Integrate network resources to optimize genetic association studies

Yuanlong Liu

UMR-946, INSERM
University Paris Diderot, Paris, France

21 May, 2016
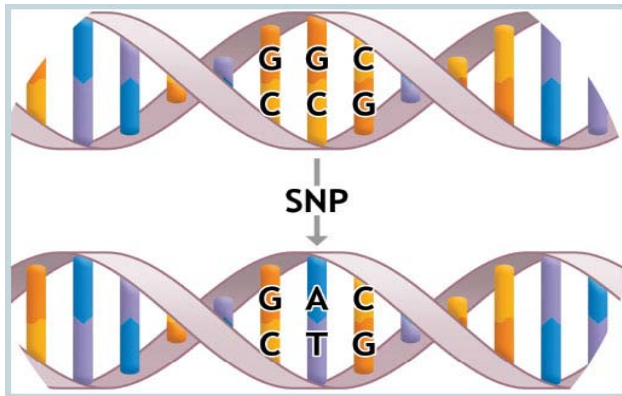yuanlong.liu@inserm.fr

Outline

1. GWAS

2. Integrate GWAS with gene network
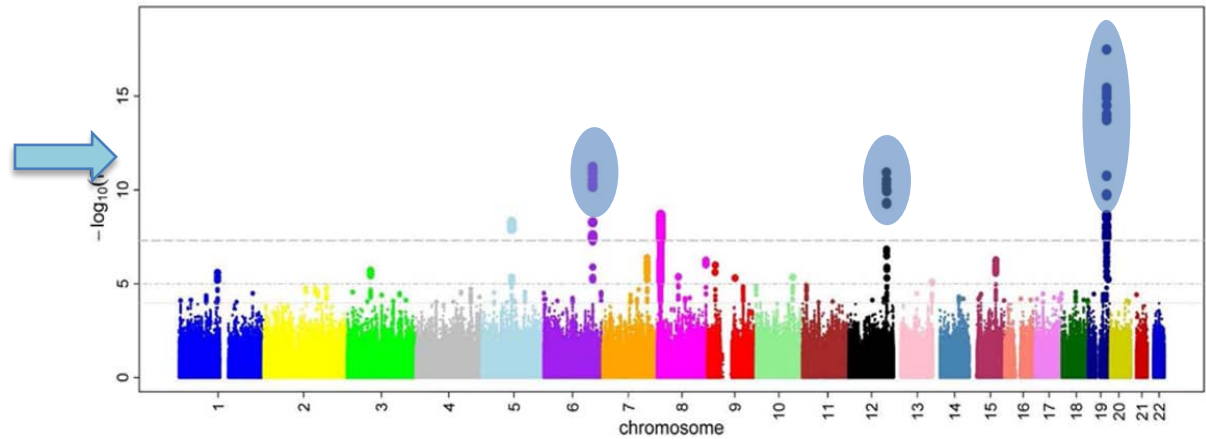
3. Practical issue

4. A successful example

5. Conclusion

❑ Genome-wide association study

- ■ Association test for each SNP(single nucleotide polymorphisms)

