NGS DATA ANALYST
Professional Overview

10+ years' research experience in academia and industry.

Concentrating in data analysis and related areas including: data mining, linear/logistic regression, predictive modeling.

Core Qualifications

Proficient in Python, SQL, Shell Script and Linux/HPC environment. Â

Excellent hands-on skills in statistical analysis using SAS and R.

Systematic approach with analytical, reasoning, and problem solving skills to develop and modify software tools.

Expertise in experiment design , exploratory data analysis, reproducible research, machine learning.

Experience
Unitypoint Health Norwalk , IA NGS Data Analyst 12/2014 to Current

- Developing statistical and analytical methods to interpret the genotype data for a variety of clinical phenotyping assays.
- Lead team effort of developing strategies and software to validate, store, visualize and interpret data in order to ensure quality control of the next generation sequencing data.
- Optimized the sequencing protocol and decreased the cost by 66.7%.
- Developed and automated the data analysis pipeline for Illumina and PacBio sequencing Platform on Amazon Web Services.

Verizon Media (Former Oath) San Francisco , CA Research Fellow (Dr.Leslie Baier's Lab) 10/2010 to 08/2014

- Independently developed and published a pipeline (NGSPE) for Next Generation Sequencing data analysis which initialized the bioinformatics research in the branch..
- Analyzed terabytes of whole exome & genome sequencing data (from alignment, recalibration and genotype calling to variant annotation) under Linux/High Performance Computing environment.
- Data mining of the genotypes and type 2 diabetes related clinical traits using SAS/SQL and R.
- Developed and deployed a database system in MySQL to provide genetic data support for the branch.
- Customized the Pima Indian common variant (MAF*5%) chips by selecting tag variants from the whole genome sequencing data.

Arizona State University City , STATE Adjunct Faculty 06/2011 to 06/2014

- Collaborated with Prof. Valentin Dinu on the Next Generation Sequencing Project
- Mentored graduate students.

Bio-X Center, Shanghai Jiao Tong University City Research Scientist 04/2009 to 01/2010

- Leaded the human genome wide association (Affymatrix SNP 6.0 Array) studies from genotyping to data analysis using the Affymatrix Genotyping Console and Plink.
- Identified several susceptibility genes of schizophrenia, diabetic retinopathy using linkage disequilibrium mapping, case-control and functional studies.
- Mentored Ph.D candidate Students.

Education
Certificate : Data Science 2015 John Hopkins University , City
Ph.D : Statistical Genetics 2009 Shanghai Jiao Tong University , City , China
B.S : Biological Sciences 2004 Huazhong Agriculture University , City , China

Publications

Huang K , Nair AK, Muller YL, Piaggi P, Bian L, Del Rosario M, Knowler WC, Kobes S, Hanson RL, Bogardus C, Baier LJ. Whole exome sequencing identifies variation in CYB5A and RNF10 associated with adiposity and type 2 diabetes. ***Obesity*** . 2014 Apr;22(4):984-8.

Huang K, Venkata Yellapantula, Leslie Baier, Valentin Dinu. A pipeline for end-to-end analysis of DNA sequencing data. ***Comput Biol Med*** . 2013 Sep;43(9):1171-6.

Huang K , Tang W, Tang R, Xu Z, He Z, Li Z, Xu Y, Li X, He G, Feng G, He L, Shi Y. Positive association between OLIG2 and schizophrenia

in the Chinese Han population. ***Hum Genet*** . 2008 Jan;122(6):659-60.

Zhang M, **Huang K** , Zhang Z, Ji B, Zhu H, Zhou K, Li Y, Yang J, Sun L, Wei Z, He G, Gao L, He L, Wan C. Proteome alterations of cortex and hippocampus tissues in mice subjected to vitamin A depletion. ***J Nutr Biochem*** . 2011 Nov;22(11):1003-8. (**Co-first author**)

Baier LJ, Muller YL, Remedi MS, Traurig M, Piaggi P, Wiessner G, Huang K, Stacy A, Kobes S, Krakoff J, Bennett PH, Nelson RG, Knowler WC, Hanson RL, Nichols CG, Bogardus C. ABCC8 R1420H loss-of-function variant in a Southwest American Indian community: association with increased birth weight and doubled risk of type 2 diabetes. ***Diabetes*** . 2015 Aug 5. pii: db150459.

