* **Deciphering Regional Variations using SNPs**

By: Yingxi Yang, and Hanli Huang

1. Summary: The project studied PCA, MDS, t-SNE and Autoencoder based dimensionality reduction technique on SNPs dataset to infer the different regions the people come from.
2. Strengths: Well motivated problem with clear visualizations graphics depicting the evolution of humans from different regions.
3. Weakness: The project chooses optimum number of principal components by computing the accuracy of random forest classification algorithms. Horn’s parallel analysis could have been explored to find out how many principal components are needed.
4. Evaluation on Clarity: 5 points. The report is clearly written with necessary background information and easy to follow.
5. Evaluation on Technical Quality: 4.5 points. Mathematical description of PCA and Random forest is not presented. The performance of random forest is evaluated on the transformed data using PCA only. The performance of classification algorithm on the transformed data by MDS, t-SNE and autoencoder could have also been performed.
6. Overall Rating: 4.5
7. Confidence: 3