Visualizations of Genotype Datasets

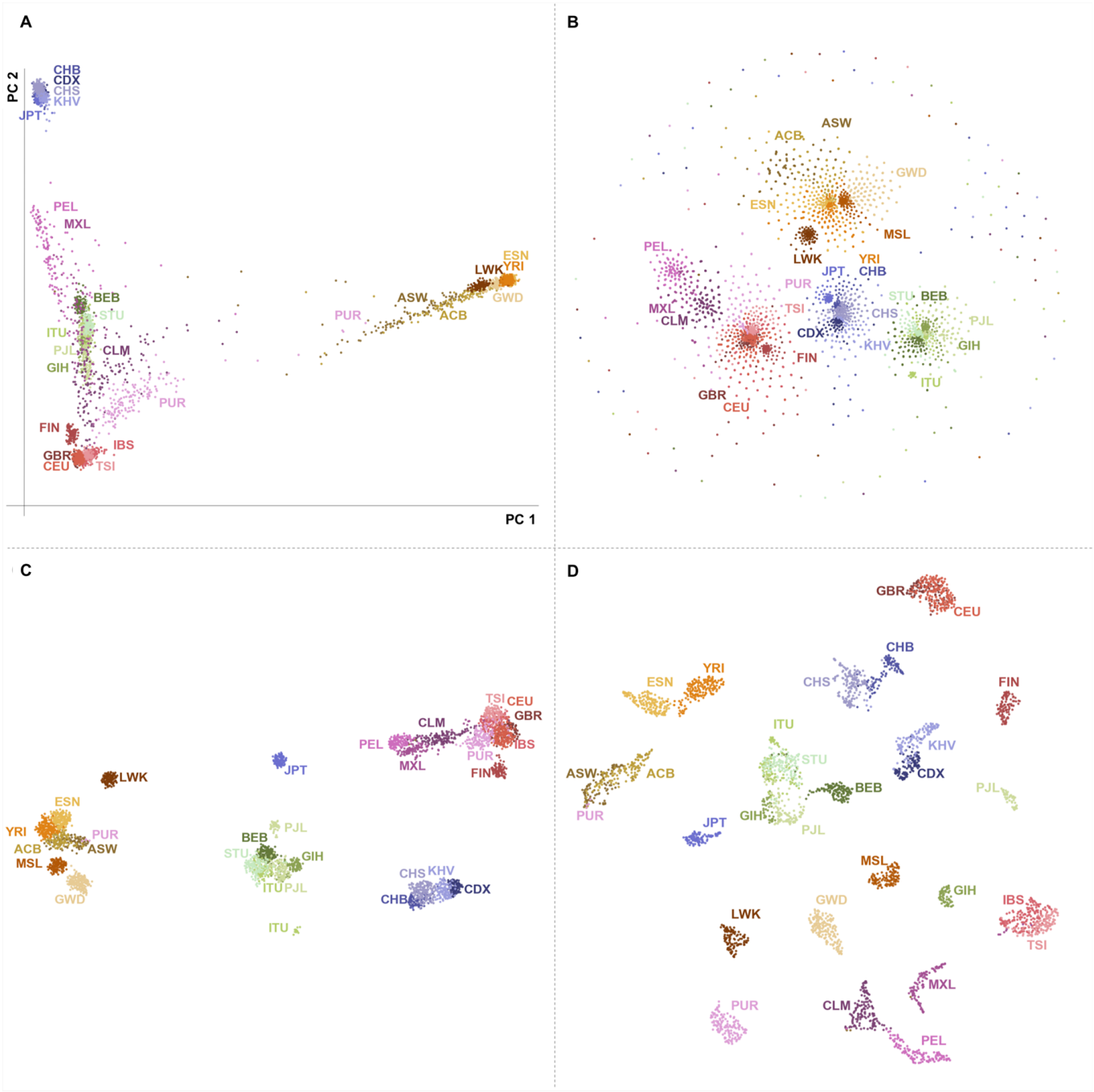
This repository contains the visualizations of 1000 Genomes dataset (1000 Genomes Project Consortium) and the modern Eurasian population dataset (Lamnidis et al.). The dimensional reduction techniques used include PCA, t-SNE, UMAP, and Adversarial Autoencoders (Makhzani et al.)

All visualizations can be founded in the form of interactive htmls in the `htmls` folder. The interactive htmls let you pan, zoom, hover on data points, and also hide data of particular labels by clicking on the legend.

We also hand pick some visualizations and display them as static images below.

