GENETIC, EPIGENETIC AND TRANSCRIPTOME STUDIES OF TOURETTE SYNDROME AND TIC DISORDERS

PhD thesis

LUCA PAGLIAROLI

Doctoral School of Molecular Medicine

Semmelweis University

Supervisor: Csaba Barta, MD, PhD

Official reviewers: Xénia Gonda, PhD
Bálint László Bálint, MD, PhD

PhD Final Theoretical Exam Board: Illona Kovalszky, MD, PhD, DSc
Beáta Törőcsik, PhD
Zoltán Gáspári, PhD

Budapest 2019

INTRODUCTION

Gilles de la Tourette Syndrome (TS) is a neurodevelopmental disorder characterized by the appearance of a vocal and multiple motor tics (DSM-5). To be diagnosed with TS, considered rare disease affecting 1% of the world population, the symptoms need to manifest for at least 1 year. It usually last only during childhood and adolescence and it affects most the boys than girls (4:1). The probability of transmitting the disorder to the next generation is be between 0.58 and 077 with a very high risk of inheritance in case of a first-degree relative (father-mother). To complicate the situation, TS rarely manifests alone: in 45-60% of the cases goes along with obsessive-compulsive disorder (OCD) while in 60% of the cases appears with attention deficit hyperactivity disorder (ADHD). Other type of behavioral disorders such as autism and learning disabilities are also usual comorbidities in individuals with TS. In order to gain more information on such rare disease, usually less funded than major diseases, the European Union established the Marie Curie Initial Training Network (ITN, 2012–2016). This network was composed by 12 PhD students spread across Europe aiming to study TS from different angles (bioinformatics, animal model, genetic, neuroimaging) and to combine all the new discoveries to have a clear picture of the pathology.

A tic is usually defined as the appearance of an involuntary locomotor activity or vocal sound that manifests occasionally for a short amount of time with different frequency, intensity, duration or anatomical localization.

In general, tics are more prone to manifest if the patients are in a situation of emotional distress while they are reduced when the patients are focused/involved on daily activities. The patients usually feel when the tic is coming, a phenomenon called “premonitory urge”, and after the tic is released the patients experience a sense of relief. The symptoms usually manifest during childhood and disappear at the end of adolescence following the typical wax and waning pattern that characterize Tourette Syndrome. In few cases, the symptoms manifest also during adulthood even

though they do not affect too much the daily life of the patients since they learnt very well how to control and live with them.

Tourette Syndrome has a complex etiology. Indeed, several environmental factors (see below) may interact with a lot of the underlying genetic components. The linkage and candidate gene association studies discovered several chromosomal regions and genetic variation that may be associated with Tourette Syndrome. The genetic variations can be classified based on the extent of DNA alteration, such as single nucleotide polymorphisms (SNPs), as well as shorter or longer repeat variants, such as variable number of tandem repeats (VNTRs) and copy number variations (CNVs). In the majority of the cases, these small-scale studies focused on variation in genes of the dopaminergic and serotonergic neurotransmitter systems.

So far, most of the TS genetic studies have been hinder by lack of statistical power, small sample sizes and clinical heterogeneity. The fragmentation of these findings resembles the genetic complexity of Tourette Syndrome and supports the theory that a combination of small effects leads to the manifestation of the phenotype. Indeed, it is very unlikely that one single variation might cause all the variety of symptoms that appear in the patients.

It is very difficult to study and to identify genetic associations in a complex disorder like TS. It is one of the reasons that lead to the establishment of an extensive network (TS-EUROTRAIN) able to dig in the genetic background of Tourette Syndrome with sufficient tools and resources to overcome issues like stratification population, sample size and statistical power.

Tourette Syndrome is not only about genetics.

In the last years there has been a growing consensus that also epigenetics might contribute to the development of the pathology. Epigenetic usually refers to those changes, such as DNA methylation, histone modification and microRNA, in the gene expression or cellular phenotype that do not alter the sequence of the DNA. Since these modifications are both cells and time specific it becomes extremely important to study not only the exact tissue involved in the pathogenesis of the disorder but

also at the correct time. Moreover, these modifications are usually dynamic and influenced by both internal and external stimuli.

