Maintenance Guide

Objective: Guide for maintaining, updating, and troubleshooting the codebase to ensure continued functionality.

System Requirements

- **Software:**
  - Python 3.x
  - Libraries: pandas, scipy, tqdm, matplotlib, seaborn, scikit-learn.
- **Hardware:**
  - 8 GB RAM (minimum recommended).
  - Google Colab or equivalent computing environment.

Maintenance Tasks

1. **Updating Libraries:** Periodically update Python libraries to their latest versions to ensure compatibility:

```
pip install --upgrade pandas scipy tqdm matplotlib seaborn scikit-learn
```

2. **Modifying the Code:**
   - **Adding Tissues:** To analyze new tissues, add their names and folder paths to organ_names and folder_names.
   - **Changing Statistical Thresholds:** Modify the p-value threshold in the t-test logic for different sensitivity levels:

```
diff_genes_df = res_df[res_df['p_value'] < 0.02].sort_values('p_value')
```

3. **Pathway Analysis Updates:**
   - Adjust visualizations by editing the pathway filtering logic.
4. **Random Forest Model:**
   - Experiment with model parameters like n_estimators or random_state in the RandomForestClassifier.

Troubleshooting

- **Memory Errors:** If Colab crashes, consider splitting the dataset into smaller chunks or upgrading the environment to Colab Pro.
- **Missing Outputs:**
  - Check if file paths are correctly defined.
  - Ensure all required files exist in the specified directory.

Backup and Recovery

- **Backup:** Save processed data and generated CSV files to Google Drive or GIT.
- **Recovery:** Keep a backup of the code and data files. Restore from the backup in case of accidental deletions.