# Warehousing Clinical & Genomics Biomedical Data

## A Challenges & Needs Analysis Report

Table of Contents – May 2021

enlightenbio™ LLC
info@enlightenbio.com
650/ 204.1938
enlightenbio.com

# Table of Contents


CONTENTS 2

OBJECTIVES 3

LEARNINGS 4

TABLE OF CONTENTS 5

TABLE OF FIGURES 6

TABLE OF TABLES 6

THE CLINICAL DIAGNOSTICS & MOLECULAR PROFILING WORKFLOW 8

Clinical Genomics Workflow End Users/Personas 8

The Computational Scientist/Bioinformatician 8

The Data Scientist 9

The Non-Computational Collaborator 9

The Genetic Counselor 9

The Clinical Geneticist 10

The Physician 10

The Patient 10

Mapping End Users to the Clinical Genomics Workflow 14

Clinical Information Delivery 14

MEDICAL INDUSTRY NEEDS & CHALLENGES 17

Clinical Genomics - Unmet Needs and Challenges 19

Technical Challenges Associated with Scaling Clinical Genomics Applications 19

Scientific Challenges Associated with the Implementation of Clinical Genomics 23

Non-Technical / Scientific Challenges Associated with the Implementation of Clinical Genomics Applications 27

BIOMEDICAL – CLINICAL & GENOMIC - DATA WAREHOUSING 28

Rethinking Data Platforms to Become Data-Driven 28

Structures of Successful Big Data Platforms 29

Scalable Data Generation and Data Flow Lack Data Standards 29

Slow Integration of Clinical Genomics Data with Other Clinical/Patient Data 30

Current EMR/EHR Systems Do not Support Genomics Data 32

Variant Data Warehousing for Data Analysis, Mining, and Querying 33

Enterprise Data Platform Architecture 34

Integration of Various Types of Data 35

A Data Warehouse That Supports Clinical Research 37

A Scalable Infrastructure is required for Data Warehousing and Fast Data Processing 37

The Ideal Variant Data Analysis and Query Platform 38

Data Warehouse versus a Data Lake 40

Cloud: The Preferred Infrastructure 42

ABBREVIATIONS 44

ABOUT ENLIGHTENBIO LLC 47

# The Warehousing Clinical & Genomics Biomedical Data – A Challenges & Needs Analysis Report

Medical organizations understand the benefit of being empowered by data-driven approaches to reduce operational costs and time and to provide researchers and clinicians the necessary tools to decipher critical research data and data for clinical-decision making. Innovative technological advancements have allowed us to sequence and uncover mutational events at an unprecedented scale, while facilitating linking genomic data to high quality clinical data and diagnosis. ***Implementation of clinical genomics applications and enterprise-wide clinical data warehouses are fundamental for successful genomic medicine programs***. As such, leading medical organizations have established precision medicine programs that support personalized patient treatment. However, technical and scientific limitations still need to be addressed for optimized and universal use of various data sources for both clinical and research purposes.

