

Εφαρμογή Streamlit για Ανάλυση scRNA-seq Δεδομένων

Ομάδα: Γεώργιος Τσαούσης, Φωτεινή Τσιανάκα, Παρασκευή Σάλλια
ΑΜ: Π2019032, Π2020052, Inf2022179

Άνοιξη 2025

Περίληψη

Αναπτύξαμε μια διαδραστική εφαρμογή με Streamlit για την ανάλυση δεδομένων single-cell RNA-seq. Η εφαρμογή υποστηρίζει upload αρχείων τύπου .h5ad, preprocessing με παραμετροποίηση από τον χρήστη και προβολή αποτελεσμάτων μέσω UMAP. Η υλοποίηση έχει Dockerized μορφή για φορητότητα και είναι διαθέσιμη σε δημόσιο GitHub repository.

1 Εισαγωγή

Η ανάλυση single-cell RNA sequencing (scRNA-seq) έχει μεταμορφώσει τη βιολογία. Ωστόσο, απαιτεί σύνθετο προγραμματισμό. Στόχος μας είναι η δημιουργία μιας εύχρηστης Streamlit εφαρμογής για preprocessing και visualization τέτοιων δεδομένων.

2 Σχεδιασμός Υλοποίησης

Η εφαρμογή αναπτύχθηκε με γνώμονα τη modular αρχιτεκτονική. Χωρίζεται σε τρεις κύριες λογικές ενότητες (tabs):

- **Analysis:** Ασχολείται με τη φόρτωση δεδομένων και εκτέλεση του preprocessing pipeline (filtering, normalization, PCA, UMAP).
- **Visualization:** Επιτρέπει στον χρήστη να δει δυναμικά τις οπτικοποιήσεις UMAP και τα μεταδεδομένα.
- **Team Info:** Παρουσιάζει τα μέλη της ομάδας και τις συνεισφορές τους.

Η εφαρμογή χρησιμοποιεί τη βιβλιοθήκη Streamlit, και η δομή των αρχείων ακολουθεί τη λογική ενός MVC-like διαχωρισμού:

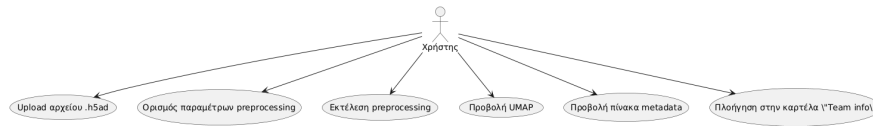
- Το αρχείο `main.py` χειρίζεται την πλοήγηση μεταξύ των tabs.
- Κάθε tab έχει το δικό του module (π.χ. `analysis.py`) και λειτουργεί με συνάρτηση `run()`.

- Τα δεδομένα μεταφέρονται μέσω `st.session_state`, ώστε να είναι προσβάσιμα μεταξύ tabs.

Το περιβάλλον υλοποιήθηκε ώστε να μπορεί να γίνει containerized μέσω Docker και να εκτελείται παντού με μία εντολή.

