Brochure

The QIAGEN Knowledge Base

The most trusted information for research and clinical genomics

Sample to Insight

The most comprehensive Knowledge Base, founded on curation

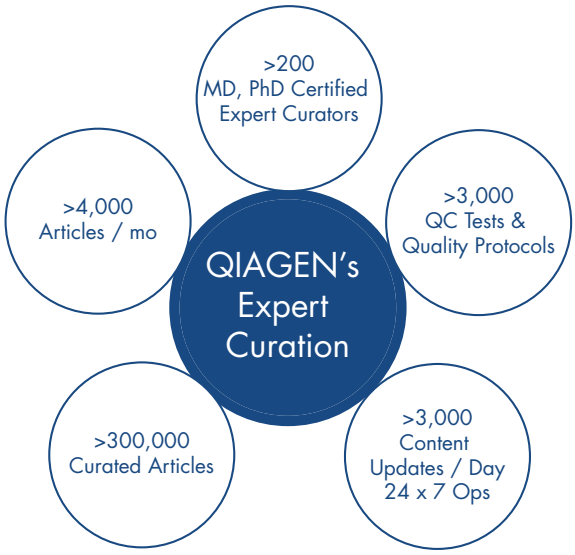

Figure 1. The industry-leading and most scalable capability for information sourcing, curation and management.

Advanced curation processes ensure complete and up-to-date biological and clinical relevance

QIAGEN Bioinformatics is the leading provider of genomic content knowledge, based on advanced curation methods to ensure relevance and accuracy of insights. The QIAGEN Knowledge Base is an industry-leading database composed of proprietary, open source, and licensed biological content that has been aggregated, integrated, and curated for clinical relevance. It has been used by researchers, clinicians, and pharmaceutical companies for more than 16 years and has been cited in more than 20,000 scientific publications.

QIAGEN Bioinformatics creates structured vocabulary systems called ontologies to define the data sources and types that are relevant for a particular biological, disease, or clinical use case. It has established and implemented novel curation processes for acquiring, processing, and integrating clinically relevant information according to the ontologies. Specifically, these curation processes employ both manual and computational methods to iteratively review data sources and compile findings under a rigorous quality management system.

Using a database founded on manual curation instead of computational prediction tools and raw public data reduces the risk of misclassifying rare or novel variants. Constructing such a database is time-consuming and not feasible for individual research or clinical laboratories, especially once multiple diseases need to be considered. By investing the time and resources into the QIAGEN Knowledge Base, we have made it possible for you to significantly reduce the time needed to interpret genomic variants and other data without compromising quality.

A Knowledge Base built with a manually curated ontology

The QIAGEN Knowledge Base is built on our own comprehensive ontology, which is manually curated. Using an ontology within the Knowledge Base helps to uniformly model relationships between different entities, such as the relationship between a variant, the gene that it resides in, and the observed phenotype. Our data interpretation products make use of the underlying ontology system and curated findings to identify new and existing relationships between identified variants, disease phenotypes, and biological processes. A dedicated team keeps the ontology up-to-date as new concepts are identified and described.

The content in our Knowledge Base meets the highest quality standards

QIAGEN Bioinformatics has unparalleled experience in content curation. We employ more than 200 people

worldwide to manually extract insights from 4,000-plus scientific journals (Table 1), ensuring comprehensive information coverage. All curators have an MD or PhD background with experience in both genetics and clinical research. To operate at such a large scale without compromising quality, we have developed sophisticated and streamlined curation processes and protocols. New curators receive detailed training and must complete a rigorous certification program that ensures consistency between individual curators and the highest-quality curation for our Knowledge Base (Figure 1).

Human Mutation	PNAS	BMC Cancer
American Journal of Human Genetics	Molecular Vision	Investigative Ophthalmology and Visual Science
Journal of Medical Genetics	Journal of Human Genetics	Atherosclerosis
American Journal of Medical Genetics	Journal of Inherited Metabolic Disease	Clinica Chimica Acta
Human Molecular Genetics	Journal of Clinical Investigation	Diabetes
PLoS One	BMC Medical Genetics	Familial Cancer
Journal of Clinical Endocrinology and Metabolism	Journal of Investigative Dermatology	Journal of the Neurological Sciences
Blood	British Journal of Cancer	Oncogene
Human Genetics	Breast Cancer Research and Treatment	Clinical Endocrinology
Nature Genetics	Neuromuscular Disorders	Brain
Clinical Genetics	Journal of Biological Chemistry	Cancer
European Journal of Human Genetics	British Journal of Haematology	Journal of Clinical Oncology
Molecular Genetics and Metabolism	Carcinogenesis	Biochemical and Biophysical Research Communications
Cancer Research	New England Journal of Medicine	British Journal of Dermatology
Neurology	Gene	
Hemoglobin	Cancer Epidemiology, Biomarkers and Prevention	
Clinical Cancer Research	Annals of Neurology	
International Journal of Cancer	Haematologica	

Table 1. Representative set of journal titles that are manually curated by the QIAGEN Bioinformatics internal curation team.

