Màster Universitari
Anàlisi de Dades Òmiques / Omics Data Analysis
FACULTAT DE CIÈNCIES I TECNOLOGIA
UVIC | UVIC·UCC

Master of Science in Omics Data Analysis

Master Thesis

Analysis of the shared genetic architecture of schizophrenia and fertility

by

Óscar Baeza Francisoud

Supervisor: Elena Bosch Fusté

Co-supervisor: Gerard Muntané Medina

Academic tutor: Josep M. Serrat

Biosciences Department

University of Vic – Central University of Catalonia

15/09/2020

Acknowledgements

Throughout the writing of this Final Master Thesis I have received a great deal of support and assistance. I would like first to thank my supervisor, Dr. Gerard Muntané Medina, whose expertise was invaluable in the formulating of the research topic and methodology in particular. Also, to Dra. Elena Bosch Fusté, who trusted me and opened the doors of her research unit to a master student in search of an engaging research project.

During this time I spent doing this thesis the COVID-19 epidemic was in its first wave, so I will also thank all the workers who made possible that while we were at home just bored, we still had food in the market, running water and many other thigs we have nowadays as perennial, until something takes us out of the routine.

Since I did work from home, I must also thank my parents for their wise counsel and sympathetic ear. You are always there for me. Finally, there are my friends, who were of great support in deliberating over our problems and findings, as well as providing happy distraction to take my mind away from my research.

Thanks to all of you.

Óscar Baeza

Abstract

Schizophrenia (SCZ) patients present an evolutionary paradox, as they exhibit strong negative effects on individual fitness, yet they persist in a worldwide prevalence of 1% approximately. Molecular mechanisms affecting fertility, which may be widely common among complex diseases with fitness effects, can be studied by the integrated analysis of data from genome-wide association studies (GWAS) of human fecundity together with any disease of interest. Here, we used GWAS summary statistics from SCZ and fertility traits to investigate SNPs significantly conjoined between both SCZ and these fertility traits. We integrated genome-wide association study (GWAS) data on SCZ (n = 105318) and GWAS data on fertility (Number of brothers and sisters, n = 380062; children fathered, n = 176258; pregnancy terminations, n = 65225; live births, n = 208434; children ever born, n = 343072). By conditioning the FDR on overlapping associations, this statistical approach increases power to discover genetic loci. We found 29 loci shared between SCZ and at least one fertility trait. The fertility trait with the most loci shared is the number of brothers.

Table of contents

1) Introduction

Schizophrenia (SCZ) is a severe psychiatric disorder that has a profound effect on both the individuals affected and society. Although outcomes might not be as uniformly negative as is commonly believed, more than 50% of individuals who receive a diagnosis have intermittent but long-term psychiatric problems, and around 20% have chronic symptoms and disability [1]. Unemployment is staggeringly high at 80–90%, and life expectancy is reduced by 10–20 years [2–5].

Schizophrenia is characterized by diverse psychopathology. The core features are positive symptoms (delusions and hallucinations; so-called psychotic symptoms in which contact with reality is lost), negative symptoms (particularly impaired motivation, reduction in spontaneous speech, and social withdrawal), and cognitive impairment (patients had poorer performance than controls over a wide range of cognitive functions, although much individual variability was reported) [6]. These positive symptoms tend to relapse and remit, but some patients have permanent or residual psychotic symptoms. The negative ones tend to be permanent and hinder the social and personal abilities of the patient [7].

Diagnosis is made clinically based on the history and by examination of the mental state; no diagnostic tests or biomarkers are available. Schizophrenia usually presents with psychosis; according to the fifth edition of Diagnostic and Statistical Manual of Mental Disorders [8] ,the main differential diagnoses are affective psychoses like bipolar disorder, other closely related non-affective psychoses like delusional disorder, brief psychotic disorder and other major psychotic disorders caused by alcohol or other general illness. Schizophrenia is typically diagnosed in the late teen years to the early thirties and tends to emerge earlier in males (late adolescence – early twenties) than females (early twenties – early thirties) [9,10].

Many genetic epidemiological studies have shown, for more than 50 years, that genetic factors contribute substantially, but not exclusively, to the underlying cause of schizophrenia. Schizophrenia is among the most heritable psychiatric disorders (heritability ~ 70%). This high heritability points to a major role for inherited genetic variants in the etiology of schizophrenia [11].

Schizophrenia is highly polygenic, as predicted in the 1960s on the basis of genetic epidemiological findings. [12] Genome-wide association studies have identified more than 100 distinct genetic loci containing fairly common alleles of small effect and the en-masse effects of many hundreds of such loci, suggesting that single-nucleotide polymorphisms (SNPs) with a range of population frequencies contribute to risk of developing the disease. This risk seems to be highly pleiotropic (one gene or allele can affect multiple seemingly unrelated phenotypic traits) and does not map onto existing definitions of disease. A study showed significant sharing of common risk variants between schizophrenia and bipolar disorder, depressive disorder, autism spectrum disorder and other psychotic alterations [13].

Despite this genetic complexity, they converge into a set of biological processes being affected. These elements have been reported in genes encoding a range of synaptic proteins like glutamate receptors, the voltage-dependent calcium channel family of proteins, and dopamine receptor D2 (DRD2), which is the principal target of antipsychotic drugs [4,14,15].

The neurobiology of schizophrenia is still poorly understood, but strong evidence implicates dysfunction of dopaminergic neurotransmission in the genesis of psychotic symptoms, and that abnormalities of glutamate signaling might be responsible for the negative and cognitive symptoms [4,16]. Disturbances of synaptic function might underlie abnormalities of neuronal connectivity, possibly through effects on interneurons, but not more is known. Progression towards schizophrenia can be triggered by postnatal environmental exposures—which might be modulated by genetic factors and environmental factors in early development—and in some cases by oxidative and inflammatory mechanisms [17].

Since the accidental discovery of chlorpromazine more than 50 years ago, almost all antipsychotic drugs available in the clinical setting for schizophrenia work via DRD2 blockade [14]. Other antipsychotics, of which clozapine is the most potent, binds and affects not only DRD2 but also other neurotransmitter receptors, such as serotonin receptors 2 (5HT-2R). These drugs relatively effective, compared with placebo, in reducing positive symptoms and are the backbone of all SCZ treatment. However, they are not effective negative symptoms and cognitive dysfunction, clinical features which are more strongly associated with functional impairment than are positive symptoms.

Although antipsychotic drugs remain the main treatment approach, a good management will require pharmacotherapy to be embedded within a framework of strong psychological and social support, like approaches aimed at improving adherence, vocational and educational support, and rehabilitation [18,19].

The relative risk for suicide is increased by 12 times, with a lifetime risk of roughly 6~5%, [20] but mortality from most natural causes, especially cardiovascular disorders, is the strongest contributor to the 10–20 year reduction in life expectancy. According to evolution theory the process of natural selection preserves genetic variants associated with survival and reproductive advantage (fitness), while genetic variants associated with low fitness are eliminated from the gene pool [21]. We still do not know why a psychiatric disease with a high heritability and associated with a large fitness reduction like schizophrenia has not been erased from the gene pool by natural selection, and this question is known as the evolutionary paradox of psychiatric disorders [22].

There are some answers proposed to this question, one of them being the antagonistic pleiotropy hypothesis. This hypothesis was proposed by George C. Williams, an evolutionary biologist, as an explication for senescence. This hypothesis is based in pleiotropy, where one gene is controlling more than one trait and one of these traits is beneficial to the organism early in it's life, and later on, another linked trait is detrimental to the individual fitness due to the loss of force of natural selection. However, this hypothesis does not explain the fact that SCZ appears early in life, and thus could affect greatly the fertility or reproduction of the individual [23–25] .

Here, we aim to investigate the genetic trade-off between SCZ and fitness by integrating data from genome-wide association studies (GWAS) on the disease and several fertility traits. Our hypothesis is that an advantage in fertility could offset the negative effects of the disease and be one of the reasons why SCZ still has a high prevalence and their genetic elements have not been eliminated from the human genetic pool in previous times [26–28].

