**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jing Huang Date

Identification of Modifying Genes Associated with Fragile X Disorders Using TWAS

By

Jing Huang
Master of Science

Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology

_____
Michael P. Epstein, Ph.D.
Advisor

_____
Jian Hu, Ph.D.
Committee Member

_____
Jingjing Yang, Ph.D.
Committee Member

Accepted:

_____
Kimberly Jacob Arriola, Ph.D., MPH
Dean of the James T. Laney School of Graduate Studies

_____
Date

# Identification of Modifying Genes Associated with Fragile X Disorders Using TWAS

By

Jing Huang
B.Sc., Zhejiang University, 2021

Advisor: Michael P. Epstein, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology
2023

Abstract

Identification of Modifying Genes Associated with Fragile X Disorders Using TWAS
By Jing Huang

Fragile X disorders are a group of genetic conditions that include a neurodegenerative disorder of FXTAS, a fertility disorder of FXPOI, and an epileptic disorder of FXS seizures. To identify the potential genetic modifiers of Fragile X disorders, we proposed to leverage the framework of Transcriptome-Wide Association Studies (TWAS) that integrates transcriptomics data with genetic data to increase the power of identifying genes related to a disorder through gene regulation. Even though only 530 individuals were collected, our study successfully identified 68 genes significantly associated with Fragile X disorders by utilizing reference data of the brain cortex, cerebellum, cerebellar hemisphere, and ovary tissues from the Genotype-Tissue Expression Consortium V8. Many TWAS identifications, such as *RGL2*, *PPP1R12C*, and *SYNGAP1*, were supported by previous studies as they have been found to be associated with relevant clinical features of Fragile X disorders. Moreover, we further validated two genes (*CFB* and *AIF1*) for FXPOI in independent TWAS studies of age at menopause from the UK Biobank. These identified significant genes are worthy of further functional validations under the Drosophila or mice model to elucidate their underlying molecular mechanisms.

# Identification of Modifying Genes Associated with Fragile X Disorders Using TWAS

By

Jing Huang
B.Sc., Zhejiang University, 2021

Advisor: Michael P. Epstein, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology
2023

# Table of Contents

# 1 Introduction

Fragile X disorders are a group of genetic conditions that include a neurodegenerative disorder, Fragile X-associated tremor/ataxia syndrome (FXTAS), a fertility disorder, Fragile X-associated primary ovarian insufficiency (FXPOI), and an epileptic disorder, Fragile X syndrome related seizures (FXS-Seizure). The pathology of Fragile X disorders is related to an expansion of the CGG triplet repeat within the Fragile X mental retardation 1 (*FMR1*) gene on the X chromosome.[1] Normal individuals have on average less than 55 CGG repeats in the *FMR1* gene. Individuals who carry between 55-200 repeats are classified as premutation carriers, while full mutation individuals carry more than 200 CGG repeats. Expanded CGG repeats differentially impact mRNA and protein levels based on the size of the repeat. The premutation leads to increased levels of mRNA and slightly decreased levels of proteins, while the full mutation leads to defective transcription and translation.[2] These differential molecular changes of *FMR1* premutation and full mutation result in different clinical phenotypes. Specifically, full mutation carriers develop Fragile X syndrome (FXS) and some of the patients also experience a comorbid health condition of seizures (FXS-Seizure), while premutation carriers are at increased risk of ataxia (FXTAS) and ovarian insufficiency (FXPOI).[3,4,5] However, CGG repeat length at the *FMR1* locus cannot explain all the increased risk of Fragile X disorders since incomplete penetrance was observed among premutation carriers. Some premutation carriers develop FXTAS or FXPOI very early on, while others do not develop any relevant symptoms in their lifetime.[6] Based on such phenomenon, we hypothesize that additional modifier genes exist that influence variable expression of clinical phenotypes of Fragile X disorder.

Given the limited sample size collected for rare genetic diseases, conventional genome-wide association studies (GWAS) on Fragile X disorders may suffer from insufficient statistical power.

Recent studies have shown that GWAS associations were enriched for expression quantitative trait loci (eQTLs), suggesting that integrating genetic and transcriptomic data could help identify key molecular mechanisms underlying complex traits.[7] One such integrative method is transcriptome-wide association study (TWAS), which leverages a reference panel with profiled transcriptomic and genetic data from the same individuals to boost identification power. In this study, we proposed to perform a TWAS on Fragile X disorders using a combination of whole-sequenced subjects and reference data of the brain cortex, cerebellum, hemisphere cerebellum, and ovary tissues from the Genotype-Tissue Expression (GTEx) consortium.[8] Even though only a small sample size of 530 individuals was collected, we successfully identified 68 genes significantly associated with Fragile X disorders. Among these genes, many identifications were found to associate with relevant clinical features of Fragile X disorders from previous studies collected in Open Targets Genetics, which collects existing SNP-gene-trait associations from UK Biobank (UKB), GWAS catalog, and many other resources.[9,10] Moreover, two genes (*CFB* and *AIF1*) for FXPOI were validated in independent TWAS studies of age at menopause and thus are worthy of further functional validations to elucidate their underlying biological mechanisms.

# 2 Methods

## 2.1 Datasets

In this study, participant recruitment and sample acquisition were coordinated through the National Fragile X Center at Emory University. Once a participant was screened for eligibility and provided consent, a blood or saliva sample was collected, and a general medical history and standardized reproductive questionnaire were completed. 530 participants were recruited in total, and the criteria used to define cases and controls of Fragile X disorders are summarized as follows. As for FXS-Seizure, cases are FXS individuals who have a history of seizures, which involve recurrent and spontaneous episodes of abnormal electrical discharges in neural networks, before age 50 (*N*=67), while controls are those with no history of seizures before age 17 (*N*=82). As for FXTAS, cases are subjects with onset of FXTAS motor symptoms, which include Parkinsonism and progressive intention tremor, before age 65 (*N*=94), while controls are the ones without symptoms of FXTAS by age 68 (*N*=56). As for FXPOI, cases are women who have amenorrhea for at least 4-6 months before age 35 because of FXPOI (*N*=95), while controls are the ones who go through natural menopause for 1 year after age 50 (*N*=95). Patients with FXPOI normally experience irregular menstrual cycles, premature menopause, reduced fertility, or even infertility.