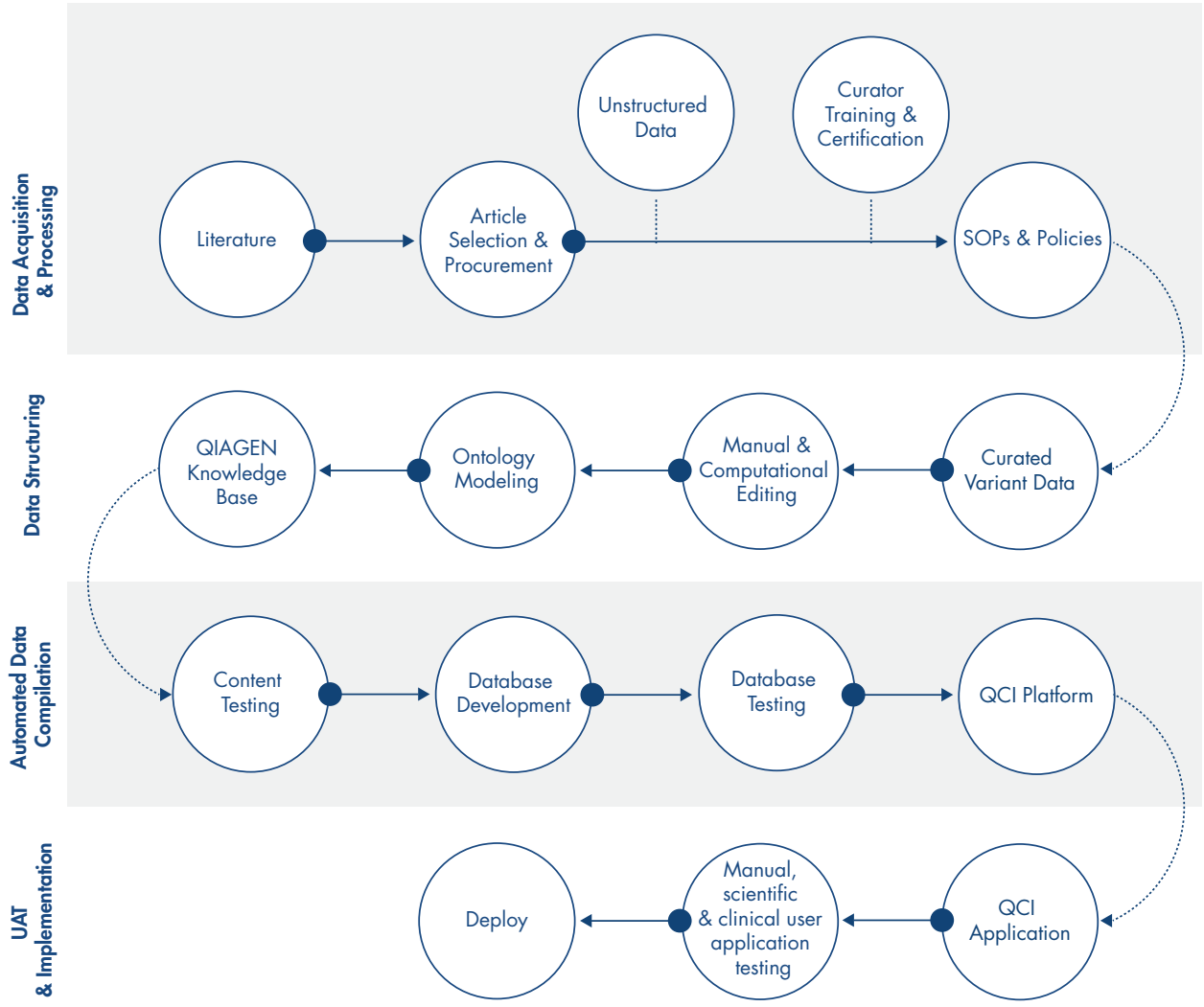

Figure 2. The QIAGEN Bioinformatics curation process.