With everything exposed above, we will try to first analyze the shared genetic architecture –or pleiotropy– between human fertility and the schizophrenia and secondly to dissect the evolutionary forces affecting all potential agonistic and antagonistic pleiotropic variants between human fertility and schizophrenia. To test for this shared genetic architecture, we will use GWAS summary statistics data and the pleiotropy-informed conjunctional false discovery rate approach to identify which SNPs can be jointly associated to the two phenotypes. Regarding fertility, we will use several indicators such as number of children ever born (NEB) n = 343,072 [13], number of live births (BIRT, female) N = 208,434, number of children fathered (FATH, male) N = 176,258, number of full brothers (BROT) and number of full sisters (SIST) with N = 380,062, and finally number of pregnancy terminations (TERM, female) with an N = 65,225 [12]. To get the schizophrenia SCZ statistics, we will use another GWAS [29] with N = 105,318. These datasets are the most recent ones we found, having the biggest N in its data type.

2) Methods

Due to the COVID-19 pandemic ravaging the world during the time this Final Master Project was conducted, we had to work from home. This meant downloading a VPN known as FortiClient, and from there connect to the UPF Marvin cluster via PuTTY, a free implementation of SSH and Telnet for Windows developed by Simon Tatham.

The Marvin cluster has a total of 27 computing nodes, 720 cores, 1,4 TB of total RAM and about 7,4 TB of disk space. The Global Disk Status is divided in three main groups, homes with 20 TB, scratch with 283 TB and projects with 357 TB. We worked in our directory in group scratch. The OS running in the computing nodes is CentOS 7.x x86_64. We used this high computational power cluster due to the massive scale of data we were processing, and because the software we used required also great quantities of storage and RAM. In order to submit our analysis to the queue system we had to build a script and run it through sbatch. The script we used to run our analysis is available on the GitHub https://github.com/obaeza16/TFM .

The software we used to run the analysis was pleioFDR, developed by Andreassen et al. [30], which will allow us to identify genetic loci jointly associated with two phenotypes, in this case, SCZ and fertility. The code can be accessed at https://github.com/precimed/pleiofdr and we cloned and used the code as it was stored in GitHub on 15/05/2020.

The Summary Stats of the SCZ GWAS was downloaded from the Walters Group Data Repository (https://walters.psycm.cf.ac.uk/) , and the data from the fertility traits was stored basically in the GWAS Atlas (https://atlas.ctglab.nl/) , except for the number of children ever born pooled, which we downloaded from the Data Sharing portal of the Social Science Genetic Association Consortium (https://www.thessgac.org/data) . The GWAS atlas trait number for our traits are 3341, 3282, 3283, 3326 and 3378 for N live births, N full brothers, N full sisters, N children fathered and N pregnancy terminations, respectively.

As we were working in a cluster, before using any external program we had to load it, and it was done with both Matlab and Python 3.6, with this command:

```
module load {program you want to use}
```

These datasets were in a `.txt` format, and we needed to convert them to csv and later to .mat in order to run the analysis script. This was done via the pleioFDR software too, with the following command:

```
python3.6 python_convert/sumstats.py csv --auto --sumstats traitfolder/trait.txt.gz --n-val {n-val of the original GWAS} --out traitfolder/trait.csv --force
```

From the above command we already got a `.csv` file with the columns we needed to do the analysis correctly, but now we faced another important problem, the lack of quality on some SNP, which can make our script more time consuming but at the same time not giving any more relevance to the results. To do this, we will download and compare with our full SCZ SNPs the names of the high-quality imputed (INFO > 0.9) markers, in LD-Score. We will do this with this command in awk:

```
awk 'FNR==NR{a[$1];next}($1 in a){print}' good.quality.SNPs SCZ.csv > SCZ.goodinfo.csv
```

Now we had the good quality SNPs alone, and we could continue. Now we needed to export the SNPs names in the rs format from the SCZ data to the other datasets, at the same time keeping only the SNPs from which we have good info. To do this, we had to do several steps towards having the appropriate data formats everywhere. To start, we sorted by base pair (third column) the trait dataset:

```
sort -k3 -n trait.csv > trait.sorted.csv
```

Now we used another awk command, this time to select and keep only the SNPs present in both datasets, and we did this comparing the location in the chromosome, the base pair and the ancestral and derived allele, with this command:

```
awk '{idx = $2 SUBSEP $3 SUBSEP $5 SUBSEP $6}FNR==NR
{f1[idx]=$1;next}(idx in f1) && (!seen[$1,$2,$3]++)
{sub(/^[^[:space:]]+/,f1[idx]);print}' SCZ.sorted.goodinfo.csv
trait.sorted.csv > SCZ.vs.trait.csv
```

After this, we needed to extract the SNP rs number from this file, and we did it with these commands. The first extracts the rs list and the second one adds the header “SNP” to this list:

```
tail -n +2 SCZ.vs.trait.csv | cut -d':' -f1 > trait.rs.list.csv

sed '1 i\SNP' trait.rs.list.csv > trait.SNP.list.csv
```

Now we copied the SNP list onto the file which had the good data but with the wrong format:

```
awk 'BEGIN {OFS="\t"}; FNR==NR{a[NR]=$1;next}{$1=a[FNR]}1'
trait.SNP.list.csv SCZ.vs.trait.csv > SCZ.vs.trait.ready.csv
```

Now we had the good data and in a good format, the only thing left is adding the z-score to the dataset, which is needed to transform into a `.mat` object. It was done doing this command:

```
python3.6 python_convert/sumstats.py zscore --sumstats
SCZ.vs.trait.ready.csv --out trait.z.ready.csv
```

Now we were able to transform the `.csv` files with the good data and format into a `.mat`, which is the data type Matlab uses. We did that with this command:

```
python3.6 python_convert/sumstats.py mat --sumstats trait.z.ready.csv
--ref 9545380.ref --out trait.mat
```

In the command above we must consider that the `.ref` must be downloaded from the pleioFDR GitHub and it is basically a reference file of 9545380 human SNPs.

Once all the above is done, we had one .mat for every trait, and we were able to continue and run the pleioFDR software. To configurate this software we had to copy the file `config_default.txt` and copy and edit this second file so that `reffile` points to the `ref9545380_1kgPhase3eur_LDr2p1.mat` file and `traitfolder` points to folder containing all the `.mat` files.

With this, we wrote our slurm script, named `oscar_script.sh` and we send it to the jobs queue. Inside this script we told the cluster what to run, how to run it, how much memory to ask for, where to store the results, etc. We ran a conjunction FDR with a threshold of 0.01. We excluded the MHC and chr8 inversion. Once the script had been sent, and after waiting about 20 minutes to complete, we had the results, available in the next section of this report.

The `config.txt` that we used is available on the GitHub https://github.com/obaeza16/TFM , alongside the unedited final results and the slurm script we used.

3) Results

After performing the bioinformatical analysis described above, we have obtained several results that we will present in the form of the following Tables and Figures.