P-value threshold: 5×10^{-8}

❑ **Why GWAS + network knowledge**

❑ **Why GWAS + network knowledge**

▪ GWASs Identified small amount of loci

→explain a small part of genetic risk

▪ GWASs focused on single SNPs analysis

→underpowered to detect SNPs with small marginal effect

❑ **Why GWAS + network knowledge**

▪ GWASs Identified small amount of loci

→explain a small part of genetic risk

▪ GWASs focused on single SNPs analysis

→underpowered to detect SNPs with small marginal effect

Need to analyze multiple genetic variants jointly

❑ **Why GWAS + network knowledge**

- GWASs Identified small amount of loci

    →explain a small part of genetic risk

- GWASs focused on single SNPs analysis

    →underpowered to detect SNPs with small marginal effect

Need to analyze multiple genetic variants jointly
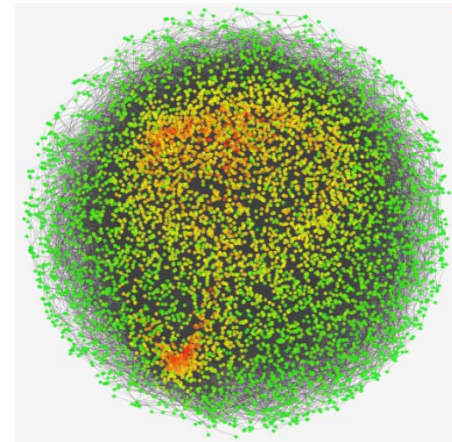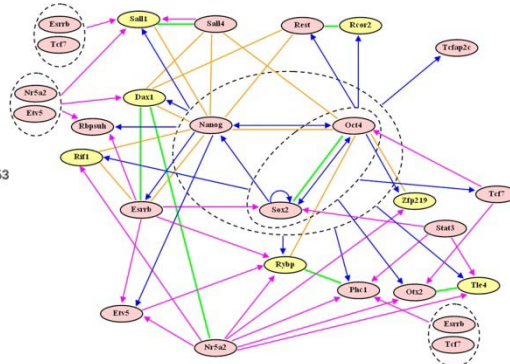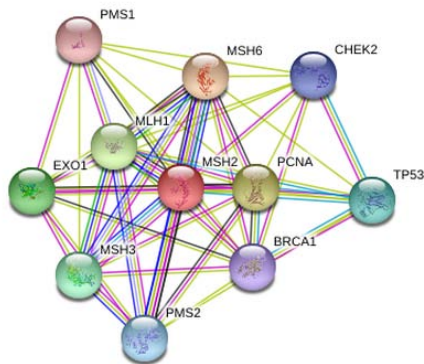
Network-based analysis:

GWAS results + Gene Network
= Disease-associated gene module

☐ **Abundance of network resources**

Either experiment verified or computationally-predicted

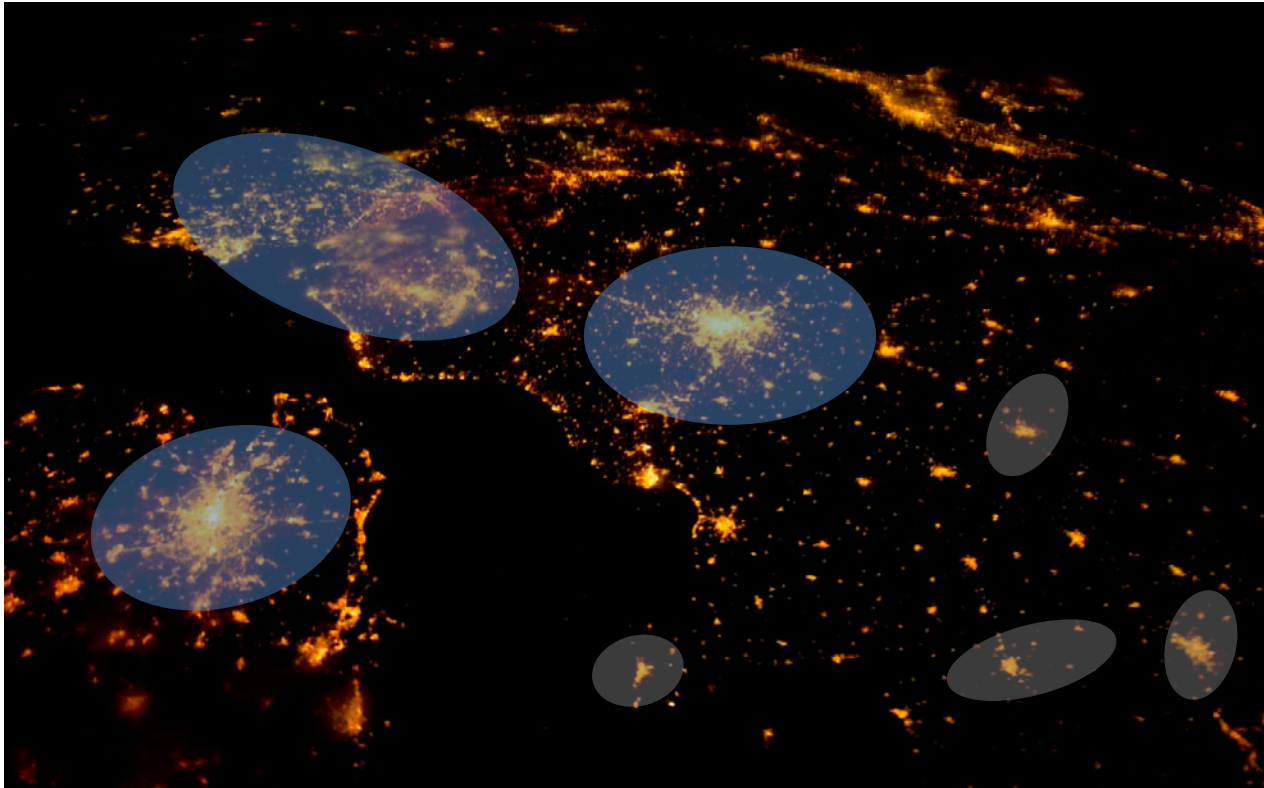- Gene regulatory network

- Protein (gene product) interaction network

- Gene co-expression network

- ....

❑ **Hypothesis**

- ▪ Guilt-by association principle:

  Genes (or its products) connected in a network are usually participating in the same, or related, cellular function

❑ General objective

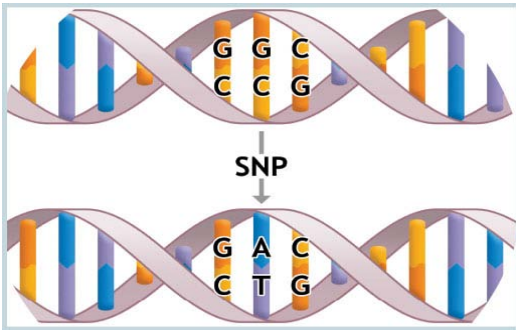- Find gene modules (connected genes) enriched in high association signals
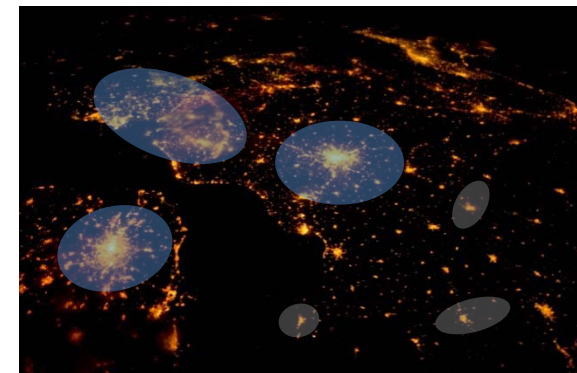
❑ **Workflow**

