

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:

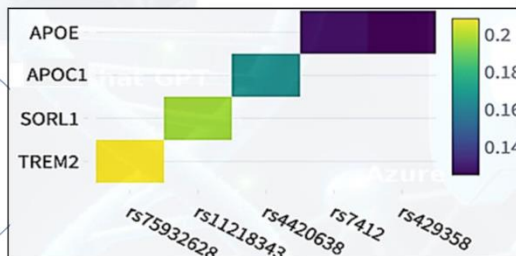
```
import plotly.graph_objects as go

# 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

Heat Map Data

rsid	APOC1(dist=345)	APOE	SORL1	TREM2
rs11218343	None	None	0.197	None
rs143332484	None	None	None	0.0729
rs429358	None	0.125	None	None
rs4420638	0.1667	None	None	None
rs7412	None	0.1285	None	None
rs75932628	None	None	None	0.2083
rs769449	None	0.0625	None	None

Visualize custom database

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.

Thought 1

To plot a heat map with variants against ranks for the top 5 ranked genes in Alzheimer, we need to retrieve the necessary data from the pgs_snpmeta table. We will filter the data based on the reported_trait column to select only the variants associated with Alzheimer. Then, we will group the data by the gene column and calculate the average rank for each gene. Finally, we will select the top 5 genes based on their average rank and retrieve the variants associated with these genes.

Action:

```
# Import necessary libraries
import plotly.express as px
import numpy as np

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

# Replace NaN with 0
top_genes_df['rank'] = top_genes_df['rank'].replace(np.nan, 0)

# Observe the query result
observe("Top Genes Data", top_genes_df)
```

Top Genes Data

	gene	rank
0	TREM2	0.2083
1	SORL1	0.197
2	APOC1(dist=345)	0.1667
		0.1285
		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.