Tourette Syndrome clearly appears to be a good candidate for epigenetic mechanisms: symptoms appear and disappear based on the emotional status of the patients with a different intensity and frequency, and the time lapse of the disorders is scattered and different for every patient. It is known that external factors like psychological stress, pre-maternal smoking, paternal age, birth weight and exposure to medication in utero represent a risk for TS development. The appearance of TS might be a consequence of a pediatric autoimmune neuropsychiatric disorder associated with streptococcal infection (PANDAS). Some direct link to Tourette Syndrome and DNA methylation or microRNA regulation has been recently reported.

Epigenetic changes are cells specific and Tourette Syndrome is a neurodevelopmental disorder. The difficulty to access human brain samples represents a major drawback in the understanding of TS pathophysiology and it makes the availability of animal models a key component in scientific studies. For these reasons, TS-EUROTRAIN Network had 2 PhD projects focused only on the establishment of new animal models of TS in order to have a direct access to brain tissues, during childhood, involved in the pathology (see below).

The use of an appropriate animal model to study Tourette Syndrome still represents a big challenge. Indeed, it is very difficult to reproduce in a single animal model all the different clinical manifestations observed in TS patients. However, they represent the only option to perform biological studies on brain tissues.

The model used in this PhD project is called abnormal involuntary movements (AIMs) and it was established by our collaborators at Boehringer Ingelheim, Germany.

This type of model is widely used to study Parkinson's disease (PD) where the chronic administration of L-DOPA often results in the development of AIMs, an

important clinical problem that affects almost 90% of PD patients. However, the pathogenic mechanisms underlying AIMs are not yet understood.

In this PhD project we decided to use the AIMs animal model to study both RNA and DNA of brain to increase the to shed some light on the speculated relevance of DNA methylation in the development of locomotor activity and to possibly identify drug targets by looking at gene expression changes after chronic administration with L-DOPA followed by treatment with Riluzole.

OBJECTIVES

1) Identify new genetic variations that may represent risks for Tourette Syndrome by focusing on the 3'UTR of TS candidate genes.

2) Shed some light on the role of DNA methylation in the development of Tourette Syndrome by performing the first ever Epigenome Wide Association Study (EWAS) on tic disorders benefiting from the TS-EUROTRAIN network resources.

3) Study the transcriptome of striatum of an AIMs animal model to build a molecular landscape aimed to unravel the molecular mechanism of action of Riluzole.

4) Study the methylation pattern of striatum of AIMs animal model at the genome level to identify which changes occur after the administration of L-DOPA or the injection of 6-OHDA.

METHODS

Genetic of Tourette Syndrome

In silico work

I originally obtained a list of TS candidate genes by performing a literature research on PubMed using the terms “Tourette Syndrome” or “TS” combined with “gene” or “genetics” and also by using a software HugeNavigator. Subsequently, I focused the attention on SNPs not only located in the 3’UTR regulatory sequences of TS candidate genes but also predicted (PolymiRTS database) to alter the seed sequence of miRNAs binding site. Thirty SNPs were successfully genotyped using our customized TaqMan® OpenArray® Genotyping chips.

Statistical Analysis

Standard case-control association analysis was conducted in PLINK using $\chi 2$ test comparing SNP frequencies within cases to that of controls. The individual population association results were further meta-analyzed using the METAL software in a case-weighted fixed-effect model on p-values. To correct our results for multiple testing 100000 permutations were applied with the Monte Carlo Permutation (MCPerm method).

Family-based association was performed using the family-based Transmission Disequilibrium Test (TDT) option in PLINK software. Empirical significance levels were generated with PLINK using max (T) permutation methods set to 1000 permutations.

Epigenetic of Tourette Syndrome (EWAS)

The analysis was comprehensive of 1678 individuals (twins, siblings, and parents) from 1057 families. DNA methylation was assessed with the Infinium HumanMethylation450 Bead Chip Kit Array. The EWAS was performed using

linear regression under an additive model correcting for principal components and covariates and CpG sites with p-value < 1.2*10-7 were considered statistically significant. The gene ontology enrichment analysis was performed by ranking all the methylation sites based on their p-value and the resulting ranked gene list was supplied to the online software tool Gorilla.

Abnormal Involuntary Movements (AIMs) animal model

The animal model was established by our collaborators at Boehringer Ingelheim, Biberach, Germany. The animals were randomly located into three groups (n = 12). The first group was chronically administered with a solution of L-DOPA. The second group (control) was administered with only saline. The third group was chronically administered with L-DOPA + Riluzole.

