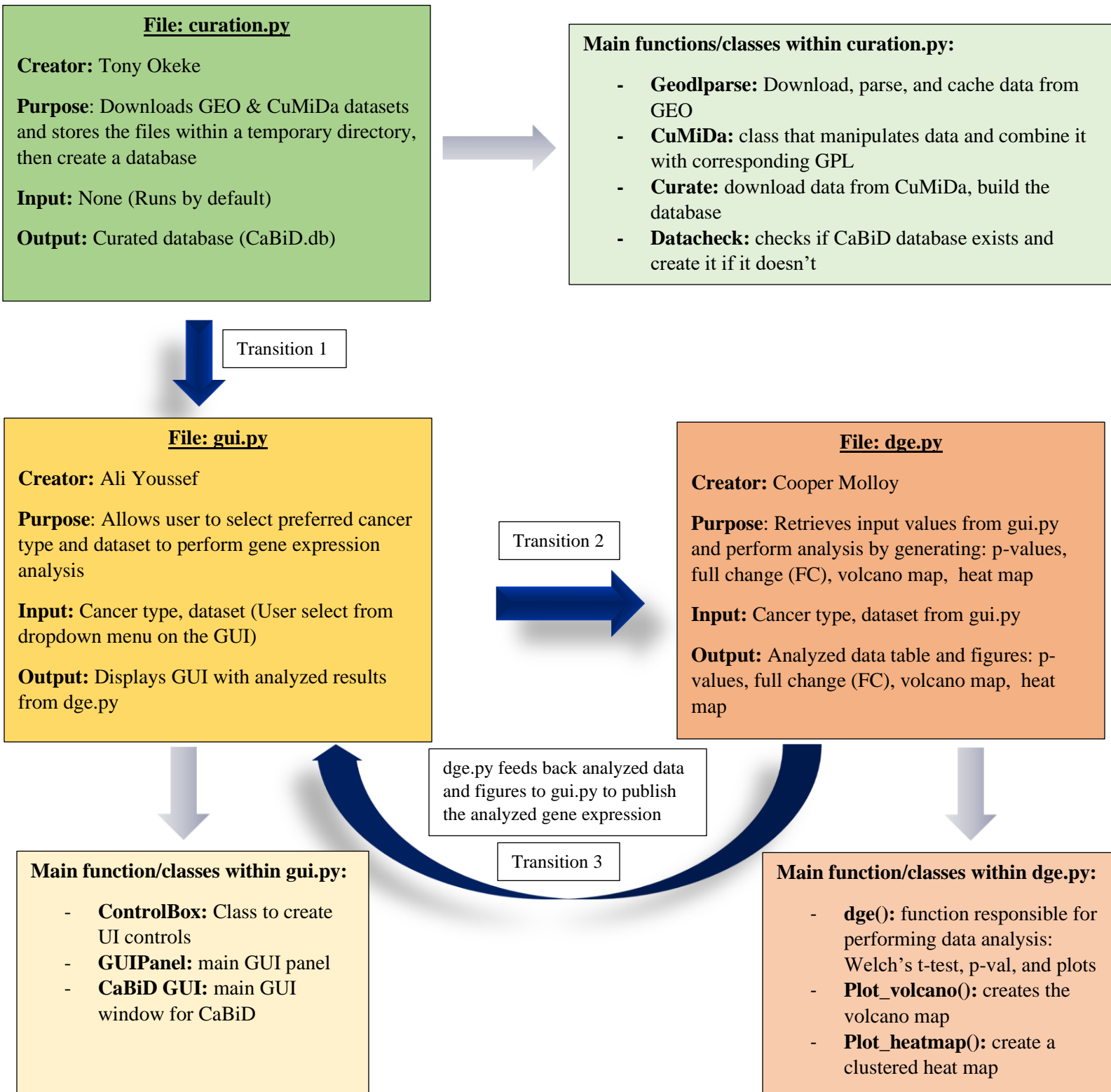BMES 550

Group 1: Tony Okeke, Cooper Molloy, Ali Youssef

Project Name: Cancer Biomarker Discovery (CaBiD)

Design & Schema

Flowchart:

File: curation.py

Creator: Tony Okeke

Purpose: Downloads GEO & CuMiDa datasets and stores the files within a temporary directory, then create a database

Input: None (Runs by default)

Output: Curated database (CaBiD.db)

Main functions/classes within curation.py:

- **Geodlparse:** Download, parse, and cache data from GEO
- **CuMiDa:** class that manipulates data and combine it with corresponding GPL
- **Curate:** download data from CuMiDa, build the database
- **Datacheck:** checks if CaBiD database exists and create it if it doesn't

Transition 1

File: gui.py

Creator: Ali Youssef

Purpose: Allows user to select preferred cancer type and dataset to perform gene expression analysis

Input: Cancer type, dataset (User select from dropdown menu on the GUI)

Output: Displays GUI with analyzed results from dge.py

Transition 2

File: dge.py

Creator: Cooper Molloy

Purpose: Retrieves input values from gui.py and perform analysis by generating: p-values, full change (FC), volcano map, heat map

Input: Cancer type, dataset from gui.py

Output: Analyzed data table and figures: p-values, full change (FC), volcano map, heat map

dge.py feeds back analyzed data and figures to gui.py to publish the analyzed gene expression

Transition 3

Main function/classes within gui.py:

- **ControlBox:** Class to create UI controls
- **GUIPanel:** main GUI panel
- **CaBiD GUI:** main GUI window for CaBiD

Main function/classes within dge.py:

- **dge():** function responsible for performing data analysis: Welch's t-test, p-val, and plots
- **Plot_volcano():** creates the volcano map
- **Plot_heatmap():** create a clustered heat map

Summary Table:

Function/Feature Name	Description	Creator
curation.py	Downloads GEO & CuMiDa datasets and stores the files within a temporary directory, then create a database	Tony Okeke
geodlparse: Function	Download, parse, and cache data from GEO	Tony Okeke
CuMiDa: Class	Class that manipulates data and combine it with corresponding GPL	Tony Okeke
curate: Function	Download data from CuMiDa	Tony Okeke
datacheck: Function	Checks if CaBiD database exists and create it if it doesn't	Tony Okeke
gui.py	Allows user to select preferred cancer type and dataset to perform gene expression analysis	Ali Youssef
ControlBox: Class	Class to create UI controls for	Ali Youssef
GUIPanel: Class	Main GUI panel	Ali Youssef
CaBiD GUI: Class	Main GUI window for CaBiD	Ali Youssef
dge.py	Retrieves input values from gui.py and perform analysis by generating: p-values, full change (FC), volcano map, heat map	Cooper Molloy
dge(): Function	Function responsible for performing data analysis: Welch's t-test, p-val, and plots	Cooper Molloy
Plot_volcano(): Function	Creates the volcano map that have threshold for both the FC (x-axis)and adjusted p-values (y-axis)	Cooper Molloy
Plot_heatmap(): Function	Create a clustered heat map that have genes (x-axis) and samples (y-axis)	Cooper Molloy

Schema (ER Diagram):