Table 3.1 Conjunctional FDR<0.01 by loci of the relation between schizophrenia vs N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations.

locusnum	snpid	geneid	chrnum	chrpos	pval_SCHI	fdr_SCHI	conjfdr_SCHI_FATH	conjfdr_SCHI_BORN	conjfdr_SCHI_BROT	conjfdr_SCHI_SIST	conjfdr_SCHI_BIRT	conjfdr_SCHI_TERM	min_conjfdr
1	rs1289870	LINC01525	1	117864047	7.386652e-05	2.899757e-01	1.416000e-01	NaN	3.265214e-03	8.108029e-01	3.972575e-03	1.000000e+00	3.265214e-03
1	rs1289873	LINC01525	1	117864775	7.952722e-05	2.964266e-01	1.368318e-01	7.020194e-03	3.208022e-03	6.492621e-01	4.230493e-03	1.000000e+00	3.208022e-03
2	rs10737693	DENND1B	1	197599030	1.626841e-04	3.779461e-01	3.382004e-01	8.936103e-02	3.590492e-03	4.618029e-01	5.025218e-01	8.044280e-01	3.590492e-03
3	rs1996610	VSNL1	2	17754316	6.899313e-05	2.840259e-01	1.945539e-01	9.995175e-01	7.316542e-03	8.541433e-01	5.140070e-01	8.040049e-01	7.316542e-03
4	rs3856505	KIAA2012	2	203028129	1.052539e-03	6.852207e-01	6.905799e-01	NaN	9.204251e-03	9.999506e-01	1.000000e+00	6.858585e-01	9.204251e-03
5	rs10511078	CADM2	3	85578035	6.736412e-04	6.231124e-01	1.449134e-02	2.886832e-02	6.768797e-03	8.224621e-01	2.468404e-02	6.835429e-01	6.768797e-03
6	rs6438253	AC026341	3	115115932	9.823619e-05	3.173366e-01	1.778975e-01	NaN	2.947363e-03	6.864736e-01	5.137695e-01	9.976309e-01	2.947363e-03
7	rs711945	AC092969	3	135111050	3.329277e-04	4.872502e-01	9.919736e-03	6.846862e-01	2.311569e-01	9.216480e-01	6.564631e-01	6.931879e-01	9.919736e-03
8	rs4683831	CLSTN2	3	140114278	3.826317e-04	5.121660e-01	5.672025e-01	1.000000e+00	4.543780e-03	8.732382e-01	5.335101e-01	8.046393e-01	4.543780e-03
9	rs4859871	AC104701	4	78628215	9.332012e-04	6.714735e-01	8.032913e-01	9.996350e-01	8.479972e-03	9.987903e-01	7.916507e-01	8.425132e-01	8.479972e-03
10	rs4302468		4	105124943	8.848636e-04	6.649932e-01	8.620143e-01	NaN	8.179752e-03	6.039880e-01	7.341732e-01	6.909415e-01	8.179752e-03
11	rs67928080		4	137433296	2.114690e-05	2.114077e-01	1.157734e-03	NaN	5.516346e-03	8.621643e-01	5.901735e-01	9.926793e-01	1.157734e-03
11	rs4479761		4	137448587	2.520201e-05	2.173284e-01	7.737934e-04	9.734934e-01	9.251014e-03	9.781736e-01	5.327670e-01	1.000000e+00	7.737934e-04
12	rs3778607	IRF4	6	403799	5.726873e-05	2.685998e-01	5.284021e-01	1.000000e+00	6.486111e-03	6.549940e-01	7.875967e-01	7.136096e-01	6.486111e-03
12	rs10900950	AL512308	6	462968	2.121693e-07	2.295475e-02	3.341673e-01	NaN	1.032905e-05	9.997580e-01	1.000000e+00	9.714628e-01	1.032905e-05
13	rs12206548	EXOC2	6	506231	1.420860e-04	3.601616e-01	5.381715e-01	3.992604e-01	6.027433e-03	7.786576e-01	6.904381e-01	6.624610e-01	6.027433e-03
14	rs6557170	ESR1	6	152203104	2.748396e-07	2.384916e-02	2.101518e-01	1.471119e-03	1.290410e-03	5.909152e-01	3.119591e-05	7.726548e-01	3.119591e-05
14	rs9479138	ESR1	6	152215199	1.681706e-05	2.042170e-01	6.838552e-03	NaN	4.755708e-04	6.901343e-01	9.829734e-04	9.810727e-01	4.755708e-04
14	rs4869748	ESR1	6	152225383	1.461123e-05	1.994413e-01	7.688721e-03	6.917429e-04	4.275107e-04	6.929180e-01	8.581876e-04	9.876257e-01	4.275107e-04
14	rs2347923	ESR1	6	152227421	8.952384e-06	1.763692e-01	1.623226e-02	7.106757e-03	2.922478e-04	7.193662e-01	5.321359e-04	8.269381e-01	2.922478e-04
15	rs717698		7	8394248	7.776312e-05	2.944571e-01	5.834549e-01	8.834286e-01	2.638429e-03	6.719327e-01	1.000000e+00	7.396930e-01	2.638429e-03
16	rs61585310	LHFPL3	7	104006510	8.453057e-05	3.024652e-01	3.095334e-01	NaN	6.318550e-03	7.443075e-01	7.028653e-01	8.566100e-01	6.318550e-03
17	rs10112601	ASPH	8	62494439	3.647288e-04	5.035856e-01	7.634892e-01	NaN	7.304636e-03	7.984954e-01	1.000000e+00	9.830777e-01	7.304636e-03
18	rs2030485		8	84178565	1.123624e-03	6.920157e-01	9.999736e-01	2.527067e-01	9.630599e-03	6.698877e-01	9.457603e-01	7.404123e-01	9.630599e-03
19	rs4742488	PTPRD	9	8371405	1.587194e-04	3.744521e-01	5.484943e-01	5.384713e-01	2.422096e-03	8.114604e-01	6.724505e-01	1.000000e+00	2.422096e-03
20	rs7924036	JMJD1C	10	65191645	3.169782e-05	2.277143e-01	9.997056e-01	NaN	7.547900e-03	6.994863e-01	5.591526e-01	8.430031e-01	7.547900e-03
21	rs915345	NRG3	10	84459894	2.349424e-05	2.149372e-01	5.849653e-01	NaN	4.642878e-03	9.956660e-01	8.035147e-01	4.871535e-01	4.642878e-03
22	rs7926362	BDNF	11	27707486	2.885244e-04	4.633870e-01	7.088636e-01	NaN	3.722944e-03	1.264266e-01	6.039294e-01	8.962790e-01	3.722944e-03
23	rs6484580		11	32515108	7.536007e-05	2.917207e-01	9.318954e-01	5.925778e-01	9.896911e-03	9.123168e-01	5.696283e-01	9.049674e-01	9.896911e-03
24	rs7957408	RN7SKP11	12	96821296	2.621700e-04	4.480831e-01	9.997475e-01	9.492158e-01	3.473633e-03	7.848888e-01	3.496341e-01	6.526806e-01	3.473633e-03
25	rs11618532	MYO16	13	109809622	5.085924e-05	2.592453e-01	6.355906e-01	9.755030e-01	8.462310e-03	9.999227e-01	1.000000e+00	6.854156e-01	8.462310e-03
26	rs2457192	SNX29	16	12197441	2.402673e-05	2.156887e-01	4.543605e-01	NaN	1.553866e-03	6.694732e-01	5.077996e-01	5.582456e-01	1.553866e-03
27	rs1035578	SNX29	16	12531365	1.155356e-04	3.350678e-01	4.212862e-01	6.006234e-01	5.624239e-03	6.321435e-01	8.833573e-01	6.652525e-01	5.624239e-03
28	rs4792267	VAMP2	17	8061337	1.645072e-04	3.795243e-01	9.993569e-01	NaN	7.248794e-02	8.236236e-01	8.055354e-03	8.349310e-01	8.055354e-03
29	rs157580	TOMM40	19	45395266	5.113185e-04	5.693796e-01	9.515384e-01	1.000000e+00	5.572074e-03	9.998106e-01	9.949259e-01	6.934765e-01	5.572074e-03

Table 3.2 Allelic variance in loci between schizophrenia vs N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations.

