PUMCH datasets

Four rare disease datasets are filed: PUMCH-S.json, PUMCH-L.json, PUMCH-MDT.json, and PUMCH-ADM.json. Those files be fed into the PhenoBrain to calculate results.

For convenience to present, we compiled those four datasets into a single excel file (PUMCH_dataset.xlsx).

Those files were complied by Xiaohao Mao(Tsinghua University), Yu Huang(Tsinghua University & Tencent Jarvis lab) and Lun Wang(Peking Union Medical College Hospital) after removing patients' information, including a total of 1,149 rare disease cases spanning over 100s diseases.

Dataset	Number of cases	Number of rare diseases
PUMCH-S	86	39
PUMCH-L	961	87
PUMCH-MDT	27	23
PUMCH-ADM	75	16