**Aim 1 Process**

1. Identify outcome (e.g., pancreatic cancer [157])
2. Identify cases
   1. MGI + UKB
3. Match cases with controls
   1. MGI + UKB
      1. MGI
         1. ID, age, age\_at\_first\_diagnosis, age\_at\_last\_diagnosis, sex, length\_of\_follow\_up, number\_of\_encounters, encounter\_density
4. Truncate phenome
   1. Use date of case assignment in matched groups
5. Weights
   1. Pruning and thresholding – NEED SELECTION WEIGHTS
   2. Pruning and thresholding with poststratification – NEED SELECTION AND POSTSTRATIFICATION WEIGHTS
   3. Pruning and thresholding with ridge regression – NEED SELECTION WEIGHTS
   4. Group LASSO

File structure

* .gitignore – includes data/ folder!
* data/
  + raw/
    - DATE/
  + processed
    - DATE/
  + weights/
    - DATE/
      * selection/
        + PHECODE/
      * poststratification/
        + PHECODE/
      * disease/
        + PHECODE/
* results/
  + DATE/
* fn/
  + functions
* r/
  + scripts
* inbox/
  + temporary or scratch files to be deleted or moved
* renv/
  + created by renv R package – maintained via renv::snapshot()