

Human genotyping

Axiom™ UK BiobankPro Microarray: Advancing human health research with powerful genotyping technology

The Applied Biosystems™ Axiom™ UK BiobankPro (UKBBPro) solution is a comprehensive platform for high-throughput analysis of selected genetic variants. The solution includes Applied Biosystems™ SwiftArray™ Assay reagents and Applied Biosystems™ Axiom™ UK BiobankPro (UKBBPro) Microarrays in a 96-peg format with flexible sample throughput options. The array content was developed in collaboration with UK Biobank Array Design Group and other scientists to facilitate populational genomic health research. Additional single-nucleotide polymorphisms (SNPs) and copy number variation (CNV) markers have been incorporated to enhance translational research, population studies, and routine genotyping applications.

The UK Biobank network has been a major contributor to understanding genetic variations in populations, and how these can contribute to pathologies [1]. The UK Biobank (UKBB) is a prospective cohort of over half a million individuals collected between 2006 and 2010, with extensive phenotypic data related to lifestyle, health, and other biometric values [2]. In addition, the UKBB collected genomic information on these individuals using custom Axiom genotyping microarrays [2]. Over 500 peer-reviewed GWAS studies have been published using this resource [1], including studies characterizing genetic contributors to treatment-resistant ulcerative colitis [3], loci involved in heart development and failure [4], severe mental illness [5] and the development of polygenic risk scores for congenital heart disease [6] and endometrial cancer [7].

Highlights

- Comprehensive genomic coverage: Includes over 820,000 probesets, providing extensive coverage for populations of European ancestry.
- Imputation-aware design: Proprietary imputation algorithms maximize genomic coverage.
- CNV detection: Includes probes targeting CNV regions that assess impact on developmental delay and other pathologies.
- Disease markers: Incorporates markers that have been implicated in inherited syndromes.
- GWAS Focus: Includes extensive coverage of non-coding loci, facilitating imputation and eQTL analysis
- Inheritance markers: Markers include extensive coverage of Y and mitochondrial chromosomes, and Neanderthal ancestry.
- High-throughput workflow: Capable of generating up to 629 million SNP determinations for 768 samples in five days using a single Applied Biosystems™ SwiftArrayStudio™ Microarray Analyzer system

applied biosystems

Content of the array

Array design

This array was developed by leading researchers in epidemiology, human disease research, and population genetics for use in the UK Biobank. The UK Biobank's prospective genotyping study of 500,000 individuals aims to research the complex interactions among genetic factors, lifestyle, and environmental exposures that contribute to human health and disease.

To maximize genomic coverage, proprietary imputation-based marker selection algorithms were employed. Reference panels from the 1000 Genomes Project (KGP) and European ancestry populations, combined with advanced imputation methods, enabling broader genomic coverage than conventional pairwise tagging approaches alone.

Copy number variant content

The array includes copy number variation (CNV) content covering 67 loci associated with developmental delay phenotypes [8-9]. Additional markers were incorporated for dense coverage across CNV regions and, where possible, at CNV breakpoints. Further CNV markers were selected to capture regions with suggestive evidence of association with lung function, as well as candidate CNV regions of interest for evaluating relationships with lung function and related phenotypes.

Rare coding variant content

Rare coding variants were selected based on estimated minor allele frequencies (EMAF) in UK and other European populations, using content derived from the Applied Biosystems™ Axiom™ Exome Genotyping Arrays and exome sequencing studies. The array includes protein-truncating variants (PTVs), such as premature stop codons, frameshift mutations, and loss-of-start-codon variants, because these alterations are strong candidates for causing loss of gene function. Additional rare coding variants consist primarily of missense mutations that may affect protein structure or function.

Table 1. Categories of markers for the European ancestry population* on the Axiom UK BiobankPro Microarray.