locus	snpid	geneid	chrn	chrpos	conjfdr_SCHI_FAT	conjfdr_SCHI_BO	conjfdr_SCHI_BROT	conjfdr_SCHI_SIST	conjfdr_SCHI_BIRT	conjfdr_SCHI_TERM
1	rs1289870	LINC01525	1	117864047	1.416000e-01	NaN	C a T	8.108029e-01	C a T	1.000000e+00
1	rs1289873	LINC01525	1	117864775	1.368318e-01	T a C	T a C	6.492621e-01	T a C	1.000000e+00
2	rs10737693	DENND1B	1	197599030	3.382004e-01	T a C	T a C	4.618029e-01	5.025218e-01	8.044280e-01
3	rs1996610	VSNL1	2	17754316	1.945539e-01	9.995175e-01	A a G	8.541433e-01	5.140070e-01	8.040049e-01
4	rs3856505	KIAA2012	2	203028129	6.905799e-01	NaN	A a G	9.999506e-01	1.000000e+00	6.858585e-01
5	rs10511078	CADM2	3	85578035	G a A	G a A	G a A	8.224621e-01	G a A	6.835429e-01
6	rs6438253	AC026341	3	115115932	1.778975e-01	NaN	A a G	6.864736e-01	5.137695e-01	9.976309e-01
7	rs711945	AC092969	3	135111050	A a G	6.846862e-01	2.311569e-01	9.216480e-01	6.564631e-01	6.931879e-01
8	rs4683831	CLSTN2	3	140114278	5.672025e-01	1.000000e+00	G a T	8.732382e-01	5.335101e-01	8.046393e-01
9	rs4859871	AC104701	4	78628215	8.032913e-01	9.996350e-01	T a C	9.987903e-01	7.916507e-01	8.425132e-01
10	rs4302468		4	105124943	8.620143e-01	NaN	G a A	6.039880e-01	7.341732e-01	6.909415e-01
11	rs67928080		4	137433296	A a G	NaN	A a G	8.621643e-01	5.901735e-01	9.926793e-01
11	rs4479761		4	137448587	T a C	9.734934e-01	T a C	9.781736e-01	5.327670e-01	1.000000e+00
12	rs3778607	IRF4	6	403799	5.284021e-01	1.000000e+00	A a G	6.549940e-01	7.875967e-01	7.136096e-01
12	rs10900950	AL512308	6	462968	3.341673e-01	NaN	C a T	9.997580e-01	1.000000e+00	9.714628e-01
13	rs12206548	EXOC2	6	506231	5.381715e-01	3.992604e-01	G a A	7.786576e-01	6.904381e-01	6.624610e-01
14	rs6557170	ESR1	6	152203104	2.101518e-01	A a G	A a G	5.909152e-01	A a G	7.726548e-01
14	rs9479138	ESR1	6	152215199	T a G	NaN	T a G	6.901343e-01	T a G	9.810727e-01
14	rs4869748	ESR1	6	152225383	T a G	6.917429e-04	T a G	6.929180e-01	T a G	9.876257e-01
14	rs2347923	ESR1	6	152227421	C a A	7.106757e-03	C a A	7.193662e-01	C a A	8.269381e-01
15	rs717698		7	8394248	5.834549e-01	8.834286e-01	G a A	6.719327e-01	1.000000e+00	7.396930e-01
16	rs61585310	LHFPL3	7	104006510	3.095334e-01	NaN	T a G	7.443075e-01	7.028653e-01	8.566100e-01
17	rs10112601	ASPH	8	62494439	7.634892e-01	NaN	T a C	7.984954e-01	1.000000e+00	9.830777e-01
18	rs2030485		8	84178565	9.999736e-01	2.527067e-01	C a T	6.698877e-01	9.457603e-01	7.404123e-01
19	rs4742488	PTPRD	9	8371405	5.484943e-01	5.384713e-01	T a G	8.114604e-01	6.724505e-01	1.000000e+00
20	rs7924036	JMJD1C	10	65191645	9.997056e-01	NaN	G a T	6.994863e-01	5.591526e-01	8.430031e-01
21	rs915345	NRG3	10	84459894	5.849653e-01	NaN	G a A	9.956660e-01	8.035147e-01	4.871535e-01
22	rs7926362	BDNF	11	27707486	7.088636e-01	NaN	A a C	1.264266e-01	6.039294e-01	8.962790e-01
23	rs6484580		11	32515108	9.318954e-01	5.925778e-01	G a T	9.123168e-01	5.696283e-01	9.049674e-01
24	rs7957408	RN7SKP11	12	96821296	9.997475e-01	9.492158e-01	C a A	7.848888e-01	3.496341e-01	6.526806e-01
25	rs11618532	MYO16	13	109809622	6.355906e-01	9.755030e-01	A a C	9.999227e-01	1.000000e+00	6.854156e-01
26	rs2457192	SNX29	16	12197441	4.543605e-01	NaN	C a A	6.694732e-01	5.077996e-01	5.582456e-01
27	rs1035578	SNX29	16	12531365	4.212862e-01	6.006234e-01	G a A	6.321435e-01	8.833573e-01	6.652525e-01
28	rs4792267	VAMP2	17	8061337	9.993569e-01	NaN	G a A	8.236236e-01	G a A	8.349310e-01
29	rs157580	TOMM40	19	45395266	9.515384e-01	1.000000e+00	G a A	9.998106e-01	9.949259e-01	6.934765e-01

Table 3.3 Odds Ratio of schizophrenia related to conjoined SNPs.

locusnum	snpid	geneid	chrnum	chrpos	OR SCZ
1	rs1289870	LINC01525	1	117864047	1.02
1	rs1289873	LINC01525	1	117864775	1.0203
2	rs10737693	DENND1B	1	197599030	1.0326
3	rs1996610	VSNL1	2	17754316	0.99065
4	rs3856505	KIAA2012	2	203028129	0.99923
5	rs10511078	CADM2	3	85578035	0.98955
6	rs6438253	AC026341	3	115115932	1.0178
7	rs711945	AC092969	3	135111050	0.98886
8	rs4683831	CLSTN2	3	140114278	0.99168
9	rs4859871	AC104701	4	78628215	1.0053
10	rs4302468		4	105124943	1.0229
11	rs67928080		4	137433296	1.009
11	rs4479761		4	137448587	1.0049
12	rs3778607	IRF4	6	403799	1.0201
12	rs10900950	AL512308	6	462968	1.0023
13	rs12206548	EXOC2	6	506231	1.0118
14	rs6557170	ESR1	6	152203104	1.0266
14	rs9479138	ESR1	6	152215199	1.0183
14	rs4869748	ESR1	6	152225383	1.0181
14	rs2347923	ESR1	6	152227421	1.0161
15	rs717698		7	8394248	1.0187
16	rs61585310	LHFPL3	7	104006510	1.0152
17	rs10112601	ASPH	8	62494439	0.98857
18	rs2030485		8	84178565	0.97871
19	rs4742488	PTPRD	9	8371405	1.0137
20	rs7924036	JMJD1C	10	65191645	0.98379
21	rs915345	NRG3	10	84459894	0.99648
22	rs7926362	BDNF	11	27707486	0.95757
23	rs6484580		11	32515108	0.99276
24	rs7957408	RN7SKP11	12	96821296	1.0153
25	rs11618532	MYO16	13	109809622	1.0008
26	rs2457192	SNX29	16	12197441	0.97904
27	rs1035578	SNX29	16	12531365	1.0211
28	rs4792267	VAMP2	17	8061337	0.9895
29	rs157580	TOMM40	19	45395266	0.99812

In Table 3.3 we can observe how the odds ratio of schizophrenia is related with the SNPs found conjoined with any fertility trait. We have 21 agonistic SNPs for SCZ and 14 antagonistic. The first are marked in red and the second in green.

Other results that support these tables are these following figures:

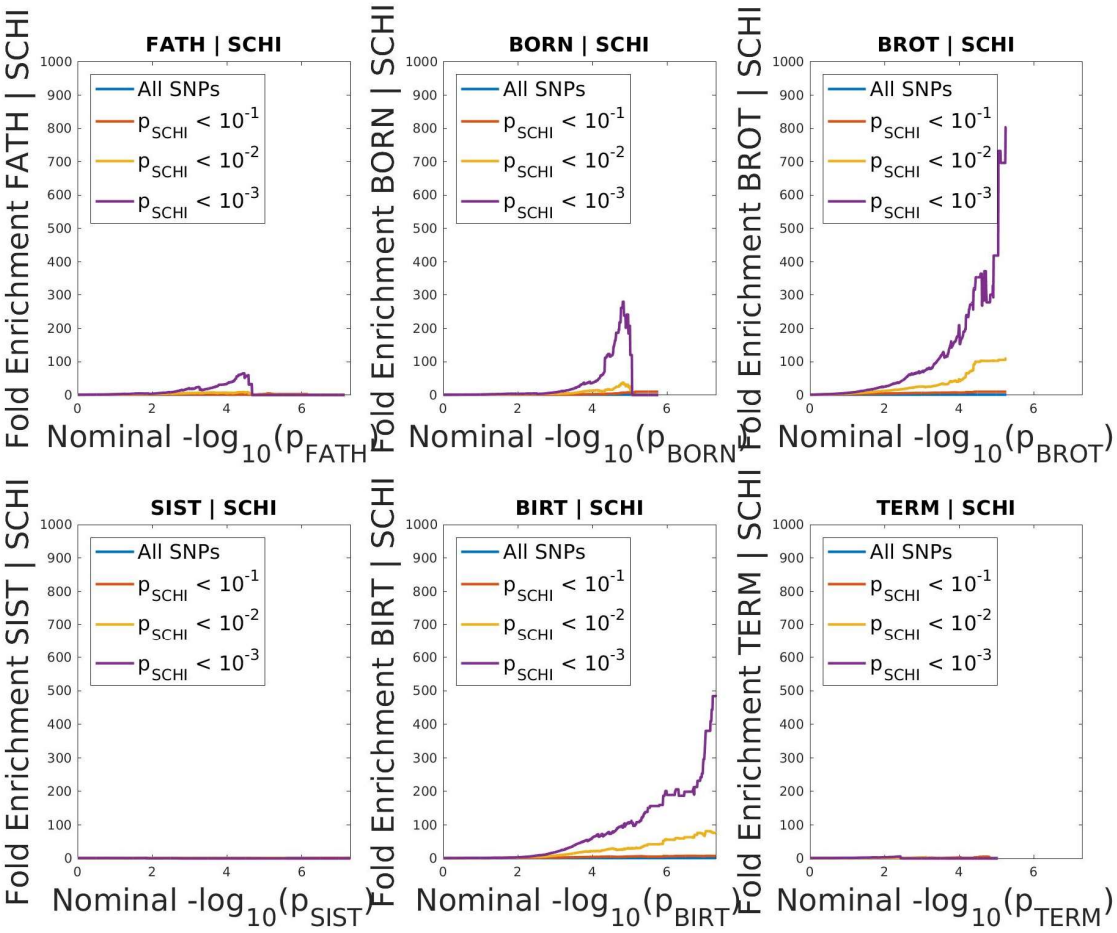

Figure 3.1. Fold Enrichment between N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations vs schizophrenia.

In this figure we can see the fold enrichment given by each relation. The greatest value and inclination we find it in the N of brothers graph, as expected following the Table 3.1. On the other hand, no enrichment is seen in both N of full sisters and N of pregnancy terminations, with N of children fathered and N of children ever born showing a minor enrichment.

The QQ plot with all the individual graphs for each conjunctional FDR is shown in Figure 3.2.

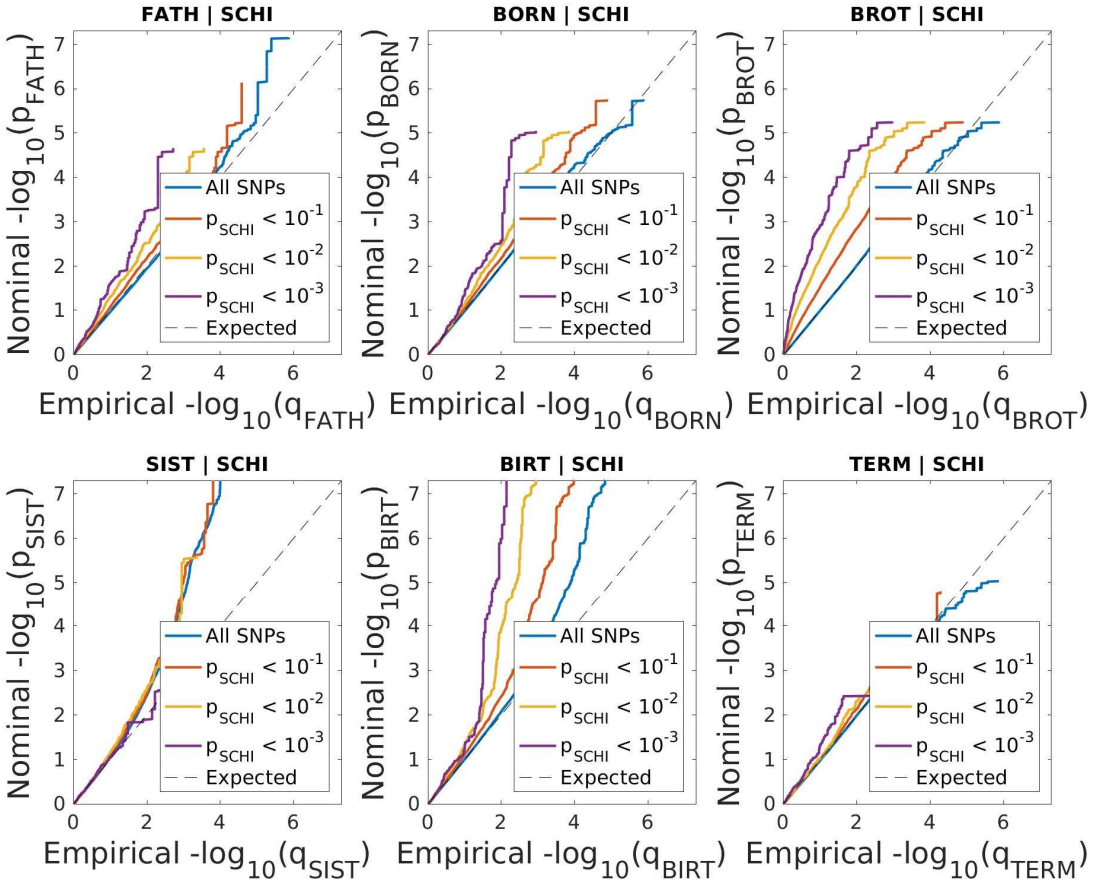

Figure 3.2 QQ plots of between N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations vs schizophrenia.

In agreement with the above results, we have the greatest diversion from the expected in the graphs relating N of brothers and N of children birthed with SCZ. The deviation we observe in N of full sisters and N of pregnancy terminations is only of those SNP with less p-value, so it is not a significant one. On the N of children fathered ad N of children ever born in conjunction, the significant SNPs are mildly more deviated and thus gives us the few significant ones in the Table 3.1.

The True Discovery Rate in each conjunctional analysis is plotted in Figure 3:

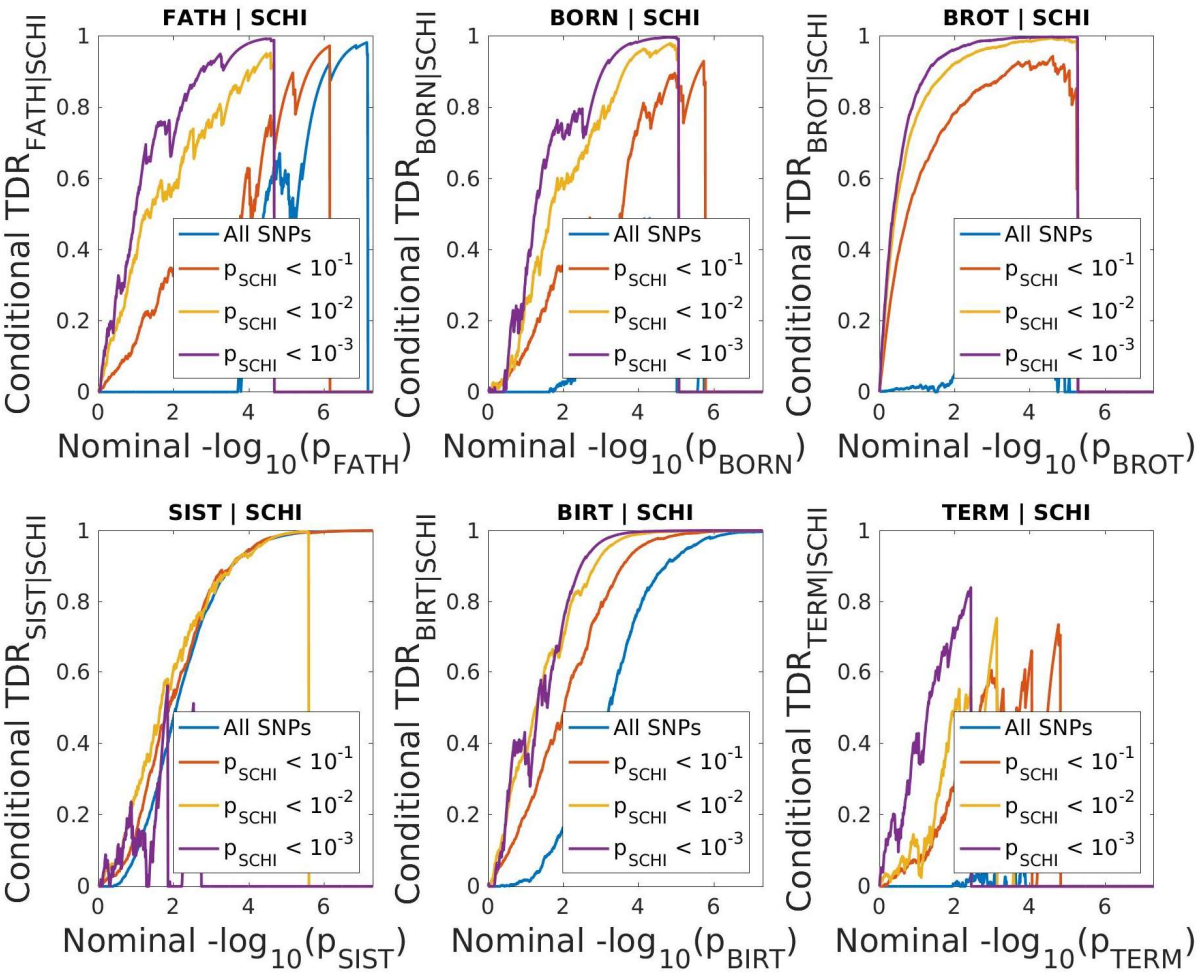

Figure 3.3 Conditional TDR between N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations vs schizophrenia.

Interpreting the graph, we obtain the same explanation as the Table 3.1 and other figures above. In N of children fathered vs SCZ we can see the TDR goes to value 1 on the most significant p-values, having at the same time lots of noise beneath due to the other SNPs. The shape of the N of children ever born vs SCZ graph is almost identical, although with less noise than the previous one. The best one is the N of brothers vs SCZ, going straight to 1 and with no noise. The N of full sisters vs SCZ does not get past 0.5 and the less significant p-values are the ones growing. On the N of children birthed vs SCZ we have lots of noise, but we get to 1 on the significance and finally the N of pregnancy terminations does not get to an acceptable threshold.

In addition to these figures, we also have the contrary, this is, the SCZ vs all the fertility terms. These figures are presented next:

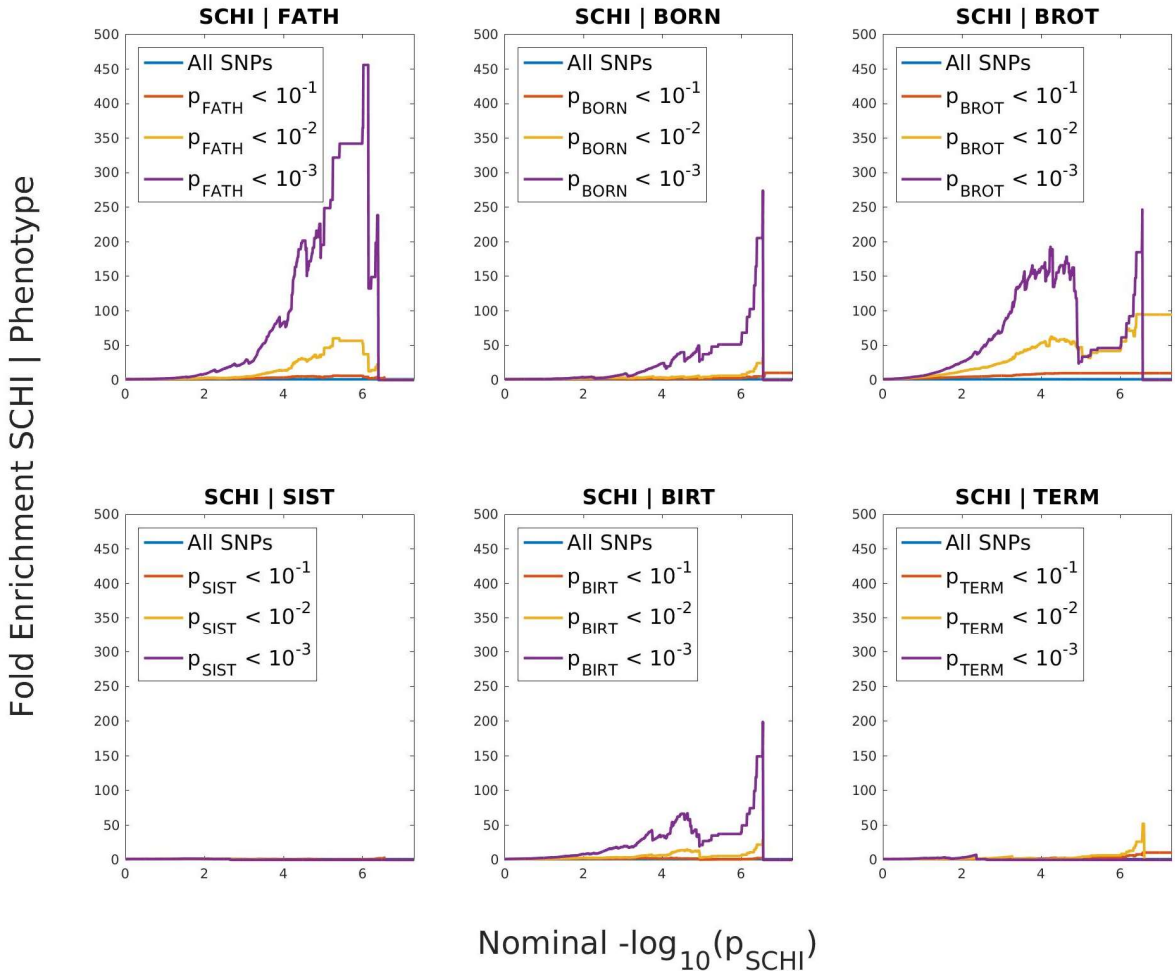

Figure 3.4 Enrichment between schizophrenia vs N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations.

The enrichment of the fertility terms in relation to SCZ is presented in Figure 3.4. As has been the norm, the enrichment of both N of full sisters and N of pregnancy terminations is nonexistent, and therefore their SNPs do not have any conjunction with SCZ. On the contrary, here we can see more clearly than in Figure 3.2 how much the terms PATH, n of children ever born , N of brothers and N of children birthed are affected by SCZ. An anomaly is presented that the highest peak is in relation to N of children fathered , this being not the trait with the most SNPs connected with SCZ.