1000 Genomes

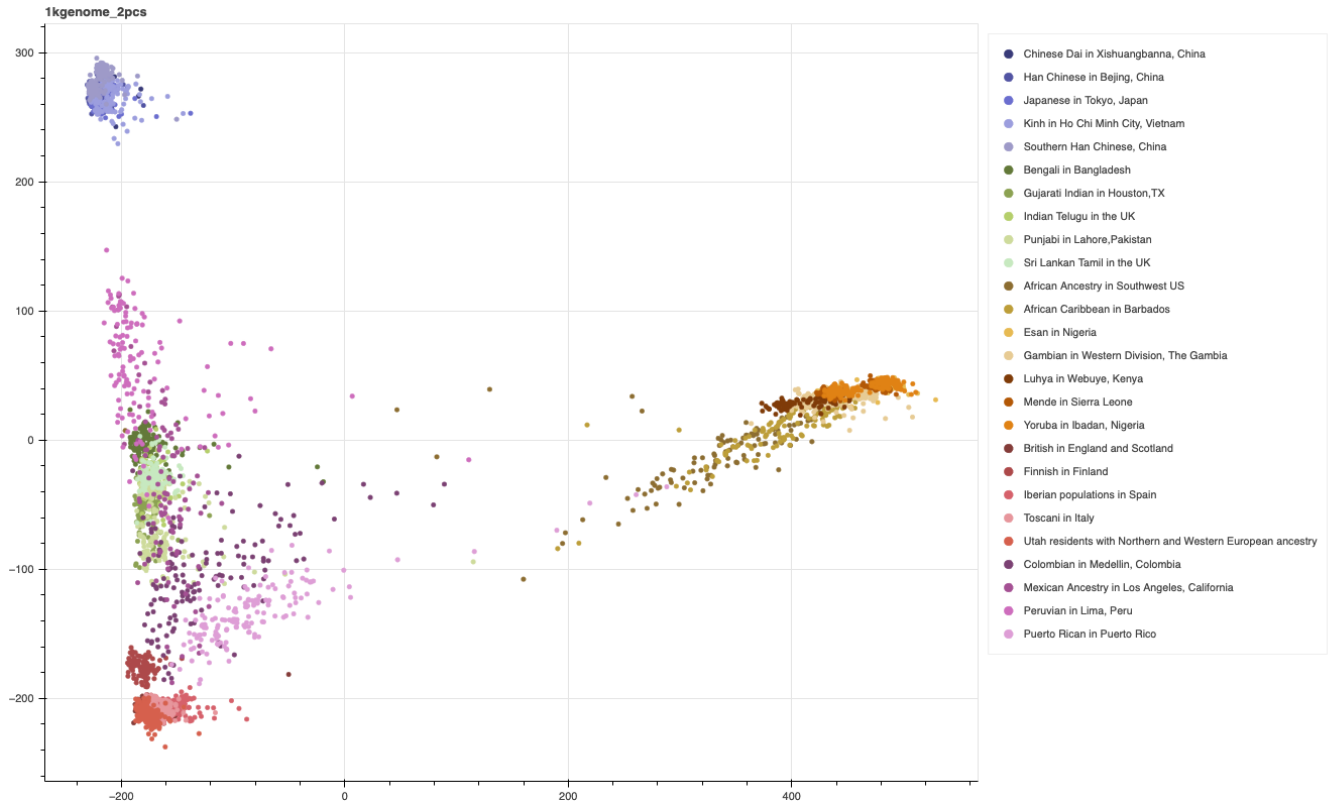
For reference, Diaz-Papkovich et al. have applied PCA, t-SNE, and UMAP to the 1000 Genomes dataset. The following figure is the first figure in the paper



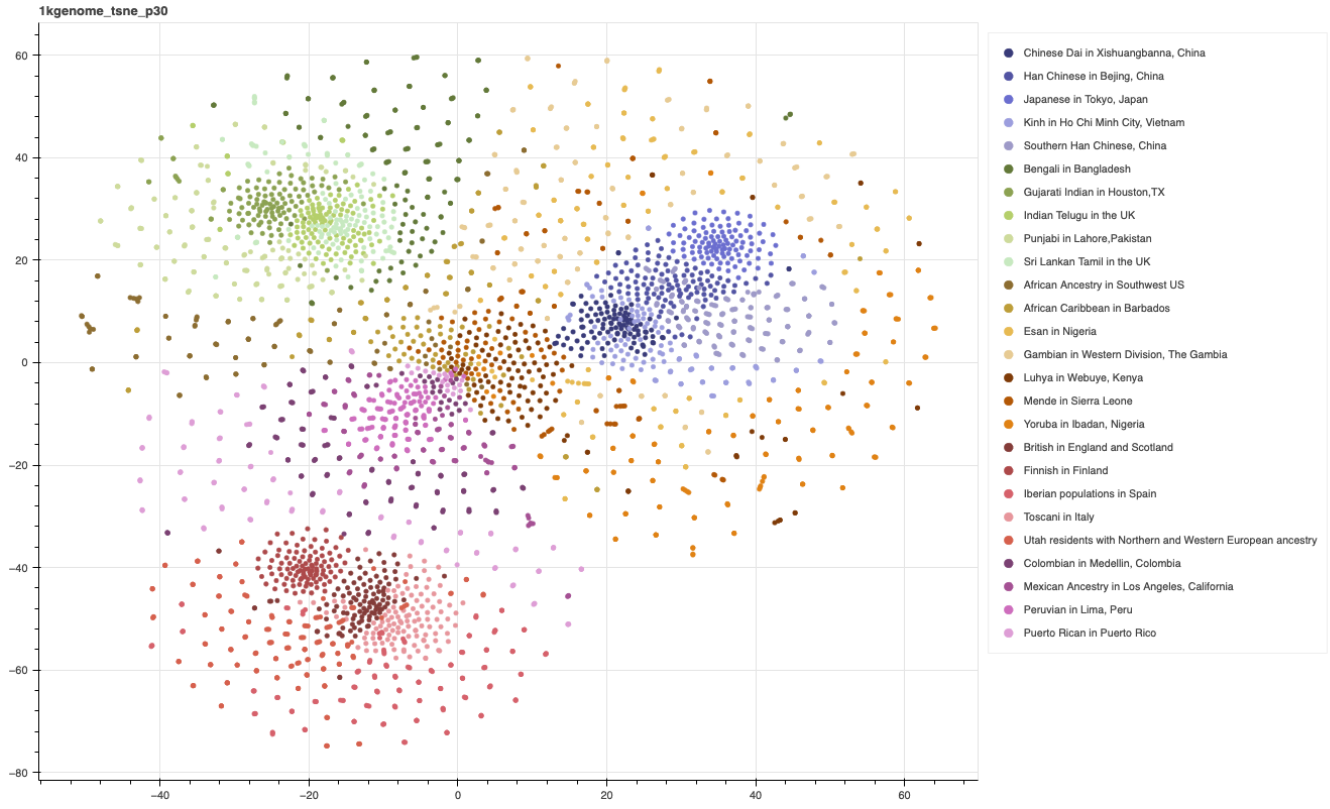
A corresponds to PCA outputs. **B** corresponds to t-SNE outputs. **C** corresponds to UMAP outputs. **D** correponds to the outputs of first applying PCA to reduce the dimension down to 15 and then applying UMAP

