University of Milano-Bicocca
Master Degree in Data Science
Academic Year 2023/2024

Comprehensive study of schizophrenia research and genetics

Presented by:
Fabio Marini, Paola Maria Cavana and Yuliia Tsymbal

Introduction

Schizophrenia is a severe mental disorder that impacts an individual's thoughts, emotions, and actions.

Those with schizophrenia may appear disconnected from reality, causing distress for themselves and their family and friends.

- affects approximately 4 per 1000 of the population
- occurs between the ages of 18 and 28
- the percentage of men and women affected is equal.

Positive symptoms

- Hallucinations
- Delusions
- Inability to organize one's thinking logically
- Bizarre behavior

Negative symptoms

- Little or no search for social relationships
- Apathy
- Impoverishment of language
- Inability to take decisions
- Difficulty of concentration

Goals

- **Characterize** schizophrenia
- **Analyse** research trends
- **Investigate** genetic links

Pipeline

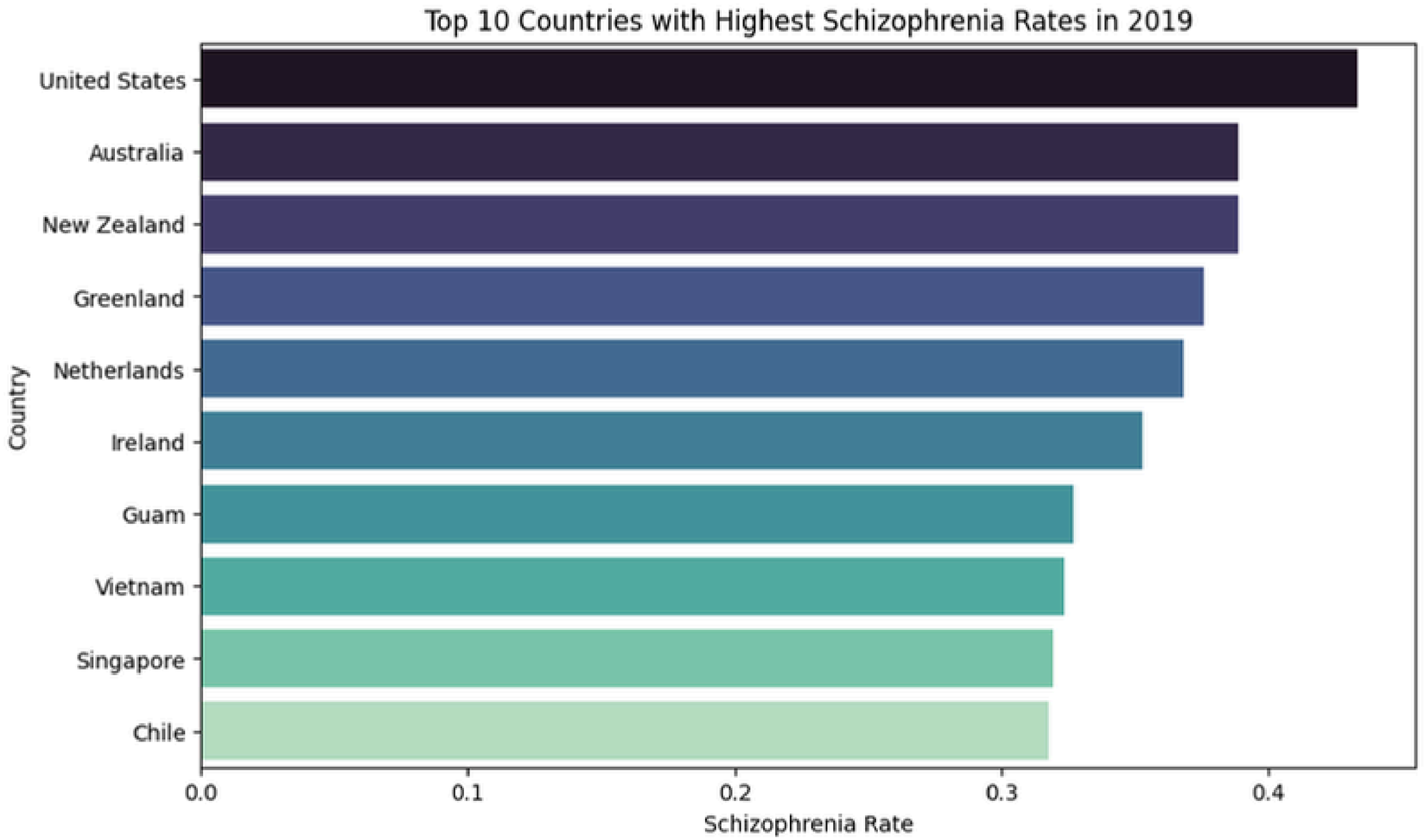
1. Introduction
2. Literature analysis
3. Topic modeling
4. Gene analysis
5. Gene study

Schizophrenia in the world

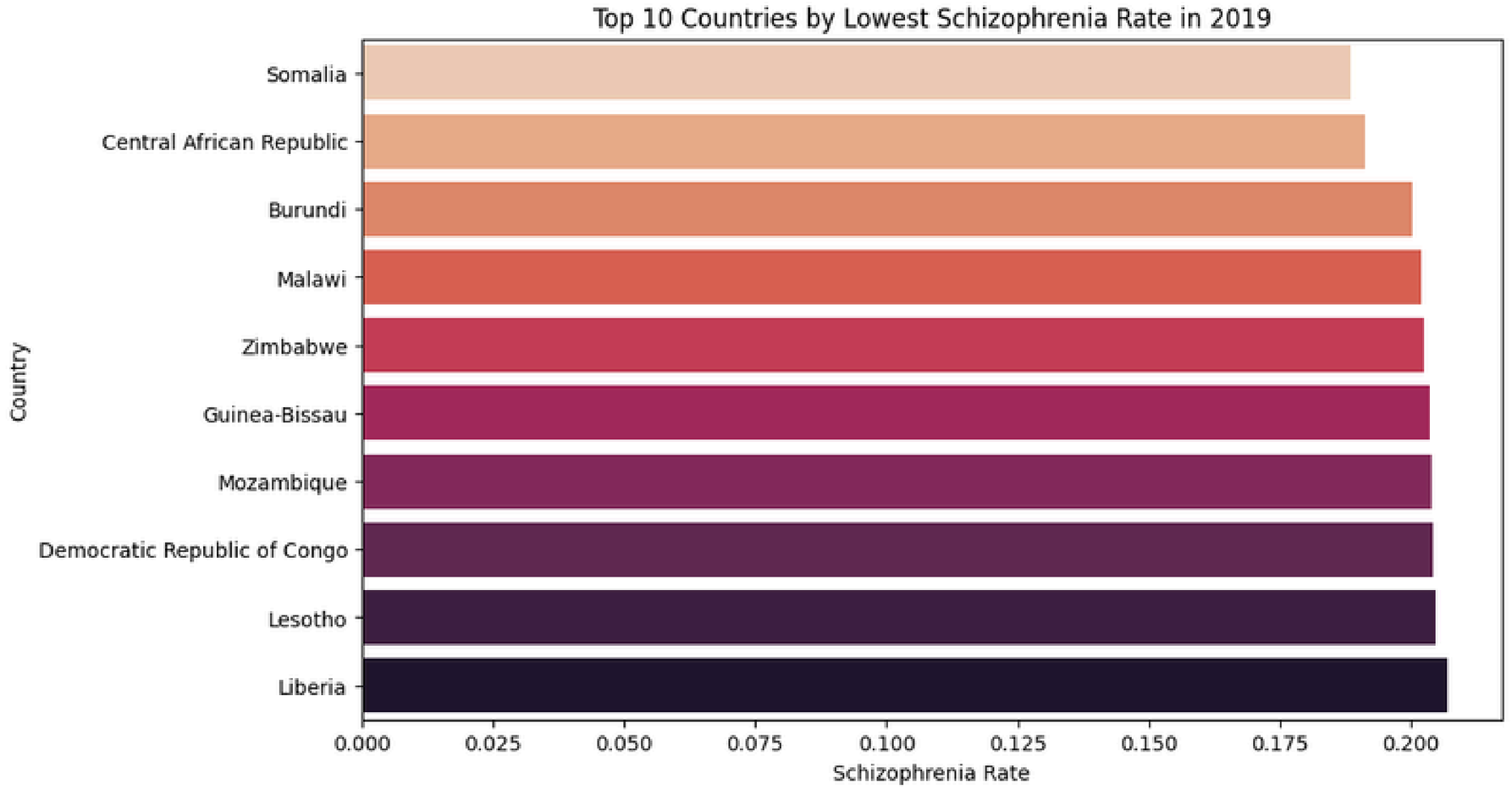
We used the "Mental Health" dataset from Kaggle as our primary data source.

As a first step, we filtered the dataset to focus specifically on schizophrenia data. We then identified and presented the top 10 countries with the highest rates of schizophrenia and the top 10 countries with the lowest rates.

