JG (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 103:4034–9.

152. Piao X, Zhang X, Wu L, Belasco JG (2010) CCR4-NOT deadenylates mRNA associated with RNA-induced silencing complexes in human cells. *Molecular and cellular biology* 30:1486–94.

153. Liu J et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science (New York, NY)* 305:1437–41.

154. Meister G et al. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell* 15:185–97.

155. Brennecke J et al. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science (New York, NY)* 322:1387–92.

156. Aravin A et al. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442:203–7.

157. Aravin AA et al. (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular cell* 31:785–99.

158. Kuramochi-Miyagawa S et al. (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & development* 22:908–17.

159. Van Leeuwen S, Mikkers H (2010) Long non-coding RNAs: Guardians of development. *Differentiation; research in biological diversity* 80:175–83.

160. Mehler MF, Mattick JS (2006) Non-coding RNAs in the nervous system. *The Journal of physiology* 575:333–41.

161. Rinn JL et al. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129:1311–23.

162. Wang X et al. (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454:126–30.

163. Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445:666–70.

164. Beltran M et al. (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes & development* 22:756–69.

165. Mercer TR, Dinger ME, Mattick JS, Rnas L (2009) Long non-coding RNAs: insights into functions. *Nature reviews Genetics* 10:155–9.

166. Costa V, Esposito R, Aprile M, Ciccodicola A (2012) Non-coding RNA and pseudogenes in neurodegenerative diseases: "The (un)Usual Suspects". *Frontiers in genetics* 3:231.

167. Cesana M et al. (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147:358–69.

168. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146:353–8.

169. Zumwalt M, Ludwig A, Hagerman PJ, Dieckmann T (2007) Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. *RNA biology* 4:93–100.

170. Lomniczi A et al. (2012) A single-nucleotide polymorphism in the EAP1 gene is associated with amenorrhea/oligomenorrhea in nonhuman primates. *Endocrinology* 153:339–49.

171. Lee J, Li Z, Brower-Sinning R, John B (2007) Regulatory circuit of human microRNA biogenesis. *PLoS computational biology* 3:e67.

172. D'Hulst C et al. (2009) Expression of the GABAergic system in animal models for fragile X syndrome and fragile X associated tremor/ataxia syndrome (FXTAS). *Brain research* 1253:176–83.

173. Handa V et al. (2005) Long CGG-repeat tracts are toxic to human cells: implications for carriers of Fragile X premutation alleles. *FEBS letters* 579:2702–8.

174. Tassone F et al. (2012) FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome medicine* 4:100.

175. Takahashi K, Niidome T, Akaike A, Kihara T, Sugimoto H (2009) Amyloid precursor protein promotes endoplasmic reticulum stress-induced cell death via C/EBP homologous protein-mediated pathway. *Journal of neurochemistry* 109:1324–37.

176. Scaglione KM et al. (2011) Ube2w and ataxin-3 coordinately regulate the ubiquitin ligase CHIP. *Molecular cell* 43:599–612.

177. Lin MT, Beal MF (2006) Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* 443:787–95.

178. Laplante M, Sabatini DM (2012) mTOR signaling in growth control and disease. *Cell* 149:274–93.

179. David G et al. (1997) Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nature genetics* 17:65–70.

180. Hoshikawa S, Ogata T, Fujiwara S, Nakamura K, Tanaka S (2008) A novel function of RING finger protein 10 in transcriptional regulation of the myelin-associated glycoprotein gene and myelin formation in Schwann cells. *PloS one* 3:e3464.

181. Heger S et al. (2007) Enhanced at puberty 1 (EAP1) is a new transcriptional regulator of the female neuroendocrine reproductive axis. *The Journal of clinical investigation* 117:2145–54.

182. Dissen G a, Lomniczi A, Heger S, Neff TL, Ojeda SR (2012) Hypothalamic EAP1 (enhanced at puberty 1) is required for menstrual cyclicity in nonhuman primates. *Endocrinology* 153:350–61.

183. Espinosa F et al. (2001) Alcohol hypersensitivity, increased locomotion, and spontaneous myoclonus in mice lacking the potassium channels Kv3.1 and Kv3.3. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 21:6657–65.

184. Alvarez-Mora MI et al. (2013) MicroRNA expression profiling in blood from fragile X-associated tremor/ataxia syndrome patients. *Genes, brain, and behavior*.

185. Harder JM, Libby RT (2011) BBC3 (PUMA) regulates developmental apoptosis but not axonal injury induced death in the retina. *Molecular neurodegeneration* 6:50.

186. Kim E-M, Hwang O (2011) Role of matrix metalloproteinase-3 in neurodegeneration. *Journal of neurochemistry* 116:22–32.

187. Zongaro S et al. (2013) The 3'UTR of FMR1 mRNA is a target of miR-101, miR129-5p and miR-221: implications for the molecular pathology of FXTAS at the synapse. *Human molecular genetics*:1–41.

188. Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C (2010) Characterization of the RNA content of chromatin. *Genome research* 20:899–907.

189. Khalil AM et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 106:11667–72.

190. Clemson CM et al. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular cell* 33:717–26.

191. Fox AH et al. (2002) Paraspeckles: a novel nuclear domain. *Current biology : CB* 12:13–25.

192. Prasanth K V et al. (2005) Regulating gene expression through RNA nuclear retention. *Cell* 123:249–63.

193. Tassone F, Iwahashi C, Hagerman PJ (2004) FMR1 RNA within the Intranuclear Inclusions of Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS). *Rna Biology* 1:103–105.

194. Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nature reviews Genetics* 11:597–610.

195. Pilotte J, Dupont-Versteegden EE, Vanderklish PW (2011) Widespread regulation of miRNA biogenesis at the Dicer step by the cold-inducible RNA-binding protein, RBM3. *PloS one* 6:e28446.

196. Hawkins PG, Morris K V (2008) RNA and transcriptional modulation of gene expression. *Cell cycle (Georgetown, Tex)* 7:602–7.

197. Ritchie ME et al. (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics (Oxford, England)* 23:2700–7.

198. Bolstad B (2001), pp 1–8.

199. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* 3:Article3.

200. Society RS (1995) Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 ,

No . 1 Published by : *Journal of the Royal Statistical Society: Series B* 57:289–300.

201. Gentleman RC et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* 5:R80.

202. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif)* 25:402–8.

203. Steibel JP, Poletto R, Coussens PM, Rosa GJM (2009) A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. *Genomics* 94:146–52.

204. Matsuda T, Cepko CL (2004) Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 101:16–22.

205. Presgraves SP, Ahmed T, Borwege S, Joyce JN (2004) Terminally differentiated SH-SY5Y cells provide a model system for studying neuroprotective effects of dopamine agonists. *Neurotoxicity research* 5:579–98.