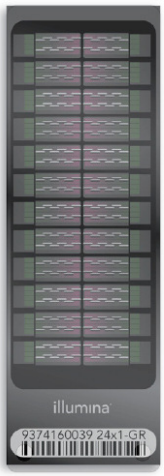
Technical overview - Infinium® Global Screening Array (GSA)

with optional Multi-disease drop in (MD)



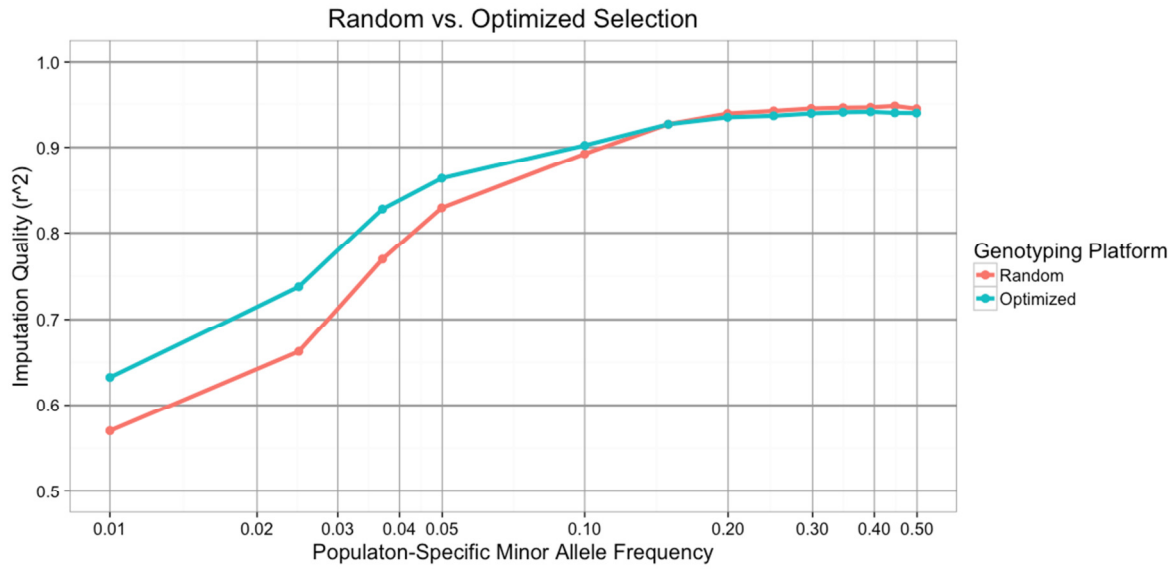
The Infinium® Global Screening Array (GSA) combines a highly optimized, universal genome-wide backbone, hand-curated clinical research variants, and sample tracking content to create an extremely powerful, highly economical array for population-scale genomics. With over 5.45 million samples committed to date globally, the GSA has become the industry standard worldwide for genomic screening.

GSA with MD contains ~50K variants, the Multi-disease drop in (MD), aimed at capturing exonic variants seen in global populations that are phenotype agnostic, putatively functional, and yield high imputation accuracy. These variants were selected from ExAC (a published database of 60K whole exome sequences) or published in recent genome-wide association studies. ExAC variants include primarily loss of function (LoF) variants with four or more copies seen in ExAC or a minor allele frequency of > 0.03 (chosen to yield high imputation accuracy). All missense mutations from ExAC were also selected that had a maximum population frequency between 0.0005-0.015 from African, American, East Asian, Finnish, and Non-Finnish Europe. The remaining genome-wide content was selected based on imputation of the 650K GSA genome-wide backbone. All variants annotated by the NHGRI-GWAS catalog or recently published by large consortia that were originally missed by imputation from the GSA genome-wide backbone were captured here. The GSA-MD is comprehensive and is representative of both rare and common functional variation that yields high imputation accuracy across populations for relevant disease loci. The GSA-MD provides an opportunity to capture additional functional variation across an entire span of diseases.

