



# Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics

## Abstract:

Clinical genetics plays a critical role in the healthcare system by providing conclusive diagnoses for a wide range of rare syndromes. It can also help patients choose the best care/treatment choices by influencing genetics prevention, disease prognosis, and helping in the selection of the best options for care/treatment. Next-generation sequencing (NGS) has transformed clinical genetics, allowing researchers to examine hundreds of genes at once. As compared to traditional Sanger sequencing, this method achieves unparalleled speed and cost savings. Despite the increasing literature on NGS in clinical settings, the goal of this review is to bridge the gap between (bio)informaticians, molecular geneticists, and clinicians by providing a broad overview of NGS technology and workflow. First, we'll take a look at the existing NGS platforms. The NGS analytical bioinformatic pipelines are then dissected, with a focus on the algorithms typically used to generate and analyse sequence variants. Finally, the main challenges around NGS bioinformatics will be put into context for future improvements. Even with the enormous advances in NGS technology and bioinformatics, more bioinformatic algorithm improvements are needed to deal with complex and genetically diverse diseases.

# Introduction:

Next-generation sequencing (NGS) is a technology for determining the sequence of DNA or RNA to study genetic variation associated with diseases or other biological phenomena. Introduced for commercial use in 2005, this method was initially called massively-parallel sequencing, because it enabled the sequencing of many DNA strands at the same time, instead of one at a time as with traditional Sanger sequencing by capillary electrophoresis (CE).

Each of these technologies has utility in today's genetic analysis environment. Sanger sequencing is best for analyzing small numbers of gene targets and samples and can be accomplished in a single day. It is also considered the gold-standard sequencing technology, so NGS results are often verified using Sanger sequencing. NGS enables the interrogation of hundreds to thousands of genes at one time in multiple samples, as well as discovery and analysis of different types of genomic features in a single sequencing run, from single nucleotide variants (SNVs), to copy number and structural variants, and even RNA fusions. NGS provides the ideal throughput per run, and studies can be performed quickly and cost-effectively. Additional advantages of NGS include lower sample input requirements, higher accuracy, and ability to detect variants at lower allele frequencies than with Sanger sequencing.

The speed, throughput, and accuracy of NGS has revolutionized genetic analysis and enabled new applications in genomic and clinical research, reproductive health, and environmental, agricultural, and forensic science, so we will focus on :

1- NGS Library

2-NGS Platforms : Second-Generation Sequencing Platforms, Third-Generation Sequencers.

3-NGS Bioinformatics : Primary Analysis ,Quality Control( Read Filtering and Trimming ),

Secondary Analysis (Sequence Alignment, Post-Alignment Processing ,Variant Calling ),

Tertiary Analysis(Variant Annotation, Variant Filtering, Prioritization and Visualization)

4- NGS Pitfalls

# Discussion:

Genetics is currently extremely important in medical practise because it allows for a conclusive diagnosis of a wide range of clinically heterogeneous diseases. As a result, it allows for a more precise disease prognosis and offers assistance in selecting the best available treatment choices for affected patients. Most of its current promise stems from its ability to probe human genome at various stages, from chromosomal to single-base changes.

so, We will discuss a few topics that show us the importance of addressing genetics issues and the interest in developing the next generation sequencing technology that has greatly contributed to the human genome project that exists in our time.

**NGS library :** A library is a series of DNA/RNA fragments that represents the entire genome/transcriptome or a target region in next-generation sequencing. Each NGS platform has its own quirks, but in general, the preparation of an NGS library begins with fragmentation of the starting content, followed by sequencing. Fragments are linked to adaptors to enable enrichment of those fragments. The fragmentation of nucleic acid is the first step in most NGS workflows to prepare libraries. Physical or enzymatic methods may be used to fragment materials. Adaptors are bound to the fragments of the starting DNA after it has been fragmented. The adaptors are used to give random sequences known beginnings and endings, enabling the sequencing process to proceed. A new technique was developed that combines fragmentation and adaptor ligation in a single step, making the process easier, quicker, and with less sample input. Tagmentation is a transposon-based technology that is used in this process.

The fragments are chosen based on the desired library size after nucleic acid fragmentation.

**NGS Platforms:** Illumina and Ion Torrent are two of the most well-known second-generation sequencing firms. Illumina is a well-known American company that sells a variety of integrated systems for the study of genetic variation and biological function in a variety of biological systems ranging from agriculture to medicine. The Illumina sequencing method is based on the sequencing-by-synthesis (SBS) concept.

**NGS Bioinformatics and analysis:** To handle, analyze, and interpret the massive

amount of NGS data, computational and bioinformatics skills are needed. As a result, significant progress in NGS (bio)informatics is being made, which can only be accelerated by growing computational capabilities (hardware) and algorithms and applications (software) to help with all of the necessary steps: from raw data collection to in-depth data analysis and variant interpretation in a clinical setting NGS bioinformatics is usually divided into three categories: principal, secondary, and tertiary research. Regardless of the NGS platform, the end objective of each review is essentially the same; nevertheless, each platform has its own unique characteristics.

