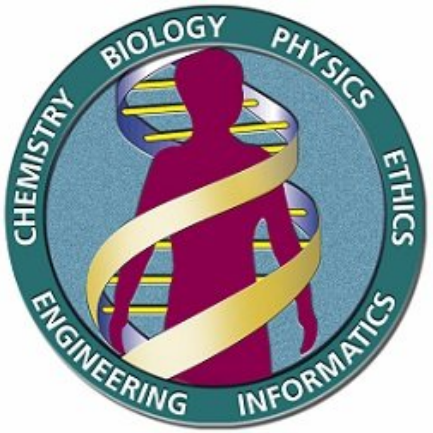
Design of A Generic System for Genomic Feature Recognition

Chuan Hock KOH (Evan)
Limsoon WONG

Human Genome Project

- A **genome** is all the DNA in an organism
- **Results**: Genome for many organisms - from microbes to plants to humans
- Next **challenge**: to derive knowledge



Reference: http://genomics.energy.gov/

Research Project

- Phase 1
  - Develop a computational method to analyze DNA sequences
- Phase 2
  - Make this method easy to use, especially for biologists with no or little computer expertise

Phase 1

- Develop a 4 step methodology
  1. Feature Generation
  2. Feature Selection
  3. Feature Integration
  4. Cascade Classification
- 18th International Conference on Genome Informatics

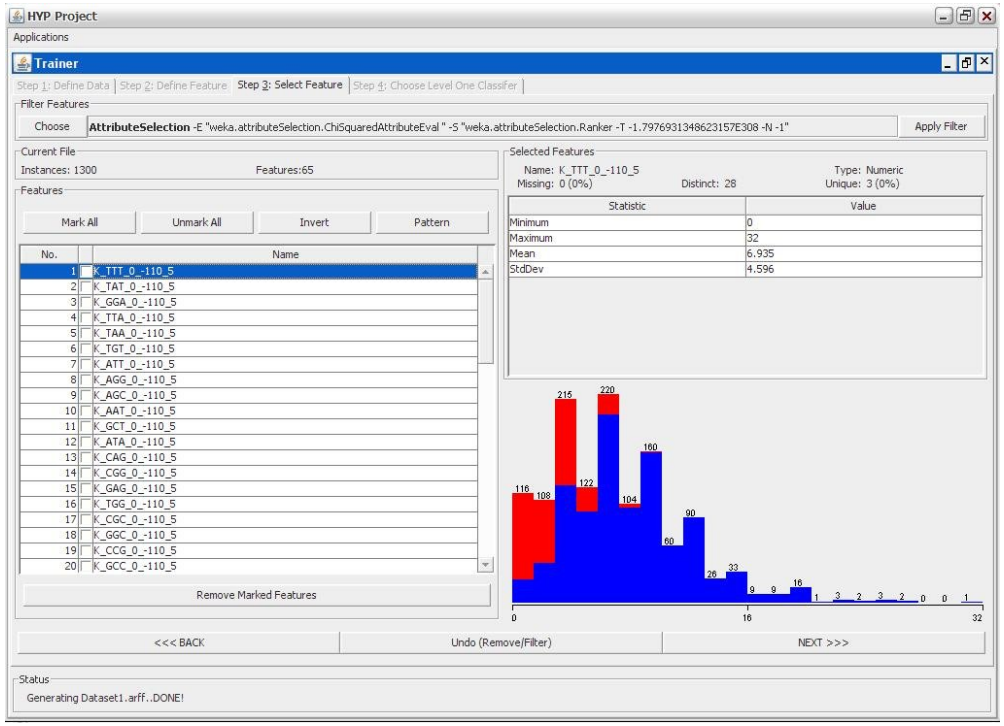
1. Feature Generation
   - Thickness of Eyebrows, Hair Length, Hair Color, Jaw Line, Eye Color etc
2. Feature Selection
   - Reject Hair Length and Hair Color
3. Feature Integration
   - Use all selected features to judge

NUS
National University
of Singapore

4. Cascade Classifier
   - Looking at its surroundings

Phase 2

- To enable the Method to be deployed with ease
  - Developed a software using java and WEKA



Conclusion

- Current Status
  - Completed initial goals of this project
- If you have any DNA sequences you would like to do prediction
  - kohchuanhock@nus.edu.sg