Schizophrenia in the world

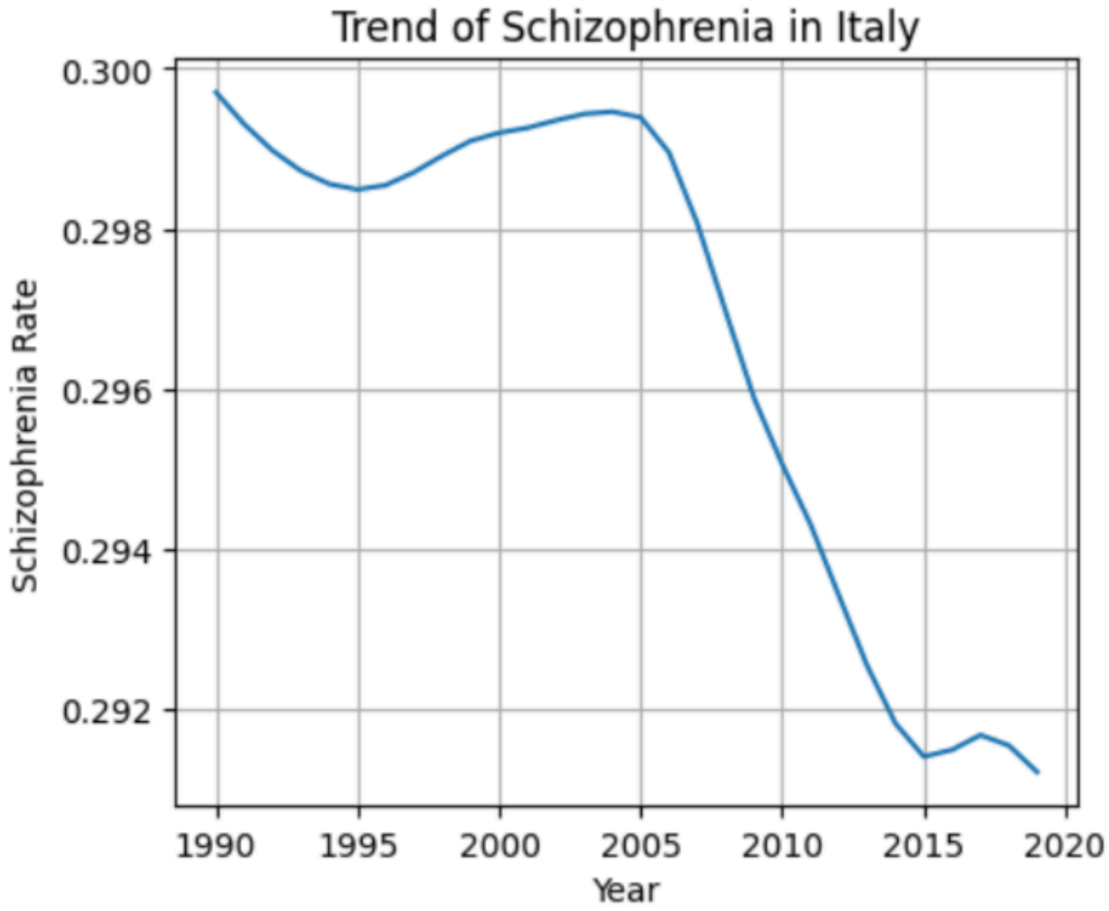


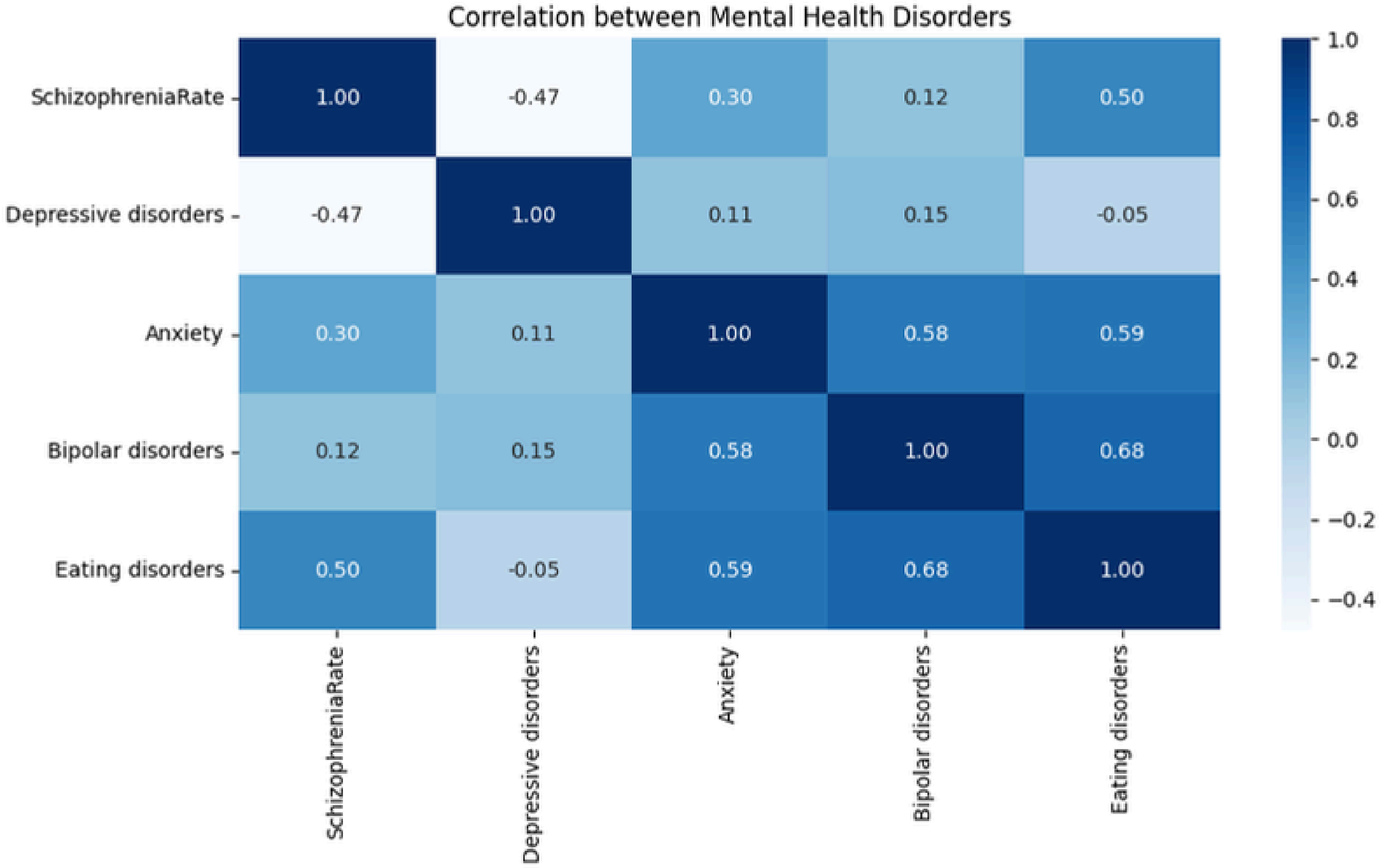
Schizophrenia in the world



Schizophrenia in the world

Then we visualized the **trends in schizophrenia rates** over the years across European countries. As an example, we showcase the data trend for Italy, illustrating how schizophrenia rates have evolved over time.

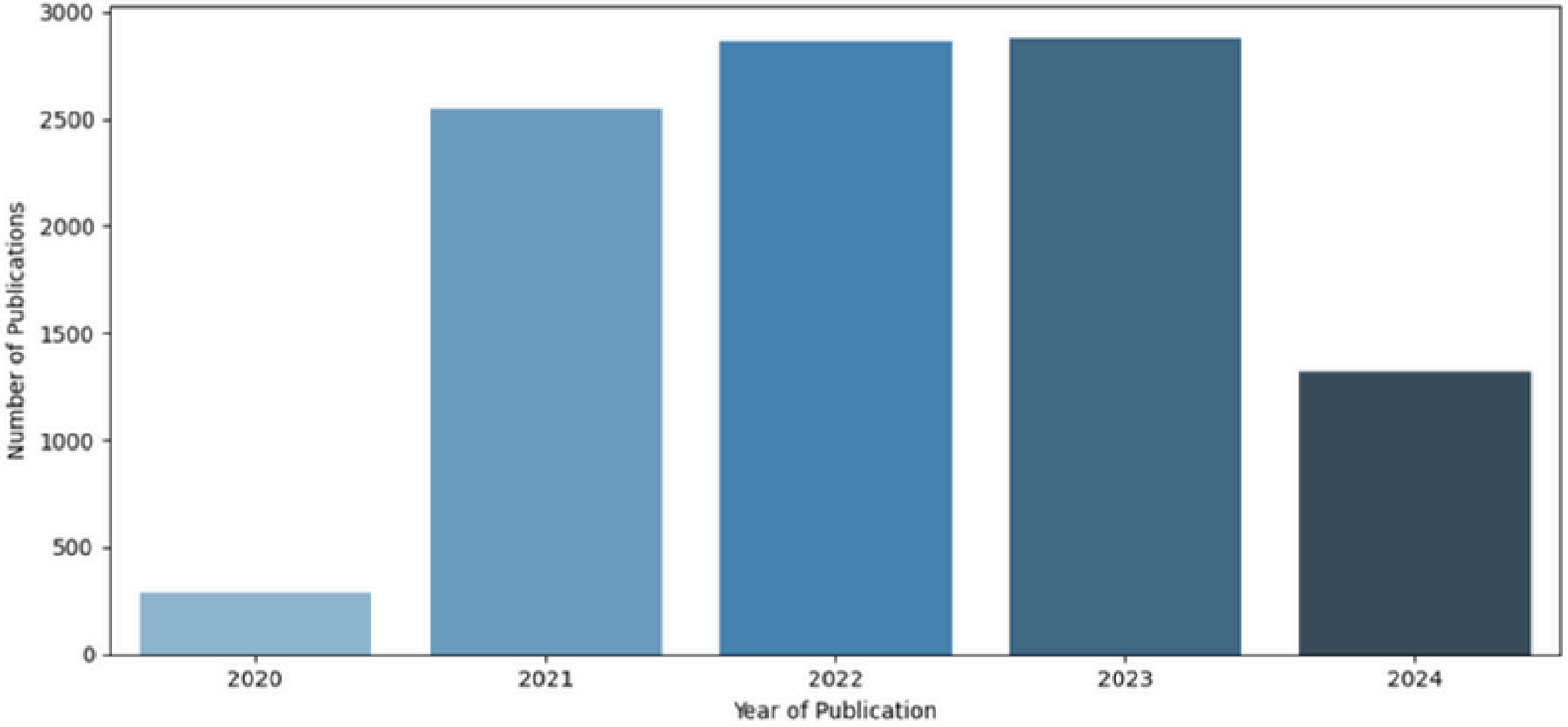



Correlation between Mental Health Disorders
SchizophreniaRate
Depressive disorders
Anxiety
Bipolar disorders
Eating disorders
1.00 -0.47 0.30 0.12 0.50
-0.47 1.00 0.11 0.15 -0.05
0.30 0.11 1.00 0.58 0.59
0.12 0.15 0.58 1.00 0.68
0.50 -0.05 0.59 0.68 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Literature analysis

In the next phase of our project, we employed the **BioPython** library to retrieve relevant publications from PubMed. Our initial query was "**schizophrenia[Title]**".

