SNP and complex traits: where is the hidden heritability?

Albert García López

Genomics
Advanced Genetics Master (2015-2016)
Overview

1. What are single nucleotide polymorphisms (SNPs)?
   1.1 SNP and disease-causing mutations: the same?
   1.2 How SNPs are detected?
   1.3 What can we do with the SNPs data?

2. GWAS: Genome-Wide Association Study
   2.1 GWAS Catalog

3. Complex traits and diseases
   3.1 Complex traits and diseases: heritability
   3.2 Studying the heritability of complex traits with GWAS
   3.3 Multi-stage disease model
   3.4 The role of epistasis in complex traits

4. Concluding remarks

5. References
What are single nucleotide polymorphisms (SNPs)?

- Single-nucleotide substitutions of one base for another
- **84,7M SNP** (88M variants, The 1000 Genome Project Consortium, 1 October 2015)
- Most common type of genetic variation (polymorphism) among individuals
- **Goal**: Identify SNP correlated with particular effects in patients.
  - **Biomarkers** of specific diseases → **Asses risk**!
  - **Effectiveness** of a drug/s
  - **Susceptibility** to environmental factors.
  - **Adverse effects** of a given drug
1. SNP ≠ disease-causing mutation → point mutations

2. SNP are present in at least 1% of the general population

Any disease-causing mutation is this common

3. Most disease-causing mutations occur within a coding or regulatory region of a gene affecting the function of the protein

SNP are not necessarily located within genes

Two main categories

- Linked SNPs outside of gene
  - no effect on protein production or function

- Causative SNPs in gene
  - Non-coding SNP: changes amount of protein produced
  - Coding SNP: changes amino acid sequence

**How SNPs are detected?**

- **SNP genotyping**: determine the number of SNP in a DNA fragment by examining the DNA with several methods and comparing it with a reference sequence.

### A Typical Genotyping Protocol

1. **Target fragment amplification**
   - PCR or strand displacement amplification

2. **Allelic discrimination reaction**
   - Primer extension
   - Pyrosequencing
   - Structure specific cleavage
   - Ligation
   - Hybridization

3. **Allele specific product identification**
   - Fluorescence intensity / FRET / FP
   - Mass spectrometry
   - Electrophoresis
   - Microarray

### Different approaches:
- Low - high throughput (next generation sequencing)

### How Do Scientists Identify SNPs?
- SNPs are first identified when scientists sequence DNA samples from multiple people.
- Because DNA sequencing is relatively expensive and time consuming, scientists have come up with other methods for detecting SNPs.
- Primer extension is one method scientists use to determine which version of a known SNP a person has.

### Using Primer Extension to Identify SNPs:

**Version 1**

- Add synthetic complementary DNA molecule, called a “primer,” which ends at SNP position.
- Add nucleotides to extend the primer.
- Nucleotides will be added to the end of the primer ONLY if the sequence is an exact match.

**Version 2**

- Compare the lengths of the products using gel electrophoresis.
- No extension occurs
- SNP identity: Unknown

### How SNPs are sorted and catalogued in databases

- dbSNP
- HapMap project
- 1000 Genomes project...

\[ \text{rs number} \]
What can we do with the SNPs data?

1. GWAS
2. Genomic data
3. SNPs are associated to a function or response
4. Phenotypic data
5. SNPs are associated to a function or response

GWAS
GWAS: Genome-Wide Association Study

- Developed in 2007
- Based on the concept that genetic variation shows considerable linkage disequilibrium → A given SNP is strongly correlated with other SNPs
- Co-inherited more often than expected by random events
- GWAS tests a single Tag SNP from regions of LD to mark the zones in the genome showing disease association

In a typical study → 500K-1000K SNPs are tested → 0.6 – 1.2% of the SNPs already known in the human genome (2015, 1000 Genome Project) → SNP accepted = p-value ≤ 5.0 × 10⁻⁸

All SNP-trait associations with p-value $\leq 5.0 \times 10^{-8}$, published in the GWAS Catalog

5267 SNP-trait associations
The vast majority of diseases are complex:

- **Birth Defects**: cleft palate/lip, neural tube defects such as spina bifida
- **Cardiovascular conditions**: high blood pressure, some causes of heart disease, high cholesterol
- **Neurological/psychiatric conditions**: Alzheimer disease in later life, schizophrenia, bipolar disorder
- **Skin conditions**: psoriasis, moles, eczema
- **Cancer**: bowel, breast, ovarian, bowel, melanoma and prostate
- **Metabolic**: diabetes
- **Muscular/skeletal**: arthritis, rheumatic disorders, osteoporosis
- **Respiratory**: asthma, allergies, emphysema

