

PGSChat Test Case-1: Information Retrieval

Ask me a question

Show top 5 ranked variants, top to bottom, along with their other information in Alzheimer. If duplicate, show once only.

Submit

User prompt: intuitive, simple, natural language

Results from the PGS Rank database: Top 5 variants and corresponding genes for Alzheimer ranked top to bottom

SQL Code

```
SELECT DISTINCT
  snpcoord,
  gene,
  func,
  rank,
  trait
FROM
  pgs_snpmeta
WHERE
  trait = 'Alzheimer'
ORDER BY
  rank DESC
LIMIT 5;
```

Natural Language queries/prompt auto-translated to SQL code

	snpcoord	gene	func	rank	trait
0	6_41161514	TREM2	exonic	0.2083	Alzheimer
1	11_121564878	SORL1	intronic	0.197	Alzheimer
2	19_44919689	APOC1(dist=345)	downstream	0.1667	Alzheimer
3	19_44908822	APOE	exonic	0.1285	Alzheimer
4	19_44908684	APOE	exonic	0.125	Alzheimer

Download CSV

Download and save the top 10 genes information as csv file in local client location

PGSChat Test Case-2: Information Retrieval & Visualization

Ask me a question

Plot a heat map for top 5 ranked variants and their genes against ranks, ranked high to low, in Alzheimer. No duplicates.

Submit

Action:

```
import plotly.graph_objects as go

# Create the heat map
fig = go.Figure(data=go.Heatmap(
    z=data_df['rank'],
    x=data_df['gene'],
    y=data_df['rsID'],
    colorscale='Viridis'))

# Add labels and title
fig.update_layout(
    title='Heat Map of Top 5 Ranked Variants and Their Genes in Alzheimer',
    xaxis_title='Gene',
    yaxis_title='rsID')

# Show the heat map
show(fig)
```

Natural Language queries/prompt auto-translated to Python code as required to display the chart



Visualize custom database

Question: Plot a heat map for top 5 ranked variants and their genes against ranks, ranked high to low, in Alzheimer. No duplicates.

Thought 1

To plot a heat map for the top 5 ranked variants and their genes against ranks in Alzheimer, we need to retrieve the necessary data from the database. We will need information from the pgs_snpmeta table, specifically the rsID, gene, and rank columns. We will filter the data for Alzheimer and sort it by rank in descending order. We will also remove any duplicate variants.

```
# Query the necessary data
sql_query = "SELECT rsID, gene, rank FROM pgs_snpmeta WHERE trait = 'Alzheimer' GROUP BY rsID, gene, rank ORDER BY rank DESC LIMIT 5"
data_df = execute_sql(sql_query)
```

```
# Replace NAN with 0
data_df = data_df.replace(np.nan, 0)
```

```
# Observe the query result
observe("Data", data_df)
```

Data

	rsID	gene	rank
0	rs75932628	TREM2	0.2083
1	rs11218343	SORL1	0.197
2	rs4420638	APOC1	0.1667
3	rs7412	APOE	0.1285
4	rs429358	APOE	0.125

PGSChat Test Case-3: Information Retrieval & Visualization

Ask me a question

Plot a heat map with variants against ranks for the top 5 ranked genes, ranked high to low, in Alzheimer. use pgssnpmeta only. If duplicate show only once.

Submit

Thought 1

Now that we have the top 5 ranked genes associated with Alzheimer, we need to retrieve the variants associated with these genes from the pgs_snpmeta table. We will filter the data based on the selected genes and calculate the average rank for each variant. Then, we will create a heat map with variants against ranks.

Action:

```
# Filter the data based on the selected genes
selected_genes = top_genes_df['gene'].tolist()
sql_query = "SELECT rsid, gene, rank FROM pgs_snpmeta WHERE trait = 'Alzheimer' AND gene IN (%s)" % (','.join(selected_genes))
variants_df = execute_sql(sql_query)

# Calculate the average rank for each variant
variants_df['avg_rank'] = variants_df.groupby('rsid')['rank'].mean().reset_index()

# Replace the data to create a heat map
heat_data_df = variants_df[['rsid', 'avg_rank', 'gene']].reset_index()

# Create the heat map
fig = px.imshow(heat_data_df, x=heat_data_df['gene'], y=heat_data_df['rsid'], color=heat_data_df['avg_rank'])
fig.update_layout(title="Heat Map of Variants against Ranks for Top 5 Ranked Genes in Alzheimer")
fig.update_xaxes(title="Genes")
fig.update_yaxes(title="Variants")
fig.show()
```

Heat Map of Variants against Ranks for Top 5 Ranked Genes in Alzheimer

Top Genes Data

gene	rank
TREM2	0.2083
SORL1	0.197
APOC1(dist=345)	0.1667
APOE	0.1285
APOA2	0.125

Even for same prompt, the chart type or design may change as it is intuitively decided by ChatGPT, unless specifically mentioned in prompt.