Markers of specific interest	
Alzheimer disease	701
APOE	829
Autoimmune and inflammatory	250
Blood phenotypes	2266
Common cancer variants	338
Common mitochondrial DNA variants	160
Cardiometabolic	369
eQTL	18035
Fingerprint	258
HLA	7266
Killer-cell immunoglobulin-like receptor (KIR)	1470
Lung function phenotypes	8287
Neurological disorders	18014
NHGRI GWAS catalog	10310
Other rare coding variants	77029
Pharmacogenomics/ADME	2288
Rare, possibly disease-causing, mutations	11660
Rare variants in cardiac disease candidates	1397
Rare variants in cancer predisposition genes	4855
Tags for Neanderthal ancestry	10838
Y-chromosome markers	590
CNV regions for developmental delay	2369
Rare coding variants	
Protein-truncating variants	28423
Other rare coding variants	77029
Genome-wide coverage	
Genome-wide coverage for common variants	341453
Genome-wide coverage for low-frequency variants	270542
Total number of markers on array	820967

* The European ancestry population is the EUR panel defined as the GBR, CEU, FIN, IBS, and TSI samples from the 1000 Genomes Project (KGP) [4]. Markers may be selected for more than one category but only appear on the array once. When we refer to the possible disease-causing or disease-associated variants, we do not mean to imply that the status of the variant, or any definitive role in, or association with, disease has been established.

Disease research content

The array incorporates a broad range of disease-focused markers selected to support research into the genetic basis of complex human disorders. Content was chosen based on published association studies, sequencing efforts, and disease-specific research initiatives, enabling investigation of variants linked to neurological, autoimmune, inflammatory, and cancer-related phenotypes.

Alzheimer's disease and APOE-related markers were selected from variants demonstrating evidence of association with Alzheimer's disease through both candidate gene studies and genome-wide association studies (GWAS). Additional markers were contributed by the Alzheimer's Disease Research Forum [10], and include a dedicated set spanning the APOE region. Notably, markers such as rs429358 and rs7412, which define the APOE isoforms associated with Alzheimer's disease risk and other clinical conditions, are included to provide comprehensive coverage of this important locus.

To support research of immune-mediated disease, the array includes autoimmune and inflammatory disease variants with reported associations to conditions such as ulcerative colitis,

Table 2. Numbers of OMIM pathologies queried by the Axiom UK BiobankPro Microarray. Classifications were according to the OMIM Clinical Synopsis definitions.

Category/System	Count
Neoplastic	187
Metabolic/endocrine	662
Immunologic	187
Hematologic	181
Cardiovascular	146
Respiratory	73
Neurologic	382
Peripheral nervous system	60
Muscle	133
Ophthalmologic	217
Otolaryngologic	203
Genitourinary/renal	112
Gastrointestinal/hepatic	80
Skeletal	210
Skin/hair/nails	158
Male reproductive/genital	34
Female reproductive/genital	24
Psychiatric/behavioral	33
Growth/development	716
Nondisease/trait/susceptibility	524

Crohn's disease, type 1 diabetes, Graves' disease, Hashimoto's thyroiditis, and celiac disease. These markers facilitate investigation of genetic factors underlying immune regulation and disease susceptibility.

Cancer-related content consists of common variants associated with cancer phenotypes identified through GWAS and curated within the NHGRI GWAS Catalog, as well as more recently reported cancer-associated variants from the scientific literature.

The array also includes neurological disorder markers selected from multiple sources, including large-scale mutation studies, recent association findings, and exome sequencing projects. These markers support research into a variety of neurological and neurodegenerative disorders, including Alzheimer's disease, Parkinson's disease, frontotemporal dementia, progressive supranuclear palsy, and amyotrophic lateral sclerosis (ALS).

GWAS-focused content

Genome-wide association study markers were selected using imputation-aware algorithms [11-13] to provide comprehensive genome-wide coverage of common variants (estimated minor allele frequency [EMAF] ≥ 5%) in populations of European ancestry. This content includes 246,055 markers from the Applied Biosystems™ Axiom™ Biobank Genotyping Array and leverages imputation from the 1000 Genomes Project (KGP) reference panel, including European populations such as GBR, CEU, FIN, IBS, and TSI [14]. The resulting design helps ensure high genomic coverage across common variation while facilitating maximal imputation performance.

Although the array was optimized for European populations, its advanced GWAS design also supports broad genomic coverage in diverse populations. The content on the array can also be used as a starting point to extend genomic coverage for virtually any population of interest in a custom array.