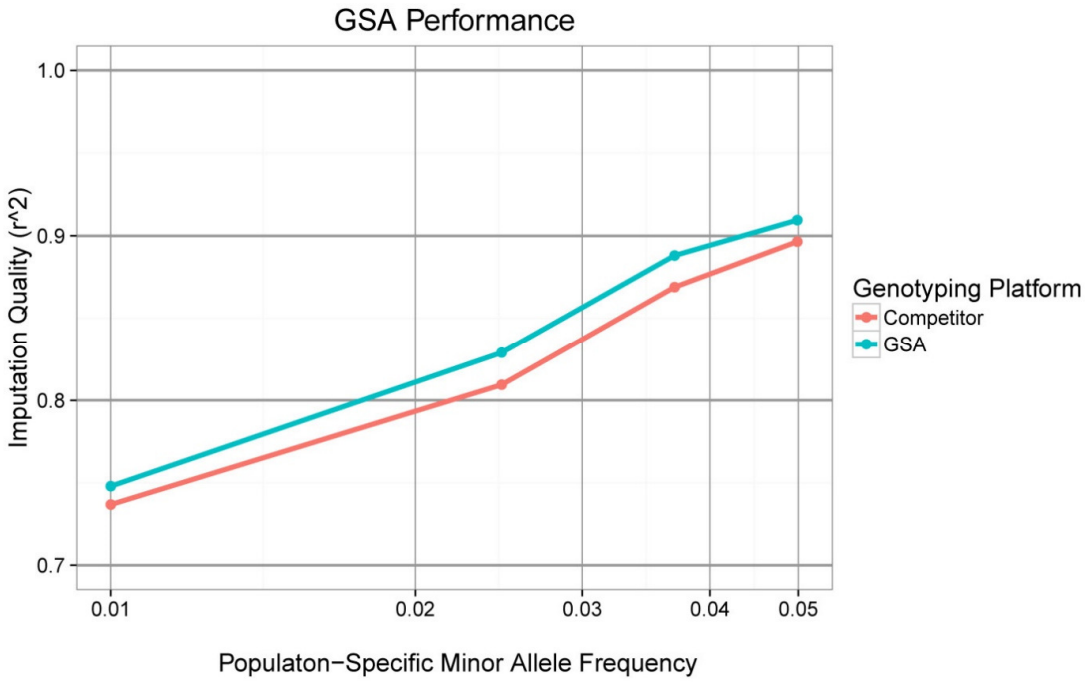
Key diffrentiators of the GSA: higher imputation coverage in Europeans, expert guided clinical research content with superior coverage of pharmacogenetic markers and clinically relevant copy number variations as well as deep coverage of pertinent quality control markers. In addition the array is already in customer hands with exceptional performance specs reported.

The GSA has been expertly designed to achieve a multi-ethnic, genome-wide backbone optimized for unparalleled genomic coverage and imputation performance in all five major super-populations. A significant portion of the GWAS backbone of GSA was derived from Illumina's Multi-Ethnic Global Array (MEG), a high-density array with > 1.7M variants. The genome-wide coverage of MEG is largely focused on variants on the low minor allele frequency variants (MAF = 1-5%) to ensure recent populations substructure is well represented and accurately imputed. The emphasis on low-frequency variation across populations is an update to previous arrays that were limited to common variants and/or outdated reference data. The MEG genome-wide backbone captures cross-population low-frequency tag SNPs optimized for imputation accuracy by leveraging the diversity from 26 populations present in Phase III of 1000 Genomes Project (1000G) in addition to a diverse set of African Americans and Hispanic Latinos from the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA). GSA genome-wide content was carefully selected from MEG and further optimized to increase imputation accuracy in Europeans and Asians.

Maximimizing Imputation Coverage: The largest challenge faced when designing GSA was maximizing imputation coverage in all populations for a low-density array with less than a third of the content (650K) compared to high-density arrays like MEG. While dense arrays such as MEG have high performance a random selection of polymorphic markers for low-density array like GSA is not suffice. Improved selection of markers for MEG that maximizes imputation accuracy and linkage disequilibrium yield an 8% increase in accuracy of low-frequency variants compared to a random selection (Figure 1).



The European optimization in the GSA yields high imputation coverage in European population's superior to recent competitive arrays particularly for the important low frequency variants (Figure 2). The high performance of the GSA genome-wide content in Europeans is a proof of principle of quality exceeds quantity as the competitor array has >300K variants compared to GSA yet lacks imputation power. This is particularly important for the population-specific low-frequency variants between 1-5% and not common variants that typically are ancient, less likely to be associated with disease, require less statistical power for detection and impute with accuracy of 90% confidence (see figure below).



GSA has a comprehensive set of markers that ultimately decrease cost and time for various applications in the era of precision medicine. Designed through collaboration with experts in medical genomics, the GSA clinical research content utilizes multiple annotation databases to create a highly relevant panel for clinical research applications.