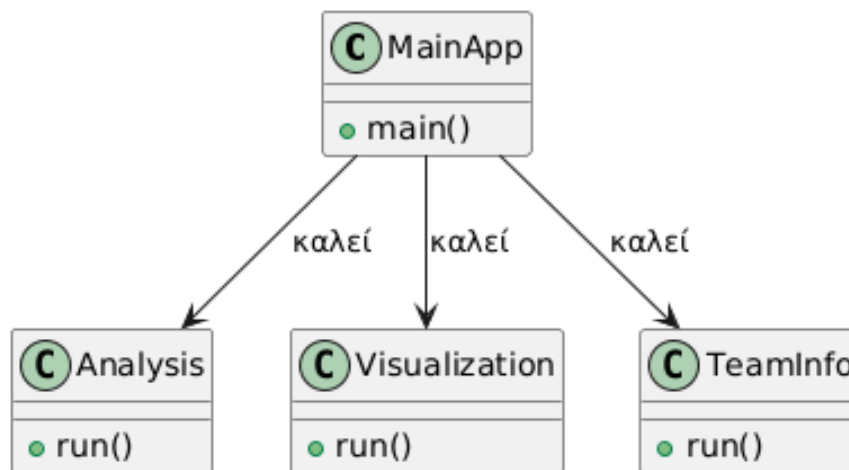
3 UML Διαγράμματα

3.1 Use Case Diagram



Σχήμα 1: Use case diagram της εφαρμογής

3.2 Class Diagram



Σχήμα 2: Class/module διάγραμμα των αρχείων Python

4 Ανάλυση Υλοποίησης

Το βασικό pipeline ανάλυσης είναι υλοποιημένο στο αρχείο `analysis.py`, και περιλαμβάνει τα εξής βήματα:

1. **Φόρτωση δεδομένων:** ο χρήστης κάνει upload αρχείο τύπου `.h5ad`, το οποίο φορτώνεται με τη βιβλιοθήκη `Scanpy`.
2. **Φιλτράρισμα:** γίνεται filtering κυττάρων με λιγότερα από `min_genes` και γονιδίων με λίγα κύτταρα.

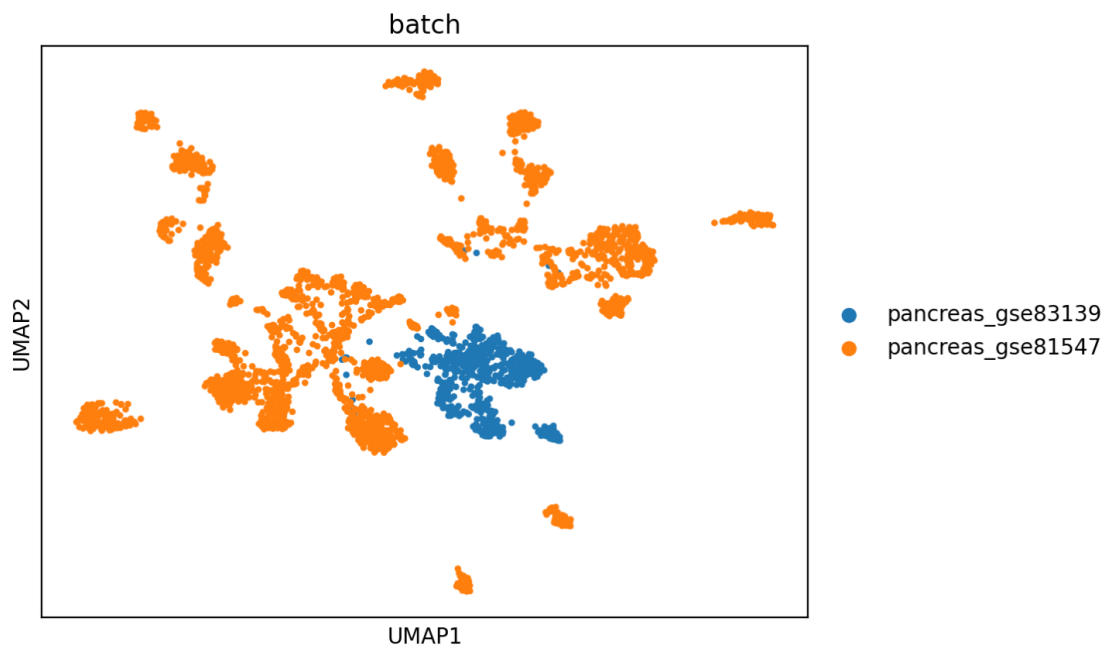
3. **Κανονικοποίηση:** εφαρμόζεται `normalize_total()` και `log1p()`.
4. **Επιλογή γονιδίων υψηλής διακύμανσης:** μέσω `highly_variable_genes()`.
5. **PCA, Neighbors, UMAP:** υπολογίζονται για χωρική προβολή των κυττάρων.

Η καρτέλα `Visualization` επιτρέπει την επιλογή μεταβλητής από το `adata.obs` για δυναμική UMAP απεικόνιση. Το plot δημιουργείται με `scanpy.pl.umap()` και επιστρέφεται μέσω `return_fig=True` ώστε να ενσωματωθεί με `st.pyplot()`.

