

# Decoding large national genetic study designs

## Guidance on marker selection

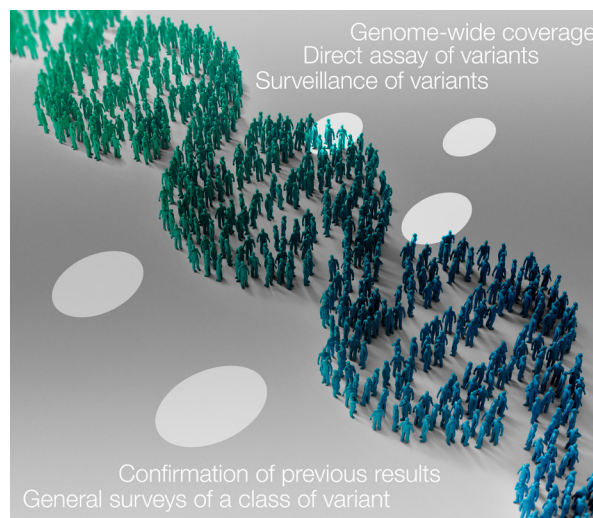
Form follows function in array design. While a generic microarray can generate powerful results for many experiments, scientists at the UK, Finnish, and Taiwan biobanks all tailored their arrays for specific research aims. Each subsequent array also incorporated new findings from the wider literature. Flexibility in array design is thus key to supporting new research directions. The well-known Applied Biosystems™ UK Biobank Axiom™ Array is an example of both incorporation of new findings and the design of an instrument for specific studies. Before looking at the specifics of the UK Biobank Axiom Array design process, we can analyze some of the major aspects of array design in general.

### Content design considerations

It is important to consider the elements of array design in light of your research goals, which could include discovery of new associations, confirmation and refinement of previous discoveries, and/or monitoring or surveillance of known, important variants. Broadly, those elements are:

#### 1. Genome-wide coverage for discovery

In the modern era of genomic research, imputation is used to leverage the power of large sequencing projects. Once sufficient genomes have been sequenced to provide a solid reference panel, a vast number of samples can be genotyped in a cost-effective manner on microarrays containing 500,000–1,000,000 markers. Imputation then allows the calling of tens of millions more variants.



Effective design of the genome-wide association study (GWAS) grid—the panel of markers selected for imputation power—is key to the success of this endeavor. For instance, the latest **Applied Biosystems™ Axiom™ Precision Medicine Diversity Array (PMDA)** was designed to provide comprehensive coverage of 1000 Genomes Project populations, and had a large fraction of available space devoted to that end. Specific research needs lead to different choices. The **UK Biobank Axiom Array** was designed specifically for high coverage of the British population by focusing on European variations. **The Finnish Biobank** designed an array for the Finnish population which, having been through a relatively recent bottleneck, did not require as many markers for high coverage. That project’s sequencing work indicated that an array designed using the 1000 Genomes Project reference panel would serve well for the study population. On the other hand, the Taiwan Precision Medicine Initiative used a unique, focused panel for the design of an array uniquely suited to that population.

## 2. Direct assay of variants critical to the study goals

Variants can include those from previous research findings and the literature or clinical practice, and variants of predicted relevance to the phenotypes of interest. The unifying factor of this class is that imputation is either not practical (e.g., for rare variants) or not fully able to support the study aims. While most such variants are accessible to ordinary probe design, some require extra effort. For example, in a study of an aging cohort, the UK Biobank wanted to analyze the *APOE* e4 variants, rs7412 and rs429358, constituting a large part of the risk for early-onset of Alzheimer's disease. Until then, the rs429358 variant had never been successfully genotyped on a microarray platform. Once it was identified as a major goal of the array design, we were able to devote extra space to many diverse probe designs, resulting in a successful assay that could subsequently be ported to future arrays.