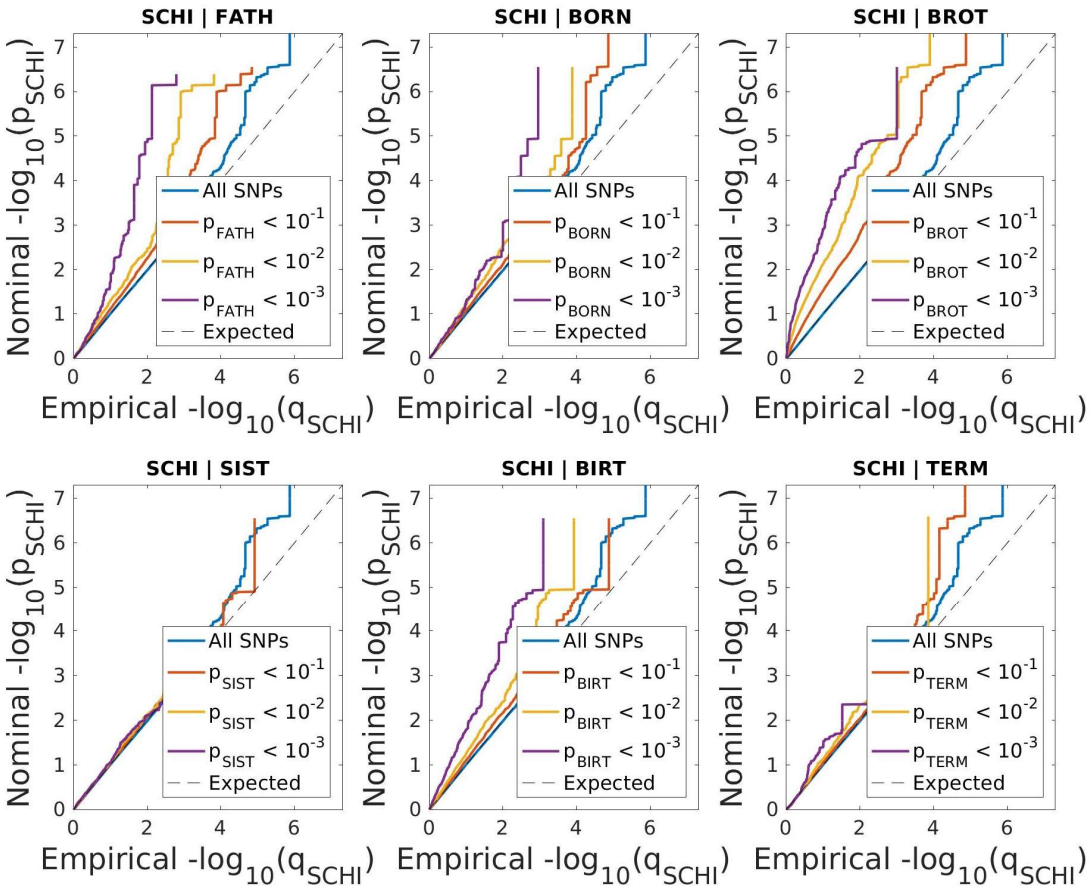

Figure 3.4 QQ plot between schizophrenia vs N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations.

In Figure 3.5 we can see the reverse of Figure 2. Here, as in Figure 4, we can see a great deviation of the expected on the relationship with N of children fathered , being this the greater deviation again in the set of plots. Again, N of full sisters and N of pregnancy terminations have no deviation of the significant SNPs and N of children ever born , N of brothers and N of children birthed have a significant deviation shown in the Table 3.1.

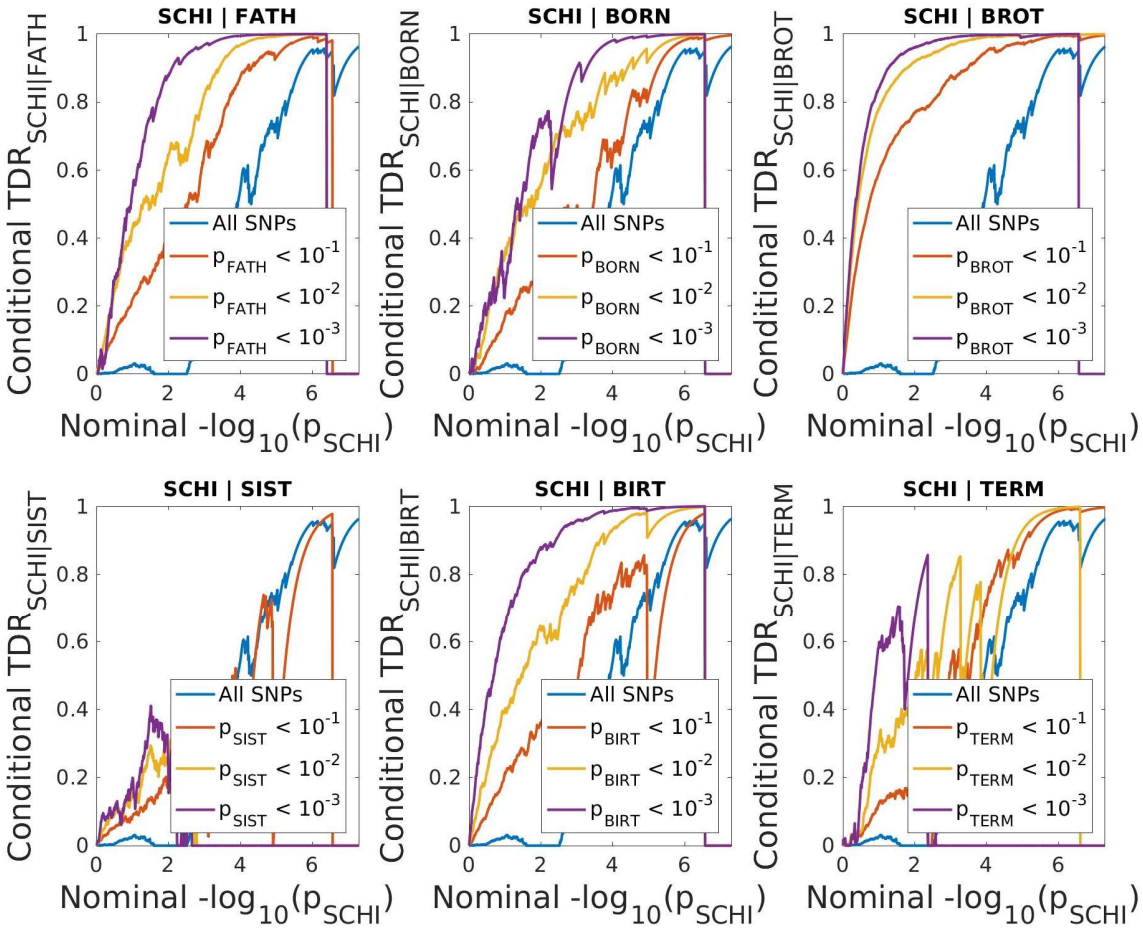

Figure 3.5 TDR between schizophrenia vs N of children fathered, N of children ever born, N of full brothers, N of full sisters, N of children birthed and N of pregnancy terminations.

In Figure 3.6 we see the true discovery rate of the conjunction between SCZ and the six fertility traits we have selected. N of full sisters and N of pregnancy terminations have no more than 0.4 the first and 0.8 the last with the most significant SNPs, but N of pregnancy terminations does not maintain this peak of 0.8. Despite this fact, N of children fathered , N of children ever born , N of brothers and N of children birthed get to an TDR of 1 and what is more important, maintain this level of TDR almost until a nominal -log10 of 6.4, which indicates that they are conjugated at least in some SNPs, as we can see in Table 3.1. Here the best TDR is disputed between N of brothers and N of children fathered , something we had not seen in the graphs depicting SCZ vs fertility, giving us an idea of how much these two traits are conjoined.

The first and most important result is the Table 3.1 of the conjunctional loci that the script has found. In it we can see the locus number associated with that SNP, the SNP id, chromosome, position in the chromosome, p-value for SCZ alone with its FDR and then we have the Conjunctional FDR for each interaction between traits. The NaN values in Tables 3.1 and 3.2 in the conjfdr_SCHI_BORN column are due to the not appearance of this SNP in the original GWAS from where we took the data.

The min_conjfdr column is made with the lower value from the conjunction FDRs done with the fertility traits. We can see we have several more SNPs than loci, because some SNPs are grouped together in the same region due to being near-independent genomic loci ($r2 < 0.2$).

We have loci from all chromosomes excepting 5, 14, 15, 18, 20, 21, 22, X and Y. Some SNPs we were not capable of mapping them to a gene, thus the white spaces. In green we have the significant p-values of the conjunctional FDR relating SCZ and each fertility trait.

We have 35 SNPs in total associated with any fertility trait. The conjunction between SCZ and N of full sisters does not have any SNP with a significant FDR, the same happens to the N of pregnancy terminations trait. The one fertility trait with the most significant SNPs is the N of brothers, with 34 out of 35 in total. The N of children fathered trait has 7 significant SNPs, one less than the N of children ever born. 6 significant SNPs are also related to the N of full brothers fertility trait. The gene with the most SNPs mapped is the ESR1. We have 4 SNPs mapped to it, remarking the importance of this gene.

