



NGS Data Analysis & Interpretation Ecosystem Analysis Report

Table of Contents – August 2016

Contents	2
Objectives	3
Learnings	4
Table of Figures	7
Table of Tables	7
The NGS Process	9
User Type	9
User Requirements	11
NGS Target Users – Unmet Needs and Challenges	12
The NGS Analysis and Interpretation Ecosystem	17
End-User Feedback on Secondary Analysis	18
End-User Feedback on Tertiary Analysis	20
NGS Utilization and Analysis Challenges	22
Secondary Data Analysis and Tertiary Interpretation Tool Companies Side-by-Side	26
Analysis and Interpretation Tool Companies – A Comparative Functionality Assessment	28
Data analysis speed, scalability, and flexibility	28
Infrastructure requirements	29
Data security and compliance features	32
Sequencing machine integration	36
Tertiary Data Interpretation Companies	37
Tertiary interpretation content	37
Partnerships	39
Overview of partnerships in the analysis and interpretation space	49
Company momentum	53
The funding situation	56
Extracted Findings	57
Concluding Notes	58
Data Explosion	58
Training Non-Computational Scientists	59
Information Management	60
Knowledge Extraction and Actionable Reporting	61
End-to-end Solutions	61
Reporting of Clinical Findings	62
Company/Product Profiles	63
Appendix	123
Companies in the Data analysis and Interpretation space	123
Abbreviations	126
About enlightenbio LLC	127

The NGS Process, its users, and unmet needs

Sequence data, analysis, interpretation, and reporting ecosystem

Concluding notes with learnings

Company profiles - funding, # of employers, product details, company vision, and more

NGS Data Analysis & Interpretation Ecosystem Analysis – August 2016

The advancements in high-throughput sequencing demand powerful, yet user-friendly bioinformatics solutions to analyze and extract meaningful knowledge from the generated sequence data. In line with this demand, the sequence data analysis and interpretation space is rich with commercial bioinformatics and/or platform solutions that provide different analysis, interpretation, and reporting solutions, and is paralleled by the ever-growing open source tools space. Commercial solutions provide a multitude of bioinformatics offerings as user-friendly stand-alone packages. They each differ in the capabilities offered, the user-friendliness they provide, the product(s) cost, and the target audience. This creates a challenging environment for the end-users faced with selecting the right offering for their specific analysis needs. Accordingly, the commercial entities need to clearly communicate their competitive advantage to customers to be among the winners in this space.

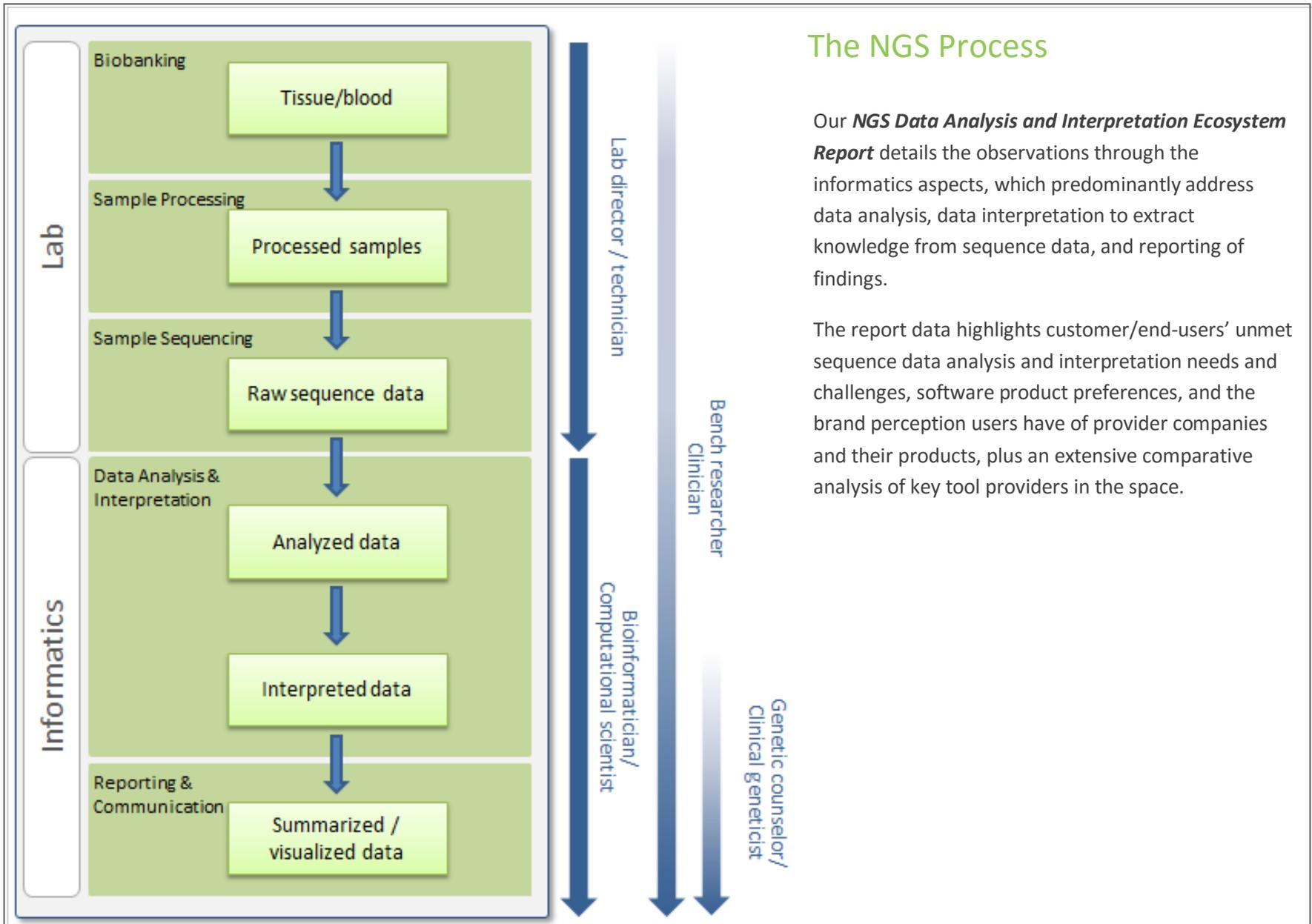
This comparative ecosystem analysis of key commercial tool providers reveals a range of capabilities that uniquely address both secondary and tertiary sequence data analysis needs. At first glance, the value propositions and communication of capabilities seem overlapping or similar. However, a deeper analysis clearly uncovers substantial differences in the product characteristics related to analysis and interpretation, data management, sharing and collaboration, reporting of findings, compliance and security mechanisms, and type of content made available for meaningful knowledge extraction.

