Global Regulatory Genomics and Transnational Education

Shenyi Zhang, Caitlin McNiff, Natasha Arzoo, Robert S. Young
Usher Institute, University of Edinburgh, U.K.
Zhejiang University – University of Edinburgh Institute, Zhejiang University, P.R. China

Research in our group focuses on evolutionary, population and functional genomics across populations. We are particularly interested in the role of promoters, the site of gene transcription initiation which frequently harbour genetic variants important for regulating medical phenotypes and disease risk. Active projects in this area are (1) studying the functional role of promoters identified within the VIKING Health Study and (2) comparing the promoter landscape in diseases regulated by divergent genetic variants across global populations. With researchers at the Centre for Cardiovascular Science, we are also investigating the role of phenotype-associated variants in post-transcriptional regulatory mechanisms.

Several group members are also involved in our affiliation at the Zhejiang University – University of Edinburgh (ZJE) Institute in Haining, China. ZJE is the largest transnational educational partnership at the University of Edinburgh and won the 'Educational Partnership of the Year' award at the 2024 China-Scotland Business Awards.

Population-specific promoters across individuals in the VIKING Health Study

The functional genomics technology Cap Analysis of Gene Expression (CAGE) was deployed to profile the promoter landscape from whole-blood samples within six individuals in the VIKING Health Study. We identified 52,633 promoters, of which 37,862 (72%) have not previously been detected in the reference FANTOM database[1].

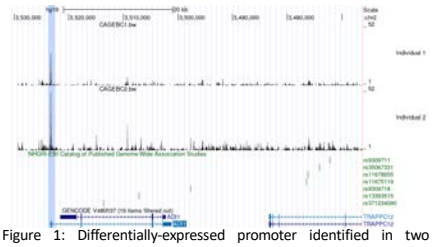

Figure 1: Differentially-expressed promoter identified in two individuals. Black histograms show the CAGE defined transcript 5' end measures for individual 1 and 2 at the ADI1 gene. Annotated genes are shown below, where exons are indicated by boxes, introns by lines and chevrons to the right denote their transcriptional orientation. Phenotype-associated genetic variants identified by genome-wide association studies (GWAS) are shown in green.

Our high-quality promoters are enriched for expression quantitative trait loci (eQTL), genetic variants which are associated with differential gene expression also identified in whole-blood samples[2].

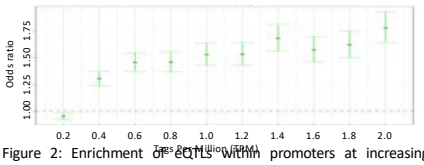

Figure 2: Enrichment of eQTLs within promoters at increasing expression as quantified by Tags Per Million (TPM). Points represent the odds ratio of genomic overlap of promoters with eQTLs relative to the genome-wide expectation while vertical lines show the 95% confidence interval of that estimate (Fisher's exact test). Odds ratios above 1 indicate a greater overlap of promoters with eQTLs.

Ongoing work is aiming to establish the role of genetic variation in regulating functional promoter output, its association with medical phenotypes and disease risk in the VIKING Health Study and the extent to which findings here can be generalised to other populations.

Population-specificity of disease susceptibility and drug response

Genome-wide association studies (GWAS) investigate the relationship between genetic variants and traits or diseases. However, it is a well-established fact that there is a European bias in these studies.

While superpopulations – defined by the 1,000 Genomes Project – share approx. 65% genetic variants, and there is some translatability in variants found between genetic variants found in GWAS, there are hundreds of thousands of variants unique to each population. The lack of inclusivity in genomic studies skews our understanding of disease biology and restricts the development of therapeutics tailored to genetically diverse populations.

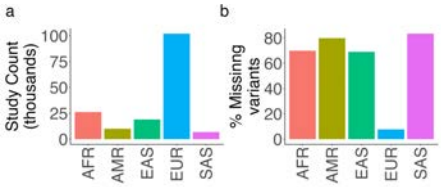

Figure 3: Summary of variant representation in GWAS across the five superpopulations – Africa (AFR), the Americas (AMR), East Asia (EAS), Europe (EUR), and South Asia (SAS). **A**. The number of GWAS performed in each superpopulation according to the NHGRI-EBI GWAS catalogue. **B**. The percentage of variants that were found in a superpopulation that the study was not performed on.