Despite the polygenic feature, heritability in several common diseases can be explained.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Heritability estimate</th>
<th>Comment (reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurocognitive disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>0.37</td>
<td>Meta-analysis (Sullivan et al. 2012)</td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>0.5–0.7</td>
<td>Review (van Dongen et al. 2012)</td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>0.58</td>
<td>Review (Sullivan et al. 2012)</td>
</tr>
<tr>
<td>Bipolar</td>
<td>0.75</td>
<td>Review (Sullivan et al. 2012)</td>
</tr>
<tr>
<td>ADHD</td>
<td>0.76</td>
<td>Review (van Dongen et al. 2012)</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>0.8</td>
<td>Review (Lundstrom et al. 2012)</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>0.81</td>
<td>Meta-analysis (Sullivan et al. 2012)</td>
</tr>
<tr>
<td>Cardiovascular disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>~0.5 (M: 0.57; F: 0.38)</td>
<td>Swedish Twin Registry (Zdravkovic et al. 2002)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>0.4–0.42</td>
<td>Rural Chinese Twin Cohort (Zhang et al. 2009)</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>0.88</td>
<td>Finnish Twin Cohort (Hyttinen et al. 2003)</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>0.64</td>
<td>Finnish Twin Cohort (Kaprio et al. 1992)</td>
</tr>
<tr>
<td>Drug efficacy and metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Half-life of dicumarol</td>
<td>0.97</td>
<td>Twin study (Vesell and Page 1968)</td>
</tr>
<tr>
<td>Platelet response to clopidogrel</td>
<td>0.73</td>
<td>Pedigree analysis (Shuldiner et al. 2009)</td>
</tr>
<tr>
<td>Renal clearance of metformin</td>
<td>0.94</td>
<td>Twin Study (Leabman et al. 2005)</td>
</tr>
</tbody>
</table>

• Genomics of complex disease remains **unresolved**

• **Genetic factors** identified only explain a small portion of heritability estimation → **Height**

• Only **20%** of estimated heritability explained by the combination of all significant SNPs → SNPs with small individual effects/ low frequent **hidden** in GWAS

3.2 **Studying the heritability of complex traits with GWAS**

• Heritability can be defined in **two ways**
  - $h^2$: Additive effect of individual alleles
  - $H^2$: Epistasis + epigenetics

**Missing/hidden heritability**

5% (50 SNPs)
Multi-stage disease model

frequent ‘normal’ variants

High/low activity SNPs

genes

epigenetics

deleterious rare variants

Wellness

(positive selection)

‘off-well’

(subclinical state)

illness

(clinical non-measurable state)

disease

(clinical measurable state)

environment

life style, behavior, drugs

rSNP

Interactions **gene-gene-environment**

- **SNP**
  - Deleterious
  - Beneficial
- **Phenotype**
- Complex traits
- Complex diseases

**Epistasis**
- Non-linear gene-gene-environment interactions
- Interactions between SNPs in the same gene

**Interactions DRD2/DAT**
- **DRD2 (Dopamine D2 receptor):** Splicing SNP increases (3X) lethal risk of cocaine abuse
- **DAT (dopamine transporter):** variants with no effects
- 7-8X increased lethal risk of cocaine abuse
- 2 interactions
- Total compensation of the disease risk

**Explains missing heritability**
- GWAS issues
- Wellcome Trust Study
Concluding remarks

- **GWAS** studies need to focus on the role of **causative SNP**, not only on **marker SNP**

  Gain or loss of function?  ⇐  Unknown in the majority of cases

- **Redefine** GWAS studies with **larger number of cases** and a **different treshold** to look for rare SNPs

- Better study of **rSNP** and its role affecting epigenetic DNA and chromatin marks → **Better understanding of epigenetics**

- **SNPs positive selected** could cause deleterious effects under certain conditions → Undetected to GWAS (low score) → **Better study of epistasis phenomena**

- How can we **quantify** the effect of environment on human health?

- **Cost of sequencing** is steadily decreasing → **Sequencing more individuals** → more SNP data of both **common** and **rare SNPs**

- **Re-estimate heritability** to contemplate the effects of environment, epigenetics, epistasis…
5 References

- **Making SNPs Make Sense.** Retrieved December 7, 2015, from http://learn.genetics.utah.edu/content/pharma/snips/
- **GWAS Catalog.** Retrieved December 8, 2015, from https://www.ebi.ac.uk/gwas/
THANK YOU VERY MUCH FOR YOUR ATTENTION!

Hidden or missing heritability?