Pharmacogenomic content: The field of pharmacogenomics is perhaps the most direct base to bedside application to date. Several research hospitals and consortia have formed to develop methods to incorporate genetic information to inform dosing and drug administration decisions. GSA contains all relevant PharmGKB variants (with over 50% more than a competitor 900K array) and all Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline variants. The GSA also contains unpublished proprietary content from the Type II Diabetes Consortia (GoT2D and T2D-GENES), private Pharmacogenomics entities, and the International Cancer Genome Consortia.

Variants of clinical relevance: Additionally, the GSA contains hand-selected variants from ClinVar known to be of the highest clinical relevance and impact. Fundamentally, the GSA targets clinically relevant variants that go beyond the standard interrogation of single variant loci and was expanded by hand-designing more than 25K variants that are traditionally difficult to design such as indels, multi-allelic markers, and copy number variants unique to Illumina's Infinium technology. With respect to clinical CNVs from ClinVar, GSA targets substantially more (>10X) than the competitor 900K array (Figure 3). GSA also contains twice as many indels and structural variants with known clinical relevance from ClinVar (Figure 3). In this manner, the Infinium® Global Screening Array has achieved the most clinically relevant variants, efficiently packaged into a powerful microarray for maximal genomic coverage

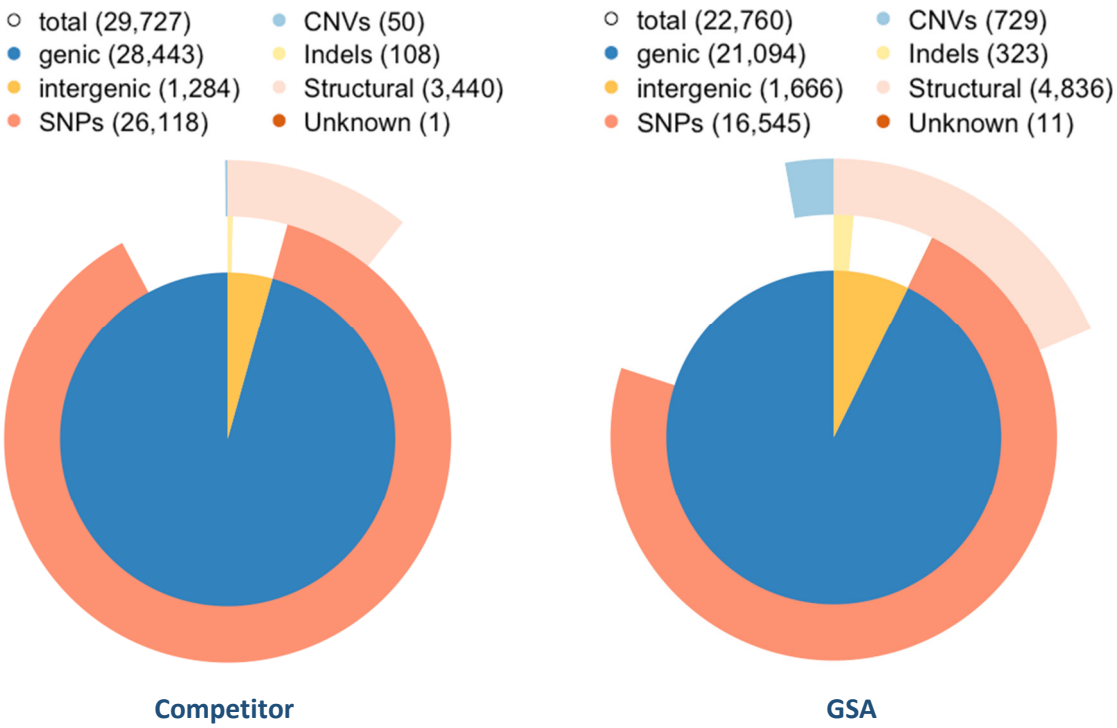


Figure 3. Types of ClinVar variants proposed by competitor (left) vs. designed and converted ClinVar variants on GSA (right).

Ancestry and QC markers: The GSA has high-value markers enable sample identification, tracking, ancestry determination and stratification that are highly valuable for large-scale genetics and screening applications. This content was derived to for high throughput genomics, Biobanking, and any other application that requires validation of sample identity and quality. With >9,000 markers, the GSA QC content is focused on high-value functional content that enables high accuracy and high-throughput genotyping required for quality control, tracking, and stratification applications. GSA is exceeds the

competitor array in regards to coverage of relevant QC content for ancestry informative markers, fingerprinting, and blood phenotypes (Figure 4).