RNA sequencing and data analysis

Total RNA was extracted from striatum tissue from eight animals of each group, the RNA was quantified and only samples with RIN value above 8.0 were used for transcriptome analysis. Statistical analyses were performed using R (www.r-project.org) and the Bioconductor package limma-voom. The Benjamini-Hochberg method was used to correct for multiple testing, and only protein-coding genes with adjusted P-value < 0.01, were used for further analyses. The upstream regulator and gene enrichment analyses were done using Ingenuity Pathway Analysis (IPA). To check the reliability of the RNA sequencing data, we performed qPCR on 10 randomly chosen CREB1 target genes

Reduced Representation Bisulfite Sequencing

The DNA was extracted from striatum tissue, the DNA fragmentation was checked on 1% agarose gel and only high-quality DNA (100ng) was used for preparation of Reduced Representation Bisulfite Sequencing libraries using Premium RRBS kit (Diagenode). The DNA was digested with MspI restriction enzyme which cuts at

C^CGG sites regardless the methylation status. Subsequently, the samples were treated with an enzyme that adds flanking ends and during this reaction, some spike in control to check the bisulfite conversion efficiency was also added. In the next step, a specific string of adaptors (6 base pair) was added to each sample followed by a size selection of the fragments (200-1200 bp). Bisulfite conversion was performed on pooled samples. Libraries were then ready for a quantification step followed by an enrichment PCR. Finally, the enriched PCR products were recovered, and the concentration was measured by Qubit dsDNA HS Assay (Life Technologies) and the library profile was checked on Bioanalyzer 2100 (Agilent) before being sequenced on Illumina HiSeq200.

CONCLUSION

Genetic of Tourette Syndrome

In the genetic association studies, we wanted to investigate SNPs in the 3'UTR of Tourette candidate genes. We decided to focus mostly on genetic variations predicted to alter the binding site of miRNAs. Indeed, there is a chance that SNPs in that area might have a greater impact on the gene expression due to the downstream effect manifested via the regulation mediated by miRNAs.

The dataset used in the analysis was composed of 148 families and 290 case-control samples. Samples were collected by the TSGeneSEE consortium collection and the dataset at our disposal was a mix of four different populations (Hungarian, Polish, Italian, Greek). In order to proper analyze the dataset, we had to separate the families from the case-control samples and check for population stratification. We analyzed the families using a TDT test while for the case-control genetic association we used a meta-analysis. The need to separate the two analyses resulted in a low-sample size that hindered the statistical power of the genetic study.

The Real-Time PCR did not find any positive results. The meta-analysis detected a positive association between rs3750486 ($p = 0.0212$) located on *LHX6* and rs7795011 ($p = 0.0290$) located on *IMMP2L* with TS. Moreover, the TDT test also picked up an over-transmission of the A allele of rs1042201 ($p = 0.0286$, $\chi 2 = 4.787$) located on *AADAC* and TS patients. These finding corroborate the literature evidence already proposing the genes as one of the many components in the TS pathology. The genetic association study also showed how lower sample size and population stratification hamper statistical power and it emphasized the need of a large collaborative effort in order to overcome those liabilities.

Worth to mention, to solve the genetic turmoil more studies on different population are need since it is likely that many of the genetic variations propelling the development of TS might differ based also on the geographic area, as observed in many studies. This population difference might also help to explain the lack of replicating findings between different studies.

In addition, to prove the actual role of miRNAs in the abnormal regulation of those genes as consequence of genetic variation (SNP) functional validation studies are warranted. However, during the course of this PhD project, we were not able to perform or to finish any functional studies.

Epigenetic of Tourette Syndrome

In this project, we wanted to investigate genome wide DNA methylation changes on blood samples of patients with a tic phenotype.

The samples were collected in The Netherlands and we described the first ever genome wide methylation analysis (EWAS) on tic disorders. One of the difficulties we first encountered was the establishment of a good samples size since the dataset was based on a self-reported questionnaire where patients would describe the manifestation of tics. The use of self-reported questionnaires is quite common.

Indeed, in many cases, the medical centers responsible for the diagnosis and blood collection might be far from the hometown of the patients therefore it is not possible to personally recruit or examine all the participants during the manifestation of tics, and clinicians rely on questionnaire to have an general overview of the symptoms.
The second factor we had to considered during the analysis of the dataset was represented by the heterogeneity: we had to correct for several factors like age, smoking, sex, tics, cell type and the position on the array. Indeed, it is known that all of these elements, with the exception of tics, affect DNA methylation levels.
Once the dataset was cleaned, we performed the EWAS. We did not find any significant CpG site after correcting for multiple tests. However, we were able to identify 57 CpGs that had an interesting p-value (< 0.0001). The gene-ontology analysis and the subsequent literature research of the top CpGs, showed that many CpG sites are in proximity to genes previously associated with psychiatric or neurological disorders.