For initial risk genes identified in our FXS sample, we subsequently attempted to validate our findings in independent studies of related phenotypes. For FXS-Seizure, we utilized GWAS summary statistics of epilepsy from the Epilepsy Genetic Association Database, which includes about 4.9 million quality-controlled SNPs and a sample size of 82,482 individuals of European ancestry.[11] Since many FXTAS patients develop movement problems of Parkinsonism, we utilized GWAS summary statistics of Parkinson's disease from the International Parkinson's Disease Genomics Consortium (IPDGC), System Genomics of Parkinson's Disease (SGPD), and

UKB, which includes about 17.5 million SNPs and a sample size of 482,730 individuals of European ancestry.[12] For FXPOI, we utilized GWAS summary statistics of age at natural menopause from UKB, which includes about 12 million SNPs and a sample size of 143,025 women of European ancestry.[13]

## 2.2 Statistical Analysis

We performed whole genome sequencing (WGS) on recruited participants. Sequence alignment and variant calling were conducted using PEMapper/PECaller.[14] After standard quality control measures, we performed principal component analysis on a set of 567,720 independent SNPs using PLINK 1.9 to help correct for confounding due to ancestry or batch effects.[15] Based on a scree plot, the top 2 principal components were retained for covariate adjustment in subsequent analyses.

We used the TIGAR-V2 software package to perform TWAS.[16] TIGAR-V2 first trains prediction models of gene expression using reference transcriptomic data and genetic data from various tissues collected by the GTEx project V8.[8] Given the neurological and ovarian nature of Fragile X disorders, we focused on expression prediction models derived from the brain cortex (*N*=184), cerebellum (*N*=189), cerebellar hemisphere tissues (*N*=158) for FXTAS and FXS-Seizures, and ovary tissue (*N*=140) for FXPOI. For each tissue, TIGAR-V2 started by performing preprocessing of the tissue expression and genetic data from GTEx using standard QC pipelines. Specifically, TIGAR-V2 considered only genes with gene expression of transcripts per million (TPM) $>0.1$ in $\geq 10$ samples and variants with minor allele frequencies $>0.01$, missing rates $<20\%$, and Hardy-Weinberg equilibrium p values $>10^{-5}$ for fitting gene expression prediction models. For each gene, TIGAR-V2 adjusted gene expression for age, body mass index, top five

genotype principal components, and top probabilistic estimation of expression residuals factors. TIGAR-V2 then fit prediction models on the adjusted gene expression data treating *cis*-SNPs (within $\pm 1$ Mb region of transcription start sites of the target gene) as predictors. Assuming an additive genetic model for the expression quantitative trait ($\boldsymbol{E_g}$) of target gene $g$, the prediction model can be represented as follows

$$\boldsymbol{E_g} = \boldsymbol{G_{ref}}\boldsymbol{w} + \boldsymbol{\varepsilon};\ \varepsilon \sim N(0, \sigma_\varepsilon^2 \boldsymbol{I}) \quad (1)$$

where $\boldsymbol{G_{ref}}$ represents the vector of *cis*-SNP genotypes for gene $g$ in the GTEx data with corresponding effect sizes $\boldsymbol{w}$. Within the model, TIGAR-V2 estimates $\boldsymbol{w}$ for each gene using a popular nonparametric Bayesian Dirichlet process regression (DPR) procedure.[17]

Using estimates of *cis*-eQTL effect sizes $\widehat{\boldsymbol{w}}$ from each tissue, we then applied TIGAR-V2 to impute the genetically regulated gene expression (GReX) of gene $g$ in our Fragile X sample as $\widehat{GReX} = \boldsymbol{G_{test}}\widehat{\boldsymbol{w}}$, where $\boldsymbol{G_{test}}$ represents the corresponding *cis*-SNP genotypes for gene $g$ in the Fragile X dataset. We finally tested for association between $\widehat{GReX}$ and Fragile X case/control status using a logistic regression model that further adjusts for the covariates of *FMR1* repeat length (both linear and quadratic terms) as well as the top 2 principal components of ancestry.

With summary-level GWAS data available, we further applied TIGAR-V2 to construct S-PrediXcan test statistics (shown in Equation 2) for the gene-based association test.

$$\tilde{Z}_{g,S_{PrediXcan}} = \frac{\sum_{l=1}^{m}(\hat{w}_l \hat{\sigma}_l Z_l)}{\sqrt{\widehat{\boldsymbol{w}}'\boldsymbol{V}\widehat{\boldsymbol{w}}}}, \widehat{\sigma_l^2} = Var(\boldsymbol{G}_{0,\boldsymbol{1}}), \boldsymbol{V} = Cov(\boldsymbol{G}_0). \quad (2)$$

where $Z_l$ denotes the *Z*-score statistic value of genetic variant $l$ by single-variant GWAS test and $\boldsymbol{V}$ denotes linkage disequilibrium (LD) covariance matrix obtained from the GTEx project V8 for the *cis*-SNPs tested.[8] Additionally, the genotype matrix of test *cis*-SNPs is also obtained from the GTEx project V8.[8]