First, we visualized the distribution of these publications over the years using a bar plot. This helped us understand the trends in research activity related to schizophrenia over time.

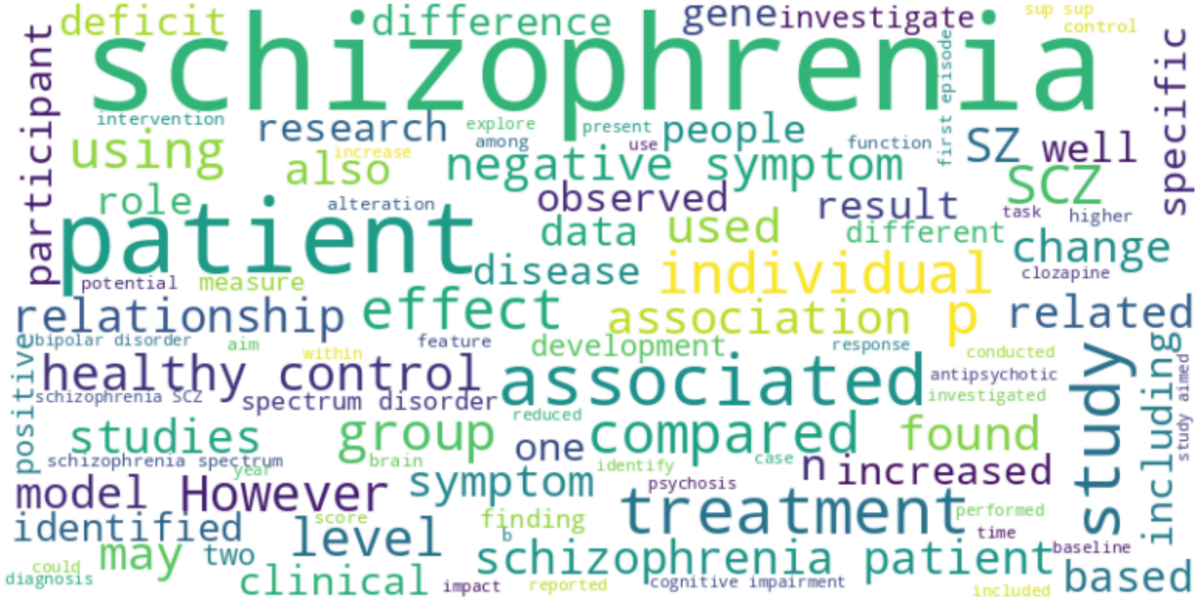


Literature analysis

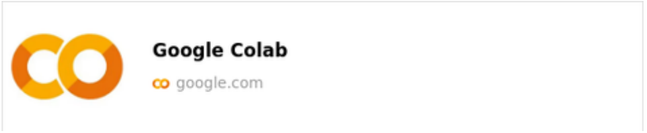
Additionally, we generated word clouds for the words found in the titles and abstracts of these articles. These visual representations highlighted the most frequently occurring terms.

Word Cloud for Titles

Word Cloud for Abstracts



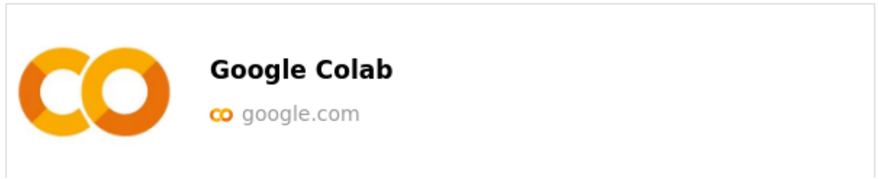
Word cloud with “schizophrenia” query:



Topic modeling

To delve deeper into the themes present in the research abstracts, we performed topic modeling using natural language processing (NLP) techniques. The steps involved in this process included:

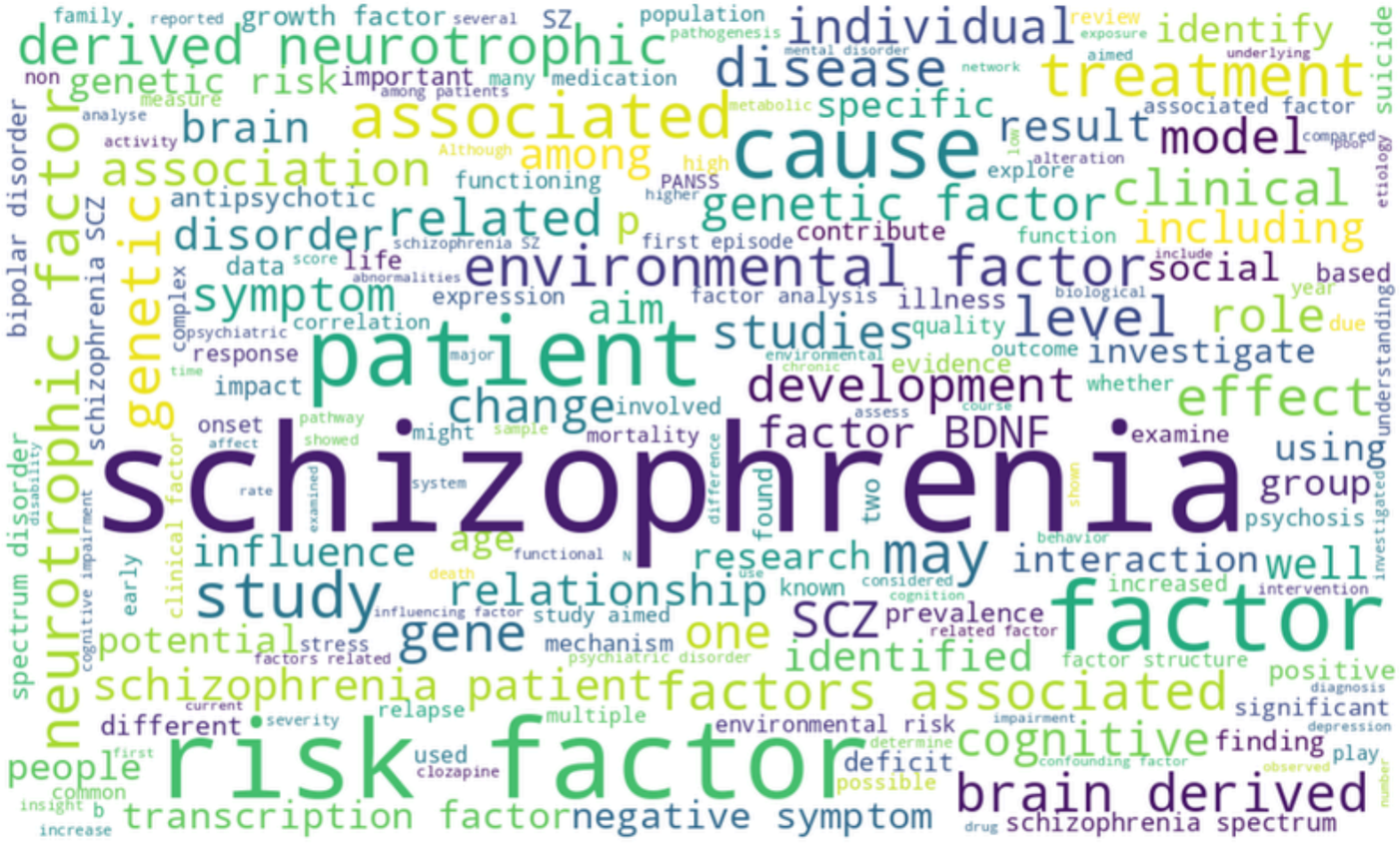
1. **Preprocessing the abstracts**
2. **Tokenizing and Lemmatizing**
3. **Building n-grams**
4. **Creating a dictionary and corpus**
5. **Training the LDA model**
6. **Evaluating model coherence**
7. **Visualizing topics**



Factors and causes

To further our research, we focused on identifying the causes and factors associated with schizophrenia. Using the search term **"schizophrenia[Title] AND (causes OR factors)"**, we retrieved relevant publications from PubMed.

We extracted sentences from the abstracts that mentioned causes or factors related to schizophrenia by searching for keywords like "cause", "causes", "factor", and "factors".