To enhance the array's utility for complex trait and disease research, additional markers were selected from studies that identified expression quantitative trait loci (eQTLs). These markers were drawn from resources including the MuTHER Project [15], the ALSPAC Project [16] and the GENCORD Project [17], enabling investigation of genetic variants that influence gene expression.

The array further incorporates markers reported directly in, or used as proxy variants for, associations cataloged by the NHGRI-EBI GWAS Catalog. Inclusion of these validated disease- and trait-associated variants supports replication studies, fine-mapping efforts, and exploration of established genotype–phenotype relationships.

Table 3. Imputed genome-wide coverage in European ancestry populations (EUR panel).

Category	EMAF* range	Mean r^2	% markers with $r^2 > 0.8$
Common	5% ≤ EMAF ≤ 50%	0.920	90.1%
Low frequency	1% ≤ EMAF < 5%	0.785	67.1%

* EMAF: estimated minor allele frequency

Genetic risk research content

The array includes a comprehensive set of risk genotype markers designed to support investigations of disease susceptibility, trait variation, and clinically relevant genetic variation. Content was selected from established disease databases, large-scale sequencing efforts, and population genetic resources to enable detailed characterization of both rare and common variants across multiple disease areas.

Rare variants within disease predisposition genes were selected to capture missense mutations in genes associated with predicted loss of function and known disease-causing mutations. Markers linked to cardiac disease predisposition include rare variants in genes such as *MYBPC3* and *MYH7*, as well as variants curated from the *ARVD/C* locus-specific database [9] for genes including *DES*, *DSC2*, *DSG2*, *DSP*, *JUP*, and *PKP2*. Additional markers were incorporated from human disease mutation databases covering genes associated with cardiac disease and hemochromatosis. To facilitate assessment

Table 4. Imputed genetic coverage metric (mean r^2) of the Axiom UK BiobankPro Microarray across 1000 Genomes Project populations at ranges of estimated allele frequencies (EMAF).

Population	EMAF range	Imputed coverage (r^2)
CEU	0.05 < MAF < 0.5	0.925
	0.01 < MAF < 0.5	0.875
	0.01 < MAF < 0.05	0.767
GBR	0.05 < MAF < 0.5	0.917
	0.01 < MAF < 0.5	0.859
	0.01 < MAF < 0.05	0.738
FIN	0.05 < MAF < 0.5	0.927
	0.01 < MAF < 0.5	0.871
	0.01 < MAF < 0.05	0.764
TSI	0.05 < MAF < 0.5	0.915
	0.01 < MAF < 0.5	0.845
	0.01 < MAF < 0.05	0.711
CHB	0.05 < MAF < 0.5	0.877
	0.01 < MAF < 0.5	0.776
	0.01 < MAF < 0.05	0.548
CHS	0.05 < MAF < 0.5	0.887
	0.01 < MAF < 0.5	0.785
	0.01 < MAF < 0.05	0.572
JPT	0.05 < MAF < 0.5	0.880
	0.01 < MAF < 0.5	0.777
	0.01 < MAF < 0.05	0.543
MXL	0.05 < MAF < 0.5	0.897
	0.01 < MAF < 0.5	0.844
	0.01 < MAF < 0.05	0.735
CLM	0.05 < MAF < 0.5	0.902
	0.01 < MAF < 0.5	0.850
	0.01 < MAF < 0.05	0.745
PUR	0.05 < MAF < 0.5	0.909
	0.01 < MAF < 0.5	0.860
	0.01 < MAF < 0.05	0.772
YRI	0.05 < MAF < 0.5	0.812
	0.01 < MAF < 0.5	0.741
	0.01 < MAF < 0.05	0.643
LWK	0.05 < MAF < 0.5	0.808
	0.01 < MAF < 0.5	0.727
	0.01 < MAF < 0.05	0.636
ASW	0.05 < MAF < 0.5	0.809
	0.01 < MAF < 0.5	0.752
	0.01 < MAF < 0.05	0.659

of variant frequency and clinical significance, the array also includes markers relevant to lung function and other phenotypes, along with variants identified in the Exome Aggregation Consortium (ExAC) database. This content supports improved discrimination between pathogenic and non-pathogenic variation within large, unselected populations.