## 3. Surveillance of variants of known significance

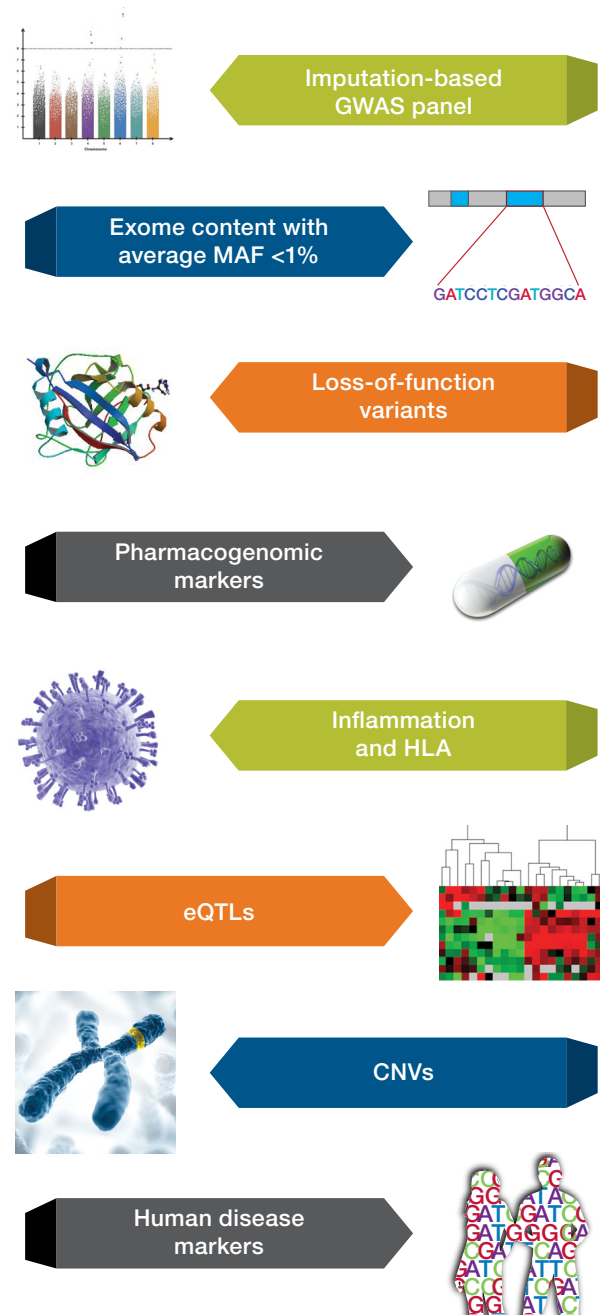
A research project aimed at new discoveries will still frequently need to assay known variants, such as pathogenic variants drawn from ClinVar or other databases, or variants affecting drug metabolism from the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines. Expertly curated modules play a role here. The inclusion of pharmacogenetic variants from the Applied Biosystems™ PharmacoScan™ array, or a module designed for the study of hereditary cancers from the UK Biobank array, etc., provides a solid basis for new array design, and allows the researchers to focus on newer findings from the literature and their own work. The latest Axiom PMDA incorporated all of these to provide a broad and up-to-date platform for many study aims, while custom arrays typically focus on specific areas of interest.

## 4. Confirmation of previous results

As noted above, this can take the form of directly evaluating specific variants. Another approach is to fine-map the region around a previous GWAS hit, directly analyzing many variants near the single-nucleotide polymorphisms (SNPs) and insertion/deletions (indels) in order to more closely pin down a causal variant. This frequently focuses on variants with known or predicted loss-of-function impact; missense and nonsense variants; variants in regulatory regions, at splice sites, etc.

## 5. General surveys of a class of variant

This is most commonly aimed at exome variants, especially with expected loss-of-function effects. For example, the UK Biobank Axiom Array included tens of thousands of predicted loss-of-function variants, ranging down to very low frequency in the British population, since the potentially large effect size could support significant associations even if only 10–100 alternate alleles were found in the 500,000 participants in the study. Regulatory elements such as micro RNA (miRNA) binding sites are also of interest in this regard.



**Figure 1. Axiom array content modules.** MAF: minor allele frequency; HLA: human leukocyte antigen; eQTLs: expression quantitative trait loci; CNVs: copy number variants.

## UK Biobank Axiom Array serving as a backbone for other research studies

The development of the **UK Biobank Axiom Array** captures many of these elements of array design. The array drew some of its content from a more generic biobank-style array, focused on coverage of common European and African-American variants and direct assays of hundreds of thousands of rare loss-of-function and nonsynonymous exome variants. The UK Biobank Axiom Array design was begun by trimming the rare content down to variants expected to be found in the British population, some at quite low frequency. Genomic coverage of African variation was reduced as well. In the space thus freed up, genomic coverage of British and European variation was greatly increased, down to a lower frequency range of 1% and up. Modules of markers for phenotypes of interest were compiled by the project's researchers and collaborators, including neuropsychiatric, cancer, cardiovascular, and many other conditions, as well as markers for pharmacogenomics and HLA imputation. The resulting array provided content specific to immediate study aims, but also a rich set of genotypes for future investigations.

## Expanded research content on the latest Axiom PMDA

Generally available arrays such as **Axiom PMDA** are designed to satisfy the most general research needs, such as global imputation coverage; known pathogenic variants in the genes recommended for reporting of incidental findings by the American College Medical Genetics; and pharmacogenomic variants including those in the CPIC guidelines and beyond. The PMDA design can in many ways be seen as an evolution of the UK Biobank Axiom Array, broadening the scope to a global population and refining the more medically interesting variants with newer findings from the literature. A microarray designed for a particular cohort or research project can be specifically aimed at the project goals. With the full flexibility to customize the Applied Biosystems™ Axiom™ Genotyping Solution, microarray uniquely suited to each research project can be designed—from adding a single, critical variant to PMDA, to adding a UK Biobank Axiom Array-derived module to an Asian population-focused array, to building a 100% custom array.



### Scientist profile: Dr. Jeremy Gollub

Dr. Gollub has been involved in microarray research and design for two decades. He received his PhD in biophysics from UC San Francisco in 1999, followed by postdoctoral work in plant genomics and microarray gene expression research at Stanford University. He has worked at Thermo Fisher Scientific in microarray genotyping and design since the introduction of the Axiom product line.

Have a question for our technical specialists?

Contact us at [thermofisher.com/genotyping-microarray-contact](https://www.thermofisher.com/genotyping-microarray-contact)

For information on genotyping strategies adopted by other researchers, go to [thermofisher.com/scientistspotlight](https://www.thermofisher.com/scientistspotlight)

**ThermoFisher**  
SCIENTIFIC