The curation process includes multiple layers of manual and automated quality control to maintain accuracy. All curation is performed using custom-developed tools that work with the Knowledge Base ontology, ensuring consistent representation of phenotypes. All submitted curation undergoes an extensive review process, allowing curators to improve with feedback. In addition, automatic QC steps detect systematic errors during the curation process and after completion. Scientific checks made during the review process and through sample data analyses performed via our interpretation products round out the review process.

Our Knowledge Base is the most comprehensive database of biological data available

Our Knowledge Base (Figure 3.) includes more than 13 million biological findings, with thousands of findings added with each new release. These findings describe:

- Variants that relate to disease
- Pharmacogenomic variants
- Variants present in healthy populations
- Gene – phenotype relationships
- Gene – drug relationships
- Gene – pathway relationships
- Gene – gene interactions
- Gene – transcript information
- Clinical trial information

We have manually curated more than 5.3 million variants and include publicly available datasets such as 1000 Genomes, ExAC, and NHBLI's ESP databases. In addition, we have licensed proprietary databases like COSMIC and OMIM as well as HGMD and PGMD, two manually curated variant databases that we exclusively distribute (Table 2). To provide the highest quality standards, we use the principles from our manual curation efforts during integration of these databases. As a result, many inconsistencies in these datasets can be identified and eliminated. Other computationally extracted information is used from public databases such as ENCODE to ensure comprehensive variant content within regulatory regions.

Curated variant-disease associations are provided in the context of information about the prevalence of the variant in healthy versus affected individuals. As an example, to

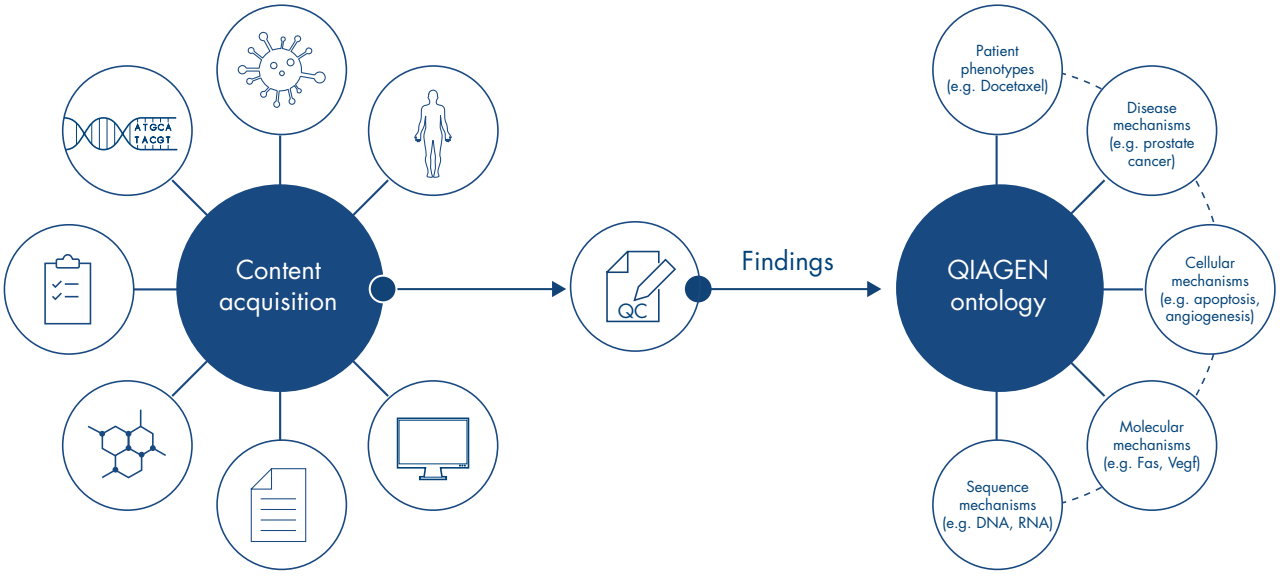

Figure 3. The QIAGEN Knowledge Base overview.