To support studies of complex disease risk, the array incorporates cardiometabolic variants associated with traits such as coronary artery disease, lipid metabolism, anthropometric measures, glycemic traits, and blood pressure. These markers enable investigation of the genetic architecture underlying metabolic and cardiovascular health.

The array also contains a targeted set of lung function–related markers selected based on established or putative associations with pulmonary phenotypes. Included variants are relevant to lung function measures, respiratory diseases such as asthma, cystic fibrosis, chronic obstructive pulmonary disease (COPD), idiopathic pulmonary fibrosis, and lung cancer, as well as smoking behavior.

Ancestry marker content

The Axiom UKBiobankPro Microarray includes a comprehensive set of ancestry-informative markers designed to support population genetics research, lineage tracing, and the study of human evolutionary history. These markers provide insight into both recent and ancient population structure, enabling analyses of ancestry, migration patterns, and population diversity.

Y-chromosome markers were selected to define lineages within the male-specific region of the Y-chromosome and to identify major branches of the Y-chromosome phylogeny. Additional markers representing rarer sub-lineages found within the UK population were incorporated to improve resolution and helps provide information on the frequency of higher-order Y-chromosome haplogroups.

To complement paternal lineage analysis, the array includes common mitochondrial DNA (mtDNA) variants spanning both coding and non-coding regions of the mitochondrial genome. These markers support the construction of mitochondrial phylogenetic networks and facilitate comparisons of mitochondrial haplogroups across populations. The content also includes variants associated with common mitochondrial disorders, extending the utility of the array for disease-related investigations.

The array further incorporates Neanderthal ancestry markers selected to capture genetic segments and haplotypes inherited from archaic human populations. These variants include Neanderthal-derived SNPs that remain present in modern human populations and may contribute to functional biological effects. Additional markers were chosen from sets of Neanderthal-derived variants that occur at substantially higher frequencies in modern European populations than in African populations, providing enhanced resolution for studies of archaic introgression and human evolutionary history.

Pharmacogenomic genotypes

Pharmacogenomics/ADME content consists of markers for genetic variants related to absorption, distribution, metabolism, and excretion (ADME) from the Pharmacogenomics Knowledgebase [18] with known relevance to drug metabolism and some markers from the Applied Biosystems™ DMET™ Plus platform. The set of ADME markers from the Axiom Biobank Genotyping Array were selected for the UK Biobank Axiom Array.

Other markers of biological significance

Blood phenotype markers were chosen from genome-wide association studies (GWAS) and candidate gene studies that identified associations with red blood cell group determination, erythropoiesis, platelet formation, and the regulation of blood homeostasis. These variants enable the study of genetic factors influencing blood-related traits and physiological processes.

The array also incorporates a fingerprint marker set consisting of SNPs developed by the University of Washington and the Broad Institute. These markers are shared across several major genotyping platforms to facilitate sample tracking, quality control, and cross-study data integration. A subset of fingerprint markers from the Applied Biosystems™ Axiom™ Biobank Genotyping Array was selected for inclusion in the UK Biobank Array to support compatibility with existing datasets and workflows.

To support immunogenetic research, the array includes markers within the human leukocyte antigen (HLA) region and killer-cell immunoglobulin-like receptor (KIR) genes. Because HLA and KIR loci play critical roles in immune response yet are difficult to assay directly using conventional array technologies, these markers were selected to facilitate imputation of classical HLA Class I and Class II alleles as well as KIR gene variation.

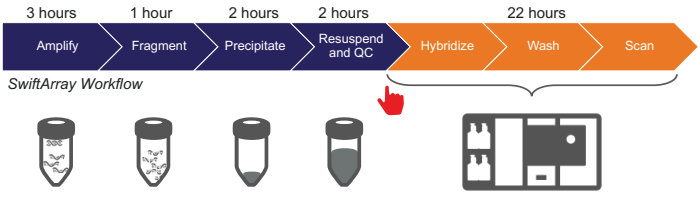


Figure 1. The SwiftArrayStudio Microarray Analysis solution. The Applied Bisystems™ SwiftArray™ Assay target preparation steps can be completed in about eight hours in a single working day. On-instrument, autonomous array processing can generate data within 22 hours of loading. Overall, a full set of genotypes and copy number calls from an Axiom plate can be generated in about 30 hours. Different colors represent workflow steps on different days; finger represents touch point on the instrument, where plates and/or reagents are loaded onto the instrument.

