QIN Pengfei

PhD in Computational Biology

Address：Room 206, 1st Research Building, SUSTC.

Email: qinpf@mail.sustc.edu.cn

Mobile: 13501638531

Education and employment

2017.2 – now	Research Scientist Biology department, Southern University of Science and Technology (SUSTech)
2015.11 - 2017.2	Technical director Awarded by Hangzhou 5050 talent plan & project leader Hangzhou Geeppies Genetic Technology Company
2014.03 - 2015.11	Post doctoral Lab of Mark Stoneking Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
2013.11 - 2014.03	Associate Researcher Lab of Shuhua Xu CAS-MPG Partner Institute for Computational Biology (PICB)
2008.09 - 2013.11	Ph.D. in Computational Biology Supervisors: Li Jin and Shuhua Xu CAS-MPG Partner Institute for Computational Biology (PICB)
2003.09 - 2007.06	B.S. in Biotechnology, Henan University

Publications

[1] ***Pengfei Qin***, *Mark Stoneking*. 2015. Denisovan Ancestry in East Eurasian and Native American Populations. **Molecular Biology and Evolution.** Doi: 10.1093/molbev/msv141.

[2] ***Pengfei Qin***, *Ying Zhou, Haiyi Lou, Dongsheng Lu, Xiong Yang, Yuchen Wang, Li Jin, Yeun-Jun Chung, Shuhua Xu.* 2015. Quantitating and Dating Recent Gene Flow between European and East Asian Populations. **Scientific Reports.** Doi:10.1038/srep09500.

[3] ***Pengfei Qin***, *Zhiqiang Li, Wenfei Jin, Dongsheng Lu, Haiyi Lou, Jiawei Shen, Li Jin, Yongyong Shi, and Shuhua Xu.* 2014. A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese. **European Journal of Human Genetics**. Doi:10.1038/ejhg.2013.111.

[4] *Wenfei Jin,* ***Pengfei Qin***, *Haiyi Lou, Li Jin, Shuhua Xu.* 2012. A Systematic Characterization of Genes Underlying Both Complex and Mendelian Diseases. **Human Molecular Genetics**. Doi: 10.1093/hmg/ddr599.

[5] *Qianqian Peng, Jinxi Li, Jingze Tan, Yajun Yang, Manfei Zhang, Sijie Wu, Yu Liu, Juan Zhang,* ***Pengfei Qin***, *Yaqun Guan, Yi Jiao, Zhaoxia Zhang, Pardis C. Sabeti, Kun Tang, Shuhua Xu, Li Jin, Sijia Wang*. 2016. EDARV370A associated facial characteristics in Uyghur population revealing further pleiotropic effects. **Human Genetics**. 135(1):99-108.

[6] *Wan Isa Hatin, Ab Rajab Nur-Shafawati, Ali Etemad, Wenfei Jin,* ***Pengfei Qin***, *Shuhua Xu, Li Jin,Soon-Guan Tan, Pornprot Limprasert, Merican Amir Feisal, Mohammed Rizman-Idid, Bin Alwi Zilfalil and The HUGO Pan-Asian SNP Consortium.* 2014. A genome wide pattern of population structure and admixture in peninsular Malaysia Malays. **The HUGO Journal.** 8:5.

Awards

[1] Awarded by Hangzhou 5050 talent plan & project leader, 2016

[2] Awarded by Shenzhen Peacock talent plan, 2017

Research Interests

Genome wide association study; Population genetic structure; Gene flow among human populations; Archaic DNA analysis; Single cell sequencing; Epigenetics; Cancer Genomics

Past projects including:

Genetic stratification in Han Chinese and its potential effect in association study

Population stratification acts as a confounding factor in genetic association studies and may lead to false-positive or false- negative results. Previous studies have analyzed the genetic substructures in Han Chinese population, the largest ethnic group in the world comprising B20% of the global human population. In this study, we examined 5540 Han Chinese individuals with about 1 million single-nucleotide polymorphisms (SNPs) and screened a panel of ancestry informative markers (AIMs) to facilitate the discerning and controlling of population structure in future association studies on Han Chinese. Based on genome- wide data, we first confirmed our previous observation of the north–south differentiation in Han Chinese population. Second, we developed a panel of 150 validated SNP AIMs to determine the northern or southern origin of each Han Chinese individual. We further evaluated the performance of our AIMs panel in association studies in simulation analysis. Our results showed that this AIMs panel had sufficient power to discern and control population stratification in Han Chinese, which could significantly reduce false-positive rates in both genome-wide association studies (GWAS) and candidate gene association studies (CGAS). We suggest this AIMs panel be genotyped and used to control and correct population stratification in the study design or data analysis of future association studies, especially in CGAS which is the most popular approach to validate previous reports on genetic associations of diseases in post-GWAS era.

Systematic study of gene flow between European and East Asian Populations.