On Table 3.2 we can observe the allelic change in each comparison derived from the SNP we have. This table will later reveal to us which effect has each SNP towards SCZ and its relationship with fertility.

On Table 3.3 we can observe that a majority of the 35 associated loci at conjFDR < 0.01 (60%) harbor alleles that increase the risk for SCZ. The rest of the present alleles that oppose the risk of SCZ. We must underline that no allele has an OR of more than 1.0326 on the agonistic side and 0.97871 on the antagonistic side, so not at all a great change in SCZ risk. However small, this change and its relation to fertility must be further researched in order to determine a clearer picture and relation between the two and answer with good data and analysis the evolutionary paradox presented in the Introduction.

The number of children fathered is related to 2 antagonistic SNPs and 5 agonistic for the risk of SCZ, so it is more associated with this risk than other traits. We have another case of a suspicious relationship in the trait of N of children ever born, being related with only 1 antagonistic and 6 agonistic. This list of agonistic includes the one with the greater OR for SCZ, rs10737693, located in the gene DENND1B. This gene codes for a guanine nucleotide exchange factor (GEF) for RAB35. It has activity in the protein transport of this protein, being involved in the intracellular membrane trafficking as well as in the regulation of immunity response.

Regarding the N of full brothers, it has only one less significant agonistic SNP, making it the trait more closely related to SCZ and therefore linked in a greater amount to the alleles carrying the SCZ risk. The N of full sisters is not related with any SNP, from which we can safely say that this fertility term has no relation with SCZ and thus has no relevance in our study.

The term N of children birthed has 6 agonistic SNPs and 1 antagonistic, putting it with many of the terms, more related to a higher risk than not. It has as the antagonistic SNP the rs10511078, located in the gene CADM2. This gene has been linked in recent studies to some psychological illness like neuroticism, mood instability and risk-taking, making it a possible part of the schizophrenia puzzle, affected by this SNP [31].

The term of number of pregnancy terminations we found no relation to SCZ, as the N of full sisters. This could be due to the low N we had from this particular GWAS, or that simply there is no biological relation between the two elements.

The gene with the most loci identified in this analysis is ESR1, an estrogen receptor vital in reproduction, and linked to 4 of 6 fertility traits. There are variants of this gene that have been linked to risk for schizophrenia [32].Its four identified SNPs in this analysis are all agonistic for risk, and thus support this hypothesis and relates to the terms of N of children fathered, N of full brothers and N of children birthed. These traits have at least 3 of the 4 SNPs located in this gene.

All the original results of the analysis can be consulted on the GitHub linked in the Section 2.

4) Conclusions

We have demonstrated in this study that genetic overlap exists between some fertility traits and SCZ. Our findings suggest that this genetic relationship could be providing an evolutionary escape route to avoid natural selection. There are traits with no overlap like the number of pregnancy termination or the number of full sisters, but other traits like the number of brothers are clearly playing a role in the inheritance of the genetic elements of schizophrenia.

Not all risk of developing schizophrenia comes from the genetic side of an individual, and maybe another psychosocial trait has an effect even bigger, and they are only being reflected in this study.

We need more studies on this disease, maybe diving deep into the evolutionary drive of these and other genetic variations could help shed more light at this psychological problem and help comprehend it and maybe later fight back.

5) References

1. Barbato, A. Psychiatry in transition: Outcomes of mental health policy shift in Italy. *Aust. N. Z. J. Psychiatry* **32**, 673–679 (1998).
2. Hiroeh, U., Appleby, L., Mortensen, P. B. & Dunn, G. Death by homicide, suicide, and other unnatural causes in people with mental illness: A population-based study. *Lancet* **358**, 2110–2112 (2001).
3. Joukamaa, M. *et al.* Mental disorders and cause-specific mortality. *Br. J. Psychiatry* **179**, 498–502 (2001).
4. Moghaddam, B. & Javitt, D. From revolution to evolution: The glutamate hypothesis of schizophrenia and its implication for treatment. *Neuropsychopharmacology* vol. 37 4–15 (2012).
5. Laursen, T. M., Nordentoft, M. & Mortensen, P. B. Excess early mortality in schizophrenia. *Annual Review of Clinical Psychology* vol. 10 425–448 (2014).
6. Cotter, J., Barnett, J. H. & Granger, K. The Use of Cognitive Screening in Pharmacotherapy Trials for Cognitive Impairment Associated With Schizophrenia. *Front. Psychiatry* **10**, 648 (2019).
7. Berrios, G. E. Positive and Negative Symptoms and Jackson: A Conceptual History. *Arch. Gen. Psychiatry* **42**, 95–97 (1985).
8. DSM-5. https://www.psychiatry.org/psychiatrists/practice/dsm.
9. Rapoport, J. L. & Gogtay, N. Childhood onset schizophrenia: Support for a progressive neurodevelopmental disorder. *International Journal of Developmental Neuroscience* vol. 29 251–258 (2011).
10. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: A concise overview of incidence, prevalence, and mortality. *Epidemiologic Reviews* vol. 30 67–76 (2008).
11. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
12. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
13. Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1–7 (2016).
14. Seeman, P., Lee, T., Chau-Wong, M. & Wong, K. Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature* **261**, 717–719 (1976).
15. Creese, I., Burt, D. R. & Snyder, S. H. Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. *Science (80-.).* **192**, 481–483 (1976).
16. Howes, O. D. & Murray, R. M. Schizophrenia: An integrated sociodevelopmental-cognitive model. *The Lancet* vol. 383 1677–1687 (2014).
17. Paolicelli, R. C. *et al.* Synaptic pruning by microglia is necessary for normal brain

development. *Science (80-.).* **333**, 1456–1458 (2011).

18. Falkai, P., Malchow, B. & Schmitt, A. Aerobic exercise and its effects on cognition in schizophrenia. *Current Opinion in Psychiatry* vol. 30 171–175 (2017).

19. Simonienko, K. *et al.* The reasons for use of cannabinoids and stimulants in patients with schizophrenia. *Psychiatr. Pol.* **52**, 261–273 (2018).

20. Nordentoft, M., Mortensen, P. B. & Pedersen, C. B. Absolute risk of suicide after first hospital contact in mental disorder. *Arch. Gen. Psychiatry* **68**, 1058–1064 (2011).

21. Darwin Online: On the Origin of Species. http://darwin-online.org.uk/EditorialIntroductions/Freeman_OntheOriginofSpecies.html.

22. van Dongen, J. & Boomsma, D. I. The evolutionary paradox and the missing heritability of schizophrenia. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **162**, 122–136 (2013).

23. Otto, S. P. Two steps forward, one step back: The pleiotropic effects of favoured alleles. *Proc. R. Soc. B Biol. Sci.* **271**, 705–714 (2004).

24. Cheverud, J. M. Developmental integration and the evolution of pleiotropyl. *Am. Zool.* **36**, 44–50 (1996).

25. Williams, G. C. Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution (N. Y).* **11**, 398 (1957).

26. Srinivasan, T. N. & Padmavati, R. Fertility and schizophrenia: Evidence for increased fertility in the relatives of schizophrenic patients. *Acta Psychiatr. Scand.* **96**, 260–264 (1997).

27. Ritsner, M., Sherina, O. & Ginath, Y. Genetic epidemiological study of schizophrenia: reproduction behaviour. *Acta Psychiatr. Scand.* **85**, 423–429 (1992).

28. Power, R. A. *et al.* Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *Arch. Gen. Psychiatry* **70**, 22–30 (2013).

29. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).

30. Andreassen, O. A. *et al.* Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate. *PLoS Genet.* **9**, e1003455 (2013).

31. Morris, J. *et al.* Genetic variation in CADM2 as a link between psychological traits and obesity. *Sci. Rep.* **9**, 1–14 (2019).

32. Weickert, C. S. *et al.* Variants in the estrogen receptor alpha gene and its mRNA contribute to risk for schizophrenia. *Hum. Mol. Genet.* **17**, 2293–2309 (2008).