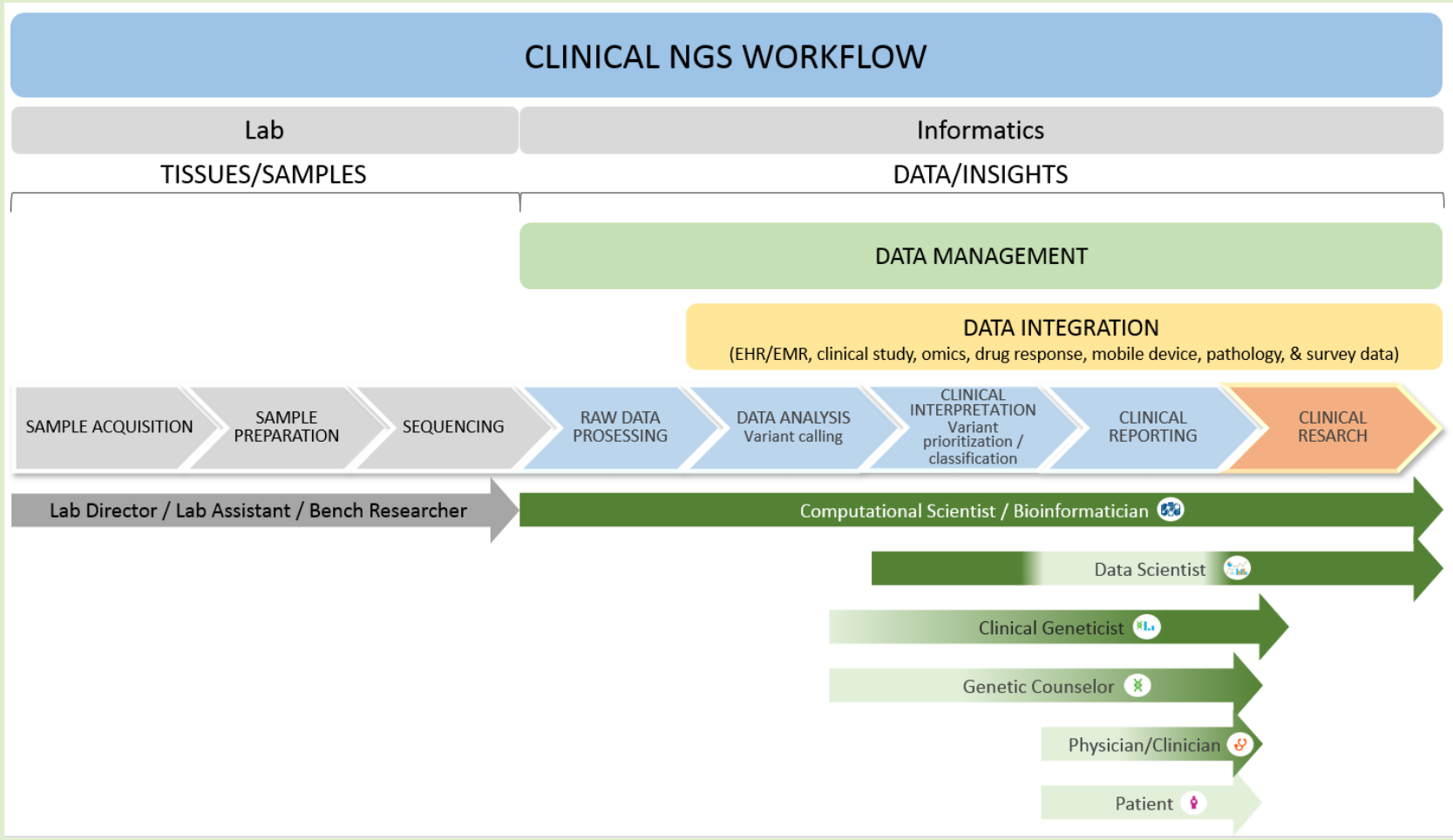
While data production is no longer a challenge, and targeted panels and whole exome sequencing are well adopted, the dramatic rise in whole genome sequencing will result in unforeseeable quantities of data at the clinical level that need to be managed, understood, and communicated. Essential to overcoming these challenges are scalable, fully automated analysis and knowledge extraction solutions incorporating rich annotation information. With massive quantities of NGS data (linked to clinical and other types of data), artificial intelligence and machine learning are hailed as pivotal solutions to address the data interpretation and knowledge extraction challenges and to advance the clinical application of genomics. Despite increasing efforts and investments in implementing clinical applications and building data solutions, ***many organizations are still challenged with the multi-faceted complexities in transforming to become data-driven***. Implementations are hindered by ineffective data sharing, scalability and automation issues, non-optimized data generation and data flow approaches, and non-standardized data from numerous sources. ***Implementing a complex clinical data warehouse*** presents many challenges starting with the various data sources it needs to support and the tools required to view the clinical information.

***Clinical interviews detailed challenges associated with creating a workflow that incorporates a clinical data warehouse connecting clinical research with clinical diagnostics and vice versa.*** Efficient clinical decision-making and reporting findings between clinical research and the clinic to optimize clinical outcome and patient treatment stem from this important workflow.