Our results are below:

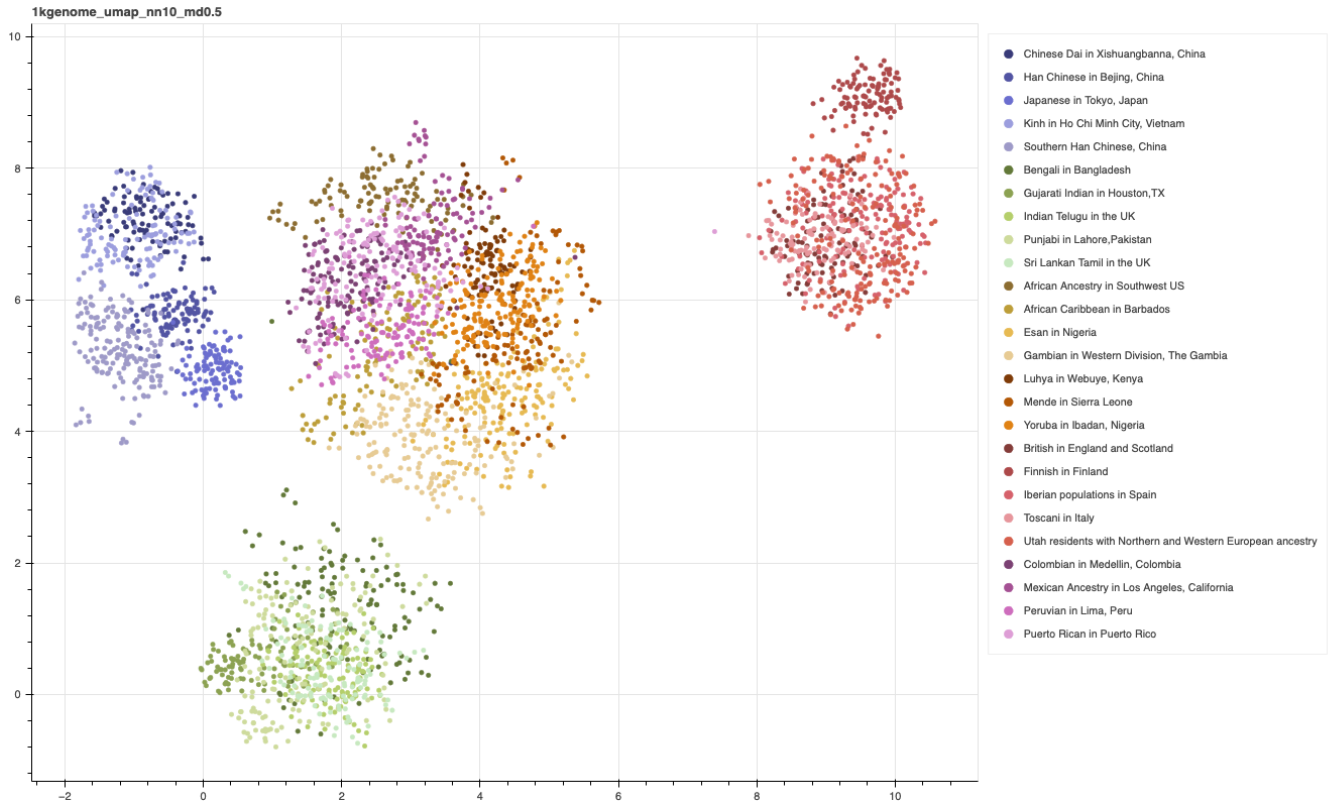
- PCA


1kgenome_2pcs
Chinese Dai in Xishuangbanna, China
Han Chinese in Bejing, China
Japanese in Tokyo, Japan
Kinh in Ho Chi Minh City, Vietnam
Southern Han Chinese, China
Bengali in Bangladesh
Gujarati Indian in Houston,TX
Indian Telugu in the UK
Punjabi in Lahore,Pakistan
Sri Lankan Tamil in the UK
African Ancestry in Southwest US
African Caribbean in Barbados
Esan in Nigeria
Gambian in Western Division, The Gambia
Luhya in Webuye, Kenya
Mende in Sierra Leone
Yoruba in Ibadan, Nigeria
British in England and Scotland
Finnish in Finland
Iberian populations in Spain
Toscani in Italy
Utah residents with Northern and Western European ancestry
Colombian in Medellin, Colombia
Mexican Ancestry in Los Angeles, California
Peruvian in Lima, Peru
Puerto Rican in Puerto Rico