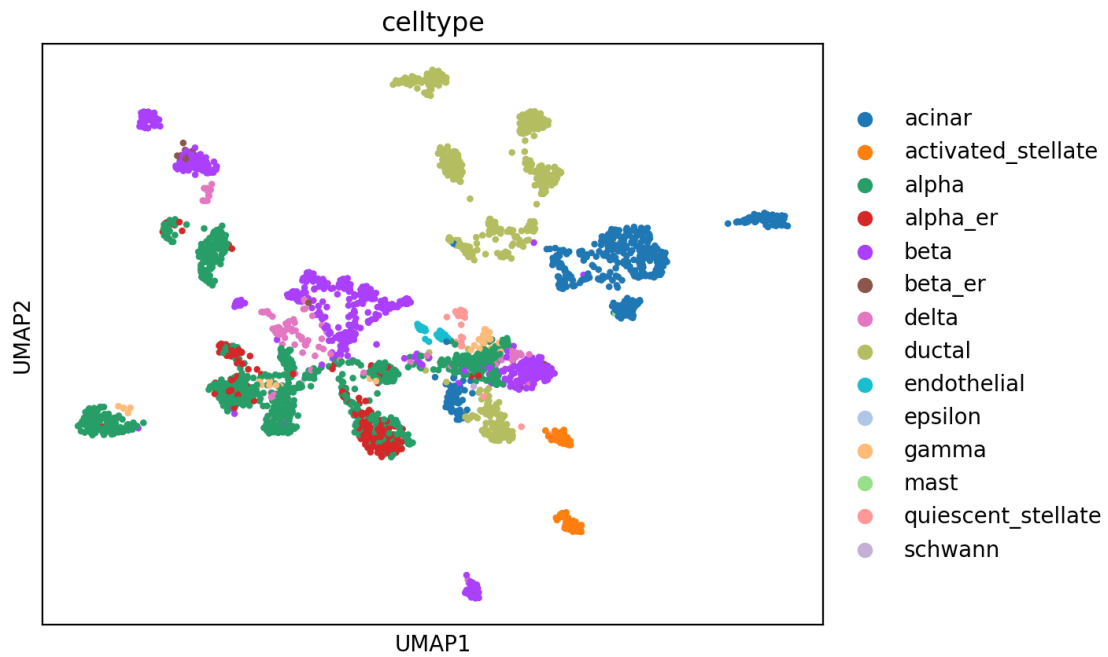
Η χρήση του `session_state` είναι κρίσιμη, καθώς διατηρεί τα δεδομένα της ανάλυσης διαθέσιμα για όλα τα tabs χωρίς επανυπολογισμό.

Ο διαχωρισμός των αρχείων ακολουθεί την αρχή *separation of concerns*, με κάθε αρχείο να έχει ξεκάθαρο ρόλο. Η εφαρμογή είναι σχεδιασμένη ώστε να μπορεί να επεκταθεί εύκολα, π.χ. με clustering ή differential expression.