The 47-page ***Warehousing Clinical & Genomics Biomedical Data - A Challenges & Needs Analysis Report*** contains ***8 Figures and 17 Tables***. An extension of this report is the full "Clinical Genomics Report – The Interplay Between Clinical Research & Clinical Diagnostics" which consists of 13 Figures, 95 Tables, 14 comprehensive Medical Organization Profiles, and 20 comprehensive Company Profiles.

For more information contact info@enlightenbio.com or visit our website, enlightenbio.com.

# The Clinical Genomics Process



# Objectives & Approach

Our critical, investigative, and qualitative "Warehousing Clinical & Genomics Biomedical Data – A Challenges & Needs Analysis Report" analyzes the observations and learnings across the complex components of biomedical data warehousing, querying, and collaboration to further clinical applications via population studies, data querying, and validation of research findings, and benefits commercial entities developing solutions to address the ever growing biomedical data needs of this expanding industry.

Twenty-one (21) end user interviews from eighteen (18) medical research organizations across the dry lab/informatics portion of the Clinical Genomics Workflow provided insights into medical industry and end users' unmet biomedical (clinical & genomics) data needs and challenges.

To create our robust analysis, our research focused on these questions:

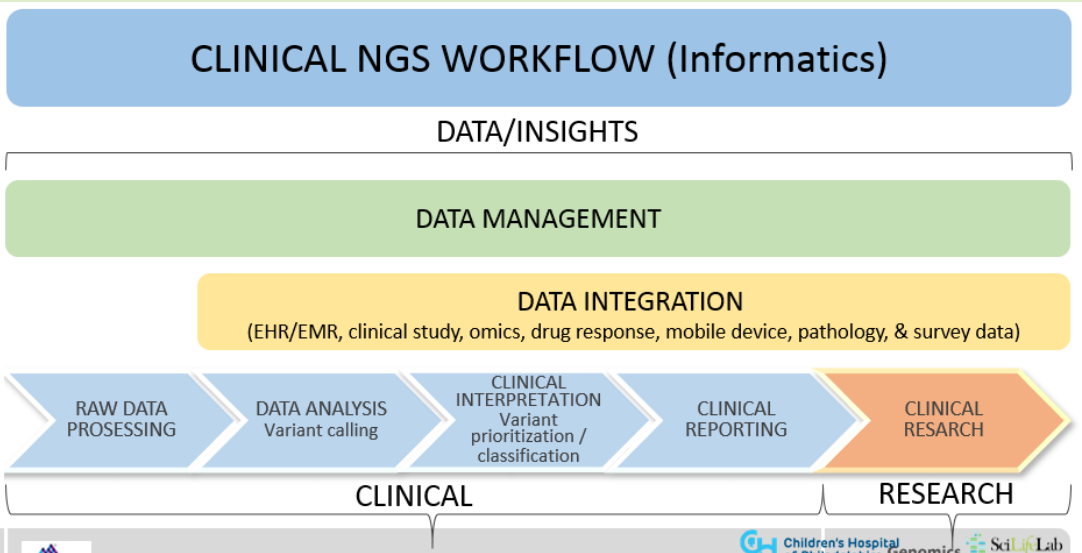
- What are the unmet needs and challenges of medical organizations/clinical end users in relation to biomedical (clinical and genomics) data implementations; and
- What innovations of emerging technologies are required for successful adoption and acceleration of clinical genomics data warehouse-based applications?

1) ***End user interviews****:* Conducted to understand medical industry and clinical end users' needs and challenges, cloud solution preferences, and challenges with clinical data and integrating and communicating findings via an electronic healthcare system with the physician and the patient.
2) ***Deep level analysis:***
    a. Researched optimal genomic data generation, data flow, and intelligence data platform requirements that support the interplay between clinical research and clinical genomics.

# Twenty-one (21) Clinical End-User Interviews

Twenty-one (21) individuals from across eighteen (18) medical and clinical research organizations, including two individuals from academic organizations who provide via their role and responsibility internal/external clinical research, were interviewed.