# 3 Results

## 3.1 FXTAS

We provide Manhattan plots of our TWAS analyses of FXTAS based on brain cortex, cerebellum, and cerebellar hemisphere tissues in **Figures 1**, **2**, and **3**, respectively. As shown in **Supplementary Figures S1**, **S2**, and **S3**, QQ plots for each TWAS present negligible systematic inflation. Even though only a small sample size of 530 individuals was collected, we observed that 35 genes reached a Bonferroni-corrected genome-wide threshold of $2.1 \times 10^{-6}$ and additional 31 genes reached a more liberal significance threshold of $10^{-4}$ in total. For these identified risk genes of FXTAS, none of these genes were validated with Parkinson's disease in independent summary-level TWAS analyses. Nevertheless, we further checked whether the risk genes we identified were previously associated with related traits using the Open Targets Genetics database.[9,10] Many TWAS identifications, such as *LSM2*, *CLIC1*, and *DXO*, were supported by previous studies as they have been found to associate with some relevant clinical features of FXTAS, e.g., Parkinsonism, multiple sclerosis, brain region volumes, mental disorder, and cognitive decline. Details of identified risk genes are shown in **Table 1**.

## 3.2 FXPOI

We provide Manhattan plots of our TWAS analyses of FXPOI based on ovary tissue in **Figure 4**. As shown in **Supplementary Figure S4**, the QQ plot presents negligible systematic inflation. Despite the limited sample size of 530 individuals recruited, we observed that 21 genes reached a Bonferroni-corrected genome-wide threshold of $2.2 \times 10^{-6}$ and additional 25 genes reached a more liberal significance threshold of $10^{-4}$ in total. Among these identified risk genes of FXPOI, *CFB* reached the genome-wide significance threshold and *AIF1* reached the replication significance

threshold (0.05 divided by the number of identified risk genes of FXPOI) for association with age at menopause in summary-level TWAS analyses in an independent GWAS dataset. We further checked risk genes' associated traits from previous studies in the Open Targets Genetics database.[9,10] Many TWAS identifications, such as *RGL2*, *PPP1R12C*, and *SYNGAP1*, were supported by previous studies as they have been found to associate with some relevant clinical features of FXPOI, e.g., age at menopause (or last menstrual period), sex hormone-binding globulin levels, and total testosterone levels. Details of identified risk genes are shown in **Table 2**.

## 3.3 FXS-Seizure

We provide Manhattan plots of our TWAS analyses of FXS-Seizure based on brain cortex tissue in **Figure 5**. As shown in **Supplementary Figure S5**, the QQ plot presents negligible systematic inflation. Even though only a small sample size of 530 individuals was collected, we observed that 12 genes reached a Bonferroni-corrected genome-wide threshold of $2.2 \times 10^{-6}$ and additional 11 genes reached a more liberal significance threshold of $10^{-4}$ in total. For these identified risk genes of FXS-Seizure, none of these genes were validated with epilepsy in independent summary-level TWAS analyses. However, we further checked whether the risk genes identified were previously associated with related traits using the Open Targets Genetics database.[9,10] Several TWAS identifications, such as *HLA-E*, *EHMT2*, and *RGL2*, turned out to be pleiotropy genes of Autism spectrum disorder (ASD), which supports the finding that FXS patients with seizures are more likely to have ASD than those without seizures.[18] Details of identified risk genes are shown in **Table 3**.

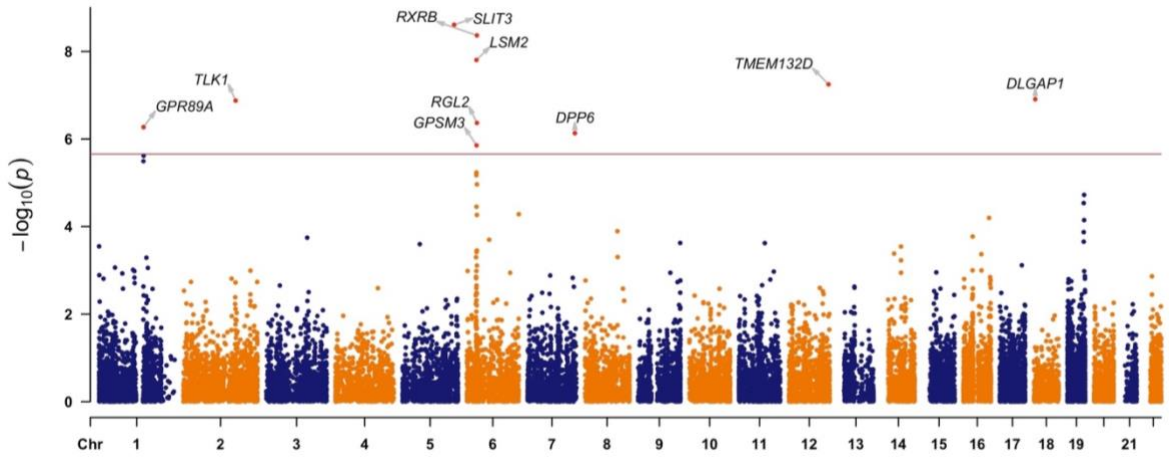


**Figure 1. Manhattan plot for FXTAS TWAS analyses based on brain cortex tissue.**
Each dot denotes the -log10(p-value) per gene by TWAS. The red line indicates the genome-wide significance threshold of 2.1 x $10^{-6}$. 10 genes reach the significance threshold.