AIMs animal model

In the transcriptome and methylome study of the AIMs animal model we wanted to investigate the area directly related to the manifestation of tic: the striatum.
The use of the animal models represents a very good option to perform molecular and biological studies in brain tissues involved in the pathogenesis of neurological disorders. Tourette Syndrome has a very complex and heterogeneous phenotype: the tics appears in situation of stress, the intensity and the severity of tics is always different, each patient manifest unique symptoms. It is impossible to have a single animal model capable of resembling all these facets. So, we decided to build the animal model based on the physiological situation observed in TS patients, a situation of hypersensitivity to dopamine. The animal model was established by our collaborators at Boehringer Ingelheim.
It was originally planned as a Tourette Syndrome animal model but after thinking about the biological background, observing the phenotype and discussing with

experts in the field, we agreed that the model would not fit Tourette Syndrome but would fall under a different classification: abnormal involuntary movements (AIMs).

The behavioral data showed that administration of L-DOPA to the hypersensitive striatum triggers the appearance of so-called tics while treatment with Riluzole reduced the appearance of involuntary movements. The transcriptome experiment showed that Riluzole treatment results into changes in the mRNA expression of those genes that are linked to the appearance of abnormal locomotor activity. The same genes which are highly expressed after the administration of L-DOPA. The molecular landscape indicates that the regulation of the apoptotic process has a pivotal role in both the appearance of AIMs after L-DOPA administration and in the reduction of locomotor activity seen after Riluzole treatment. In general, the CREB1 target genes lead to the hypothesis that L-DOPA administration leans towards increased neuronal survival while the Riluzole treatment triggers apoptotic processes.

The exact mechanism explaining how Riluzole would reduce the activity of *CREB1* is still unknown. By looking at what is reported in the literature and at the transcriptome data, we were able to propose a mechanism that might explain this mystery. In our opinion, Riluzole reduces the activity of *CREB1* leading to a modulation of the apoptotic process, and ultimately lower AIMs. To start with, Riluzole has many anti-glutamatergic properties that negatively regulate the glutamatergic neurotransmission. The release of glutamate is able to trigger the post-synaptic NMDA receptors function thus leading to the activation ERK1/2. The upregulation of ERK1/2 leads to CREB1 phosphorylation/activation. To recapitulate, Riluzole would decrease the activity of CREB1 by decreasing the ERK1/2 signaling modulated by NMDA receptor. To continue, Riluzole is a strong inhibitor of PK, a kinase directly involved in the phosphorylation/activation of CREB1. Moreover, PKC activates CREB1 also indirectly by positively modulating the NMDA receptors thus leaning on the effect mediated by active ERK1/2.

On the other side, the DNA methylation studies did not provide as exciting results as the transcriptome analysis. However, we gained knowledge on the changes in DNA methylation levels in animal model of AIMs. Indeed, we observed an overall decrease of methylation (linked to higher gene expression) and discovered that most of the changes were located outside promoters/CpG islands. Moreover, the DMSs analyses revealed that treatment administration has an influence on DNA methylation levels, mostly resulting in hypomethylation which subsequently could lead to gene expression and might result in the manifestation of locomotor activity. Moreover, the comparison between lesion and contralateral striatum clearly demonstrated that 6-OHDA injection (lesioned side) has not significant effects on dynamic methylation changes. In addition, the data clearly show that those changes in DNA methylation happen regardless the type of treatment: L-DOPA or L-DOPA + Riluzole.

To gain some more knowledge about the overall biological function of all the CpGs detected by RRBS, we performed gene ontology analysis which indicated an enrichment in genes involved either in psychiatric disorders, glutamate neurotransmitter or brain functions. It also provided different genes for follow-up studies.