Historical records indicate that extensive cultural, commercial and technological interaction occurred between European and Asian populations. What have been the biological consequences of these contacts in terms of gene flow? We systematically estimated gene flow between Eurasian groups using genome-wide polymorphisms from

34 populations representing Europeans, East Asians, and Central/South Asians. We identified recent gene flow between Europeans and Asians in most populations we studied, including East Asians and Northwestern Europeans, which are normally considered to be non-admixed populations. In addition we quantitatively estimated the extent of this gene flow using two statistical approaches, and dated admixture events based on admixture linkage disequilibrium. Our results indicate that most genetic admixtures occurred between 2,400 and 310 years ago and show the admixture proportions to be highly correlated with geographic locations, with the highest admixture proportions observed in Central Asia and the lowest in East Asia and Northwestern Europe. Interestingly, we observed a North-to-South decline of European gene flow in East Asians, suggesting a northern path of European gene flow diffusing into East Asian populations. Our findings contribute to an improved understanding of the history of human migration and the evolutionary mechanisms that have shaped the genetic structure of populations in Eurasia.

Archaic ancestry in modern human

Although initial studies suggested that Denisovan ancestry was found only in modern human populations from island Southeast Asia and Oceania, more recent studies have suggested that Denisovan ancestry may be more widespread. However, the geographic extent of Denisovan ancestry has not been determined, and moreover the relationship between the Denisovan ancestry in Oceania and that elsewhere has not been studied. Here we analyze genome-wide single nucleotide polymorphism data from 2,493 individuals from 221 worldwide populations, and show that there is a widespread signal of a very low level of Denisovan ancestry across Eastern Eurasian and Native American (EE/NA) populations. We also verify a higher level of Denisovan ancestry in Oceania than that in EE/NA; the Denisovan ancestry in Oceania is correlated with the amount of New Guinea ancestry, but not the amount of Australian ancestry, indicating that recent gene flow from New Guinea likely accounts for signals of Denisovan ancestry across Oceania. However, Denisovan ancestry in EE/NA populations is equally correlated with their New Guinea or their Australian ancestry, suggesting a common source for the Denisovan ancestry in EE/NA and Oceanian populations. Our results suggest that Denisovan ancestry in EE/NA

is derived either from common ancestry with, or gene flow from, the common ancestor of New Guineans and Australians, indicating a more complex history involving East Eurasians and Oceanians than previously suspected.

Systematic characterization of genes underlying complex and Mendelian diseases

Traditionally, genetic disorders have been classified as either Mendelian diseases or complex diseases. This nosology has greatly benefited genetic counseling and the development of gene mapping strategies. However, based on two well-established databases, we identified that 54% (524 of 968) of the Mendelian dis- ease genes were also involved in complex diseases, and this kind of genes has not been systematically ana- lyzed. Here, we classified human genes into five categories: Mendelian and complex disease (MC) genes, Mendelian but not complex disease (MNC) genes, complex but not Mendelian disease (CNM) genes, essential genes and OTHER genes. First, we found that MC genes were associated with more diseases and pheno- types, and were involved in more complex protein–protein interaction network than MNC or CNM genes on average. Secondly, MC genes encoded the longest proteins and had the highest transcript count among all gene categories. Especially, tissue specificity of MC genes was much higher than that of any other gene categories, although their expression level was similar to that of essential genes. Thirdly, evidences from different aspects supported that MC genes have been subjected to both puri- fying and positive selection. Interestingly, functions of some human disease genes might be different from those of their orthologous genes in non-primate mammalians since they were even less conserved than OTHER genes. The significant over-representation of copy number variations (CNVs) in CNM genes sug- gested the important roles of CNVs in complex diseases. In brief, our study not only revealed the character- istics of MC genes, but also provided new insights into the other four gene categories.

Report and Poster

[1] Report in IMPRS (International Max Planck Research Schools) Genetics lecture, 2015, Leipzig, Germany

Report: ***Pengfei Qin***. Ancient DNA implications in Human Evolutionary History.

[2] Poster presentation in Cold Spring Harbor meeting on the Biology of Genomes, 2015, New York, USA.

Poster: ***Pengfei Qin***, *Mark Stoneking.* Denisovan Ancestry in East Eurasian and Native American Populations.

[3] Poster presentation in IMPRS, 2014, Shanghai, China.

Poster: ***Pengfei Qin***, *Mark Stoneking.* Archaic Ancestry in out-of-African Populations.

[4] Poster presentation in IMPRS, 2013, Berlin, Germany.

Poster: ***Pengfei Qin***, *Shuhua Xu.* Quantitating and Dating Recent Gene Flow between European and East Asian Populations.

[5] Poster presentation in ASHG (American Society of Human Genetics) annual meeting, 2012, San Francisco, USA.

Poster: ***Pengfei Qin***, *Wenfei Jin, Li Jin, and Shuhua Xu.* A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese.

Referees

Prof. Dr. **Mark Stoneking**
Dept. of Evolutionary Genetics
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6
D04103 Leipzig
Germany
Phone: +49-341-3550-502
Secretary: +49-341-3550-504

Fax: +49-341-3550-555
E-mail: stonekg@eva.mpg.de

Prof. Dr. **Shuhua Xu**
CAS-MPG Partner Institute for Computational Biology
Shanghai Institutes for Biological Sciences
Chinese Academy of Sciences
320 Yueyang Road.
Shanghai 200031, China
Tel: 86-21-54920479
Fax: 86-21-54920451
E-mail: xushua@picb.ac.cn

Prof. Dr. **Li Jin**
School of Life Sciences
Fudan University
220 Handan Road.
Shanghai 200433, China
Tel: +86-21-65642800
Fax: +86-21-55663714
Email: lijin.fudan@gmail.com