Muller YL, Piaggi P, Hanson RL, Kobes S, Bhutta S, Abdussamad M, Leak-Johnson T, Kretzler M, **Huang K** , Weil EJ, Nelson RG, Knowler WC, Bogardus C, Baier LJ. A cis-eQTL in PFKFB2 is associated with diabetic nephropathy, adiposity and insulin secretion in American Indians. ***Hum Mol Genet*** . 2015 May 15;24(10):2985-96.

Muller YL, Thearle MS, Piaggi P, Hanson RL, Hoffman D, Gene B, Mahkee D, **Huang K** , Kobes S, Votruba S, Knowler WC, Bogardus C, Baier LJ. Common genetic variation in and near the melanocortin 4 receptor gene (MC4R) is associated with body mass index in American Indian adults and children. ***Hum Genet*** . 2014 Nov;133(11):1431-41.

Muller YL, Piaggi P, Hoffman D, **Huang K** , Gene B, Kobes S, Thearle MS, Knowler WC, Hanson RL, Baier LJ, Bogardus C. Common genetic variation in the glucokinase gene (GCK) is associated with type 2 diabetes and rates of carbohydrate oxidation and energy expenditure. ***Diabetologia*** . 2014 Jul;57(7):1382-90.

Hanson RL, Muller YL, Kobes S, Guo T, Bian L, Ossowski V, Wiedrich K, Sutherland J, Wiedrich C, Mahkee D, **Huang K** , Abdussamad M, Traurig M, Weil EJ, Nelson RG, Bennett PH, Knowler WC, Bogardus C, Baier LJ. A genome-wide association study in American Indians implicates DNER as a susceptibility locus for type 2 diabetes. ***Diabetes*** . 2014 Jan;63(1):369-76.

Bian L, Traurig M, Hanson RL, Marinelarena A, Kobes S, Muller YL, Malhotra A, **Huang K** , Perez J, Gale A, Knowler WC, Bogardus C, Baier LJ. MAP2K3 is associated with body mass index in American Indians and Caucasians and may mediate hypothalamic inflammation. ***Hum Mol Genet*** . 2013 Nov 1;22(21):4438-49.

Zhao Q, Li T, Zhao X, **Huang K** , Wang T, Li Z, Ji J, Zeng Z, Zhang Z, Li K, Feng G, St Clair D, He L, Shi Y. Rare CNVs and Tag SNPs at 15q11.2 Are Associated With Schizophrenia in the Han Chinese Population. ***Schizophr Bull*** . 2013 May;39(3):712-9.

Chen P, Pan D, Fan C, Chen J, **Huang K** , Wang D, Zhang H, Li Y, Feng G, Liang P, He L, Shi Y. Gold nanoparticles for high-throughput genotyping of long-range haplotypes. ***Nat Nanotechnol*** . 2011 Sep 4;6(10):639-44.

Li Z, Qu J, Xu X, Zhou X, Zou H, Wang N, Li T, Hu X, Zhao Q, Chen P, Li W, **Huang K** , Yang J, He Z, Ji J, Wang T, Li J, Li Y, Liu J, Zeng Z, Feng G, He L, Shi Y. A genome-wide association study reveals association between common variants in an intergenic region of 4q25 and high-grade myopia in the Chinese Han population. ***Hum Mol Genet*** . 2011 Jul 15;20(14):2861-8.

Skills

Statistical analysis, R, SAS, Python, C/C++, MySQL , Database, Linux/HPC/AWS, Machine learning, Classification, Cluster, Predictive modeling, Experiment design, Management, Genetics, Quality Control, Scientific research, Software development,.