Factors and causes

('schizophrenia', 'patients'): 2089
('healthy', 'controls'): 1382
('negative', 'symptoms'): 1311
('bipolar', 'disorder'): 1301
('schizophrenia', 'spectrum'): 976
('people', 'schizophrenia'): 941
('risk', 'factors'): 883
('individuals', 'schizophrenia'): 853
('spectrum', 'disorders'): 778
('schizophrenia', 'scz'): 748
('positive', 'negative'): 714
('schizophrenia', 'bipolar'): 697
('study', 'aimed'): 679
('prefrontal', 'cortex'): 664
('symptoms', 'schizophrenia'): 662
('95', 'ci'): 658
('cognitive', 'impairment'): 654
('cognitive', 'deficits'): 653
('gene', 'expression'): 576
('present', 'study'): 559

Additionally, we analyzed word co-occurrences by identifying and counting the most common pairs of words that appeared together in the text.

- Genetic factors
- Symptomatology
- Comparative studies
- Risk factors
- Brain regions

Differents points of view

The **aetiology** of schizophrenia is divided into 4 main areas:

- Psychological
- Genetic
- Environmental causes
- Substance abuse

First we made an analysis to find which is the **most studied** compound of causes via some keywords to identify the category, then via **NLP** we extract a **summary** of what a category is specifically about

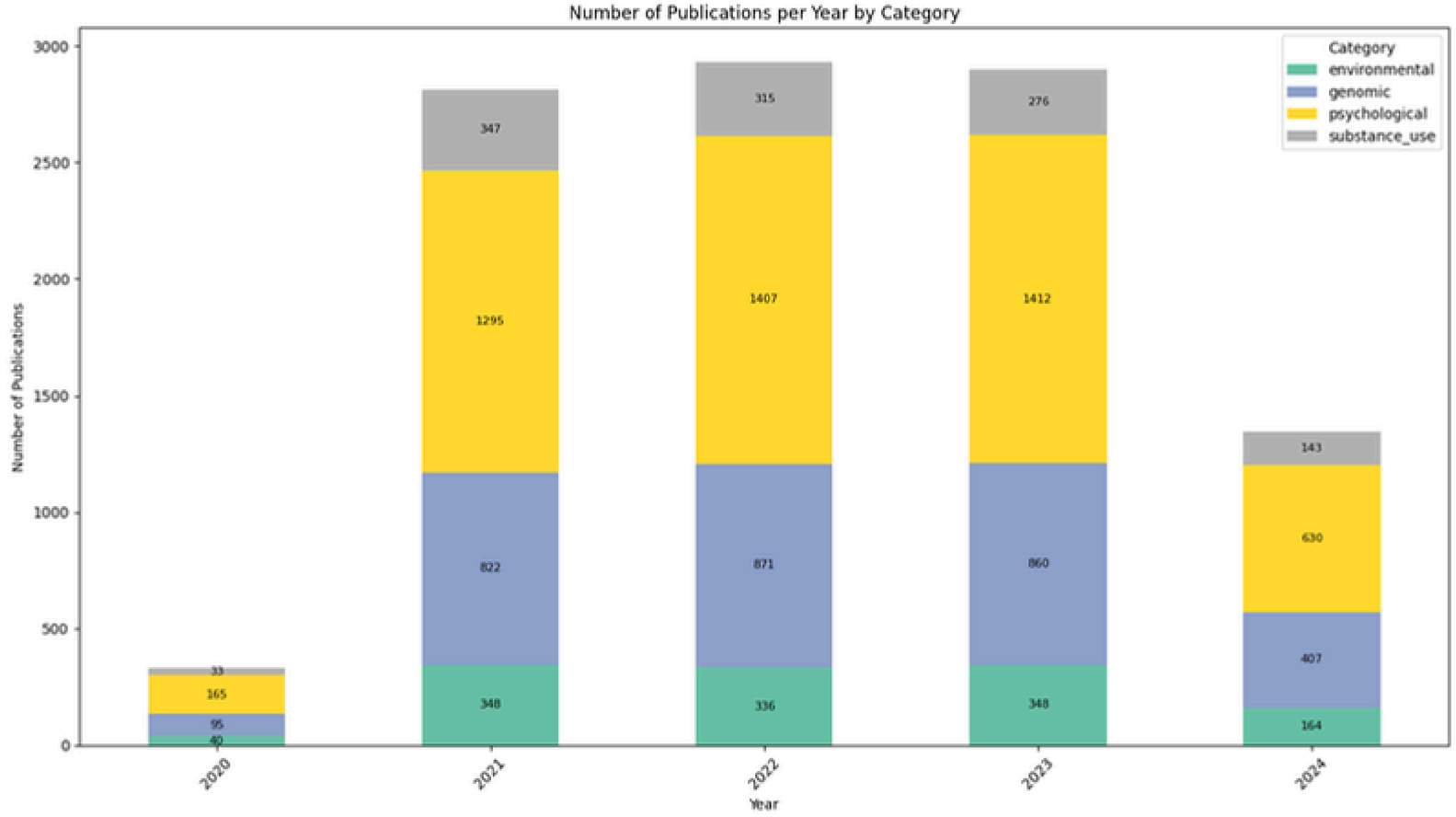
Keywords grouping

- Psychological: ["therapy", "behavioral", "psychology", "cognitive", "mental", "psychiatric", "counseling", "psychotherapy"]
- Genomic: ["gene", "genetic", "DNA", "genome", "mutation", "chromosome", "heredity", "genotyping", "allele"]
- Environmental: ["environment", "pollution", "stress", "trauma", "urban", "toxins", "climate", "exposure", "socioeconomic"]
- Substance use: ["substance", "drug", "cannabis", "alcohol", "addiction", "opioids", "nicotine", "stimulant", "dependency"]

Points of view analysis

Then we have categorized the different publications by scanning their abstracts for specific keywords related to each category: **psychological**, **genomic**, **environmental**, and **substance use**.
We achieved these results:

- Psychological: 4909
- Genomic: 3055
- Environmental:1236
- Substance use: 1114



Points of view NLP findings

- **Psychological:** The studies explore the intricate **relationships between the onset age of schizophrenia, cognitive impairments, and psychological symptoms, revealing how earlier onset is closely linked to severe cognitive deficits** and enhanced negative symptoms. The analysis delves into the role of serum BDNF as a moderating factor that influences cognitive decline associated with schizophrenia, underlining the potential of environmental enrichment to alleviate these cognitive challenges in experimental models.

- **Genomic:** The articles highlight the **pivotal role of the EGR3 gene in schizophrenia** through extensive association studies, expression analyses, and exploration of its biological functions and interactions with schizophrenia susceptibility genes. It also notes significant findings from large-scale gene expression studies which identify **lowered C1QA mRNA expression in the prefrontal cortex of schizophrenia patients**, suggesting a complex interplay between genetic factors and core schizophrenia symptoms like hallucinations and self-perception disturbances.

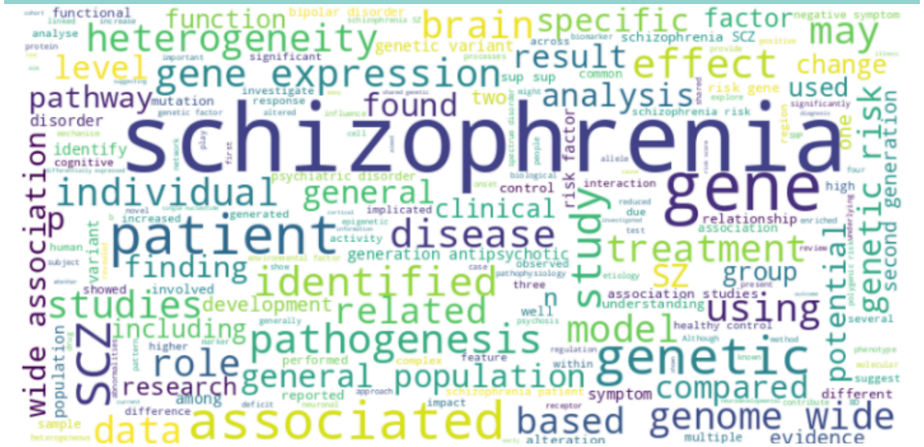
- **Enviromental:** The studies delves into how **environmental stressors such as urban exposure and socio-economic factors converge with genetic influences at the DNA methylation level within the CYP17A1 gene**, which is critical in the **stress-response pathway of schizophrenia**. It highlights the significant role of environmental conditions in modulating stress-related schizophrenia symptoms and discusses public perceptions on the interplay between environmental triggers and genetic susceptibility in the development and management of the disorder.

- **Substance abuse:** Researches have systematically reviewed the **efficacy of long-acting injectable (LAI) antipsychotics in treating schizophrenia patients with concurrent substance use disorders, covering a range of substances** including alcohol, cocaine, cannabis, opioids, MDMA, and ketamine. The findings highlight that LAI antipsychotics like paliperidone are particularly effective and safe in managing schizophrenia symptoms alongside chronic conditions like hepatitis, demonstrating **no significant drug-drug interactions** and providing a stable pharmacological option for patients facing dual challenges of schizophrenia and substance addiction.