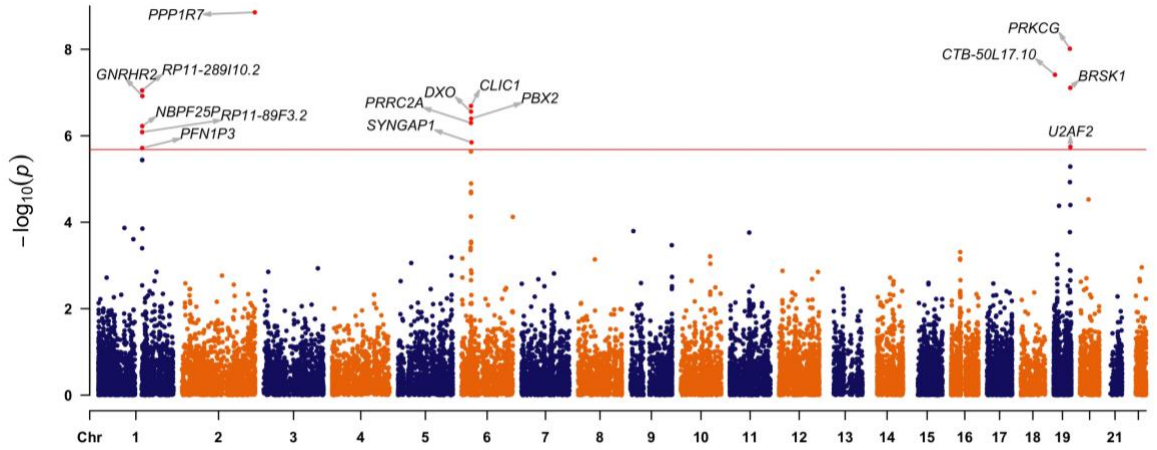


**Figure 2. Manhattan plot for FXTAS TWAS analyses based on brain cerebellum tissue.**
Each dot denotes the -log10(p-value) per gene by TWAS. The red line indicates the genome-wide significance threshold of 2.1 x $10^{-6}$. 15 genes reach the significance threshold.

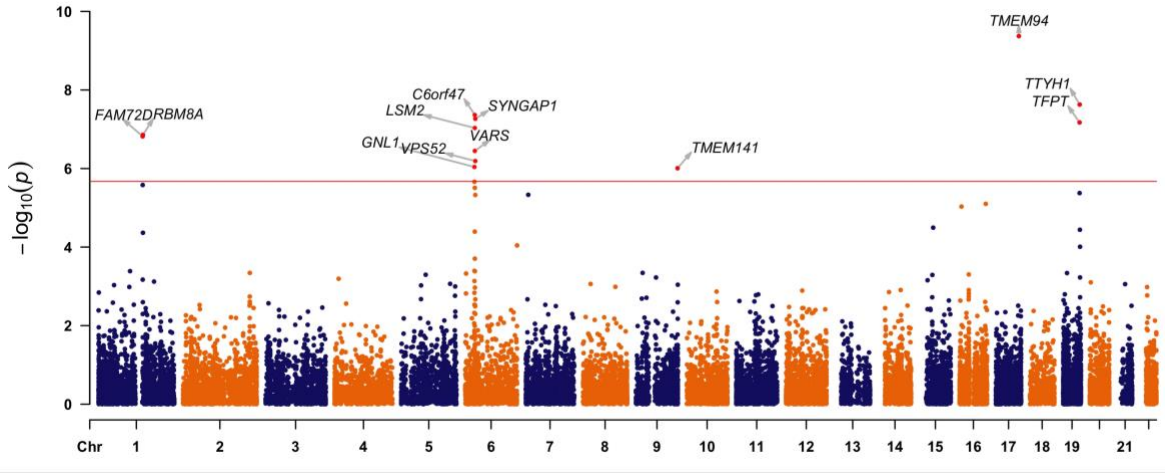


**Figure 3. Manhattan plot for FXTAS TWAS analyses based on brain cerebellar hemisphere tissue.**
Each dot denotes the -log10(p-value) per gene by TWAS. The red line indicates the genome-wide significance threshold of 2.1 x $10^{-6}$. 12 genes reach the significance threshold.

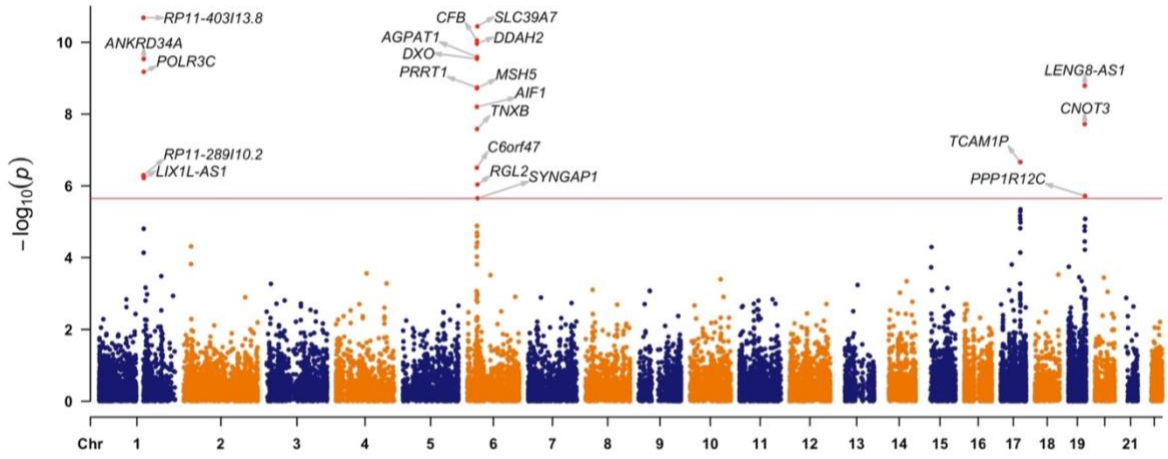


**Figure 4. Manhattan plot for FXPOI TWAS analyses based on ovary tissue.**
Each dot denotes the -log10(p-value) per gene by TWAS. The red line indicates the genome-wide significance threshold of 2.2 x $10^{-6}$. 21 genes reach the significance threshold.