Throughput

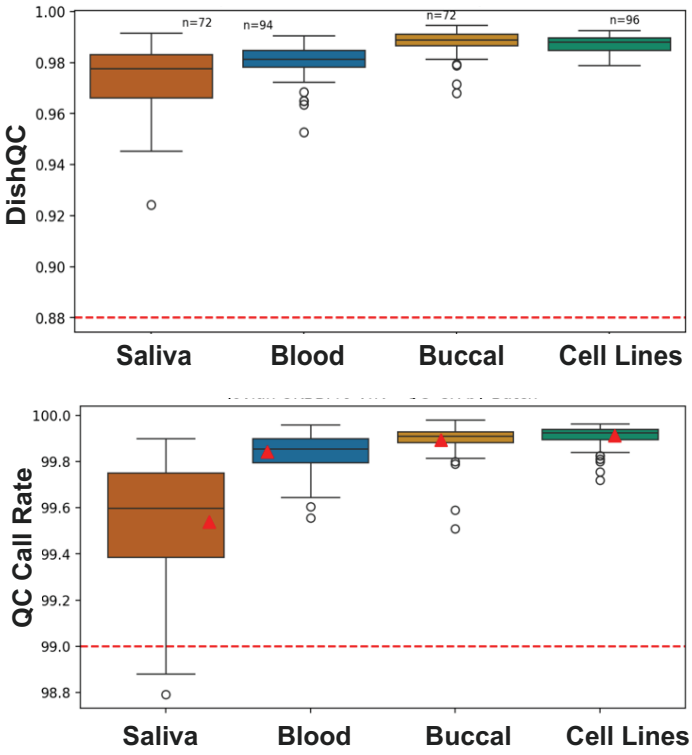
To provide greater efficiency for researchers in high-throughput microarray studies, the Axiom SwiftArray Assay with automated array processing on the Applied Biosystems SwiftArrayStudio Microarray Analyzer (Figure 1; for further information, see the **SwiftArrayStudio Microarray Analyzer webpage**). This solution consists of a target preparation that can be completed in a single work day, and the SwiftArrayStudio Microarray Analyzer that automates hybridization, washing and scanning, providing genotypes in 30 hours. This assay reduces the sample-to-answer time from five days to two days, allowing researchers to gather and interpret genetic data to keep pace with their needs. Of course, multiple runs can be initiated in the same week, increasing the throughput to 16 plates (for details, see **this application note**).

Performance

Thermo Fisher Scientific maintains rigorous performance standards for all microarray products, and the Axiom UK BiobankPro Microarray meets or exceeds these criteria. This array was extensively verified using genomic DNA purified from 335 samples, including Coriell HapMap cell lines, buccal swabs, blood collections, and saliva collections. The Applied Biosystems™ Axiom™ Analysis Suite software, which automatically calculates quality-control metrics, including dishQC, QC call rates, and copy number calling noise, was used to evaluate the performance (Figure 2).

Summary

The Axiom UK BiobankPro Microarray is a platform for comprehensive genotyping populations of European descent. Its imputation-aware design facilitates the research of disease and susceptibility genes, copy number variants that contribute to pathologies, and integration with GWAS data. The SwiftArrayStudio Microarray Analyzer and SwiftArray Assay accelerate the generation of these data with a simplified workflow and array processing that can be completed in 30 hours. Together, they present a powerful solution for translational research.