Points of view Words Cloud

Psychological

Genomic



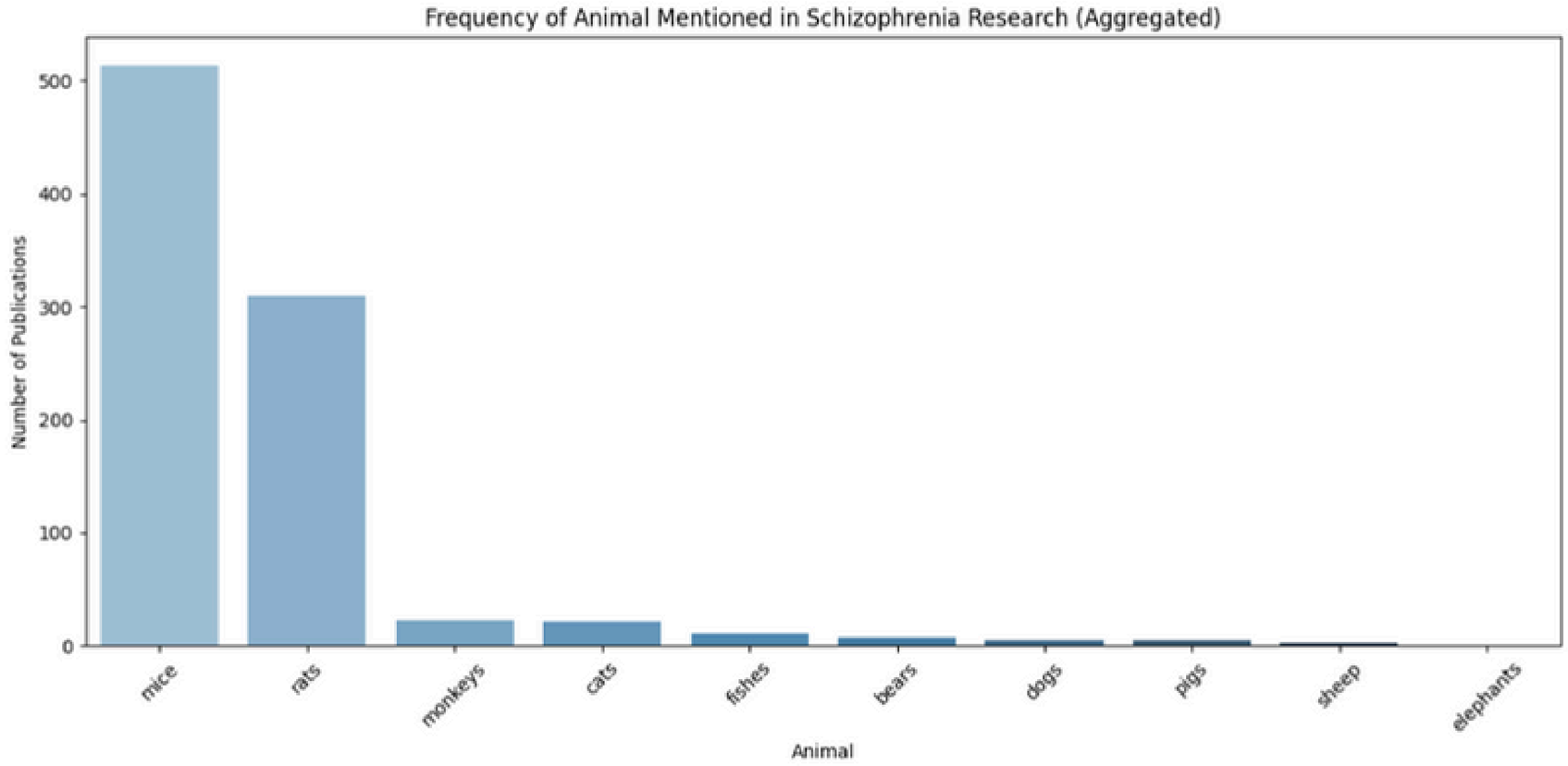
Enviromental

Substance use

Literature analysis on animals

Finally, we analysed the different publications to understand how the disease is studied among different organisms. Given a list of **animals**, we obtained the following number of articles about them:

- Mice: 513
- Rats: 310
- Monkeys: 23
- Cats: 21
- Fishes: 11
- Bears: 7
- Dogs: 5
- Pigs: 5
- Sheep: 3
- Elephants: 1



Animals observations by NLP

- **Mice:** Studies using both male and female mice demonstrate that **genetic and environmental manipulations significantly affect neural architecture and function, indicative of schizophrenia-like symptoms.** Key findings include reduced dendritic spine size in ZnT3 KO mice, alterations in excitatory postsynaptic potentials linked to specific genetic modifications, and changes in sensory gating as reflected by altered protein expressions in the prefrontal cortex.

- **Rats:** The researches employ various genetically-selected rat strains to simulate schizophrenia-relevant neurobehavioral traits. These include Apomorphine-susceptible (APO-SUS) rats, Low-prepulse inhibition rats, and others, each showcasing distinct behavioral anomalies linked to schizophrenia. **Findings indicate significant gender-specific behavioral differences in MAM-treated rats, with altered social interactions and dopaminergic neuron activity, highlighting the impact of genetic and environmental factors on schizophrenia's neurobiological and social manifestations.**

- **Monkeys:** The studies utilize for large part rhesus monkeys with specific neonatal lesions **to evaluate the impacts on acoustic startle responses and pre-pulse inhibition, critical in schizophrenia research. Monkeys with amygdala lesions exhibited heightened startle responses with minimal PPI disruption, contrasting with those having orbital frontal cortex lesions who showed reduced startle responses and significantly impaired PPI.** Interestingly, monkeys with hippocampal lesions displayed normal startle and PPI responses at short intervals but demonstrated enhanced prepulse facilitation at longer intervals, suggesting differentiated impacts of regional brain disruptions on sensory gating mechanisms.

Animals observations by NLP

- **Cats**: **Researches reveal significant associations between cat exposure and Toxoplasma seropositivity in individuals with psychiatric disorders, notably schizophrenia.** Statistical analyses identified age, contact with cats, and soil as correlated with increased Toxoplasma seropositivity. Furthermore, a case study highlights potential zoonotic transmission, including BoDV-1 infection, from stray cats to a long-term schizophrenia patient, underscoring a higher likelihood of cat ownership during childhood among schizophrenia patients compared to controls. This suggests a **possible environmental link involving pet exposure in the etiology and risk factors of schizophrenia.**

- **Fishes**: Machine learning analysis linked differences in gut microbiota to psychosocial stress factors, dietary habits including fish consumption, and cognitive impairments in schizophrenia. **The studies also explored the effects of ketamine on fish, noting that lower doses increased erratic movements while higher doses disrupted navigational abilities, mirroring schizophrenia-like symptoms.** These findings underscore the relevance of dietary and environmental factors, along with novel animal model findings, in understanding the multifaceted nature of schizophrenia.

Gene analysis

External data source

	Gene Symbol	Gene Description	# Publications
72	ERBB4	erb-b2 receptor tyrosine kinase 4	57
689	CYP2D6	cytochrome P450, family 2, subfamily D, polype...	47

Top 10 genes by number of publication

Querying PubMed

ERBB4	247
CYP2D6	228
CACNA1C	167
TCF4	118
CHRNA7	115
HTR2C	70
NOTCH4	57
HTR1A	35
CHRFAM7A	22
TCF7L2	12

Analysing Top 5 genes based on PubMed buplications

{gene_symbol}[Title/Abstract] AND schizophrenia[Title/Abstract]

Gene analysis

ERBB4 (Erb-B2 Receptor Tyrosine Kinase 4)

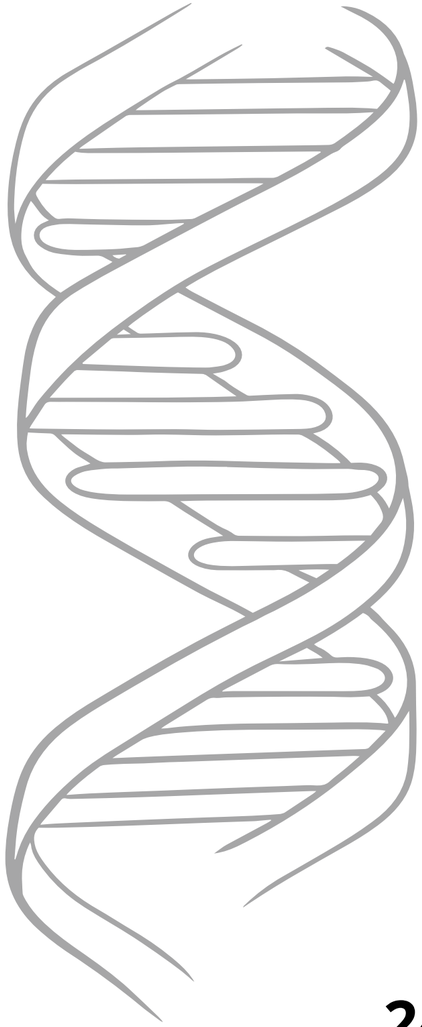
- Function: Regulates cell proliferation, differentiation, and apoptosis.
- Role: Involved in cancer development and nervous system regulation.

CYP2D6 (Cytochrome P450 2D6)

- Function: Metabolizes drugs and toxins in the liver.
- Role: Affects drug efficacy and risk of adverse effects.

CACNA1C (Calcium Voltage-Gated Channel Subunit Alpha1 C)

- Function: Forms part of L-type voltage-dependent calcium channels.
- Role: Essential for muscle contraction, neurotransmitter release, and gene expression.
- Associated Disorders: Linked to Timothy syndrome, bipolar disorder, and **schizophrenia**.