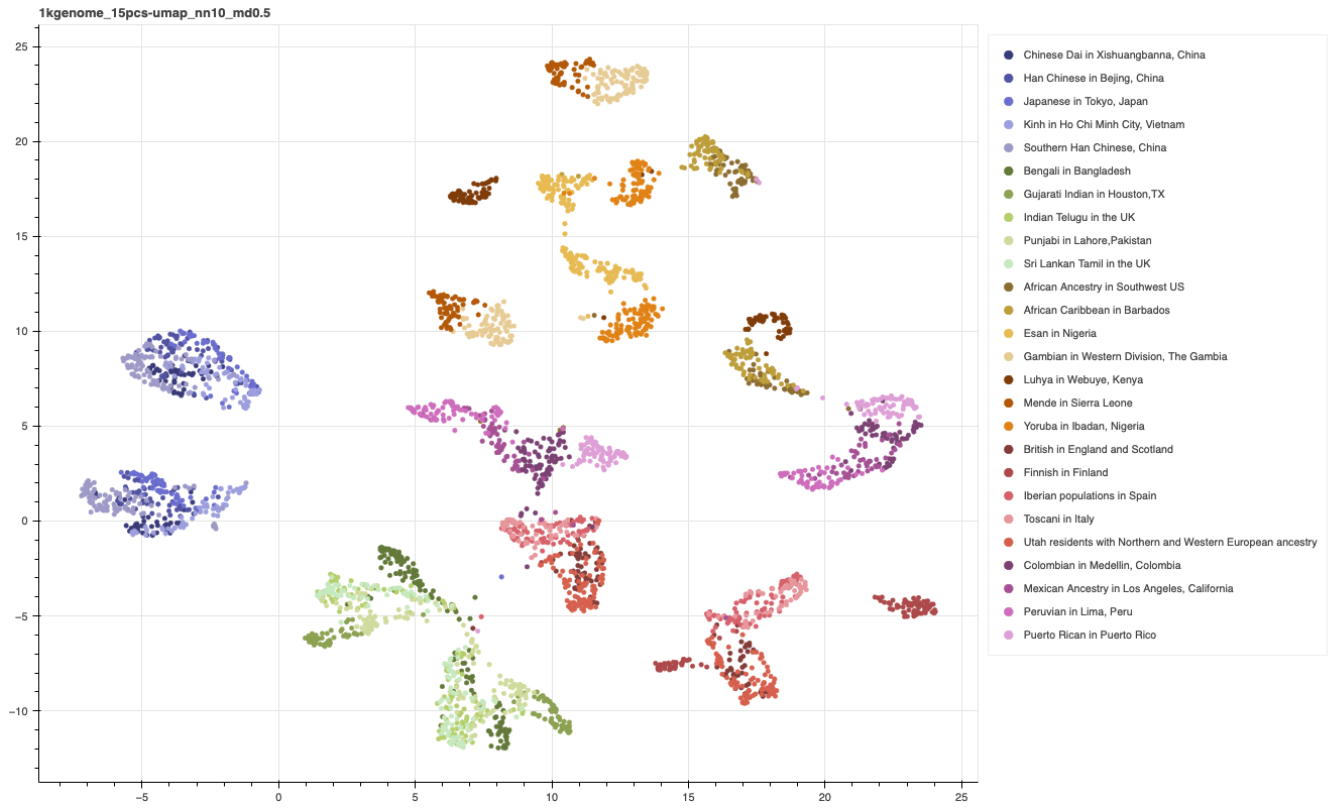
- t-SNE


1kgenome_tsne_p30
Chinese Dai in Xishuangbanna, China
Han Chinese in Bejing, China
Japanese in Tokyo, Japan
Kinh in Ho Chi Minh City, Vietnam
Southern Han Chinese, China
Bengali in Bangladesh
Gujarati Indian in Houston,TX
Indian Telugu in the UK
Punjabi in Lahore,Pakistan
Sri Lankan Tamil in the UK
African Ancestry in Southwest US
African Caribbean in Barbados
Esan in Nigeria
Gambian in Western Division, The Gambia
Luhya in Webuye, Kenya
Mende in Sierra Leone
Yoruba in Ibadan, Nigeria
British in England and Scotland
Finnish in Finland
Iberian populations in Spain
Toscani in Italy
Utah residents with Northern and Western European ancestry
Colombian in Medellin, Colombia
Mexican Ancestry in Los Angeles, California
Peruvian in Lima, Peru
Puerto Rican in Puerto Rico

