Course Wrap-up

Kenneth Lange, Project Leader.

Papp, Sinsheimer, Sobel, Suchard, and Zhou

The OpenMendel Project

1. Move from classic, closed source Mendel to open source OpenMendel. Foster a community of software developers.

2. Freeze the classic Mendel, but maintain it long-term. The same goes for MendelEnterprise (the web-based clinical version of Mendel.)

3. Design OpenMendel to allow for rapid prototyping of new computational methods.

4. Include in OpenMendel state-of-the-art base routines, such as likelihood and optimization algorithms, that can be built upon.

Software Management

1. We have suggested code style guidelines. Please adhere to them.

2. Write code that is fast and highly reliable. This takes extra effort. Exploit parallel processing whenever possible.

3. We need to put in place a structure for vetting applications. Can you help us with this?

4. You should document your code and provide test problems. Julia makes this relatively easy. This is a community effort.

Programming Issues

1. Large data sets are difficult to read and store in memory. Sequence data is particular problematic. Data files are memory hogs and contain far more information than necessary. Data should be stored as close to the bit level as possible.

2. Modern processors have multiple cores. Parallelization with a good compiler can take advantage of multiple cores. Graphics processing units (GPUs) offer even greater speedups.

3. As we have seen, Julia is a state-of-the-art language that fills our needs well.

Questions to Ponder

1. Will pedigree data and linkage studies stage a comeback?

2. Why is so little of the variance of most common traits explained by SNP GWAS?

3. SNP data is being replaced by sequence data. What is the best approach to mapping genes involving rare variants? How does one distinguish normal polymorphisms from disease predisposing mutations?

4. What is the best way to look for gene-by-gene and gene-by-environment interactions? How do we exploit pathways?

5. How can the genetics community validate software? Is commercialization of software inevitable or desirable?

6. The scale of studies has increased enormously. How will gene banks be organized, and can we avoid publishing only monolithic papers?

7. What is the best approach to meta-analysis in combining studies using different markers and different measures of the same phenotype?

8. How can we take advantage of genetic information to improve genetic counseling and personalized medicine?

In House Projects

The UCLA group are currently working on several OpenMendel projects:

1. Variance components models (aka LMMs) for GWAS (Hua Zhou).

2. Model selection in random sample GWAS (Kevin Keys and Ken Lange).

3. Trait simulation (Huwenbo Shi and Ken Lange)

4. Genotype and haplotype imputation (Janet Sinsheimer, Rory Wasiolek, and Ken Lange).

Short-Term Issues and Predictions

1. There is too little emphasis on interactions, both gene-by-gene and gene-by-environment interaction. Also, haplotypes will gradually become the unit of statistical analysis.

2. A renewed pursuit of family data and population isolates will pay dividends, particularly for rare variants, which are the vast majority.

3. Sequence data *should* allow us to assess the role of rare variants, but the results so far are inconclusive.

4. We need to bring more biology to bear in mapping. New discoveries are waiting to be found in the current databases.

What to Expect in the Next Two Decades

1. Traditional genetic epidemiology's current growth trajectory will slow. All of the low-hanging association fruit will be picked. (However, note that there are still many monogenic traits that remain unmapped.)

2. The translational aspect of genetic epidemiology will grow. People want to know how risky their genomes are and what can be done to mitigate their susceptibilities.

3. The next revolution in genetics is the CRISPR-based editing technology. (We are moving from a read-only to a read/write world.) Functional genomics will become much easier.

4. Interaction and network analysis will come to the fore. Systems biology will increase in importance.

5. Drug discovery will be become more science based.

6. Computing will grow massively parallel and cloud-based.

7. Even if traditional genetic epidemiology wanes or changes beyond recognition, jobs will be plentiful for scientists with a combination of skills in the biological and mathematical sciences.

8. The message for young scientists: be adaptable, keeping learning, and brush up on your skills in applied mathematics, statistics, and computing. Big data is here to stay.