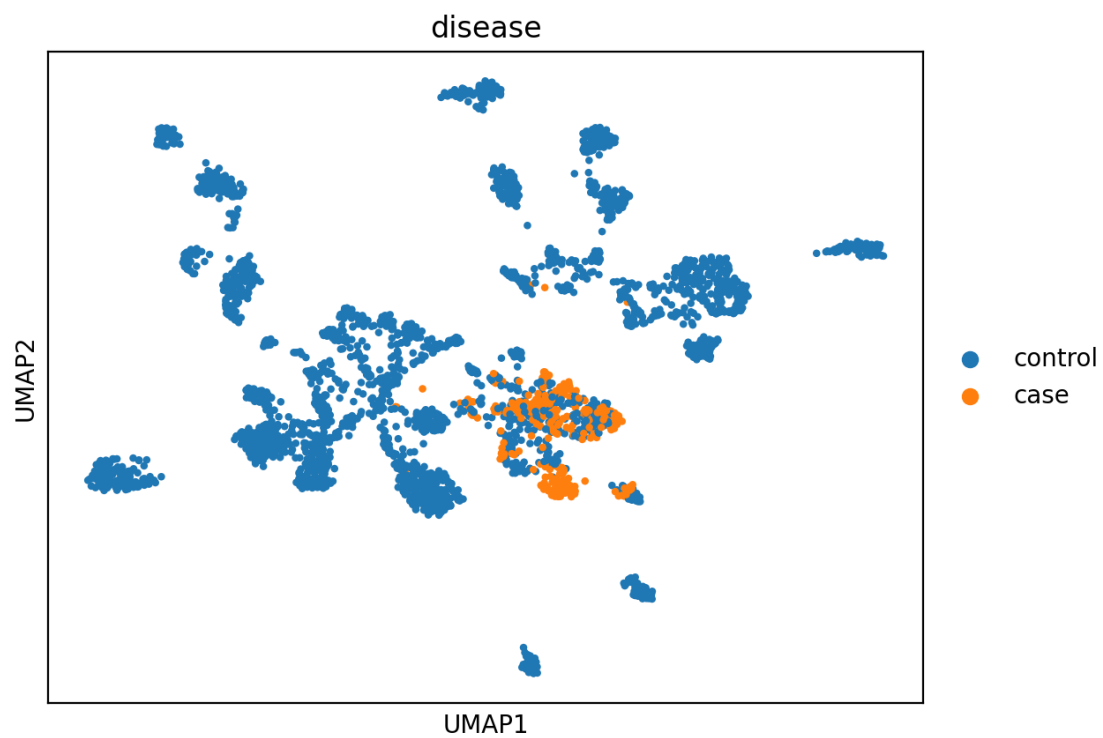
5 Οπτικοποιήσεις Αποτελεσμάτων



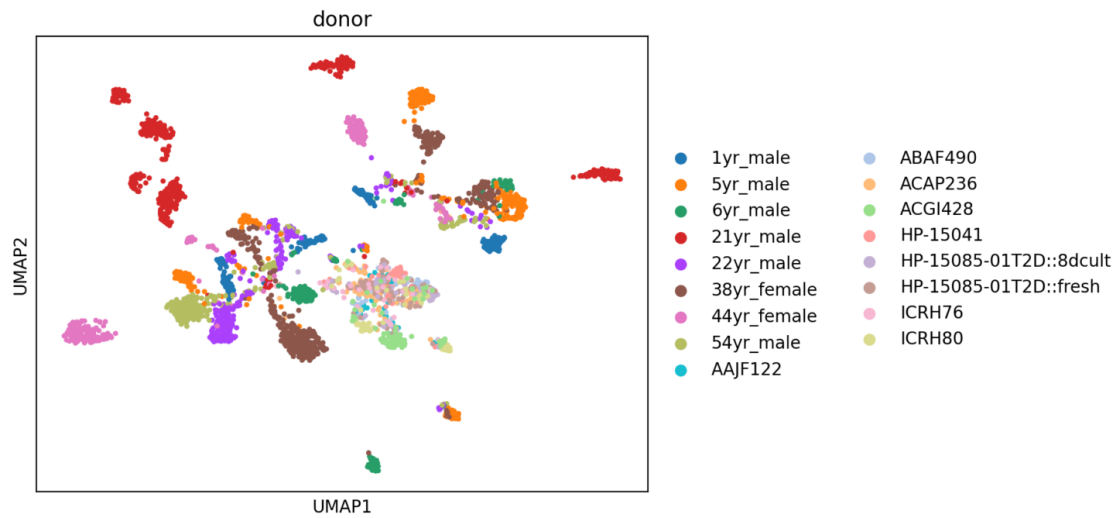
