Datasets

1. Human Microbiome Project
https://drive.google.com/file/d/1d3TXk37IslmXL1wO8QCrWBJZad7CzcUe/view?usp=sharing

2. Genetic Ancestry Inference Project
https://drive.google.com/drive/folders/1fv1AODzbUiCo2ZAXncLuUBs8L4yTO1QY?usp=sharing

3. COVID19byZip
https://github.com/Big-Bio/COVID19byCountyAndZip

4. Cyclebase 3.0: A multi-organism database on cell-cycle regulation and phenotypes
https://cyclebase.org/CyclebaseSearch

5. SARS
https://www.who.int/csr/sars/country/en/

6. Genotype-phenotype data
https://drive.google.com/file/d/1IqRqXFbpQBxinmCcBOQVk_7tUCJbzp5B/view?usp=drive_web

7. Single-cell dataset
https://drive.google.com/file/d/1TcQZJv69anDzCcHWD6nABEfsaajXPMWn/view?usp=drive_web

8. COVID-19 Open Research Dataset Challenge (CORD-19)
https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge

9. APTOS 2019 Blindness Detection
https://www.kaggle.com/c/aptos2019-blindness-detection/data

10. OSIC Pulmonary Fibrosis Progression
https://www.kaggle.com/c/osic-pulmonary-fibrosis-progression/data

11. OpenVaccine: COVID-19 mRNA Vaccine Degradation Prediction
https://www.kaggle.com/c/stanford-covid-vaccine/data

12. Chest X-Ray Images (Pneumonia)
https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

13. Personalized Medicine: Redefining Cancer Treatment
https://www.kaggle.com/c/msk-redefining-cancer-treatment/data

14. BioBombe gene expression dataset for cancer classification
https://greenelab.github.io/BioBombe/#resources

15. Brain MRI Images for Brain Tumor Detection
https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection

16. ProteinNet (machine learning of protein structure)
https://github.com/aqlaboratory/proteinnet

17. MIMIC EHR dataset
https://mimic.physionet.org/

18. CheXpert
https://stanfordmlgroup.github.io/competitions/chexpert/

19. KEGG Metabolic Regulatory Network
https://archive.ics.uci.edu/ml/datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29
https://archive.ics.uci.edu/ml/datasets/KEGG+Metabolic+Relation+Network+%28Directed%29

20. Single-cell expression atlas
https://www.ebi.ac.uk/gxa/sc/

21. TargetFinder (DNA-DNA interaction pairs)
https://github.com/shwhalen/targetfinder

22. METABRIC (Cancer genomics)
http://www.cbioportal.org/study/summary?id=brca_metabric

23. Leukemia image classification
https://www.kaggle.com/andrewmvd/leukemia-classification

24. Retinal OCT Images (optical coherence tomography)
https://www.kaggle.com/paultimothymooney/kermany2018

Miscellaneous
https://github.com/blengerich/CompBioDatasetsForMachineLearning
https://www.kaggle.com/tags/healthcare

25. Single cell dataset:
https://ucsf.box.com/s/nwrb07m9rp86nkcs4zo5vztqde0bdpnf
Password: eZhRCisXZ8^S41p%m

There are four different tasks with the corresponding data in the input folder.
Task description: (top directory of box): sc_hack2021_tasks.docx
Introductions can be found via Zoom recording files.
Data description can be found in hackathon slides folder.