A set of 25 commercial companies was analyzed revealing top players who are strong in different areas of data analysis and interpretation: DNAnexus and Seven Bridges Genomics have emerged as open and flexible platforms for data analysis, while Bluebee, Genallice, and Edico Genome, as alternatives, enable fast DNA processing. On the data interpretation side, notable players are Cartagenia (Agilent), Congenica, Ingenuity (Qiagen), Omicia, Station X, Tute Genomics, and WuXi NextCODE (WuXi AppTec). Content is clearly king on the interpretation side, and tertiary companies with rich content are highly valued: it is therefore not surprising to see Ingenuity with its manually curated content receiving high marks, or Omicia with the additional phenotype content being highlighted as value providers. End-user interviewees repeatedly pointed out the challenges associated with the complexity of data analysis which favors complete and integrated solutions that include secondary, tertiary, and reporting capabilities. This is particularly relevant for organizations that have a range of users at different computer sophistication levels. Tute Genomics is a company that was mentioned favorably supporting the entire workflow from data analysis to interpretation, and findings reporting.

While this report does not intend to provide direct recommendations to end-users for specific product selection, the deep-dive analysis is meant to be an insightful review to help researchers and commercial entities choose the best partner for success.

The 127 page Ecosystem Analysis Report consists of 13 Figures, 45 Tables, and 30 comprehensive Company Profiles which includes company metrics, funding details, product details, founder/executive and board information, additional notes, and respective company visions.