Σχήμα 3: UMAP χρωματισμένο ανά batch



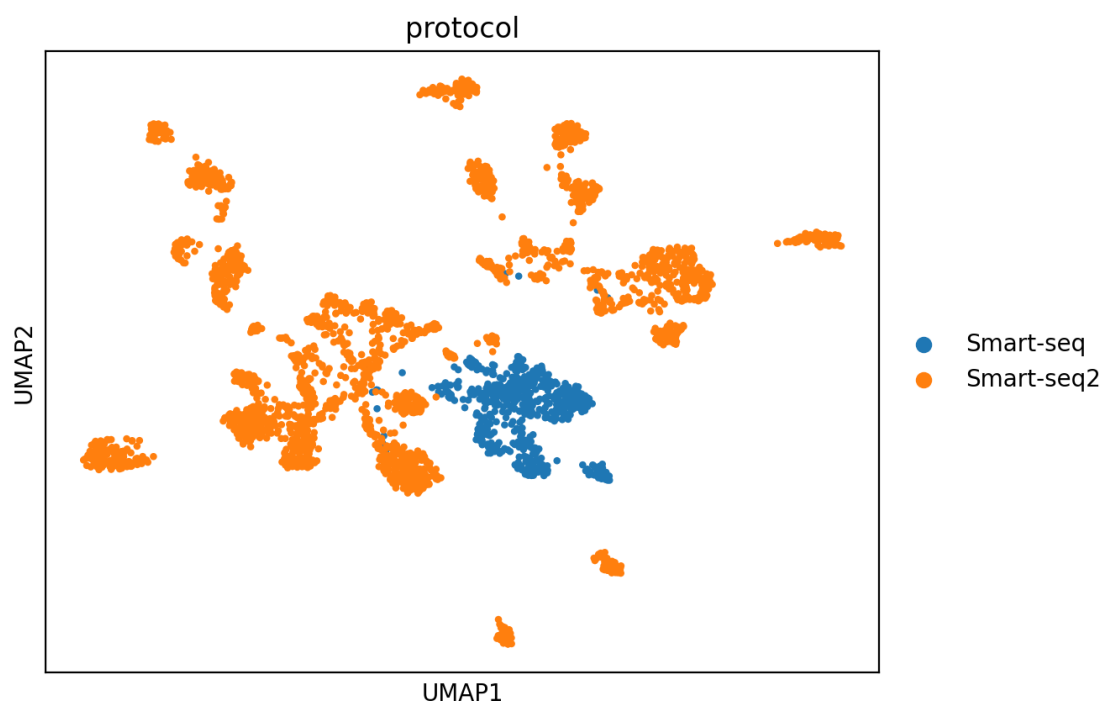
Σχήμα 4: UMAP χρωματισμένο ανά cell type