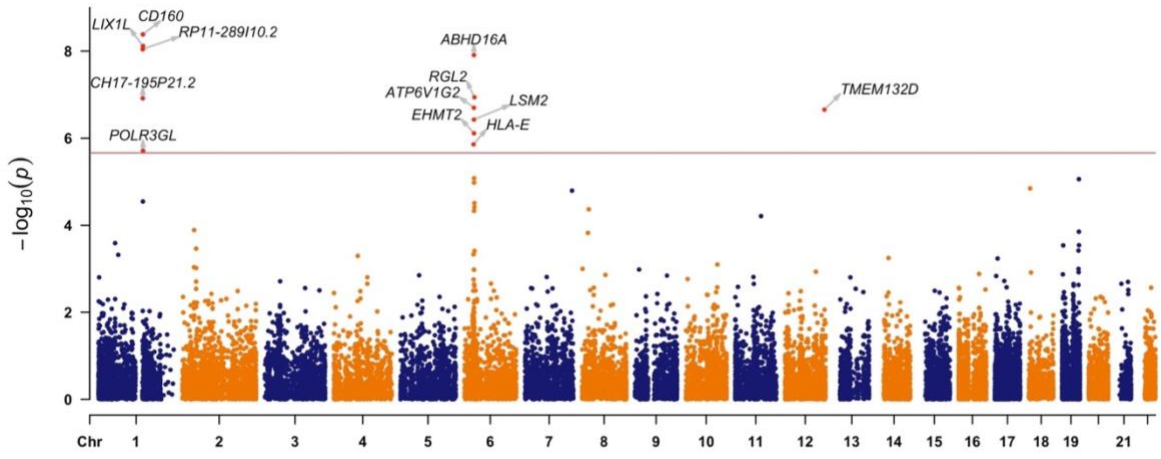


**Figure 5. Manhattan plot for FXS-Seizure TWAS analyses based on brain cortex tissue.** Each dot denotes the -log10(p-value) per gene by TWAS. The red line indicates the genome-wide significance threshold of $2.2 \times 10^{-6}$. 12 genes reach the significance threshold.

**Table 1. TWAS risk genes of FXTAS (genome-wide significance threshold)**

| Gene | Chromosome | Tissue | Associated traits in Open Targets Genetics | P value |
|---|---|---|---|---|
| *TMEM94* | 17 | Brain cerebellar hemisphere | (1) Multiple and sytemic sclerosis. | 4.25E-10 |
| *PPP1R7* | 2 | Brain cerebellum | | 1.40E-09 |
| *SLIT3* | 5 | Brain cortex | | 2.48E-09 |
| *RXRB* | 6 | Brain cortex | (1) Multiple and sytemic sclerosis. | 4.32E-09 |
| *PRKCG* | 19 | Brain cerebellum | | 9.71E-09 |
| *LSM2* | 6 | Brain cortex | (1) Parkinson's disease,<br>(2) Systemic sclerosis. | 1.57E-08 |
| *TTYH1* | 19 | Brain cerebellar hemisphere | | 2.36E-08 |
| *CTB-50L17.10* | 19 | Brain cerebellum | | 3.88E-08 |
| *C6orf47* | 6 | Brain cerebellar hemisphere | (1) Parkinson's disease. | 4.32E-08 |
| *SYNGAP1* | 6 | Brain cerebellar hemisphere | (1) Multiple and sytemic sclerosis. | 5.42E-08 |
| *TMEM132D* | 12 | Brain cortex | | 5.68E-08 |
| *TFPT* | 19 | Brain cerebellar hemisphere | | 6.71E-08 |
| *BRSK1* | 19 | Brain cerebellum | | 7.79E-08 |
| *RP11-289I10.2* | 1 | Brain cerebellum | | 8.95E-08 |
| *GNRHR2* | 1 | Brain cerebellum | | 1.21E-07 |
| *DLGAP1* | 18 | Brain cortex | | 1.24E-07 |
| *TLK1* | 2 | Brain cortex | | 1.33E-07 |
| *RBM8A* | 1 | Brain cerebellar hemisphere | | 1.39E-07 |
| *FAM72D* | 1 | Brain cerebellar hemisphere | | 1.52E-07 |
| *CLIC1* | 6 | Brain cerebellum | (1) Parkinson's disease. | 2.05E-07 |
| *DXO* | 6 | Brain cerebellum | (1) Parkinson's disease,<br>(2) Multiple sclerosis. | 2.76E-07 |
| *VARS* | 6 | Brain cerebellar hemisphere | | 3.56E-07 |
| *PBX2* | 6 | Brain cerebellum | (1) Parkinson's disease,<br>(2) Multiple and systemic sclerosis. | 4.00E-07 |
| *RGL2* | 6 | Brain cortex | (1) Systemic sclerosis. | 4.30E-07 |
| *PRRC2A* | 6 | Brain cerebellum | (1) Parkinson's disease. | 5.05E-07 |
| *GPR89A* | 1 | Brain cortex | | 5.39E-07 |
| *NBPF25P* | 1 | Brain cerebellum | | 5.98E-07 |
| *VPS52* | 6 | Brain cerebellar hemisphere | (1) Multiple and sytemic sclerosis. | 6.44E-07 |
| *DPP6* | 7 | Brain cortex | | 7.39E-07 |
| *RP11-89F3.2* | 1 | Brain cerebellum | | 8.21E-07 |
| *GNL1* | 6 | Brain cerebellar hemisphere | (1) Parkinson's disease. | 9.12E-07 |
| *TMEM141* | 9 | Brain cerebellar hemisphere | | 9.82E-07 |
| *GPSM3* | 6 | Brain cortex | (1) Parkinson's disease,<br>(2) Multiple and systemic sclerosis. | 1.40E-06 |
| *U2AF2* | 19 | Brain cerebellum | | 1.81E-06 |
| *PFN1P3* | 1 | Brain cerebellum | | 1.93E-06 |

The 'Tissue' column indicates which tissue the TWAS p value of each gene is based on.

