

Project Outline: APOL1 Heat Map Development

Project Title:

Data Collection for Identifying Geographic Regions with APOL1-Positive Patients in the UK

Objective:

To gather, organize, and validate data sources to identify regions with the highest number of APOL1-positive patients, forming the foundation for a geographic heat map.

Phase I: Data Gathering

1. Primary Hypothesis:

Regions with higher populations of individuals of African descent in the UK are more likely to have a higher prevalence of APOL1-positive patients.

2. Data Source Identification:

- **Public Healthcare Databases:**
 - NHS resources
 - UK Biobank
 - Genomics England
- **Forums and Patient Advocacy Groups:**
 - Online communities and social media platforms
 - Patient support groups for kidney diseases
- **Academic Publications and Clinical Trials:**
 - PubMed
 - ClinicalTrials.gov
- **Public Health Datasets:**
 - Office for National Statistics (ONS)
 - Public Health England (PHE)

3. Raw Data Collection:

- Compile a comprehensive list of data sources.
 - Document access links, permissions, and relevance.
 - Extract raw data from identified sources.
-

Phase II: Data Consolidation and Preprocessing

1. Data Structuring:

- Organize data into structured formats suitable for analysis (e.g., CSV, Excel).
- Annotate data with metadata (source, relevance, date).

2. Data Quality Assessment:

- Verify data consistency and reliability.

- Assess the relevance and quality of data collected.
- Document any restrictions or limitations on usage.

3. Preliminary Insights:

- Highlight regions or trends observed during the collection process.
 - Initial observations on potential high-priority regions based on data trends.
-

Phase III: Exploratory Data Analysis (EDA)

1. Data Exploration:

- Conduct descriptive statistics to understand data distribution.
- Identify patterns and correlations related to APOL1 prevalence.

2. Visualization:

- Create initial visualizations to explore geographic distribution.
 - Use charts and graphs to identify potential hotspots.
-

Phase IV: Heat Map Development

1. Data Preparation for Heat Map:

- Clean and preprocess data for geographic analysis.
- Ensure data is geocoded for mapping purposes.

2. Dashboard Creation:

- Develop a dashboard using tools like Tableau or Power BI.
- Integrate a heat map to visualize geographic regions with APOL1-positive patients.

3. Final Analysis and Reporting:

- Provide a comprehensive report with findings and insights.
 - Highlight key regions for further investigation or intervention.
-

Deliverables:

1. Data Source Index:

- Documented index of all data sources used.

2. Structured Data Files:

- Consolidated and cleaned datasets ready for analysis.

3. Exploratory Data Analysis Report:

- Summary of initial findings and visualizations.

4. **Interactive Dashboard:**

- Heat map showcasing geographic distribution of APOL1-positive patients.

5. **Final Report:**

- Detailed analysis and recommendations for next steps.
-