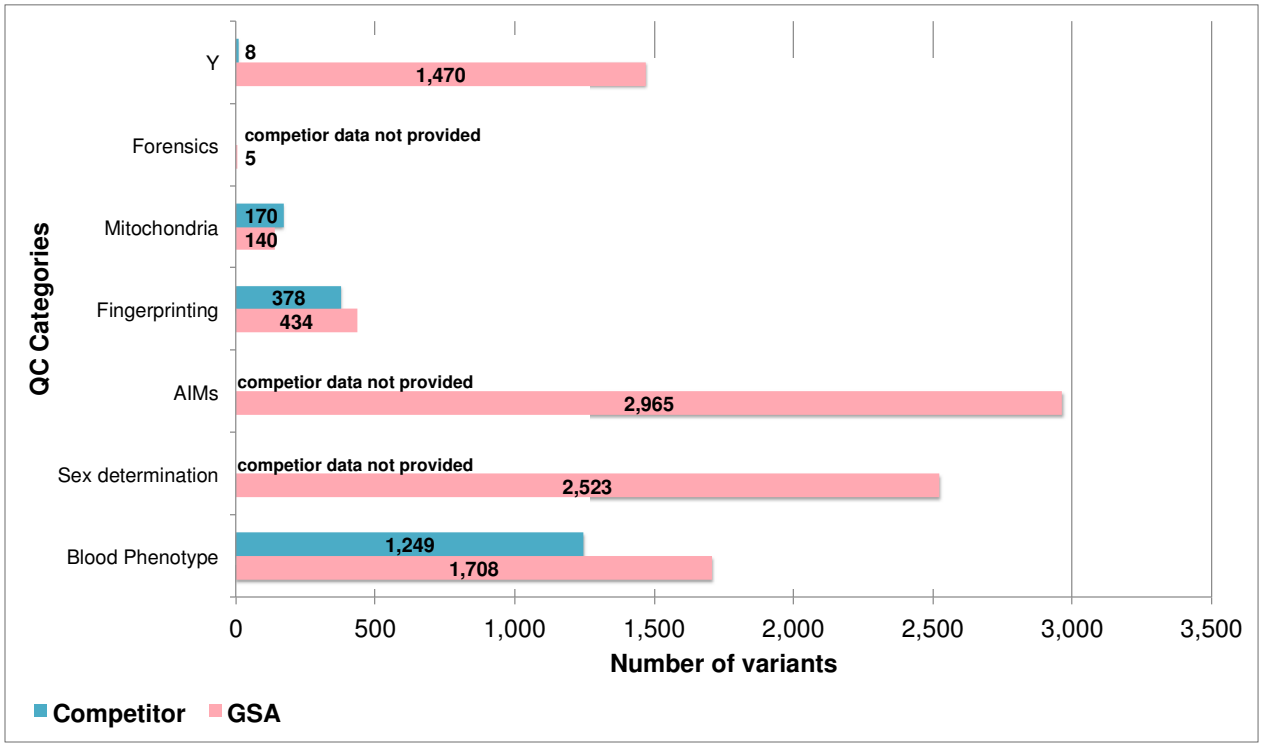


Figure 4

Built on the trusted Infinium® HTS platform, the Infinium® Global Screening Array provides the same high-quality, robust, and reproducible data Illumina genotyping arrays have been delivering for over a decade with call rates of >99% and reproducibility of >99.99%. The GSA design consideration to achieve 700,656 markers translates into a robust, 24 sample array format that preserves an average probe redundancy of 15X for each marker, thus ensuring the highest performance standard of any commercially available genotyping array for SNP and CNV calling. Since the GSA has already completed its design and production phases, the array content as described is fully converted and realized on the array and can be provided upon request.

Consortia supported development: Finally, the Infinium® Global Screening Array has been selected as the array of choice for the European GSA Consortia which is comprised of three founding member institutions – Erasmus MC (Rotterdam, Netherlands), Life & Brain (Bonn, Germany), and CNG (Paris, France). This coordinated genotyping effort has established a network of top medical geneticists throughout Europe using the GSA. With over 530,000 samples collected thus far, this consortia represents one of the largest genotyping efforts in Europe. The data collected will have important implications for future, large scale meta-analyses and a network to establish more efficient calling of rare variants due to a common, shared genotyping platform utilizing the GSA.

