Illumina’s GWAS Roadmap: next-generation genotyping studies in the post-1KGP era

Alex Helm
Product Manager
Genotyping Applications
Overview

First-gen GWAS vs. Next-gen GWAS

Next-gen Sequencing and the 1kGP Revolution

2010 – 2011 GWAS Roadmap
First-gen GWAS vs. Next-gen GWAS
The GWAS Approach is Successful in Human Genetics

<table>
<thead>
<tr>
<th>Year</th>
<th># of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>2</td>
</tr>
<tr>
<td>2006</td>
<td>8</td>
</tr>
<tr>
<td>2007</td>
<td>89</td>
</tr>
<tr>
<td>2008</td>
<td>151</td>
</tr>
<tr>
<td>2009</td>
<td>222</td>
</tr>
</tbody>
</table>

- First publications in 2005
- Almost 600 total publications since 2005
- Over 3500 associations published
- Wide-range of phenotypes and diseases
1,095,000 published genome-wide associations through 9/2009, surpassed GWA at \( p < 5 \times 10^{-8} \)

NHGRI GWA Catalog

www.genome.gov/GWAStudies

Acute lymphoblastic leukemia
- Adip
- Bladder cancer
- Age
- AID
- Alco
- Alzh
- Amy
- Asth
- Atria
- Atpa
- Basa
- Bipo

Fasting glucose
- Folate pathway vitamins
- Freckles and burning
- Galactoses
- Glaucoma
- Hair color
- Heart rate
- Height
- Hepatitis
- Hirschsprung’s disease
- HCL cholesterol
- Idiopathic pulmonary fibrosis
- Inflammatory bowel disease
- Intracranial aneurysm
- LDL cholesterol
- Liver enzymes
- Lung cancer
- Malaria
- Male pattern baldness
- MCP-1
- Mean platelet volume
- Melanoma
- Menopause & menstruation
- Multiple sclerosis
- Protein levels
- Pulmonary fibrosis
- Pulmonary function COPD
- Peripheral arterial disease
- Plasma LP(a) levels
- Primary biliary cirrhosis
- Prostate cancer
- Authentic gene
- Thyroid cancer
- Total cholesterol
- Total triglycerides
- Urinal
- Skin pigmentation by reflectance spectroscopy

Type 1 diabetes
- Type 2 diabetes
- Venous thromboembolism
- Warfarin dose
For most common diseases, the sum of individual effects found so far is much less than the total estimated heritability.
Tackling the Full Spectrum of Variants in Disease

- Very Rare Variants: Large Effect Size
- Rare/Intermediate Variants: Intermediate Effect Size
- Common Variants: Small Effect Size

Effect size:
- Large
- Small

Allele Frequency:
- Low
- High

Methods:
- Linkage
- Sequencing
- Next-gen GWAS

(NEW!)
Next-gen Sequencing and the 1kGP Revolution

a new era beyond the HapMap Project
The 1,000 Genomes Project

Sequence 2,500 genomes to complete the picture of genetic variation

Achieve a nearly complete catalog of common human genetic variants with frequency 1% or higher.

Project Goals

1. Accelerate fine-mapping efforts in gene regions identified through genome-wide association studies or candidate gene studies

2. Improve the power of future genetic association studies by enabling design of next-generation genotyping microarrays that more fully represent human genetic variation

3. Enhance the analysis of ongoing and already completed association studies by improving our ability to “impute” or “predict” untyped genetic variants
New Content for Next-gen GWAS Arrays

Rich content to explore new hypotheses and enable new discoveries

Sequence to discover SNPs >1% MAF (1000-Genomes project)

Leverage the power of LD to select tagSNPs and remove redundancy

Include progressively more SNPs at lower allele frequencies (5%, 2.5%, 1%)

<table>
<thead>
<tr>
<th>Project</th>
<th>Year</th>
<th>Approx. Cumulative SNPs found</th>
<th>Tag SNPs needed for max coverage</th>
<th>Lower limit of allele frequency targeted</th>
<th>% variation tagged (r^2&gt;0.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HapMap</td>
<td>2003-2007</td>
<td>3M</td>
<td>~0.6M</td>
<td>5%</td>
<td>&gt;90%</td>
</tr>
<tr>
<td>1kG Pilot Project</td>
<td>2008-2009</td>
<td>13M</td>
<td>~2.5M</td>
<td>2.5%</td>
<td>~80%</td>
</tr>
<tr>
<td>1kG Full Project</td>
<td>2010</td>
<td>35M*</td>
<td>~5.0M</td>
<td>1%</td>
<td>&gt;90%</td>
</tr>
</tbody>
</table>

* Estimated
Illumina’s GWAS Roadmap
GWAS Roadmap Review

Content Source
- HapMap Phase 1
- HapMap Phase 2
- HapMap Phase 3
- 1,000 Genomes Project

Future GWAS Products
- HumanOmni1-Quad & OmniExpress
- Human1M-Duo
- Human660-Quad
- HumanHap500
- HumanHap300

Array Products

Data Points per Sample
- 317K Current
- 550K Projected
- 660K Projected
- 1M Projected
- 2.5M Projected
- 5M Projected