PGSChat Test Case-4: Information Retrieval & Visualization

Ask me a question

Plot the PGS score IDs and their count in European Ancestry for Alzheimer

Submit

User prompt: intuitive, simple, natural language, even small typos lead to correct results

PGS Score IDs and Count in European Ancestry for Alzheimer

Answer:

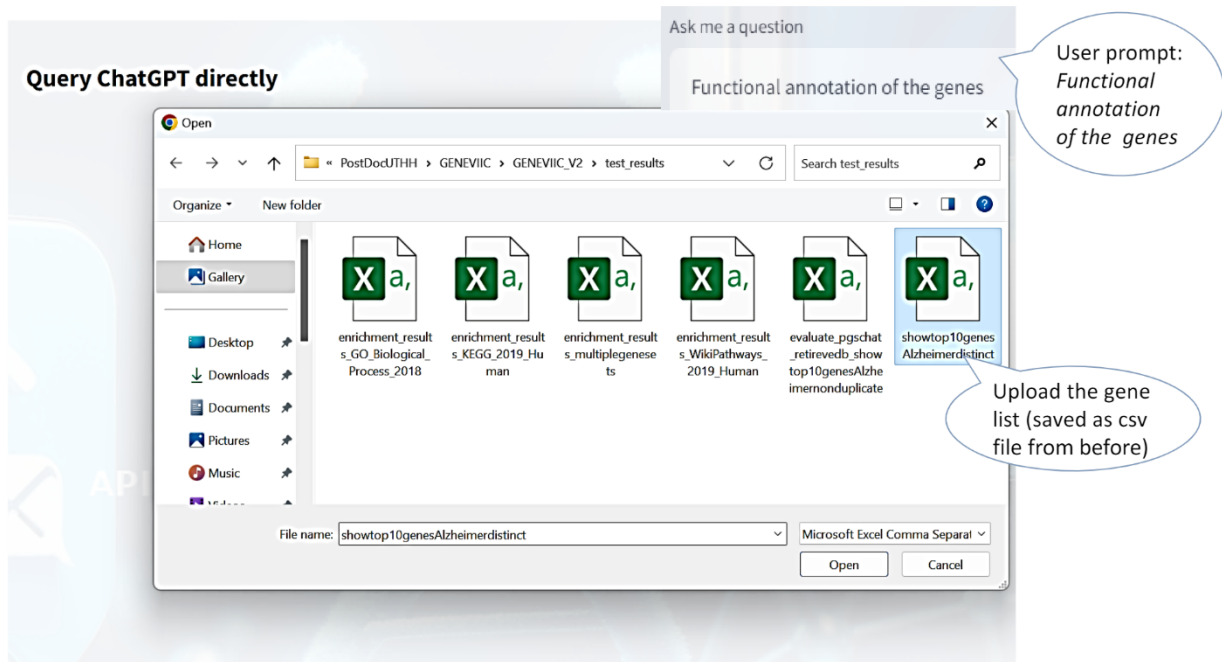
The PGS score IDs and their count in European Ancestry for Alzheimer are as follows:

- PGS000025: 3
- PGS000026: 1
- PGS000053: 1
- PGS000334: 3
- PGS000779: 1
- PGS000823: 2
- PGS000876: 2
- PGS000898: 1
- PGS000945: 2
- PGS001347: 2

Results from the PGS Rank database: Total Count of each PGS Score ID for Alzheimer in European Ancestry; auto-decides the relevant type of graph and generates the Python code for visualization as well from natural language prompts

These counts represent the number of PGS scores available for each PGS score ID in the European Ancestry for Alzheimer.

PGSChat Test Case-5: Query Chat GPT Directly



Functional annotation is the process of assigning biological functions to genes based on experimental evidence or computational predictions. Here is a brief functional annotation of the genes you mentioned:

1. TREM2 (Triggering Receptor Expressed on Myeloid Cells 2): It is involved in the regulation of immune responses and phagocytosis. Mutations in TREM2 have been associated with increased risk of Alzheimer's disease.
2. SORL1 (Sortilin-Related Receptor 1): It plays a role in the trafficking and processing of amyloid precursor protein (APP), which is involved in the production of amyloid-beta plaques in Alzheimer's disease.
3. APOC1 (Apolipoprotein C1): It is a component of lipoprotein particles and plays a role in lipid metabolism. Variations in APOC1 have been associated with increased risk of Alzheimer's disease.
4. APOE (Apolipoprotein E): It is involved in lipid transport and metabolism. APOE has been extensively studied in relation to Alzheimer's disease, as certain variants of this gene are associated with increased risk.
5. BIN1 (Bridging Integrator 1): It is involved in membrane remodeling and endocytosis. BIN1 has been implicated in Alzheimer's disease, as it interacts with tau protein and affects its aggregation.
6. CYP27C1 (Cytochrome P450 Family 27 Subfamily C Member 1): It encodes an enzyme involved in the metabolism of vitamin D. Its specific role in relation to Alzheimer's disease is not well understood.
7. ABI3 (Abl Interactor 3): It is involved in signal transduction and regulation of cell adhesion. ABI3 has been associated with immune-related disorders, but its role in Alzheimer's disease is not well characterized.
8. DSG2 (Desmoglein 2): It is a component of desmosomes, which are cell adhesion structures. DSG2 mutations have been associated with arrhythmogenic right ventricular cardiomyopathy, but its role in Alzheimer's disease is unclear.

Please note that the functional annotations provided here are a general overview and may not encompass all known functions or associations of these genes. Further research and experimental evidence are required to fully understand the functions and roles of these genes in specific biological processes or diseases.

Chat GPT/GPT-4 provides the answer to the query using its vast pre-trained knowledge