Database Source	Description	URL
QIAGEN Knowledge Base	The QIAGEN Knowledge Base (QKB), previously known as the Ingenuity Knowledge Base (IKB), contains over 13 million high-quality facts, or "findings", about the genomic basis of human disease, including 5 million findings specific to the role and impact of genetic variation on a braod range of clinical indications. The QIAGEN Knowledge Base is a publically available, commercially supported database exclusively available through QIAGEN products. The QKB is individually curated by professionals in the genetics field following strict protocols. This knowledge base contains curated data of all clinically known cases of any variants in genes included in our clinical gene list with biological significance alongside information such as pathway interactions and functional studies. The knowledge base is continually updated with new findings added to QIAGEN Clinical Insight (QCI™) on a weekly basis.	www.ingenuity.com/science/knowledge-base
HGMD	The Human Gene Mutation Database, represents an attempt to collate known (published) gene lesions responsible for human inherited disease and is exclusively licenced through QIAGEN.	www.hgmd.cf.ac.uk/ac/index.php
COSMIC	Catalog of Somatic Mutations in Cancer. COSMIC is designed to store and display somatic mutation information and related details and contains information relating to human cancers.	www.cancer.sanger.ac.uk/cosmic
ClinVar	Maintained by the NCBI. ClinVar aggregates information about genomic variation and its relationship to human health.	www.ncbi.nlm.nih.gov/clinvar
BIC	Breast Cancer Information Core, an open-access on-line breast cancer mutation database	www.research.nhgri.nih.gov/bic
OMIM	Online Mendelian Inheritance in Man is a continuously updated catalog of human genes and genetic disorders and traits, with particular focus on the molecular relationship between genetic variation and phenotypic expression. It is thus considered to be a phenotypic companion to the Human Genome Project.	www.ncbi.nlm.nih.gov/omim
1000 genomes	The goal of the 1000 Genomes Project is to find most genetic variants that have frequencies of at least 1% in the populations studied.	www.1000genomes.org
Exome Variant Server Project	The goal of the Exome Variant Server Project is to discover novel genes and mechanisms contributing to heart, lung and blood disorders by pioneering the application of next-generation sequencing of the protein coding regions of the human genome across diverse, richly-phenotyped populations and to share these datasets and findings with the scientific community to extend and enrich the diagnosis, management and treatment of heart, lung and blood disorders.	www.evs.gs.washington.edu/EVS
Allele Frequency Community	It is a freely accessible "opt-in" community resource designed to facilitate sharing of anonymized, pooled allele frequency statistics among laboratories for the benefit of patients and biomedical research. The community include many founder set including ExAc and CGI	www.allelefrequency-community.org
Jaspar	The JASPAR CORE database contains a curated, non-redundant set of profiles, derived from published collections of experimentally defined transcription factor binding sites for eukaryotes.	www.jaspardev.genereg.net
Encode	The National Human Genome Research Institute (NHGRI) launched a public research consortium named ENCODE, the Encyclopedia Of DNA Elements, to carry out a project to identify all functional elements in the human genome sequence.	www.genome.ucsc.edu/ENCODE
Vista	VISTA is a comprehensive suite of programs and databases for comparative analysis of genomic sequences.	www.genome.lbl.gov/vista/index.shtml
Predicted biochemical Impact	Polyphen, SIFT, PhyloP, Blosum and MaxEntScan	
Guidelines	NCCN, ASCO, ESMO	www.nccn.org www.asco.org www.esmo.org
Drug Labels	FDA and EMA	www.fda.gov/Drugs
Clinical trials	QIAGEN curates open/active genotype-specific trials in support of patient-to-trial matching. Matching is based on criteria including cancer/disease indication, mutation status, and geographic location.	www.clinicaltrials.gov

Table 2. Overview of biological, genomic and clinical content in QIAGEN Knowledge Base.

Disease Classes	>20,000
Ontology Components	>2.8m
Drugs	>8,000
Biological Findings	>13,000,000
Manually Curated Findings	>5,300,000
Databases (public, licensed, proprietary)	>30
Interprwtative Comments	>500
Signaling & Metabolic Pathways	>500
Proprietary Allele Frequency Data	>135,000 Genome

Table 3. Overview of biological, genomic and clinical content in QIAGEN Knowledge Base.

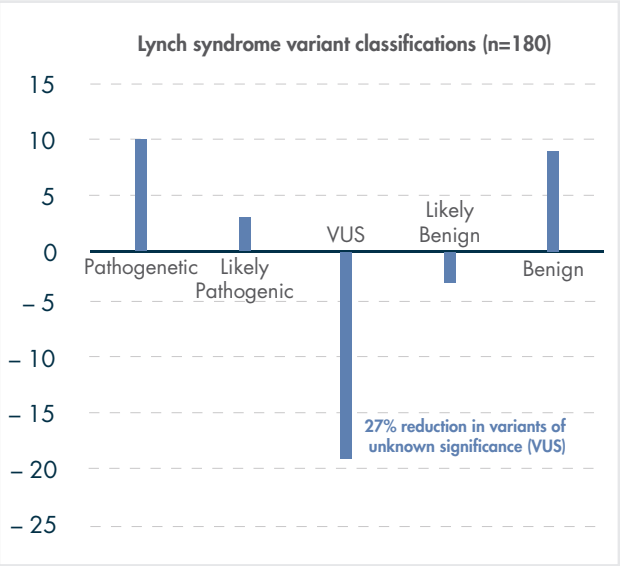

Figure 4. Variant classifications for Lynch Syndrome using QCI Interpret with and without QIAGEN curated primary literature. Data is represented as the difference between variant classifications with QIAGEN content minus variant classifications without QIAGEN content (i.e. public data sources only).