Announcement at ASHG, Oct 2009
# The Omni Family of Microarrays

*Next-generation GWAS. NOW.*

<table>
<thead>
<tr>
<th>Omni Express*</th>
<th>Omni1-Quad</th>
<th>Omni1S-8</th>
<th>Omni2.5-Quad</th>
<th>Omni2.5S</th>
<th>Omni5</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Highest-throughput array with industry-proven quality at an exceptional price.</strong></td>
<td><strong>Optimal combination of common SNPs, CNVs, and content from 1kGP.</strong></td>
<td><strong>Takes researchers from Omni1/Express to 2.5M</strong></td>
<td><strong>The most optimal and comprehensive set of both common and rare SNP content from the 1kGP</strong></td>
<td><strong>~2.5M additional markers providing rare 1kGP content</strong></td>
<td><strong>The ultimate GWAS tool providing near complete coverage of common and rare variation</strong></td>
</tr>
</tbody>
</table>

| **MAF > 5%** | **MAF >2.5%** | **MAF >1%** |
2010 Infinium Roadmap

- Content optimized from next-gen re-sequencing efforts such as 1000 Genomes.
- Pushing the boundary of GWAS content into unexplored territory
- Cost effective path for researchers that want to ride the cutting edge today

* Includes customizable OmniExpress"
## Roadmap Paths

*Multiple chips made Easier with the Multi-use Workflow*

<table>
<thead>
<tr>
<th>Path</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Total Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OmniExpress</td>
<td>Omni1S</td>
<td>Omni2.5S</td>
<td>~4.4 Million</td>
</tr>
<tr>
<td>2</td>
<td>Omni1</td>
<td>Omni1S</td>
<td>Omni2.5S</td>
<td>~5 Million</td>
</tr>
<tr>
<td>3</td>
<td>Omni2.5</td>
<td>Omni2.5S</td>
<td></td>
<td>~5 Million</td>
</tr>
<tr>
<td>4</td>
<td>Omni5</td>
<td></td>
<td></td>
<td>~5 Million</td>
</tr>
</tbody>
</table>
Enabling Discoveries with Next-Gen GWAS

![Graph showing allele frequency and effect size]

- **Very Rare Variants** (Large Effect Size)
- **Rare/Intermediate Variants** (Intermediate Effect Size)
- **Common Variants** (Small Effect Size)

Arrays:
- Omni1
- Omni2.5M
- Omni5M

 Allele Frequency

- 1% to 25%

Effect size

- 1.1 to 12.0
Omni2.5 Details
This is Not Just an Array with “New” Content!

CEU Coverage Estimates: HapMap vs. 1kGP Reference Data

- Competitor “New Array” *
- Competitor “Old 900K”
- 660W
- Omni1/ OmniExpress
- Omni2.5

% Captured at r^2 > 0.8

- HapMap 5%
- 1kGP 5%
- 1kGP 2.5%

*Base content only
Genomic Coverage Stats for African Populations

YRI Coverage Estimates: HapMap vs. 1kGP Pilot Data

% Captured at r² > 0.8

- HapMap 5%
- 1kGP 5%
- 1kGP 2.5%

Competitor
"New Array"*

Competitor
"Old 900K"*

660W

Omni1/ OmniExpress

Omni2.5

*Base content only
Genomic Coverage Stats for Asian Populations

CHB/JPT Coverage Estimates: HapMap vs. 1kGP Pilot Data

% Captured at r^2 > 0.8

- HapMap 5%
- 1kGP 5%
- 1kGP 2.5%

Competitor "New Array" *
Competitor "Old 900K"
660W
Omni1/ OmniExpress
Omni2.5

*Base content only
Summary

- First-generation GWAS has provided a foundation for beginning to understand the genetic architecture of many diseases and traits.

- However, first-generation GWAS was limited by the extent of knowledge about the spectrum of variation in humans in the HapMap era.

- NGS re-sequencing efforts, such as 1kGP, are providing a much more comprehensive catalog of common variation (>1% MAF) in diverse populations.

- Next-gen GWAS tools are leveraging this expanded catalog of variation to drive a new wave of genetic discovery by enabling exploration of the rare-variant hypothesis and higher resolution CNV research in a cost-effective tools.
Thank You!
## Illumina GWAS Portfolio at a Glance

<table>
<thead>
<tr>
<th></th>
<th>Omni2.5</th>
<th>Omni1</th>
<th>OEx</th>
<th>CytoSNP12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Markers per Sample</td>
<td>2,450,000</td>
<td>1,140,419</td>
<td>733,202</td>
<td>301,232</td>
</tr>
<tr>
<td>Number of Samples per BeadChip</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Scan Times per Sample (minutes)</td>
<td>15</td>
<td>13</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Spacing (Mean / Median / 90% percentile largest gap)</td>
<td>1.18 / 0.63 / 2.74</td>
<td>2.4 / 1.2 / 6.4</td>
<td>4.1 / 2.2 / 9.2</td>
<td>9.6 / 6.2 / 18.6</td>
</tr>
<tr>
<td>Markers Within 10 kb of a RefSeq Gene</td>
<td>1,233,932</td>
<td>618,959</td>
<td>381,329</td>
<td>148,666</td>
</tr>
<tr>
<td>Non-Synonymous SNPs§</td>
<td>49,564</td>
<td>32,110</td>
<td>12,134</td>
<td>3,480</td>
</tr>
<tr>
<td>MHC/ ADME / Indel</td>
<td>11,149 / 27895 / 0</td>
<td>19,081 / 22,429 / 459</td>
<td>7,566 / 16680 / 0</td>
<td>761 / 2,382 / 0</td>
</tr>
<tr>
<td>Sex Chromosome (X / Y / PAR Loci)</td>
<td>57,061 / 1897 / 554</td>
<td>27,493 / 2,322 / 1,157</td>
<td>18,239 / 1697 / 540</td>
<td>15,063 / 2,841 / 1,579</td>
</tr>
<tr>
<td>Mitochondrial SNPs</td>
<td>93</td>
<td>27</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>