**Table 2. TWAS risk genes of FXPOI (genome-wide significance threshold)**

| Gene | Chromosome | Associated traits in Open Targets Genetics | P value |
|---|---|---|---|
| *RP11-403I13.8* | 1 | | 2.07E-11 |
| *SLC39A7* | 6 | (1) Total testosterone levels,<br>(2) Sex hormone-binding globulin levels. | 3.59E-11 |
| *CFB* | 6 | (1) Total testosterone levels. | 8.92E-11 |
| *DDAH2* | 6 | (1) Total testosterone levels,<br>(2) Sex hormone-binding globulin levels. | 1.09E-10 |
| *AGPAT1* | 6 | (1) Total testosterone levels. | 2.56E-10 |
| *DXO* | 6 | (1) Sex hormone-binding globulin levels. | 2.87E-10 |
| *ANKRD34A* | 1 | | 2.89E-10 |
| *POLR3C* | 1 | | 6.64E-10 |
| *LENG8-AS1* | 19 | | 1.62E-09 |
| *PRRT1* | 6 | (1) Total testosterone levels. | 1.81E-09 |
| *MSH5* | 6 | (1) Total testosterone levels,<br>(2) Sex hormone-binding globulin levels. | 1.92E-09 |
| *AIF1* | 6 | (1) Total testosterone levels,<br>(2) Sex hormone-binding globulin levels. | 6.27E-09 |
| *CNOT3* | 19 | | 1.90E-08 |
| *TNXB* | 6 | (1) Total testosterone levels. | 2.61E-08 |
| *TCAM1P* | 17 | | 2.18E-07 |
| *C6orf47* | 6 | (1) Total testosterone levels,<br>(2) Sex hormone-binding globulin levels. | 3.14E-07 |
| *RP11-289I10.2* | 1 | | 5.07E-07 |
| *LIX1L-AS1* | 1 | | 6.01E-07 |
| *RGL2* | 6 | (1) Age at menopause,<br>(2) Sex hormone-binding globulin levels. | 9.14E-07 |
| *PPP1R12C* | 19 | (1) Age at menopause (last menstrual period). | 1.90E-06 |
| *SYNGAP1* | 6 | (1) Age at menopause,<br>(2) Total testosterone levels,<br>(3) Sex hormone-binding globulin levels. | 2.21E-06 |

**Table 3. TWAS risk genes of FXS-Seizure (genome-wide significance threshold)**

| Gene | Chromosome | Associated traits in Open Targets Genetics | P value |
|---|---|---|---|
| *CD160* | 1 | | 4.11E-09 |
| *LIX1L* | 1 | | 7.69E-09 |
| *RP11-289I10.2* | 1 | | 9.07E-09 |
| *ABHD16A* | 6 | | 1.23E-08 |
| *RGL2* | 6 | Autism spectrum disorder (pleiotropy). | 1.14E-07 |
| *CH17-195P21.2* | 1 | Autism spectrum disorder or schizophrenia. | 1.21E-07 |
| *ATP6V1G2* | 6 | Autism spectrum disorder or schizophrenia. | 2.01E-07 |
| *TMEM132D* | 12 | | 2.22E-07 |
| *LSM2* | 6 | | 3.74E-07 |
| *EHMT2* | 6 | Autism spectrum disorder (pleiotropy). | 7.71E-07 |
| *HLA-E* | 6 | Autism spectrum disorder (pleiotropy). | 1.39E-06 |
| *POLR3GL* | 1 | | 1.94E-06 |

# 4 Discussion

By leveraging the TWAS framework, this study has successfully identified 68 genes significantly associated with Fragile X disorders. Many TWAS identifications, such as *RGL2*, *PPP1R12C*, and *SYNGAP1*, were supported by previous studies as they were found to associate with relevant clinical features of Fragile X disorder in Open Targets Genetics. Moreover, *CFB* and *AIF1* for FXPOI were validated in independent studies of age at menopause and thus are worthy of further functional investigations (e.g., utilizing Drosophila or mice model) to elucidate the underlying biological mechanisms.[19]

With the high costs of WGS, the power to detect potential genetic modifiers was greatly limited by the sample size collected. Several approaches could be taken to increase the detection power of the TWAS framework. To begin with, we currently only utilized *cis*-eQTL information in our TWAS analyses, whereas trans-eQTL are equally important and explain a significant proportion of variation for most expression quantitative traits.[20] Therefore, other TWAS frameworks such as Bayesian Genome-wide TWAS method that incorporates both *cis*- and *trans*-eQTL can also be used to identify genetic modifiers of Fragile X disorders.[21] Furthermore, with emerging reference summary-level eQTL data generated by the eQTLGen and CommonMind consortia, TWAS frameworks that can utilize summary-level reference data to increase sample size and enhance power (e.g., OTTERS) are also worthy of future application.[22,23,24]

# 5 References

[1] Willemsen, R et al. “CGG repeat in the FMR1 gene: size matters.” *Clinical genetics* vol. 80,3 (2011): 214-25. doi:10.1111/j.1399-0004.2011.01723.x

[2] Kenneson, A et al. “Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers.” *Human molecular genetics* vol. 10,14 (2001): 1449-54. doi:10.1093/hmg/10.14.1449