We are currently identifying diseases which are regulated by variants that differ across East Asian and European populations. Chosen variants will be engineered into cell lines and the transcriptomic differences between the populations measured using CAGE.

This information can be used to determine whether known drug-gene interactions for these targets are effective across populations.

RNA-binding proteins in cardiometabolic disease

This project examines the genomic regulation of cardiometabolic traits. Our findings indicate that genetic variants associated with obesity and related traits are specifically enriched within coding regions and 3' untranslated regions (3' UTRs), suggesting a potential involvement of RNA-binding proteins (RBPs).

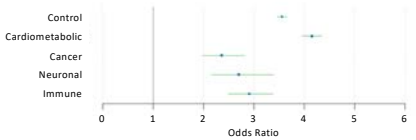

Figure 4: Enrichment of phenotype-associated genetic variants identified by across a range of medical phenotypes within protein-coding gene sequences (CDS). The control group represents all GWAS variants. Points represent the odds ratio of genomic overlap of relative to the genome-wide expectation while horizontal lines show the 95% confidence interval of that estimate (Fisher's exact test). Odds ratios above 1 indicate a greater overlap of CDS with GWAS variants.

To advance this research, we aim to identify RBPs active at various stages of adipogenesis and intersect these with the genes identified by our bioinformatics analyses. This approach will seek to validate the involvement of adipogenesis-related RBPs in the development of cardiometabolic diseases.

Biomedical Informatics at the Zhejiang University – University of Edinburgh (ZJE) Institute

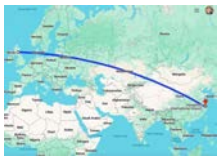

Figure 5: Global locations of the Usher Building and ZJE Institute.

Figure 6: View of the campus auditorium.

ZJE offers two undergraduate degree programmes – Integrative Biomedical Sciences (BMS) and Biomedical Informatics (BMI) – alongside several postgraduate programmes.

We deliver research-led teaching in several courses in data science, statistics, and genomics and also contribute to supervision of postgraduate students in various research groups at ZJE.

How do biomedical scientists feel about learning computer programming?

We recently surveyed our first-year undergraduate students at ZJE about their feelings around learning computer programming.

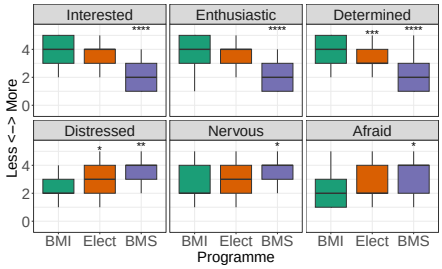

Figure 7: Student agreement with various positive and negative emotions[3] when asked how they felt about learning computer programming across the BMI, Elective and BMS cohorts. (* indicates $p < 0.05$, ** indicates $p < 0.05$, and **** indicates $p <$ 0.001 for Mann-Whitney tests relative to the BMI scores).

Incoming Biomedical Sciences students feel less positive **and** more negative about learning computer programming.

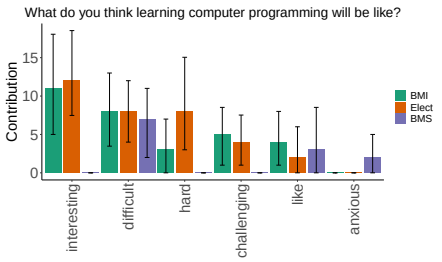

Figure 8: Sentiment analysis across student cohorts. Bars represent the contribution of each word to the responses from each cohort while error bars show the 95% confidence interval of that contribution from sampling with replacement.

All students think that learning computer programming will be difficult but biomedical scientists are not as interested in this topic.

@r0bah0lic robert.young@ed.ac.uk

1. Forrest et. al. 2014. Nature **507**: 462-70
2. Lonsdale et al. 2013. Nat Genet **45**: 580-5
3. Watson D et al. 1988. J Pers Soc Psychol **54(6)**:1063-70

浙江大学爱丁堡大学联合学院
ZJU-UoE Institute