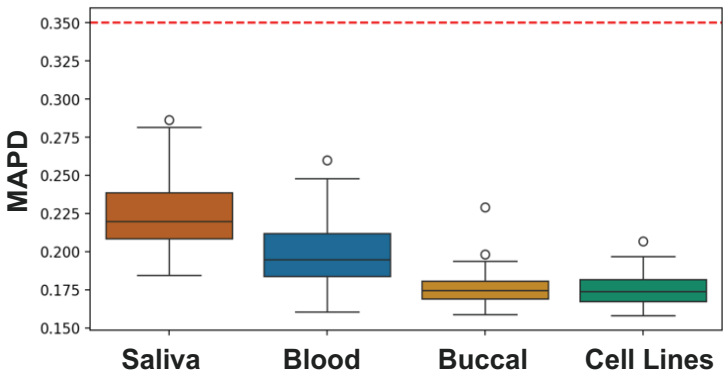


Figure 2. Quality control metrics from Axiom UK BiobankPro Microarray. Metrics were determined for the indicated sample types. Red dashed line indicates internal passing criteria. Top left: DishQC, a measure of the amount of overlap between two homozygous peaks created by non-polymorphic probes. A value of 1.0 is considered optimal. Bottom left: QC Call Rate, the average percentage of 200 autosomal SNPs with a call other than NoCall. Note that one of the saliva samples fell below the recommended threshold. Above: MAPD of copy number loci. MAPD measures the absolute difference between the log2 copy number ratios of neighboring bins and then calculates the median across all bins. Larger MAPD values indicate greater noise.

Acknowledgements

We are very grateful to the UK Biobank Array Design Group and the many scientists and research groups who provided markers and data as part of the design process for the UK Biobank Axiom Array and to the 500,000 UK participants who have generously donated time, medical information, and biological samples to better understand the complex interactions between genes, lifestyle, and environment.

Ordering information

Description	Cat. No
Axiom UK BiobankPro Microarray Kit, 1x	952694
Axiom UK BiobankPro Microarray Kit, 2x	952695
SwiftArrayStudio Microarray Analyzer	00-0471

References

1. Tan VY et al. (2022) The UK Biobank: A shining example of genome-wide association study science with the power to detect the murky complications of real-world epidemiology. Ann Rev Genom Hum Genet 23:569–589. doi.org/10.1146/annurev-genom-121321-093606
2. Bycroft C et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–209. doi.org/10.1038/s41586-018-0579-z
3. Hyams JS et al. (2021) Clinical and host biological factors predict colectomy risk in children newly diagnosed with ulcerative colitis. Inflamm Bowel Dis 28:151–160. doi.org/10.1093/ibd/izab061
4. Aung N et al. (2021) Genome-wide analysis of left ventricular image derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. Circulation 140:1318–1330. doi.org/10.1161/circulationaha.119.041161
5. Kochunov P et al. (2022) Brain-wide versus genome-wide vulnerability biomarkers for severe mental illness. Hum Brain Mapp 43:4970–4983. doi.org/10.1002/hbm.26056
6. Spendlove SJ et al. (2022) Polygenic risk scores of endo-phenotypes identify the effect of genetic background in congenital heart disease. Hum Genet Genom Adv 3:100112. doi.org/10.1016/j.xhgg.2022.100112
7. Bafligil C et al. (2022) Development and evaluation of polygenic risk scores for prediction of endometrial cancer risk in European women. Genet Med 24:1847–1856. doi.org/10.1016/j.gim.2022.05.014
8. Cooper GM, et al. (2011) A copy number variation morbidity map of developmental delay. Nature Genetics 43(9):838–846
9. ARVD/C Genetic Variants Database. www.arvcdatabase.info/.
10. Lambert JC, et al. (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics 45(12):1452–1458.
11. Hoffman TJ, et al. (2011) Design and coverage of high-throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ ethnicity using imputation and a novel hybrid SNP selection algorithm. Genomics 98(6):422–430.
12. Hoffmann TJ, et al. (2011) Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. Genomics 98(2):79–89
13. Howie BN, et al. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genetics 5(6):e1000529
14. McVean GA, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491(7422):56–65.
15. Grundberg, E, and the MuTHER consortium Mapping cis and trans regulatory effects across multiple tissues in twins Nature Genetics, 2012 Sep 2;44(10):1084-1089. doi: 10.1038/ng.2394. Epub 2012 Sep 2.
16. www.bristol.ac.uk/alspac/
17. Borel C et al. Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts. Genome Res. 2011 Jan;21(1):68-73. doi: 10.1101/gr.109371.110.
18. The Pharmacogenomics Knowledgebase. www.pharmgkb.org ARVD/C Genetic Variants Database.

Learn more at thermofisher.com/swiftarraystudio

For Research Use Only. Not for use in diagnostic procedures. © 2026 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. **DTAS-15316802 0726**