



H3Africa Chip Frequently Asked Questions

Why was the H3Africa genotyping array created?

Most genotyping arrays available at the time the array was designed (2015) did not have adequate representation of African genomic variation. Additionally, most reference panels for imputation were based on data from primarily Caucasian populations, included only a limited number of African ancestry samples and did not accurately reflect African genomic structure; therefore, H3Africa designed a more appropriate array and also generated a more useful imputation reference panel.

What data sets were used to identify African variants and select variants for the array?

There were three sources of whole genome sequence data from African populations that were used:

1. Publicly available sequence data from the 1000 Genomes Project and other sequencing efforts
2. Data generated by H3Africa for the purpose of designing the array
 - 350 samples from across Africa were selected by the H3Africa Genome Analysis Working Group. Selection was based on:
 - i. Ethnolinguistic representation, with priority given to filling gaps left by available data sets
 - ii. Consent category (i.e. No restriction consent, OR general research use with no prohibition on for-profit use)
 - Sequence data provided by the TrypanoGen H3Africa project
3. Aggregated data generated by the Wellcome Sanger Institute
 - For access controlled data, approved data access requests described the use of these data for the design of the H3Africa genotyping array
 - For the African Genome Variation Project, the secondary data analysis form described the use of these data for the design of the H3Africa genotyping array

How were variants chosen for the array?

- ~75% of the array content came from existing genotyping array content
- ~25% was selected content (existing and new) chosen to:
 - Increase tagging of African variation
 - Capture common variants seen in multiple African populations
 - Increase coverage of known disease-associated genomic regions
 - Capture known SNPs of interest for specific H3Africa projects



How does the array perform in African populations?

The H3Africa array has a dense tag-SNP coverage of genome variation in multiple African population groups and includes some novel common African variants. It also contains a comprehensive set of SNPs of potential clinical and pharmacogenomics significance derived from relevant public databases and disease-associated variants nominated for inclusion by H3Africa researchers. The performance of the array was evaluated computationally to assess coverage and imputation accuracy. The H3Africa array was shown to outperform currently available arrays of similar size in terms of genomic coverage and imputation accuracy for African populations. The array will provide researchers in Africa with a better tool for genetic research on African populations and those with high African admixture. It is a research tool and should not be used in clinical settings.

What is the array being used for?

The array is being used for research, mostly by the H3Africa consortium, to study the association of genomic variants with various health and disease states in order to understand risk factors for these diseases specifically in African and African-ancestry populations. The hope is that the findings of these studies will help scientists to eventually develop African-appropriate diagnostics, interventions, treatments, and even cures.

Why was Illumina chosen to manufacture the array?

Illumina was chosen based on their track record in developing consortium arrays with other research networks, price-point for manufacture and service, ability to include all the custom content requested by the consortium in the time-frame needed, and their commitment to increasing capacity for genomics research in Africa. As part of this commitment they have already contributed to establishing a genotyping service on the continent and intend to support others.

Is the array available for purchase?

While the array is considered an H3Africa consortium product, Illumina may sell this array at cost to others who are engaged in genomic research to benefit Africans.



Who profits from the array?

This array was designed by scientists for scientists and developed so that African scientists can do better science to ultimately benefit the people of Africa. No group, person or data provider associated with H3Africa who worked on the project has any commercial interest in the array or will receive any financial reward whatsoever from the array design, or sale of any arrays.

In order to manufacture the array, we had to engage a commercial partner with capacity to manufacture high quality arrays at reasonable prices. As explained above, Illumina was chosen as the technology partner. Through H3Africa negotiation with Illumina, competitive pricing was agreed for the array for African scientists. A standard consortium agreement was established and signed in order to give H3Africa members and other African scientists access to the array as a research tool at low cost. Illumina has provided a free array scanner for African use, and covered costs of sample shipping for genotyping for H3Africa researchers. Any profit that Illumina makes is a reasonable and appropriate profit for the manufacture of the array rather than the content of the array.

What IP is contained in the array and who owns it?

As with all genotyping arrays, the IP is contained in the technology and the probes to detect the variants and not in the variant data or content itself. Illumina does not own any IP on the content. None of the underlying individual level data has been provided to Illumina, only the reference information needed to generate the probes, which include a subset of common variants present in multiple populations derived from aggregated data. Current understanding and legal opinion are that sequence data does not contain IP (see: <https://ghr.nlm.nih.gov/primer/testing/genepatents>).

The array contains two sets of probes: (1) Probes for SNPs which researchers specifically asked for; and (2) Probes for SNPs that were chosen because of the allele frequency in African or global populations, taking into account the linkage disequilibrium structure of different populations. The result of the design process was only the list of probes that should be included in the array. Not only was no individual-level data made available to Illumina, no frequency or linkage disequilibrium data was provided nor any reason for why the probe was selected or in which population the SNP had been found. Any IP that might subsist in this sequence data therefore has not been affected by the array design process.