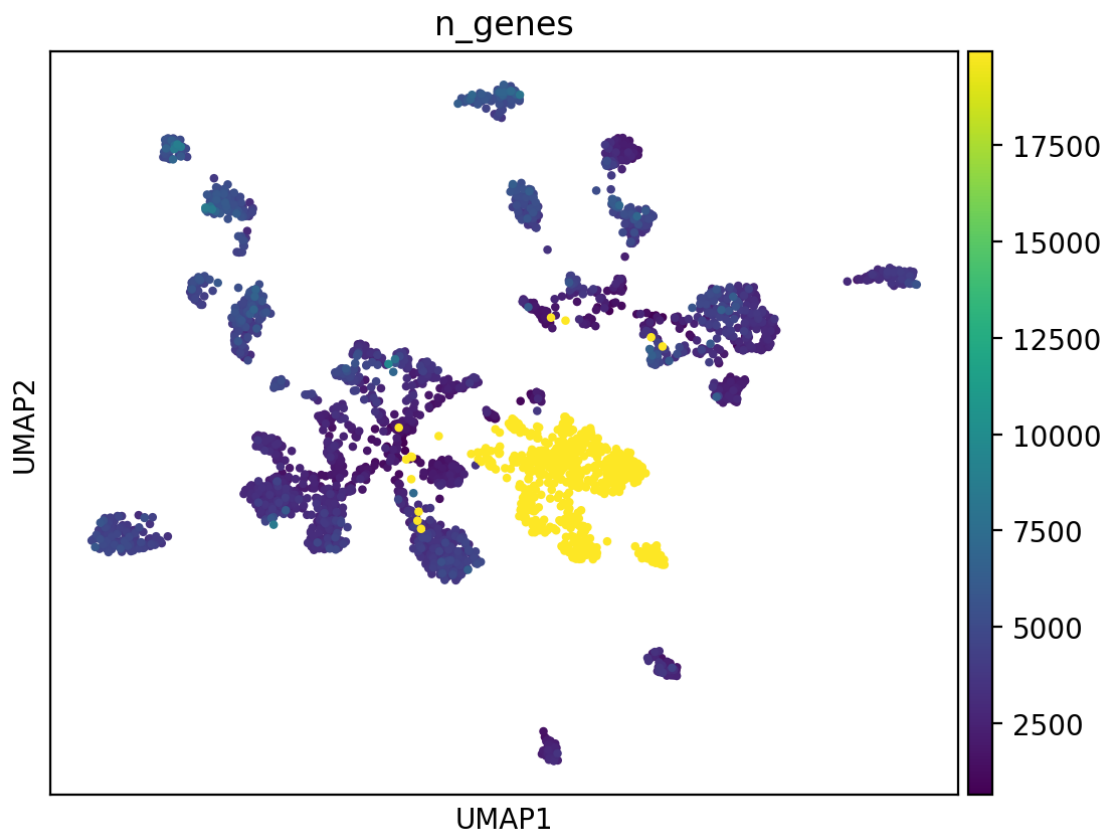
Σχήμα 5: UMAP χρωματισμένο ανά disease



Σχήμα 6: UMAP χρωματισμένο ανά donor



Σχήμα 7: UMAP χρωματισμένο ανά protocol

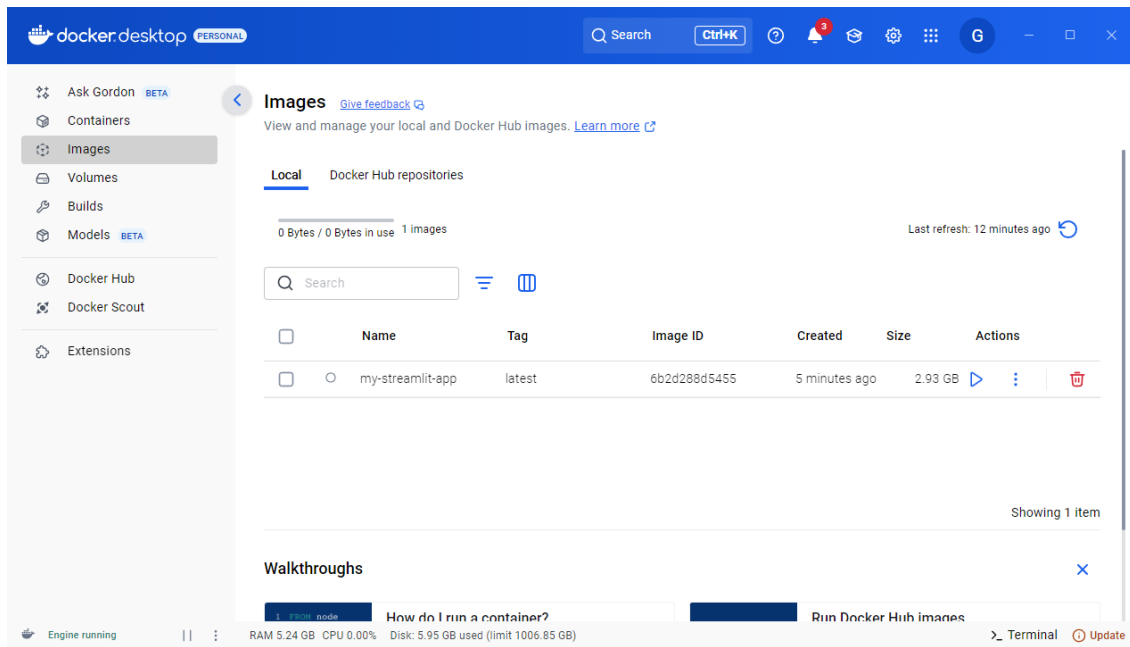


Σχήμα 8: UMAP χρωματισμένο ανά αριθμό γονιδίων

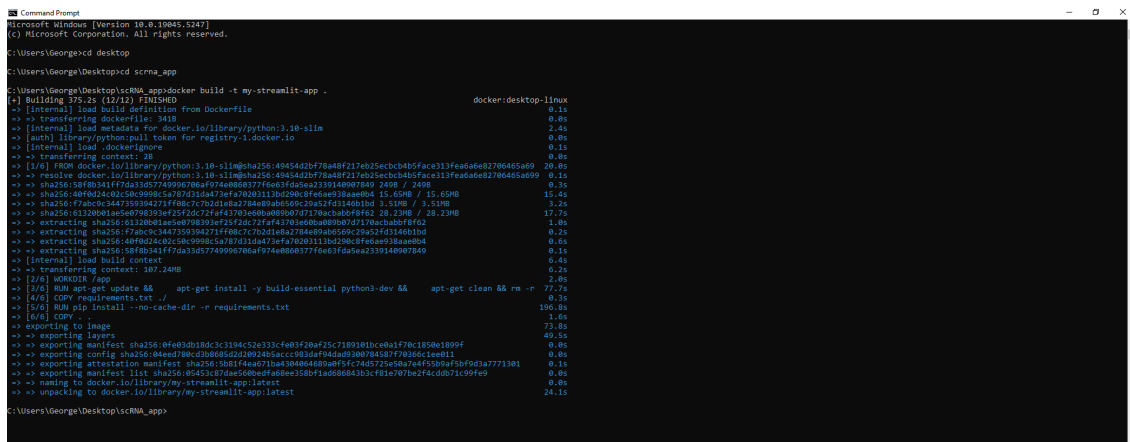
6 Dockerization

Για το dockerize της εφαρμογής, χρησιμοποιήσαμε το Docker Desktop application. Με την χρήση του .dockerfile και το Docker Desktop κάναμε "containerize" την εφαρμογή