[3] Berry-Kravis, Elizabeth et al. “Seizures in fragile X syndrome: characteristics and comorbid diagnoses.” *American journal on intellectual and developmental disabilities* vol. 115,6 (2010): 461-72. doi:10.1352/1944-7558-115.6.461

[4] Hoyos, Luis R, and Mili Thakur. “Fragile X premutation in women: recognizing the health challenges beyond primary ovarian insufficiency.” *Journal of assisted reproduction and genetics* vol. 34,3 (2017): 315-323. doi:10.1007/s10815-016-0854-6

[5] Famula, Jessica et al. “Neuropsychological changes in FMR1 premutation carriers and onset of fragile X-associated tremor/ataxia syndrome.” *Journal of neurodevelopmental disorders* vol. 14,1 23. 23 Mar. 2022, doi:10.1186/s11689-022-09436-y

[6] Rodriguez-Revenga, Laia et al. “Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families.” *European journal of human genetics : EJHG* vol. 17,10 (2009): 1359-62. doi:10.1038/ejhg.2009.51

[7] Nicolae, Dan L et al. “Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS.” *PLoS genetics* vol. 6,4 e1000888. 1 Apr. 2010, doi:10.1371/journal.pgen.1000888

[8] GTEx Consortium et al. “Genetic effects on gene expression across human tissues.” Nature vol. 550,7675 (2017): 204-213. doi:10.1038/nature24277

[9] Ghoussaini, Maya et al. “Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics.” *Nucleic acids research* vol. 49,D1 (2021): D1311-D1320. doi:10.1093/nar/gkaa840

[10] Mountjoy, Edward et al. “An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci.” *Nature genetics* vol. 53,11 (2021): 1527-1533. doi:10.1038/s41588-021-00945-5

[11] International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address: epilepsy-austin@unimelb.edu.au. “Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies.” *The Lancet. Neurology* vol. 13,9 (2014): 893-903. doi:10.1016/S1474-4422(14)70171-1

[12] Nalls, Mike A et al. “Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies.” *The Lancet. Neurology* vol. 18,12 (2019): 1091-1102. doi:10.1016/S1474-4422(19)30320-5

[13] Ruth, Katherine S et al. “Genetic insights into biological mechanisms governing human ovarian ageing.” *Nature* vol. 596,7872 (2021): 393-397. doi:10.1038/s41586-021-03779-7

[14] Johnston, H Richard et al. “PEMapper and PECaller provide a simplified approach to whole-genome sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 114,10 (2017): E1923-E1932. doi:10.1073/pnas.1618065114

[15] Chang, Christopher C et al. “Second-generation PLINK: rising to the challenge of larger and richer datasets.” *GigaScience* vol. 4 7. 25 Feb. 2015, doi:10.1186/s13742-015-0047-8

[16] Parrish, Randy L et al. “TIGAR-V2: Efficient TWAS tool with nonparametric Bayesian eQTL weights of 49 tissue types from GTEx V8.” *HGG advances* vol. 3,1 100068. 4 Nov. 2021, doi:10.1016/j.xhgg.2021.100068

[17] Zeng, Ping, and Xiang Zhou. “Non-parametric genetic prediction of complex traits with latent Dirichlet process regression models.” *Nature communications* vol. 8,1 456. 6 Sep. 2017, doi:10.1038/s41467-017-00470-2
[18] Berry-Kravis, Elizabeth et al. “Seizures in Fragile X Syndrome: Associations and Longitudinal Analysis of a Large Clinic-Based Cohort.” *Frontiers in pediatrics* vol. 9 736255. 30 Dec. 2021, doi:10.3389/fped.2021.736255
[19] Sherman, Stephanie L et al. “Use of model systems to understand the etiology of fragile X-associated primary ovarian insufficiency (FXPOI).” *Journal of neurodevelopmental disorders* vol. 6,1 (2014): 26. doi:10.1186/1866-1955-6-26
[20] Lloyd-Jones, Luke R et al. “The Genetic Architecture of Gene Expression in Peripheral Blood.” *American journal of human genetics* vol. 100,2 (2017): 371. doi:10.1016/j.ajhg.2017.01.026
[21] Luningham, Justin M et al. “Bayesian Genome-wide TWAS Method to Leverage both cis- and trans-eQTL Information through Summary Statistics.” *American journal of human genetics* vol. 107,4 (2020): 714-726. doi:10.1016/j.ajhg.2020.08.022
[22] Võsa, Urmo et al. “Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression.” *Nature genetics* vol. 53,9 (2021): 1300-1310. doi:10.1038/s41588-021-00913-z
[23] Sieberts, Solveig K et al. “Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions.” *Scientific data* vol. 7,1 340. 12 Oct. 2020, doi:10.1038/s41597-020-00642-8
[24] Dai, Qile et al. “OTTERS: a powerful TWAS framework leveraging summary-level reference data.” *Nature communications* vol. 14,1 1271. 7 Mar. 2023, doi:10.1038/s41467-023-36862-w

# 6 Supplementary Data

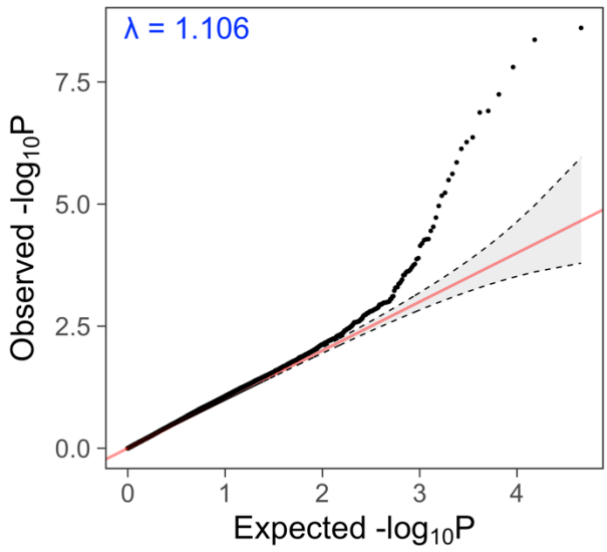


**Figure S1. QQ plot for FXTAS TWAS analyses based on brain cortex tissue.**
The genomic inflation factor $\lambda$ is close to 1 indicating no obvious evidence of inflation.