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



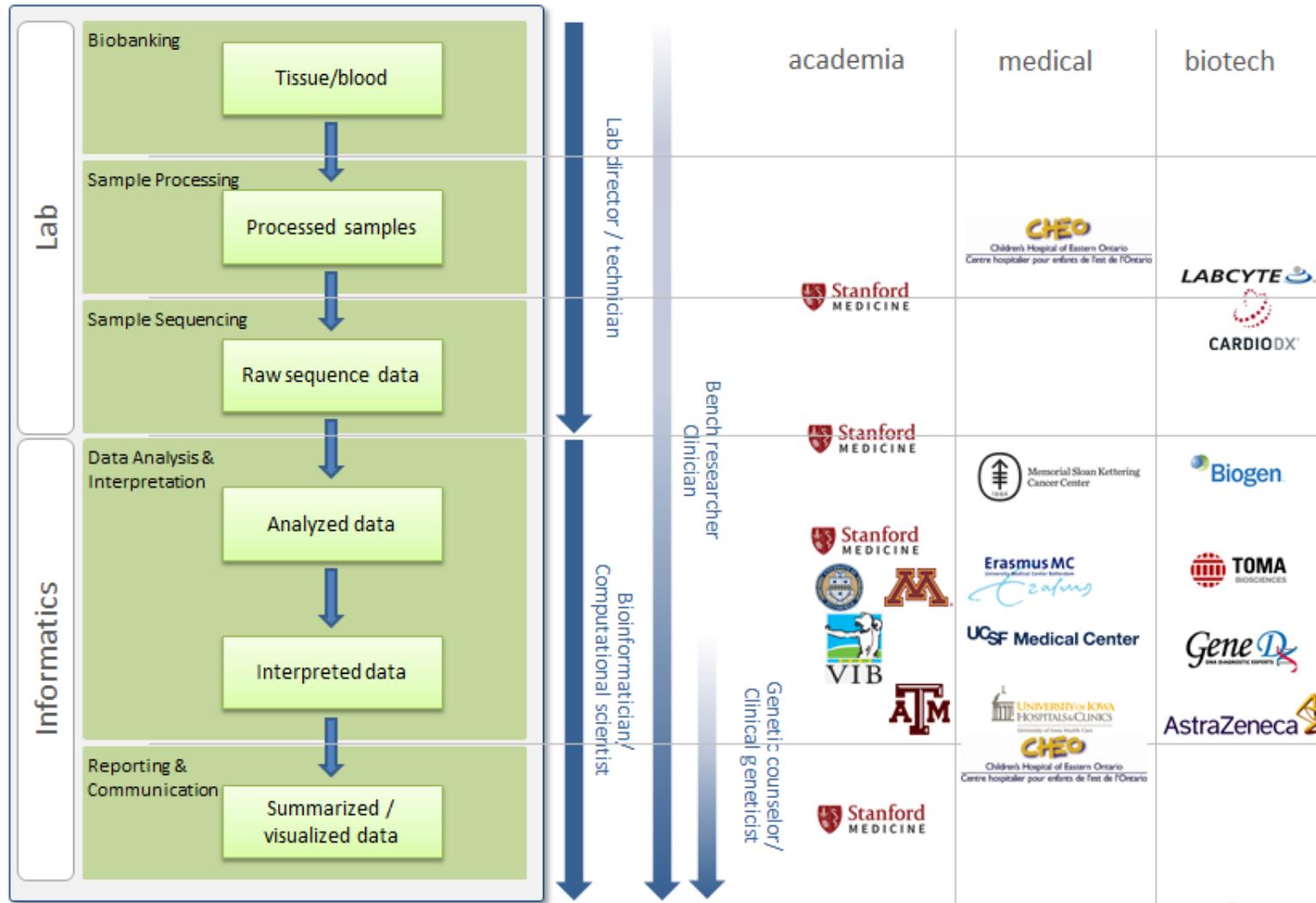
Objectives of the Report

Our NGS Data Analysis and Interpretation Ecosystem Report details the observations through the informatics aspects, which predominantly address data analysis, data interpretation to extract knowledge from sequence data, and reporting of findings. The report data highlights customer/end-users' unmet sequence data analysis and interpretation needs and challenges, software product preferences, and the brand perception users have of provider companies and their products, plus an extensive comparative analysis of key tool providers in the space.

To create a robust comparison of the companies in the data analysis and interpretation space, we conducted our research following an investigative path using the elements outlined below; elements that are rooted in the following questions: who are the key commercial companies in this space; who are the end-users and what are their unmet needs, pains, and challenges; when researching the analysis and interpretation side of the equation, what challenges do these companies need to address, and are those challenges being properly addressed?

- 1) End-user interviews seeking to understand user-analysis needs, product preferences, and challenges encountered when analyzing sequence data.
- 2) Analysis of meta-data available via the World Wide Web to perform a deep dive interrogation of individual tool providers.
- 3) A deep ecosystem level analysis of a major set of key commercial software and platform providers (N=25) in the secondary and tertiary NGS analysis space, to understand product focus, strategy to address end-user needs, and company momentum (incl. funding situation, number of employees, and more):
 - a. Advaita, Appistry, BaseSpace (Illumina), Bina Technologies (Roche), Bluebee, Cartagenia (Agilent), CLC Bio (Qiagen), Congenica, DNAnexus, DNASTAR, Edico Genome, Genalice, Genestack, Genomatix, Genospace, Golden Helix, Ingenuity Systems (Qiagen), Partek, Omicia, Seven Bridges Genomics, SoftGenetics, SolveBio, Station X, Tute Genomics, and WuXi NextCODE (WuXi AppTec).
- 4) Company profiles for the key players offering comprehensive solutions across the entire NGS process (N=29) which includes Agilent Technologies, Biomatters, Qiagen, and Thermo Fisher Scientific, besides the aforementioned companies.
- 5) Input from key representatives for established commercial software product suppliers to learn about their vision for product solutions in the current - and future - space.