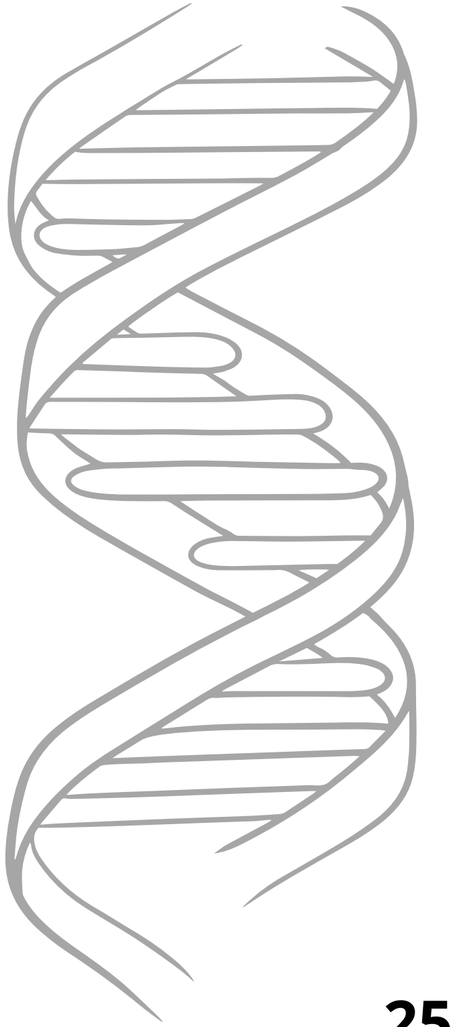
Gene analysis

TCF4 (Transcription Factor 4)

- Function: Regulates gene expression during development and differentiation.
- Role in Health: Crucial for nervous system development.
- Associated Disorders: Associated with Pitt-Hopkins syndrome.

CHRNA7 (Cholinergic Receptor Nicotinic Alpha 7 Subunit)

- Function: Part of the nicotinic acetylcholine receptor, involved in fast synaptic transmission.
- Role in Health: Key role in cognitive function and neural development.
- Associated Disorders: Linked to schizophrenia and epilepsy.

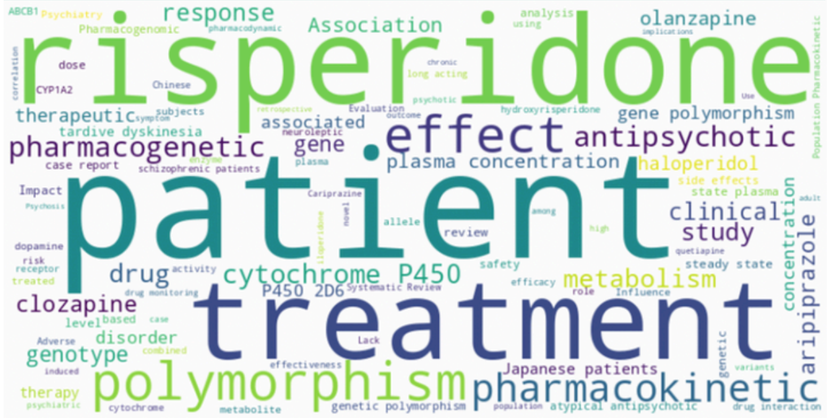


ERBB4

CHRNA7

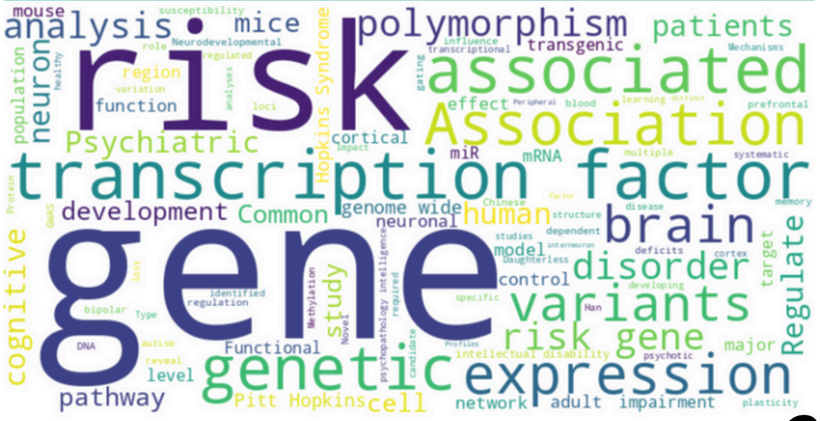
Word Cloud

CYP2D6



CACNA1C

TCF4



Querying KEGG

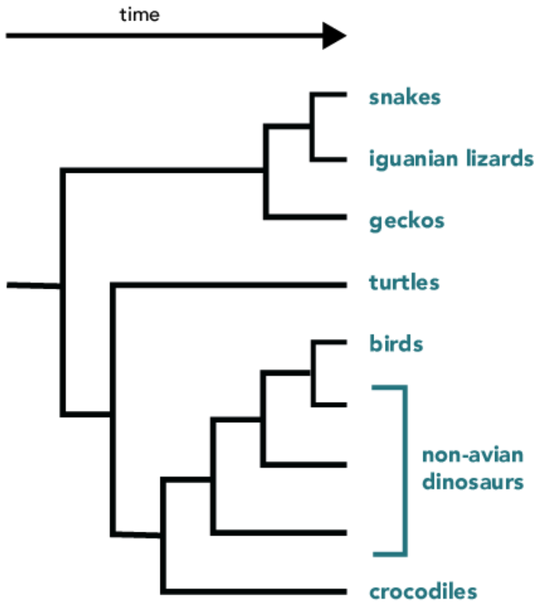
Gene study

Gene	Associated disease
ERBB4	H00058 Amyotrophic lateral sclerosis (ALS)
CYP2D6	H01171 Poor drug metabolism
CACNA1C	H00720 Long QT syndrome H00728 Brugada syndrome H02397 Neurodevelopmental disorder with movement abnormalities or hypotonia
TCF4_12	H00255 Hypogonadotropic hypogonadism H00756 Pitt-Hopkins syndrome H00960 Fuchs corneal dystrophy H02160 Craniosynostoses
CHRNA7	H01877 Chromosome 15q13.3 microdeletion syndrome

Shared pathway	Genes
'map04010' MAPK signaling pathway	ERBB4, CACNA1C
'map04020' Calcium signaling pathway	ERBB4, CACNA1C, CHRNA7
'map04726' Serotonergic synapse	CYP2D6, CACNA1C
'map04725' Cholinergic synapse, 'map05010' Alzheimer disease, 'map05207' Chemical carcinogenesis - receptor activation, 'map05022' Pathways of neurodegeneration - multiple diseases	CACNA1C, CHRNA7

Gene study

BLAST Search



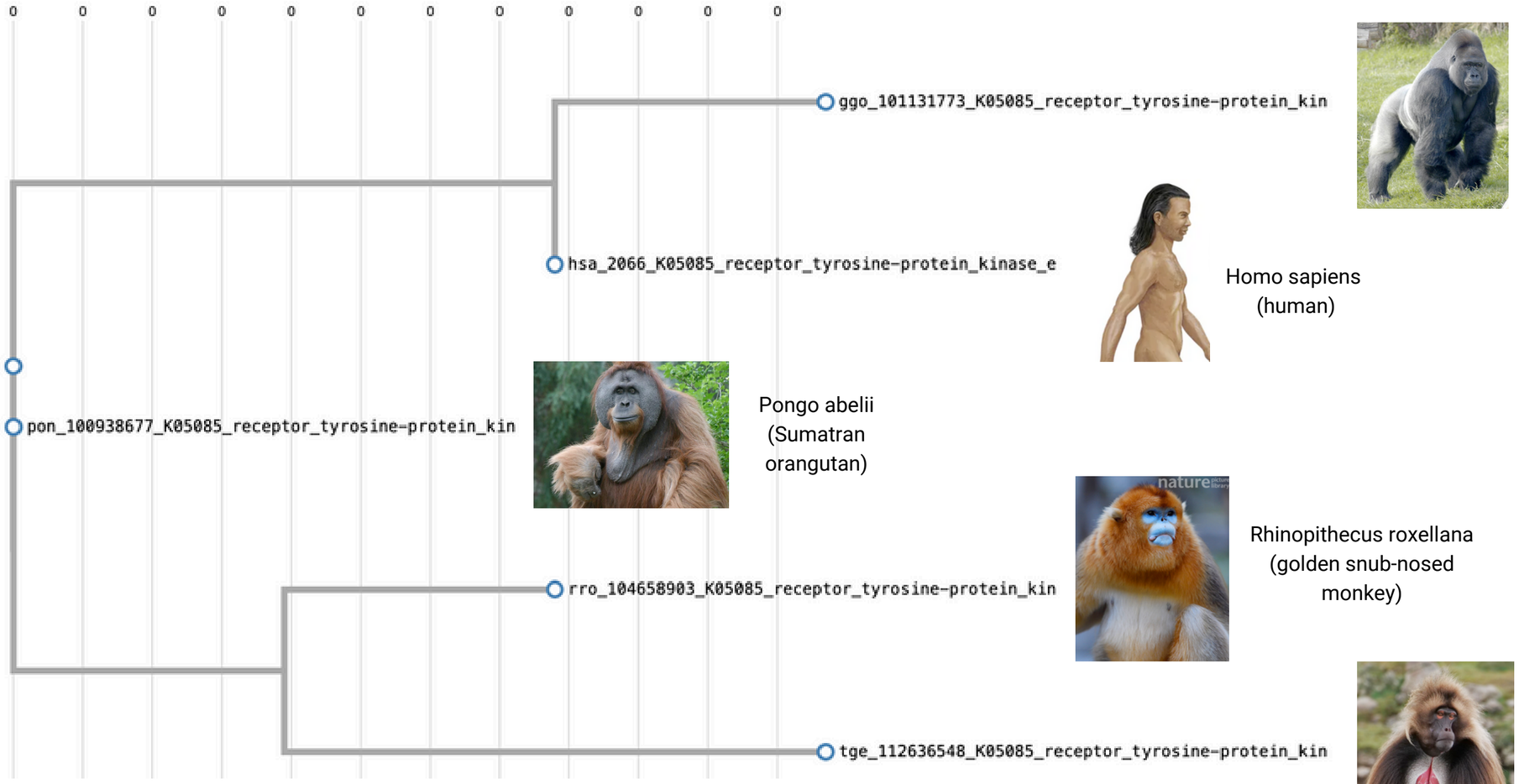
Identifying and comparing **similar genes** in other organisms using **BLAST**