Clinical interviews detailed challenges associated with creating a workflow that incorporates a clinical data warehouse connecting clinical research with clinical diagnostics and vice versa. This important workflow leads to efficient clinical decision-making and reporting findings between clinical research and the clinic, which can optimize clinical outcome and patient treatment.




| | | |
|---|---|---|
| Data scientist | Icahn School of Medicine at Mount Sinai; sema4; CHEO | Children's Hospital of Philadelphia; Genomics England; SciLifeLab; Nationwide Children's; Geisinger; Baylor College of Medicine; Stanford Health Care; University of Colorado Anschutz Medical Campus; Stanford Medicine |
| Clinical geneticist<br>Genetic counselor | | |
| Physician / clinician | School of Medicine Vanderbilt University; St. Jude Children's Research Hospital; Kaiser Permanente | Kaiser Permanente |
| Strategist / task force member | Providence St. Joseph Health; UCSF; Kaiser Permanente | Cincinnati Children's; Johns Hopkins School of Medicine |

## Interviews Revealed Medical Industry Unmet Needs & Challenges

The needs and challenges analysis conducted with various end users across the dry lab/informatics portion of the Clinical Genomics Workflow uncovered answers to the questions below.

| | Focus |
|---|---|
| *Challenges* | What are the challenges with scaling genomics applications in the clinical setting? |
| *Challenges* | What are the challenges with selecting a test, testing lab, and when clinical findings are returned to the test requestor (clinical geneticists, genetic counselors, and patients)? |
| *Needs* | What are current critical needs to scale clinical genomics data analysis and interpretation for successful and efficient diagnosis? |
| *Needs* | How is clinical data (ideally) reported to physicians, patients, and health care organizations? |
| *Needs* | What is required for successful integration of clinical research with clinical diagnostics? |
| *Outlook* | What innovations of emerging technologies are required for successful adoption and acceleration of clinical genomics applications? |

The challenges and needs identified can be divided into two major categories, the technical/infrastructure challenges and the scientific data challenges, which are discussed in detail within this report.

## Key Data from Secondary Sources

| | |
|---|---|
| Clinical genomics workflow end user challenges, & unmet needs | • End user interviews<br>• Public reports<br>• Peer reviewed publications<br>• World Wide Web |
| Data warehouse research & unmet needs | • End user interviews<br>• Peer reviewed publications<br>• Peer reviewed publications<br>• World Wide Web |

## About enlightenbio LLC

enlightenbio was founded in 2013 in the San Francisco Bay Area to provide a conduit between research and related technical and analytical resources. Our company consists of PhD level research scientists who bring decades of industry experience and expertise in the biotechnology, molecular diagnostics, pharma, and life science research markets. We are dedicated to communicating in the researcher's language, identifying unmet needs, and understanding product development. Our goals are aligned to researchers' needs to increase experiment productivity and to make sense of the resulting biological data.

In addition to our varied industry experiences - Applied Biosystems (now Thermo Fisher Scientific), DNAnexus. Iconix Biosciences, Incyte, Ingenuity Systems (now a Qiagen company), and Pfizer – we have built and maintained content curation services, defined product strategy, managed tactical product projects, performed extensive ecosystem analyses, and defined go-to-market plans.

Building on our initial success and previous experiences - microarray and next-generation sequence data analysis, toxicogenomics, solutions for sequence data management, analysis, and interpretation, drug discovery, and biochemistry - we continuously monitor worldwide market trends in healthcare information technology, life sciences, genomics, clinical diagnostics, and the medical devices space to expand our critical service offerings. Combined with our extensive global network, we can identify market pain points and unmet needs, perform detailed market and product research, and undertake horizontal and vertical ecosystem or competitive analyses, and more.

This background with our future-focus makes us a resourceful and agile alternative to traditional market research companies. Our comprehensive knowledge of the market in which we live and breathe is invaluable to our partnerships and potential.

enlightenbio, along with market research reports is managed by Brigitte Ganter, PhD, Founder & Managing Director of enlightenbio LLC.