με όλα τα απαραίτητα αρχεία και βιβλιοθήκες έτσι ώστε να μπορεί να εκτελεστεί σε οποιοδήποτε περιβάλλον χωρίς επιπλέον ρυθμίσεις.

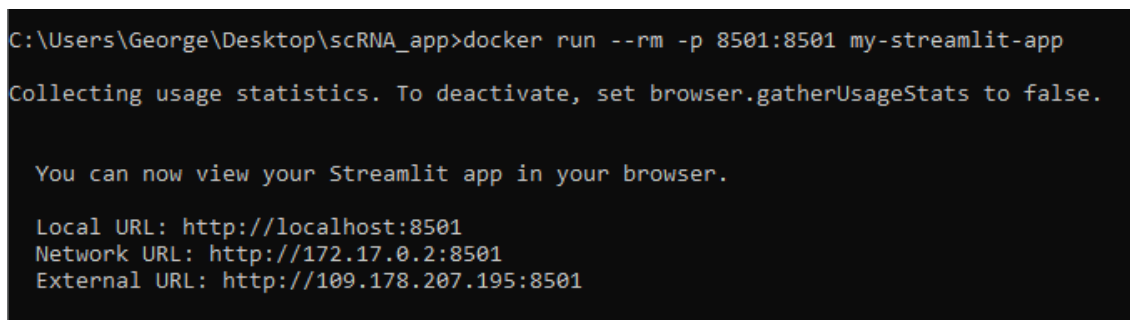
```
docker build -t scrna-app .  
docker run --rm -p 8501:8501 scrna-app
```



Σχήμα 9: Docker Desktop



Σχήμα 10: Χρήση build prompt στο CMD



Σχήμα 11: Χρήση run prompt στο CMD

7 GitHub Repository

Ο κώδικας είναι διαθέσιμος στο GitHub:
<https://github.com/GeorgeTsaousis/scRNA-streamlit-app>