Phylogenetic Tree Analysis: to visualize the **evolutionary relationships** between organisms based on the genes analyzed.

hsa:ERBB4

hsa:2066

Presence in other species: almost all Mammals, Reptiles, Amphibians, Fishes, Cartilaginous fishes, Jawless fishes, 75% Birds



Gorilla gorilla gorilla (western lowland gorilla)

Homo sapiens (human)

Pongo abelii (Sumatran orangutan)

Rhinopithecus roxellana (golden snub-nosed monkey)

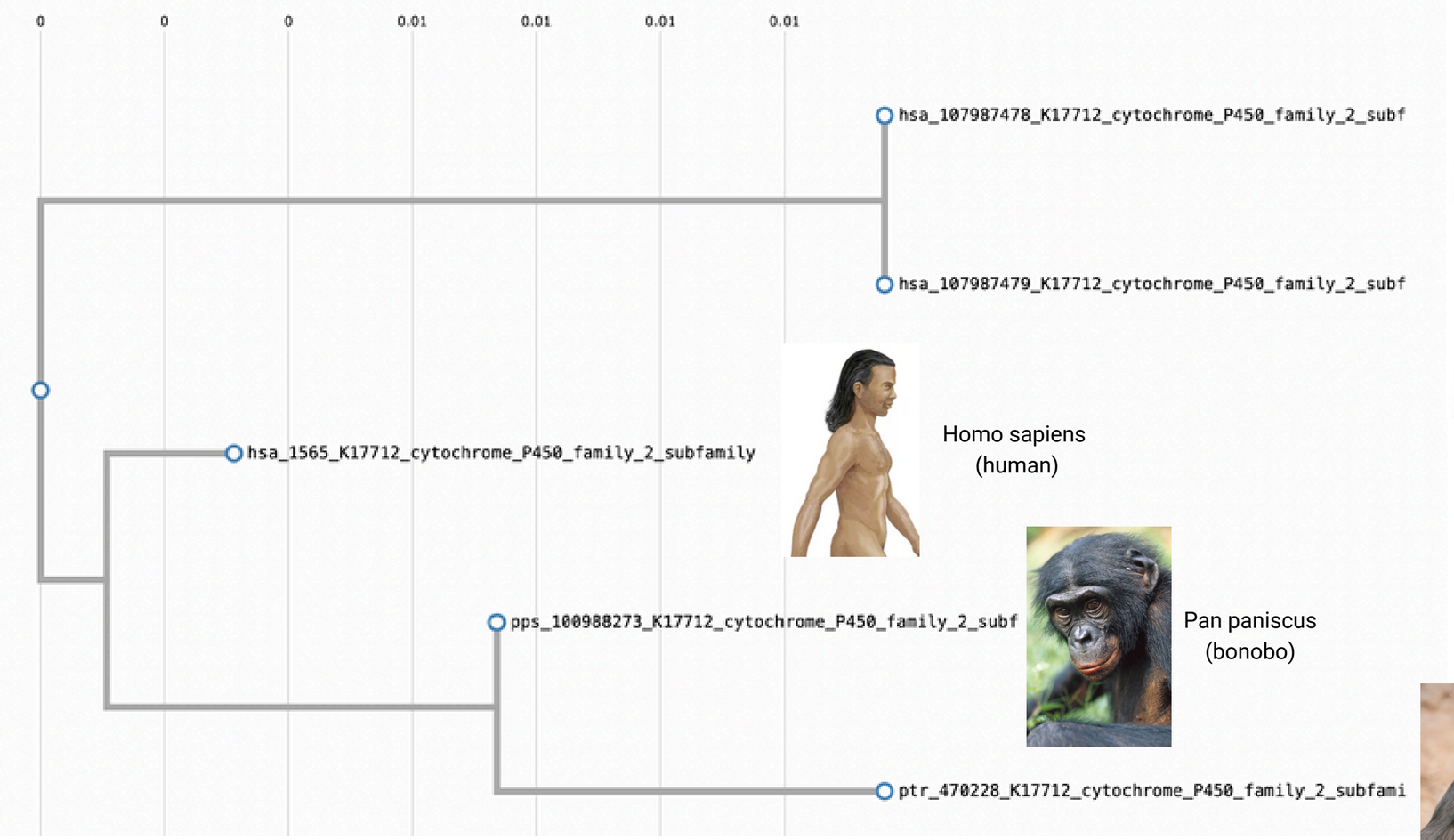
Theropithecus gelada (gelada)

Receptor tyrosine-protein kinase erbB-4

hsa:CYP2D6

hsa:107987479

Presence in other species: almost all Mammals, Amphibians, Cartilaginous fishes, 90% Birds, 53% Reptiles, no Fishes, no Jawless fishes

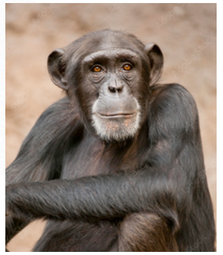


Homo sapiens (human)

Homo sapiens (human)

Homo sapiens (human)

Pan paniscus (bonobo)



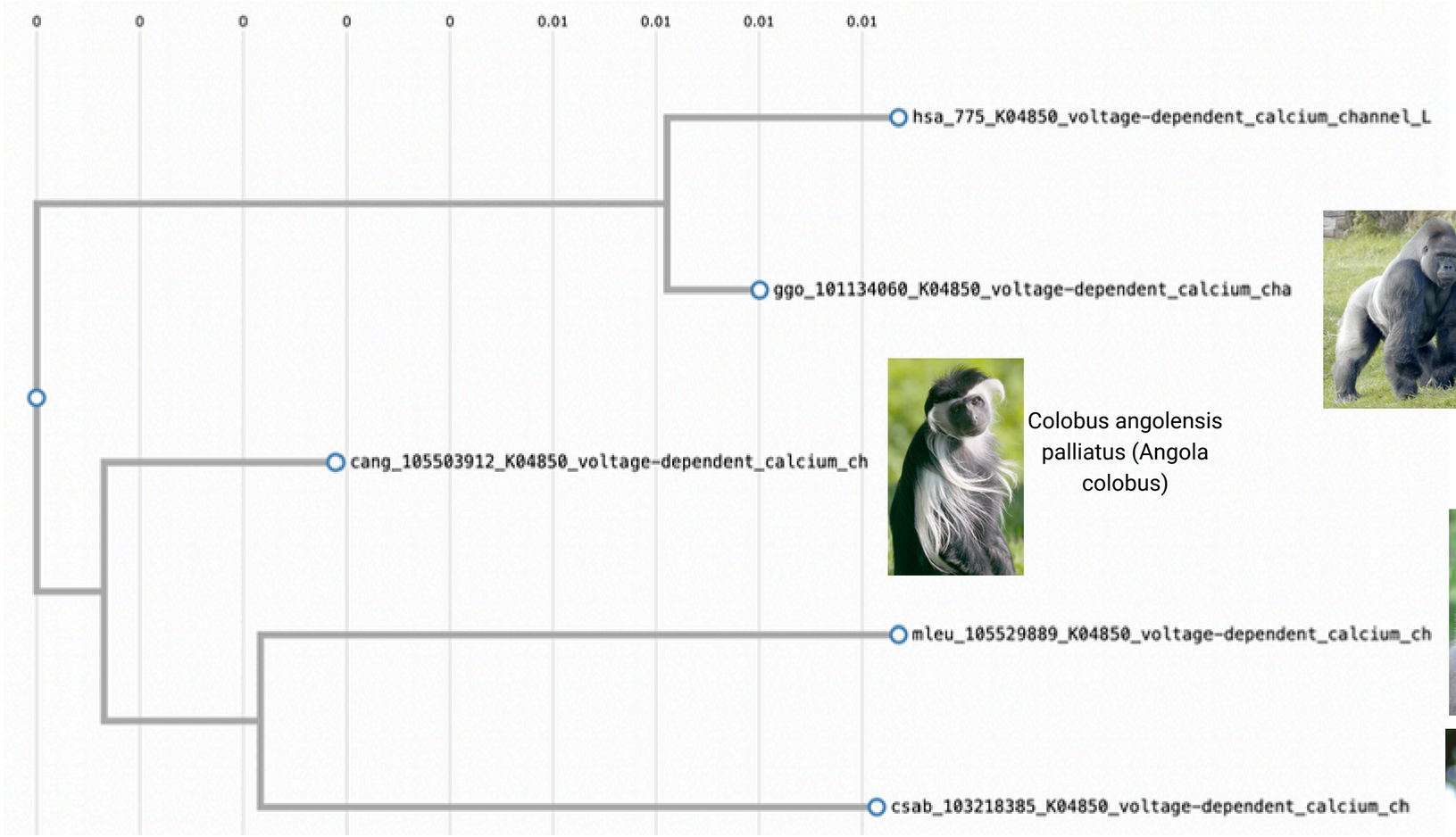
Pan troglodytes (chimpanzee)

Cytochrome P450 family 2 subfamily D6

hsa:CACNA1C

hsa:775

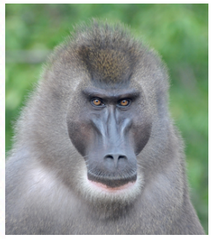
Presence in other species: almost all Mammals, Reptiles, Amphibians, 80% Cartilaginous fishes, 77% Birds, 70% Fishes, 50% Jawless fishes