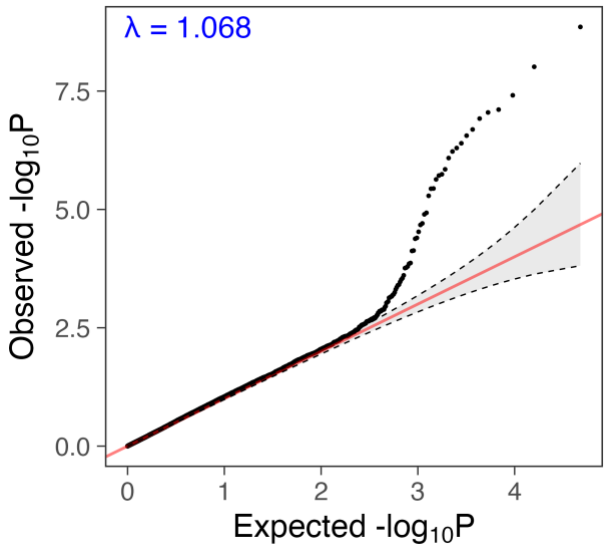


**Figure S2. QQ plot for FXTAS TWAS analyses based on brain cerebellum tissue.**
The genomic inflation factor $\lambda$ is close to 1 indicating no obvious evidence of inflation.

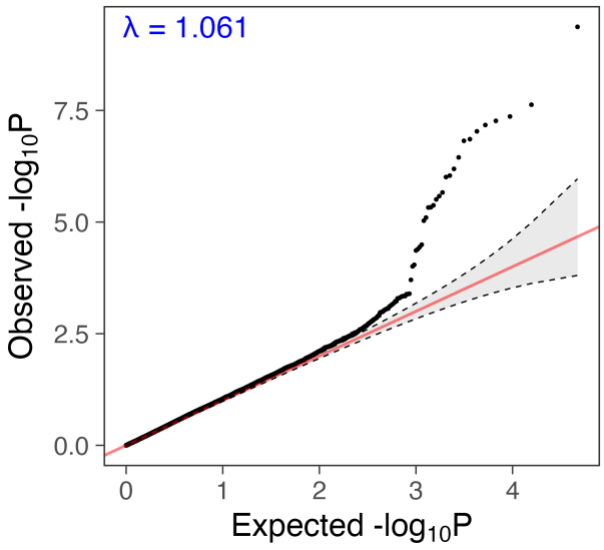


**Figure S3. QQ plot for FXTAS TWAS analyses based on brain cerebellar hemisphere tissue.**
The genomic inflation factor $\lambda$ is close to 1 indicating no obvious evidence of inflation.

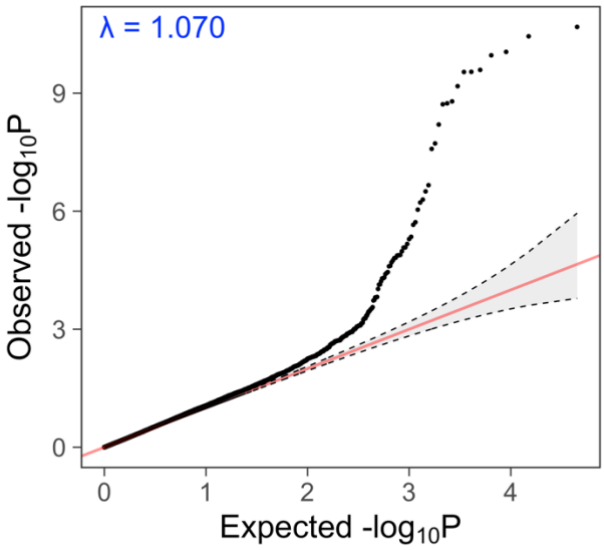


**Figure S4. QQ plot for FXPOI TWAS analyses based on ovary tissue.**
The genomic inflation factor $\lambda$ is close to 1 indicating no obvious evidence of inflation.

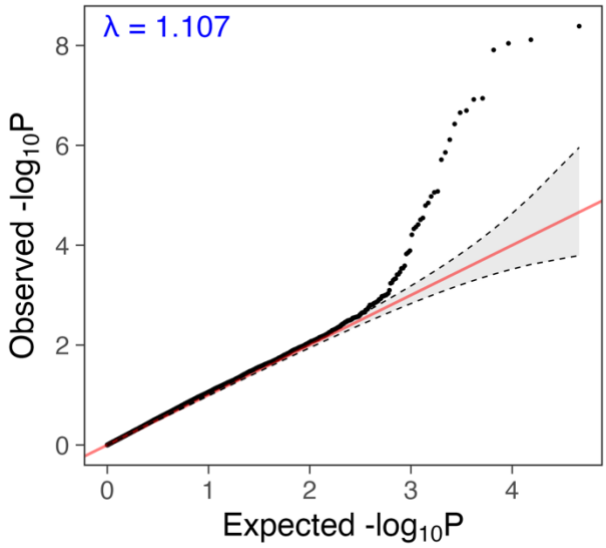


**Figure S5. QQ plot for FXS-Seizure TWAS analyses based on brain cortex tissue.**
The genomic inflation factor $\lambda$ is close to 1 indicating no obvious evidence of inflation.