End-User Interviews



19 end-user interviews conducted to elucidate processing needs and challenges, and research product preferences.

The NGS Analysis and Interpretation Ecosystem – A Deep Dive Analysis

The advancements in high-throughput sequencing demand powerful, yet user-friendly bioinformatics solutions to analyze and extract meaningful knowledge from the generated sequence data. In line with this demand, the sequence data analysis and interpretation space is rich with commercial bioinformatics and/or platform solutions that provide different analysis, interpretation, and reporting solutions, and is paralleled by the ever-growing open source tools space. Commercial solutions provide a multitude of bioinformatics offerings as user-friendly stand-alone packages. They each differ in the capabilities offered, the user-friendliness they provide, the product(s) cost, and the target audience. This creates a challenging environment for the end-users faced with selecting the right offering for their specific analysis needs. Accordingly, the commercial entities need to clearly communicate their competitive advantage to customers and end-users to be among the winners in this space.

This ecosystem comparative analysis focuses on 25 key players with a high company momentum. The companies selected have high funding power, a well-received customer perception, are strategically well connected with relevant partner companies, or enjoy a general high visibility in the sector. They either provide a solution for secondary data analysis (2^o), tertiary data interpretation (3^o), or reporting of the findings. Some of these companies offer a focused capability in one sector, whereas others attempt to address the entire analysis and interpretation section in a true end-to-end solution.



30 Company/Product Profiles

- Advaita
- Agilent Technologies
- Appistry
- BaseSpace (Illumina)
- Bina Technologies (Roche)
- Biomatters
- Bluebee
- Cartagenia
- CLC Bio (Qiagen)
- Congenica
- DNAnexus
- DNASTAR
- Edico Genome
- Genalice
- Genestack
- Genomatix
- Genospace
- Golden Helix
- Ingenuity Systems - IPA & IVA (Qiagen)
- Omicia
- Pacific Biosciences
- Partek
- Qiagen
- Seven Bridges Genomics
- SoftGenetics
- SolveBio
- Station X
- Thermo Fisher Scientific
- Tute Genomics
- WuXi NextCODE

Company/product profiles of sequencing instrument and data analysis and interpretation solution providers that supply or provide data analysis and interpretation software (and hardware) included in the deep dive ecosystem analysis.

Company profiles highlight company metrics (funding, number of employees, etc.), product details, founder/executive and board information, additional notes, and the respective company vision.

Key Data from Secondary Sources

User challenges Unmet needs	<ul style="list-style-type: none">• User interviews
Product details	<ul style="list-style-type: none">• User interviews• Company websites• Press releases• Product evaluations• Company representatives
Company details	<ul style="list-style-type: none">• Company websites• Annual reports• Press releases• Public databases• World Wide Web

Companies Mentioned in Report

10x Genomics, Advaita, Affymetrix, Agilent Technologies, Amazon, Appistry, BaseSpace (Illumina), Bina Technologies, Biomatters, Bluebee, Cartagenia, CLC Bio, Congenica, Core Informatics, DNAnexus, DNASTAR, Enlis Genomics, Edico Genome, Epicenter, Fluidigm, Genia (Roche Sequencing System), Genalice, GeneData, Genestack, GenomOncology, GenoLogic (Illumina), GenoSpace, Golden Helix, Google, Illumina, IBM, IMS Health, Ingenuity Systems, Lab7 Systems, Microsoft, Maverix Biomics, Nanostring, N-of-One, Omicia, Oxford Nanopore, Pacific Biosciences, Partek, Perkin Elmer, Qiagen, Raindance, Roche, Sapio Sciences, Seven Bridges Genomics, SolveBio, SoftGenetics, Spiral Genetics, Station X, Thermo Fisher Scientific, Tute Genomics, WuXi NextCODE Genomics.

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 to the table. We are dedicated to communicating in the researcher's language, identifying unmet needs, and understanding product development. Our goals are aligned with 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), Iconix Biosciences, Ingenuity Systems (now a Qiagen company), and DNAnexus – 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 in the medical devices space to expand our critical service offerings. Combined with our extensive global network, we can identify target market pain points and unmet needs, perform detailed market and product research, perform horizontal and vertical ecosystem or competitive analyses, and more.

This background with our future-focused outlook makes us a resourceful and exciting alternative to traditional market research companies. Our comprehensive knowledge of the market we live and breathe is invaluable to our partnerships and to the potential for our customers.

enlightenbio and associated market research reports are managed by Brigitte Ganter, PhD, Founder & Managing Director of enlightenbio LLC.