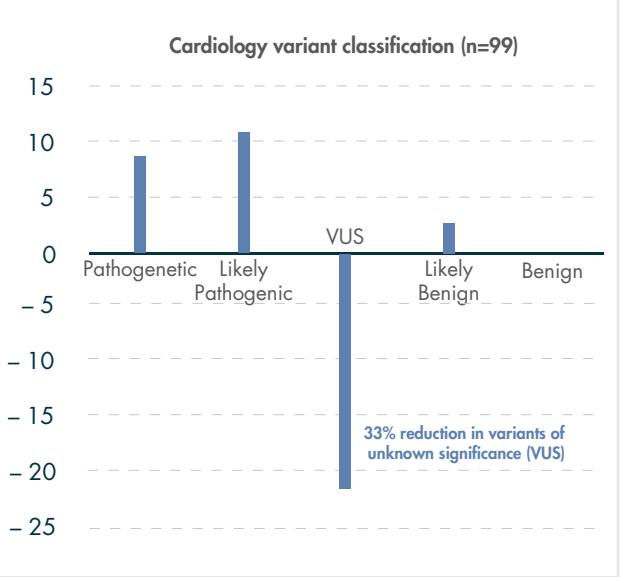

Figure 5. Variant classifications for cardiology disease using QCI Interpret with and without QIAGEN curated primary literature. Data is represented as the difference between variant classifications with QIAGEN content minus variant classifications without QIAGEN content (i.e. public data sources only).

support assessment of potential somatic cancer variants, the percentage of observed cases compiled from manual literature curation and from COSMIC is noted. More than 2.8 million ontology classes contained within the QIAGEN Knowledge Base (Table 3) provide further context by establishing relationships between variants, genes, tissue types, and pathways.

The QIAGEN Knowledge Base offers the most comprehensive reference coverage available (Table 3). Our variant bibliographies contain at least 90% of the relevant references found with Google Scholar and a significant number of relevant references that are missed by Google Scholar or manual search methods. We are able to ensure such comprehensive reference coverage by using algorithms that learn to identify useful articles from training sets of clinically relevant references.

Demonstrated value of the QIAGEN Knowledge Base - up to 33% reduction in VUS

Other clinical decision support products annotate variants using sources such as HGMD and ClinVar, but they lack curated content from the primary literature. This sticks users with the time-consuming task of searching articles and curating their own papers to fully classify variants. This is espe-

cially problematic for workflows that incorporate ACMG guidelines for variant interpretation, since these guidelines require information that public sources can't provide — examples include co-segregation data, de novo status, co-occurrence with other pathogenic variants, functional study data, and case-control study data.

To demonstrate the challenge of interpreting variants without using peer-reviewed literature, we classified 279 randomly selected variants associated with Lynch syndrome (n=180) (Figure 4) or Cardiology diseases (n=99) (Figure 5) using QCI Interpret with and without content from the curated primary literature. The number of variants classified as having unknown significance (VUS) based on ACMG guidelines was 27-33% lower in the datasets interpreted with primary literature than in the datasets relying only on public sources. That significantly increased the number of variants with clinically meaningful classifications (pathogenic, benign, likely pathogenic, and likely benign).

Conclusion

QIAGEN Bioinformatics provides the most comprehensive Knowledge Base available. It includes more than 13 million biological findings, millions of which have been curated by QIAGEN directly. All information — whether it's from content curation efforts or from publicly available or in-licensed databases — is structured into our own ontology and exhaustively reviewed. With this foundation, our data interpretation products allow users to identify new and existing connections between genotypes, phenotypes, diseases, drugs, and drug trials to pinpoint disease-causing mutations, new biomarkers, or the right treatment for the patient.

To learn more from a sales or support solution specialist, contact us using the information below:

QIAGEN Bioinformatics

EMEA
Silkeborgvej 2 · Prismet
8000 Aarhus C
Denmark
Phone: +45 8082 0167

qiagenbioinformatics.com

Americas
1700 Seaport Boulevard #3
Redwood City · CA 94063
USA
Phone (Toll Free): +1 866 464 3684

bioinformaticssales@qiagen.com