Homo sapiens (human)

Gorilla gorilla gorilla (western lowland gorilla)

Colobus angolensis palliatus (Angola colobus)



Mandrillus leucophaeus (drill)

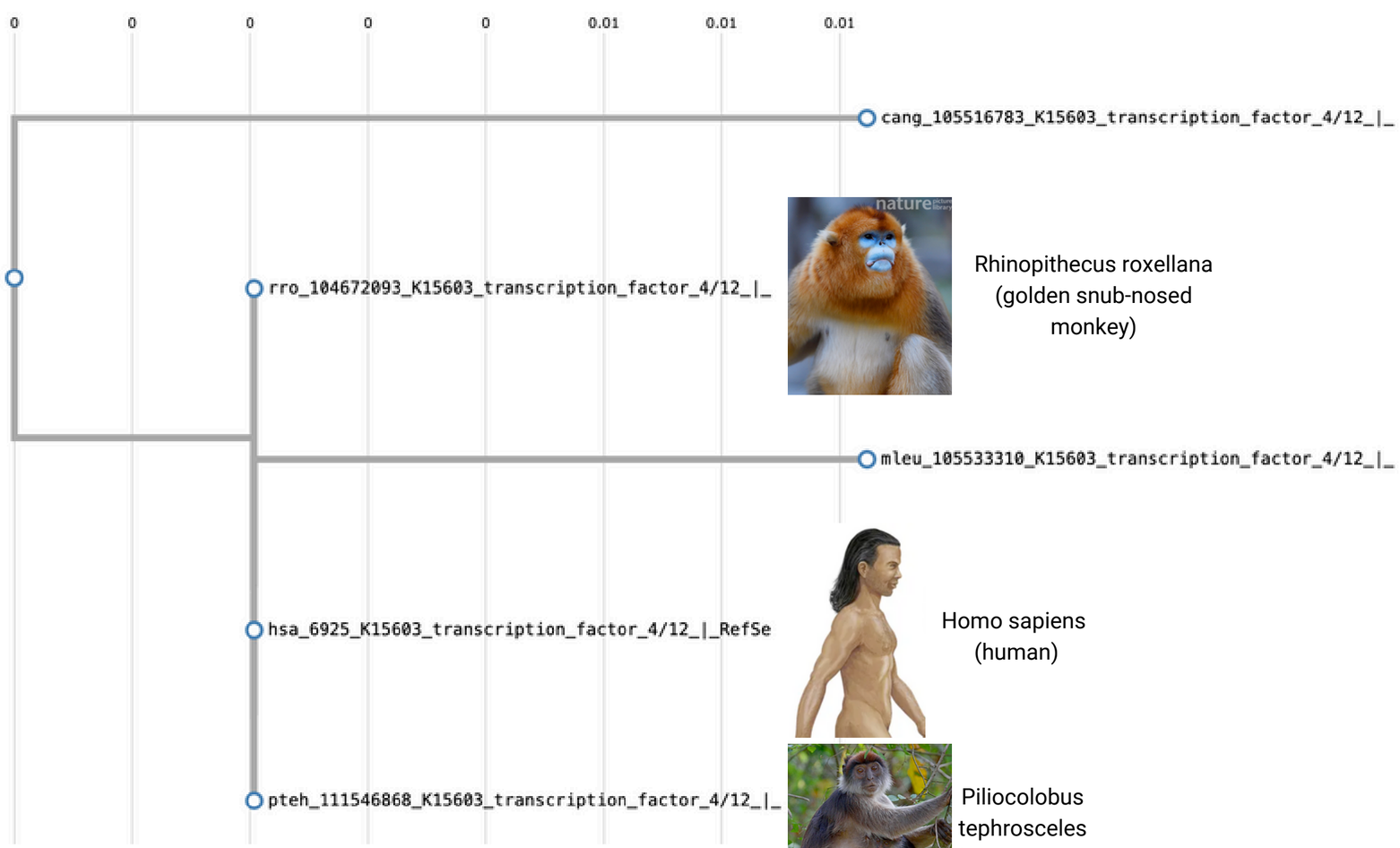
Chlorocebus sabaeus (green monkey)

Voltage-dependent calcium channel L type alpha-1C

hsa:TCF4

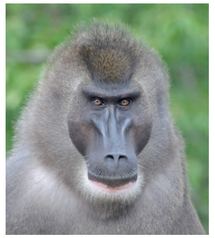
hsa:6925

Presence in other species: all Mammals, Reptiles, Amphibians, Cartilaginous fishes, 68% Birds, 82% Fishes, no Jawless fishes



Colobus angolensis palliatus (Angola colobus)

Rhinopithecus roxellana (golden snub-nosed monkey)



Mandrillus leucophaeus (drill)

Homo sapiens (human)

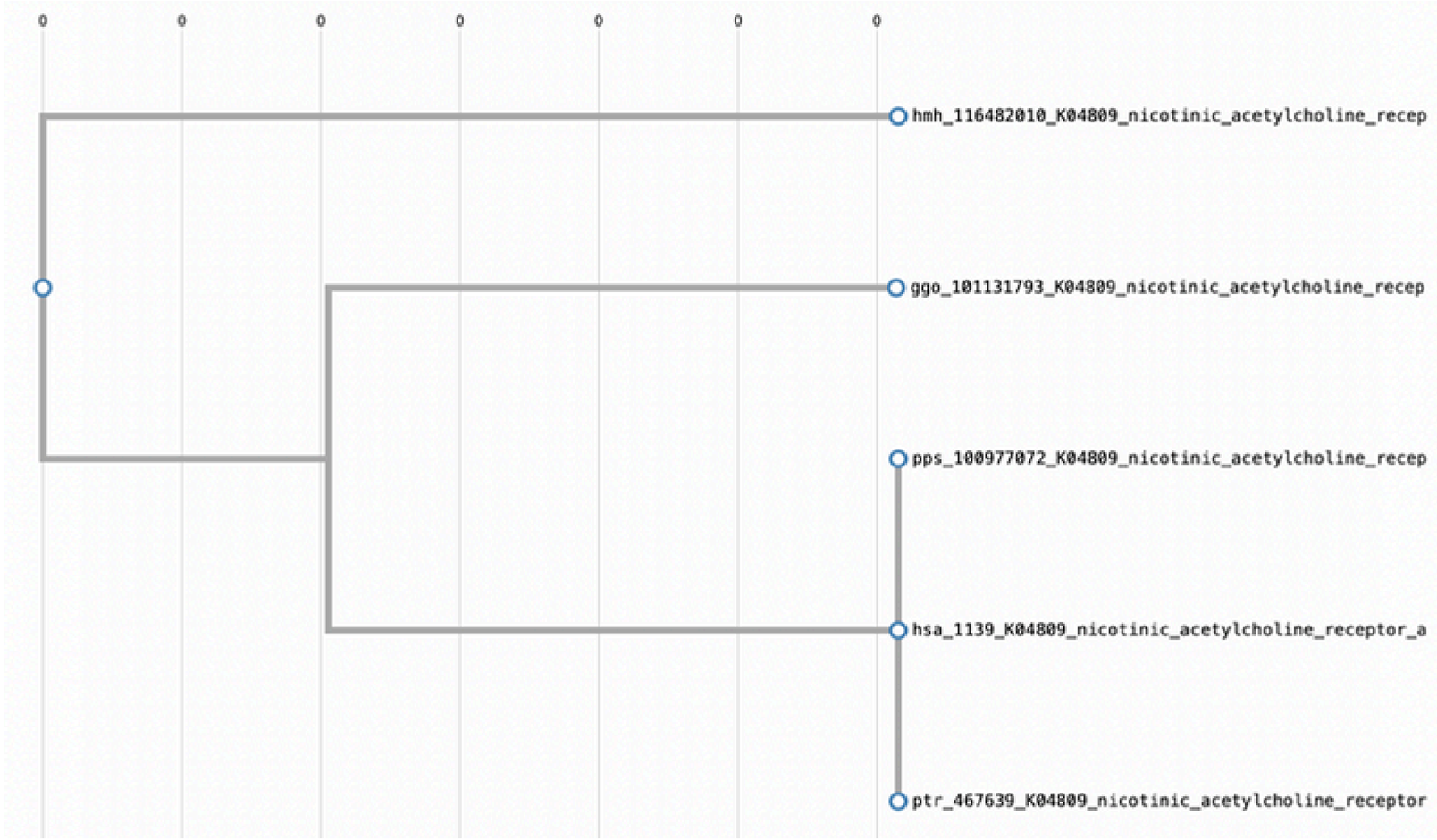
Piliocolobus tephrosceles (Ugandan red Colobus)

Transcription factor 4/12

hsa:CHRNA7

hsa:1139

Presence in other species: almost all Mammals, Reptiles, Amphibians, Birds, Fishes, Jawless fishes, 40% Cartilaginous fishes



Hylobates moloch (silvery gibbon)

Gorilla gorilla gorilla (western lowland gorilla)

Pan paniscus (bonobo)

Homo sapiens (human)

Pan troglodytes (chimpanzee)

Nicotinic acetylcholine receptor alpha-7

References

- https://www.nimh.nih.gov/health/topics/schizophrenia
- https://www.harmoniamentis.it
- https://www.kegg.jp/kegg/
- https://blast.ncbi.nlm.nih.gov/Blast.cgi
- https://it.wikipedia.org/wiki/Schizofrenia

Thank you for your attention!