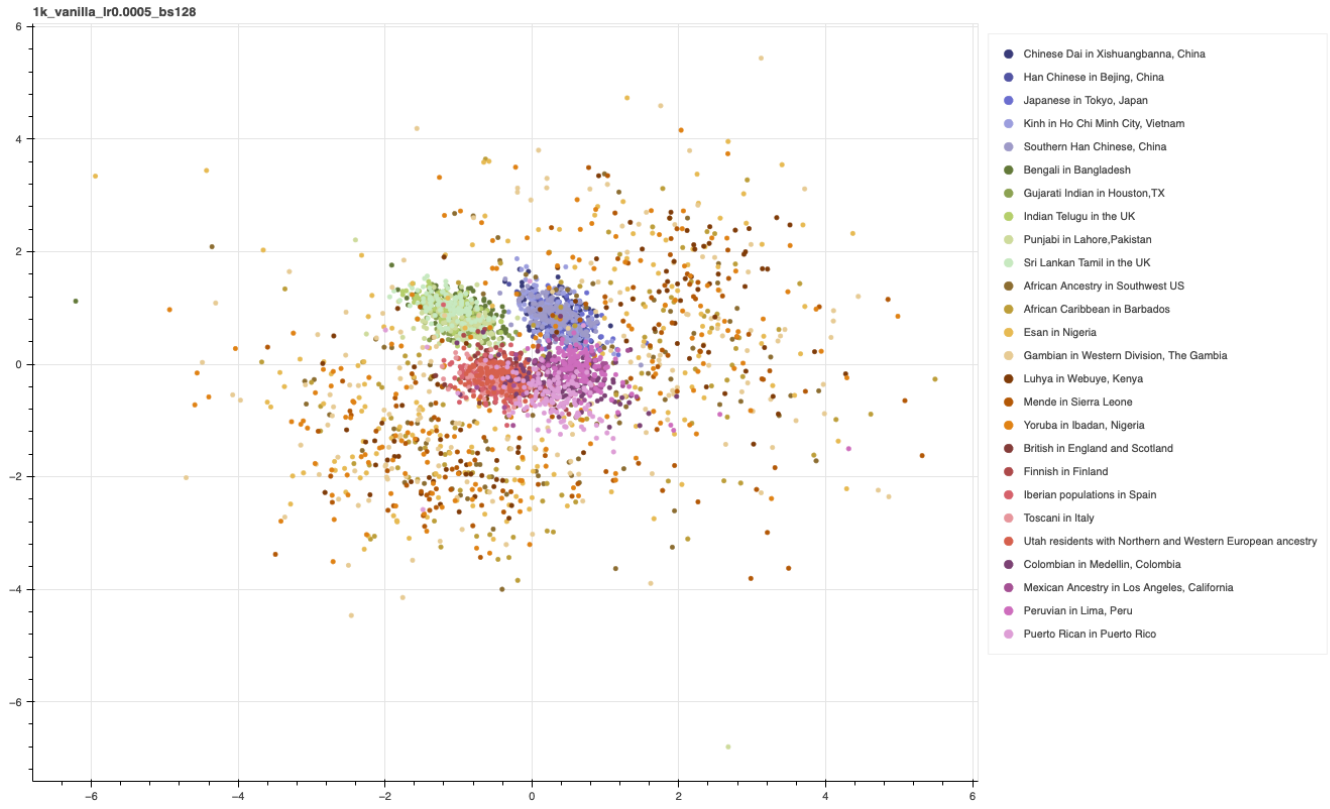
- UMAP



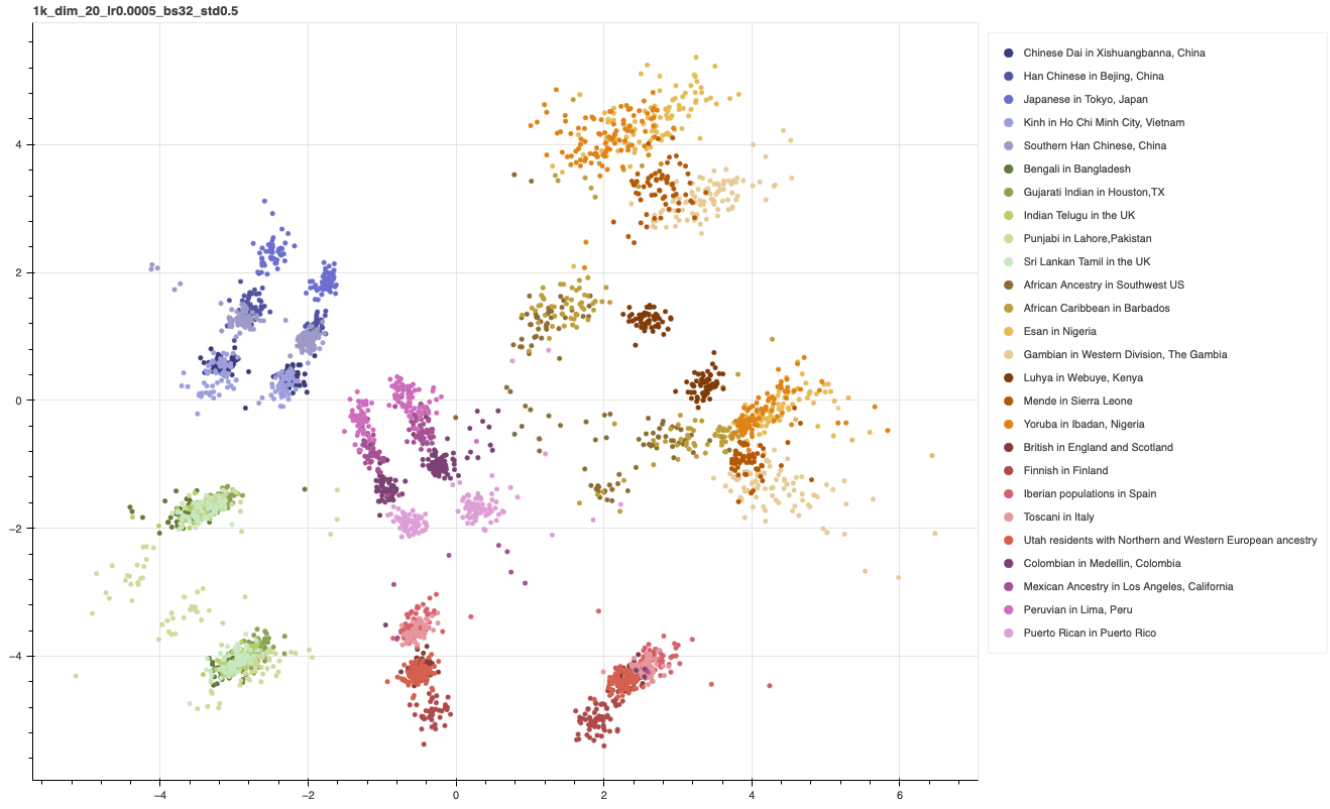
- UMAP on 15 PCs



- Vanilla Adversarial Autoencoders



- Adversarial Autoencoders with Cluster Heads

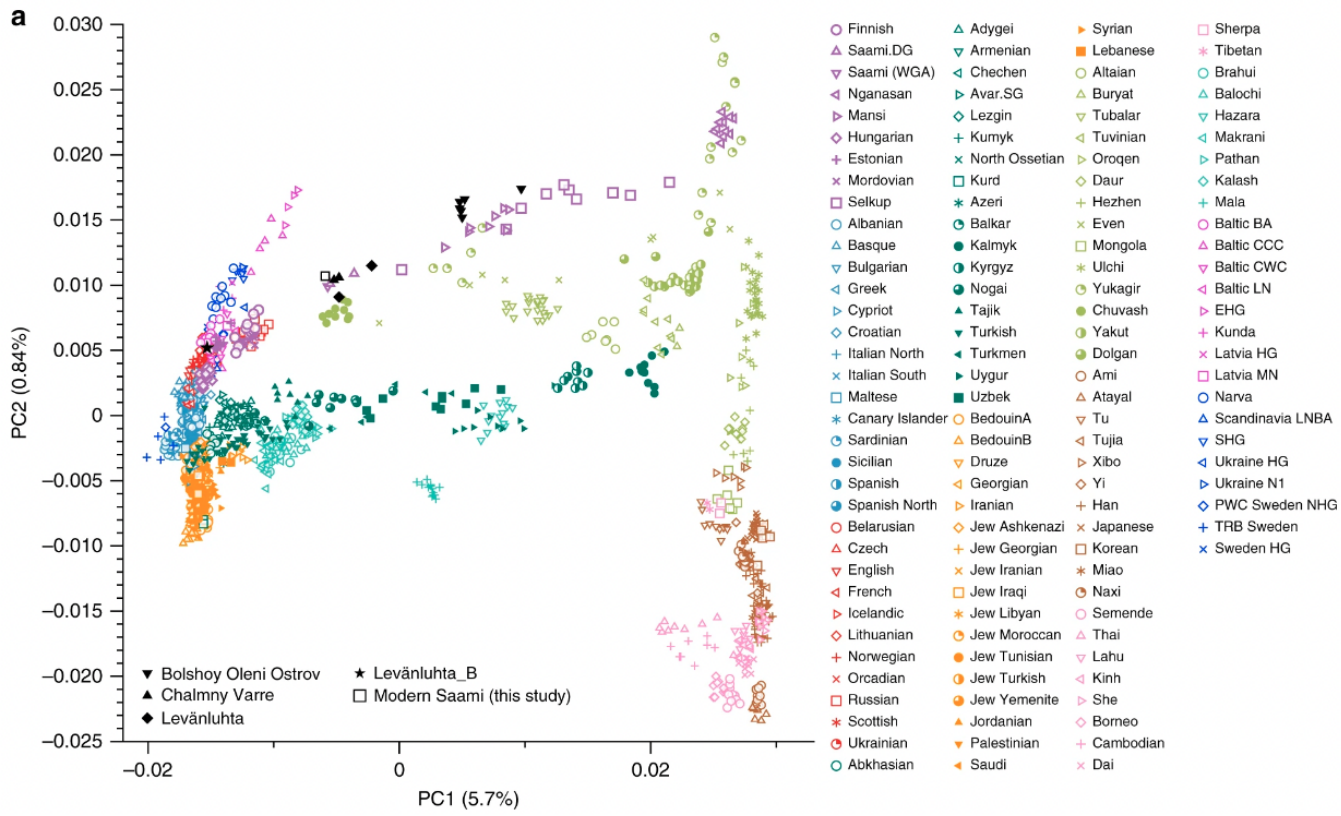


We haven't managed to exactly reproduce the results from Diaz-Papkovich et al; we're working on it.