Custom Content:
Customers can use Illumina's iSelect custom design suite of software to design and add their own custom content or alternatively work with Illumina's Concierge Custom Design Service. The Illumina Concierge Custom Design Service is a personalized consulting solution for customers who need assistance in the design and delivery of custom content for genomics projects. Under the leadership of dedicated design experts, the Illumina Concierge Service can be used to provide advanced design assistance, including in-silico design optimization and extended product capabilities, for a variety of sequencing and microarray products. Or customers work with dedicated project managers to optimize a product for their specific project needs. Product optimization support ranges from functional testing with controls, to iterative product enhancements, and shipment coordination.

If a customer selects the GSA together with the MD dropin, there is still sufficient space to add an additional 5K custom markers without removing any content. If additional custom markers are needed Illumina's Concierge Service can work with the customer to identify portions of the GSA content which can be removed to allow substitution with user defined custom content.

Additional information on Illumina microarray developments

Illumina human genotyping arrays contain the most informative content, supporting whole-genome, targeted genome, and exome analyses. Powered by industry-leading, established Infinium assay technologies, they deliver exceptional data quality and high-density genomic coverage. Supporting high-throughput, multiplex processing, Illumina microarrays are ideal for large-scale population study researchers and biobanks performing genome-wide and phenome-wide association studies (GWAS and PheWAS).

CNV, LOH and indels: The Infinium microarrays offer the high resolution required for detection of copy number variants (CNV), loss of heterozygosity (LOH), and insertions/deletions. All Illumina microarrays, including whole-genome and targeted array products are designed for continued collaboration and expansion to meet the needs of the research community

Illumina has continually developed its microarray business to incorporate the latest scientific and bioinformatics design strategies, to deliver a powerful set of genotyping capabilities for its customers. Key developments include new SNP content from multi-ethnic population groups, significant new disease and pharmacogenomics content and imputed with new WGS content from various geographic regions. This new generation of microarrays are designed and implemented using sophisticated algorithms giving superb imputation accuracy, which leverage the Phase lll content of the 1000 Genomes project.

The table below lists new Illumina microarray products developed in the last 18 months

Array	Number of SNP's	Application focus
Neuro Array	180,000	Consortium array – extensive neurodegenerative disease coverage
Oncoarray	499,170	Disease markers for a wide range of tumor types
DrugArray		Researchers investigating 'drugable' targets and DMPK
Multiethnic array- AMR & AFR	1,430,000 + custom	Genotyping of Hispanaic and American populations
Multiethnic EUR, E.& S Asian	1,430,000 + Custom	Genotyping of European, East & South East Asian populations
Multi-ethnic Global	1,756,820	A Superset of markers for global population genotyping
Global Screening array - GSA	660,000 +50,000 (Custom)	Population genotyping,
EPIC array	860,000	Latest methylation array – massively updated 450K

H3Africa Consortium Array

Requirement for a tool to address African genomic diversity, Illumina was selected to develop an array for genome-wide association studies (GWAS) by the Human Heredity and Health in Africa (H3Africa) Initiative, a partnership between NIH, the African Society of Human Genetics, and Wellcome Trust. H3Africa will use the new array to foster genomic and epidemiological research to improve the health of people in Africa. This 2.5M African-specific GWAS array and extensive African-enriched reference panel will be available for all interested researchers.

Infinium Consortium Neuro Array

This array was developed as a next generation genotyping array designed specifically by the neurodegenerative community for deep replication of GWAS and fine mapping of genetic variants in a range of diseases. It allows for cost effective high throughput characterisation of genetic mutation and risk factors associated with common neurodegenerative disease.

Infinium microarrays and Biobanking

Illumina has signed agreements to genotype the sample collections at four biobanks. Each of the projects is aiming to build a repository of linked genotypic and phenotypic data, with the genotyping results generated by Illumina being paired to electronic medical records (EMRs) possessed by the owners of the biobanks. Vanderbilt University, the University of Colorado at Denver, Partners HealthCare and Montreal Heart Institute are the four biobanks to sign up to work with Illumina. The specifics of each project vary somewhat from biobank to biobank, but each initiative is underpinned by a desire to improve healthcare and drug research by unlocking the information contained in samples. In total, the first wave of work at the biobanks will genotype in excess of 200,000 samples on the Multi Ethnic MEGA array.