Genome-wide raw data

Gene-disease association signal

Signal weighted GeneNet

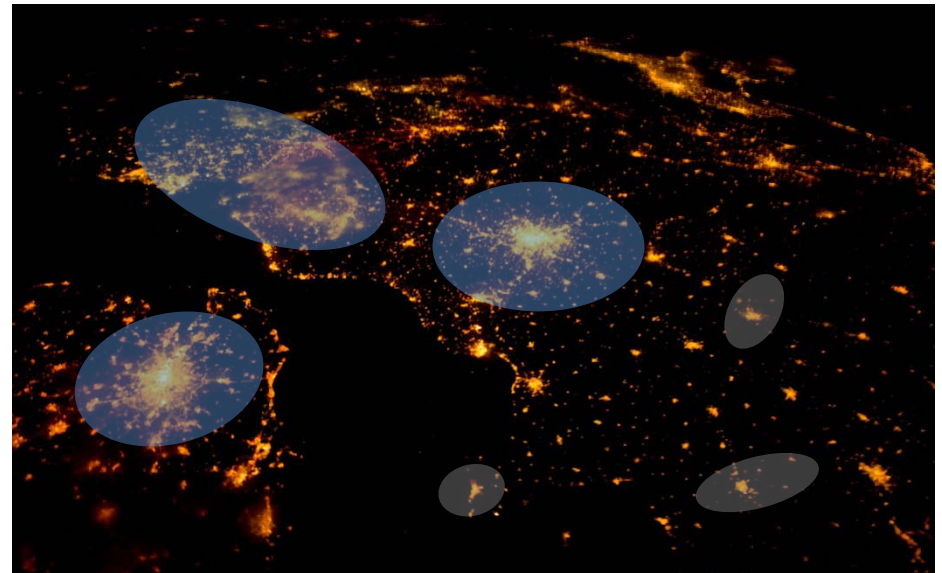**Gene modules**

❑ **Practical issues**

- **Big Data**: genome-wide study ~1M SNPs, 20K genes

- **Noise** in association signals / network information

- Method **efficiency**
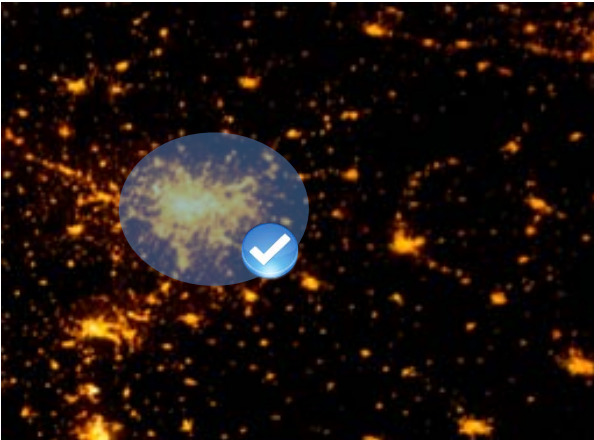
- Result **replicability** / **Interpretability**

- ….

❑ **Find gene module associated with asthma**

- asthma is an inflammatory disease of the airways of the lungs

- caused by genetic + environmental factors

❑ **Specific objective: find gene module:**

- Enriched in high signals

- **Strongly** interconnected

❑ **Specific objective: find gene module:**

- Enriched in high signals

- **Strongly** interconnected

  ✓ Robust to noise → high replicability

  ✓ Functional closely related → high interpretability

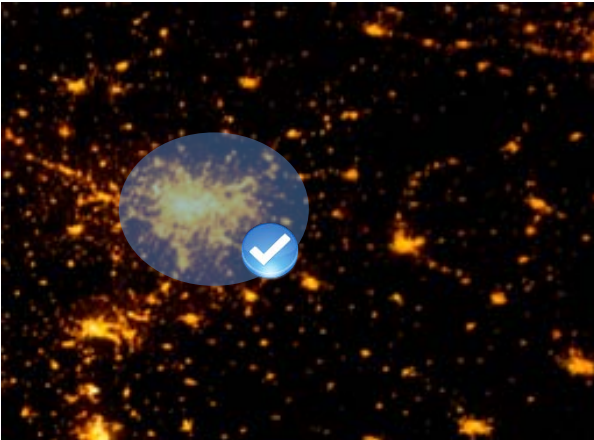❑ **Specific objective: find gene module:**

■ Enriched in high signals

■ **Strongly** interconnected

✓ Robust to noise → high replicability

✓ Functional closely related → high interpretability

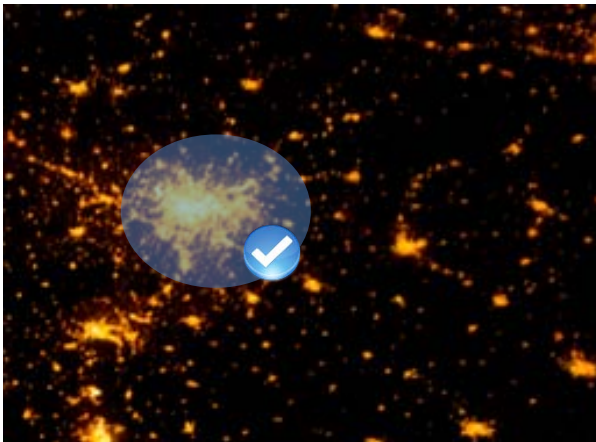Method: select module by solving maximizing:

$$g(u) = z^T u + \lambda u^T A u - \eta \| u \|_0; \lambda \geq 0, \eta \geq 0$$

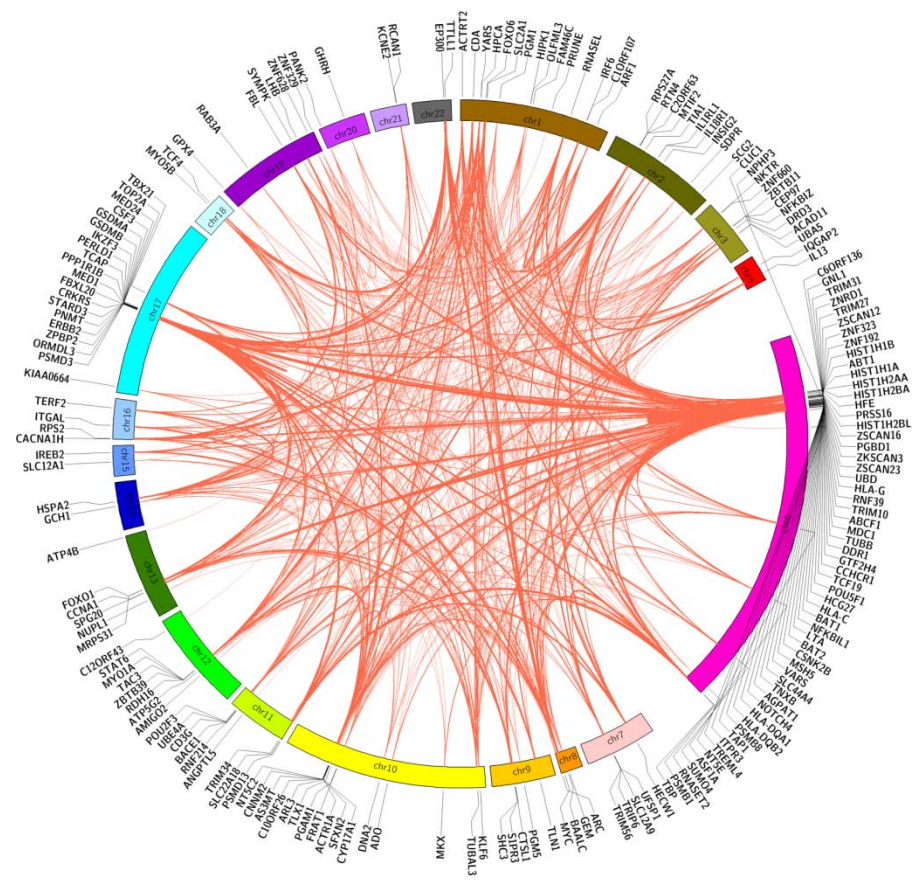**Encourage high signal**   **Encourage interconnection**   **Encourage sparsity**

☐ **Result: identified a gene module associated with asthma**

✓ Strongly interconnected

✓ Enriched in high/consistent signals

✓ Enriched in known asthma genes

✓ Genes function closely related

✓ Functional relevance to asthma

- **190 genes / 1250** connections

- **All genes** have P-value < 0.05

- Include **7 reported loci** of asthma

- **9 functionally related** gene groups

- **15 KEGG pathways** related to asthma

Conclusion

- Single marker-based test is underpowered to detect moderate effect markers

- Network-based analysis helps analyze genetic variants jointly and systematically

- Robustness, efficiency, interpretability are important aspects to be considered by network-based analysis

Acknowledgements

INSERM UMR-946 research team

- Florence Demenais
- Myriam Brossard
- Emmanuelle Bouzigon
- Chloé Sarnowski
- Patricia Margaritte-Jeannin
- Damian Roqueiro
- Chloé-Agathe Azencott
- Karsten Borgwardt

Thanks!