And feel free to explore the visualizations further using the interactive htmls that can be founded in the `htmls` folder

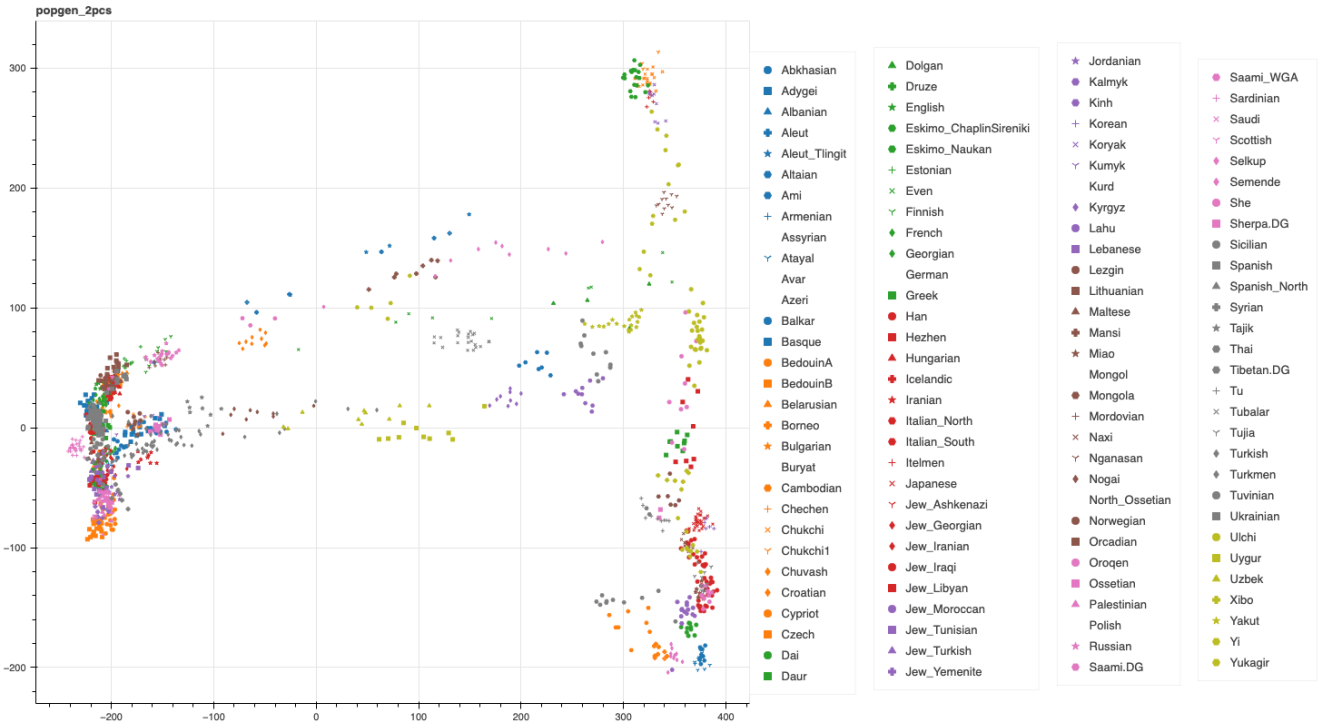
Modern Eurasian population

For reference, Lamnidis et al. have applied PCA to this dataset:

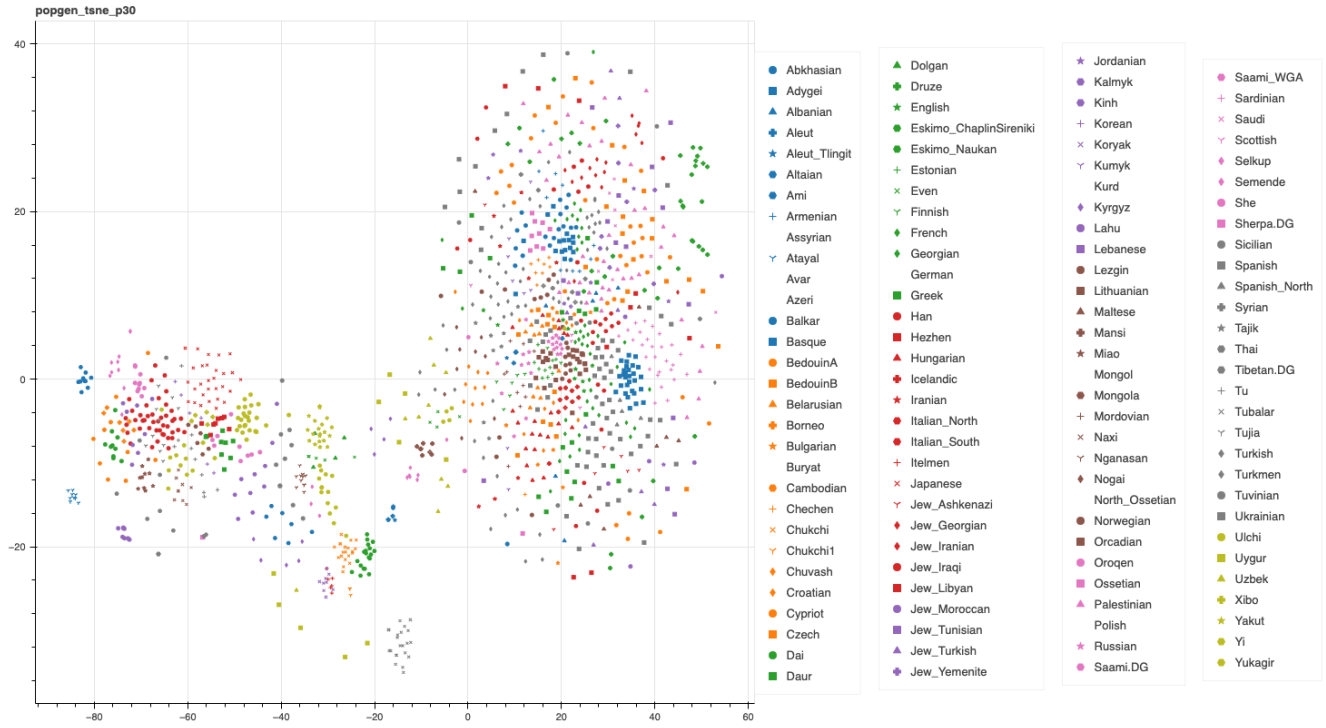


Our results are below:

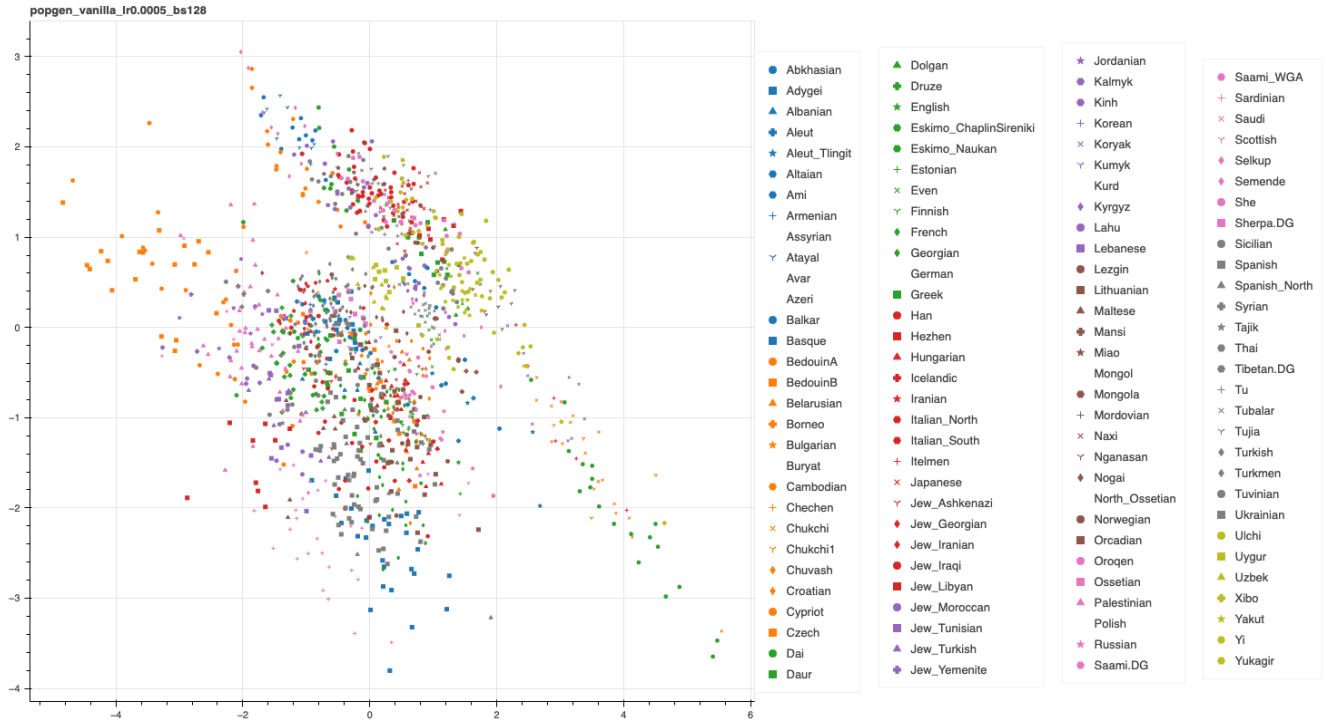
- PCA



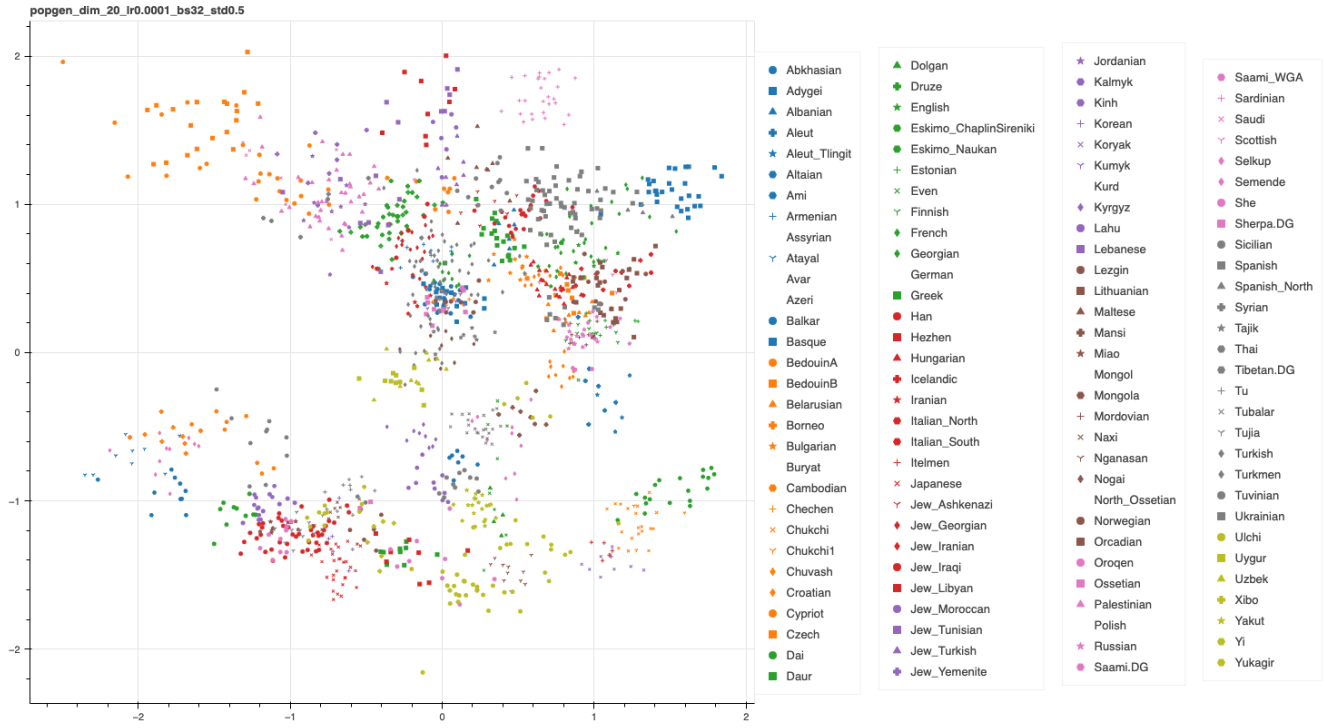
- t-SNE



- Vanilla Adversarial Autoencoders



- Adversarial Autoencoders with Cluster Heads


popgen_dim_20_lr0.0001_bs32_std0.5
Abkhasian
Adygei
Albanian
Aleut
Aleut_Tlingit
Altaian
Ami
Armenian
Assyrian
Atayal
Avar
Azeri
Balkar
Basque
BedouinA
BedouinB
Belarusian
Borneo
Bulgarian
Buryat
Cambodian
Chechen
Chukchi
Chukchi1
Chuvash
Croatian
Cypriot
Czech
Dai
Daur
Dolgan
Druze
English
Eskimo_ChaplinSireniki
Eskimo_Naukan
Estonian
Even
Finnish
French
Georgian
German
Greek
Han
Hezhen
Hungarian
Icelandic
Iranian
Italian_North
Italian_South
Itelmen
Japanese
Jew_Ashkenazi
Jew_Georgian
Jew_Iranian
Jew_Iraqi
Jew_Libyan
Jew_Moroccan
Jew_Tunisian
Jew_Turkish
Jew_Yemenite
Jordanian
Kalmyk
Kinh
Korean
Koryak
Kumyk
Kurd
Kyrgyz
Lahu
Lebanese
Lezgin
Lithuanian
Maltese
Mansi
Miao
Mongol
Mongola
Mordovian
Naxi
Nganasan
Nogai
North_Ossetian
Norwegian
Orcadian
Oroqen
Ossetian
Palestinian
Polish
Russian
Saami.DG
Saami_WGA
Sardinian
Saudi
Scottish
Selkup
Semende
She
Sherpa.DG
Sicilian
Spanish
Spanish_North
Syrian
Tajik
Thai
Tibetan.DG
Tu
Tubalar
Tujia
Turkish
Turkmen
Tuvinian
Ukrainian
Ulchi
Uygur
Uzbek
Xibo
Yakut
Yi
Yukagir