PHD CANDIDATE'S PUBLICATIONS

Publications discussed in the PhD dissertation

1) Riluzole Attenuates L-DOPA-Induced Abnormal Involuntary Movements Through Decreasing CREB1 Activity: Insights from a Rat Model. **Pagliaroli L**, Widomska J, Nespoli E, Hildebrandt T, Barta C, Glennon J, Hengerer B, Poelmans G. Mol Neurobiol. 2018 Nov 27. doi: 10.1007/s12035-018-1433-x. [Epub ahead of print] PubMed PMID: 30484112. **IF 5.026**

2) From Genetics to Epigenetics: New Perspectives in Tourette Syndrome Research.
Pagliaroli L, Vető B, Arányi T, Barta C. Front Neurosci. 2016 Jul 12;10:277. doi: 10.3389/fnins.2016.00277. eCollection 2016. Review. PubMed PMID: 27462201; PubMed Central PMCID: PMC4940402. **IF 3.556**

3) Epigenome-Wide Association Study of Tic Disorders. Zilhão NR, Padmanabhuni SS, **Pagliaroli L**, Barta C; BIOS Consortium, Smit DJ, Cath D, Nivard MG, Baselmans BM, van Dongen J, Paschou P, Boomsma DI. Twin Res Hum Genet. 2015 Dec;18(6):699-709. doi: 10.1017/thg.2015.72. Epub 2015 Oct 26. PubMed PMID: 26499864. **IF 2.339**

4) TS-EUROTRAIN: A European-Wide Investigation and Training Network on the Etiology and Pathophysiology of Gilles de la Tourette Syndrome. de la Tourette Syndrome. Forde NJ, Kanaan AS, Widomska J, Padmanabhuni SS, Nespoli E, Alexander J, Rodriguez Arranz JI, Fan S, Houssari R, Nawaz MS, Rizzo F, **Pagliaroli L**, Zilhäo NR, Aranyi T, Barta C, Boeckers TM, Boomsma DI, Buisman WR, Buitelaar JK, Cath D, Dietrich A,

Driessen N, Drineas P, Dunlap M, Gerasch S, Glennon J, Hengerer B, van den Heuvel OA, Jespersgaard C, Möller HE, Müller-Vahl KR, Openneer TJ, Poelmans G, Pouwels PJ, Scharf JM, Stefansson H, Tümer Z, Veltman DJ, van der Werf YD, Hoekstra PJ, Ludolph A, Paschou P. Front Neurosci. 2016 Aug 23;10:384. doi: 10.3389/fnins.2016.00384. eCollection 2016. PubMed PMID: 27601976; PubMed Central PMCID: PMC4994475. **IF IF 3.556**

Publications not discussed in the PhD dissertation

1) Reduced expression of MYC increases longevity and enhances healthspan. Hofmann JW, Zhao X, De Cecco M, Peterson AL, **Pagliaroli L**, Manivannan J, Hubbard GB, Ikeno Y, Zhang Y, Feng B, Li X, Serre T, Qi W, Van Remmen H, Miller RA, Bath KG, de Cabo R, Xu H, Neretti N, Sedivy JM. Cell. 2015 Jan 29;160(3):477-88. doi: 10.1016/j.cell.2014.12.016. Epub 2015 Jan 22. PubMed PMID: 25619689; PubMed Central PMCID: PMC4624921. **IF 28.710**

2) Summaries of plenary and selected symposia sessions at the XXIV World Congress of Psychiatric Genetics; Jerusalem, Israel; 30 October 2016-3 November. Ciobanu LG, Ori AP, **Pagliaroli L**, Polimanti R, Spindola LM, Vincent JB, Cormack FK. 2016. Psychiatr Genet. 2017 Apr;27(2):41-53. doi: 10.1097/YPG.0000000000000167. PubMed PMID: 28212207. **IF 1.586**

3) Summaries of plenary, symposia, and oral sessions at the XXII World Congress of Psychiatric Genetics, Copenhagen, Denmark, 12-16 October 2014. Aas M, Blokland GA, Chawner SJ, Choi SW, Estrada J, Forsingdal

A, Friedrich M, Ganesham S, Hall L, Haslinger D, Huckins L, Loken E, Malan-Müller S, Martin J, Misiewicz Z, **Pagliaroli L**, Pardiñas AF, Pisanu C, Quadri G, Santoro ML, Shaw AD, Ranlund S, Song J, Tesli M, Tropeano M, van der Voet M, Wolfe K, Cormack FK, DeLisi L. Psychiatr Genet. 2016 Feb;26(1):1-47. doi: 10.1097/YPG.0000000000000112. PubMed PMID: 26565519. **IF 1.557**