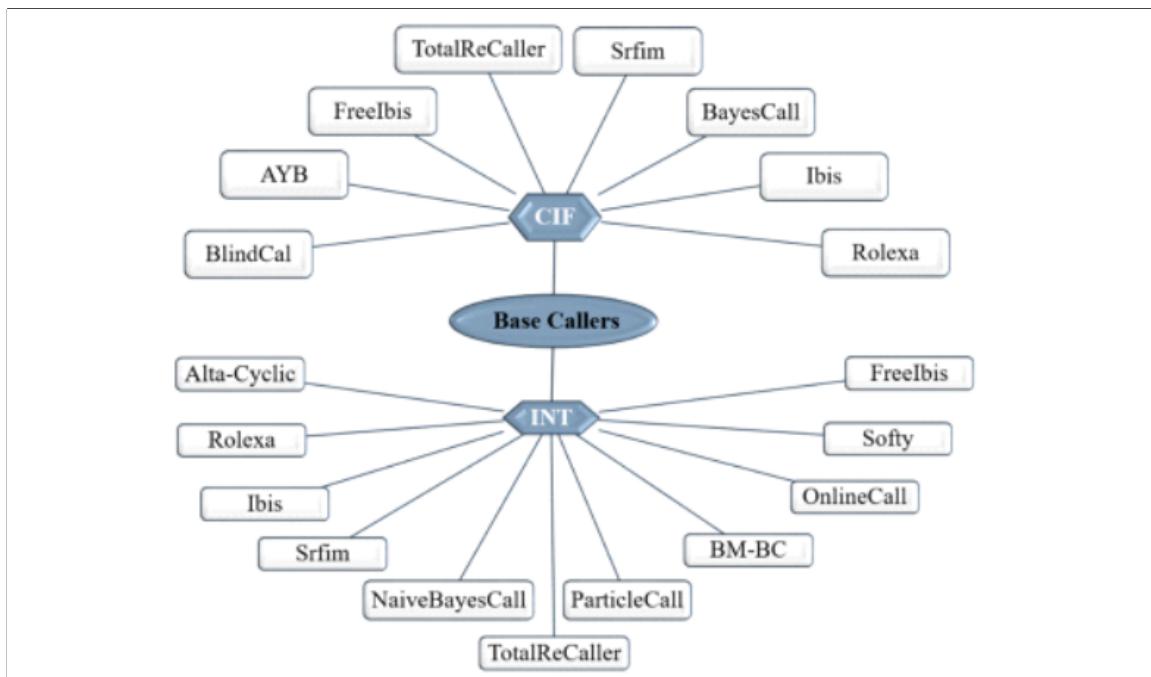
**Primary Analysis**, Quality Control (Read Filtering and Trimming),

**Secondary Analysis** (Sequence Alignment, Post-Alignment Processing, Variant Calling),

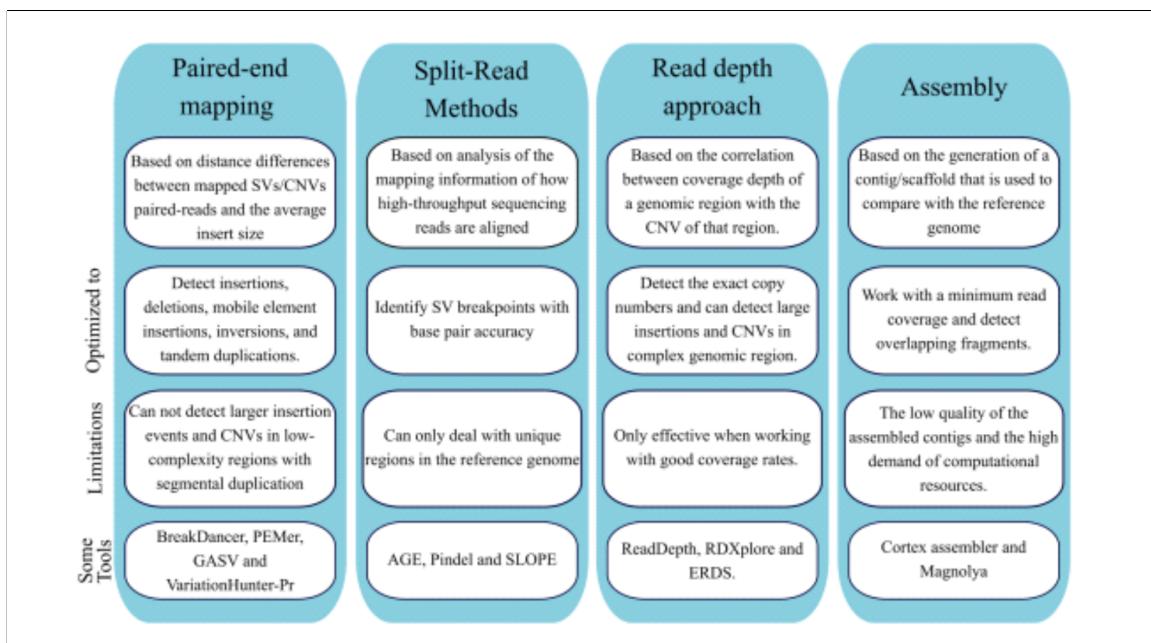
**Tertiary Analysis** (Variant Annotation, Variant Filtering, Prioritization and Visualization)

**NGS Pitfall:** It is undeniable that NGS provided a plethora of benefits and solutions for medicine as well as other fields such as agriculture, all of which helped to improve quality and productivity. It has, however, brought with it new challenges. The first issue is with the costs of sequencing. Although it is true that the total costs of NGS compared to the gene-by-gene sequence of Sanger sequencing, NGS is more expensive and not yet available to all laboratories. It has high upfront costs for the sequencing machine, which can range from thousands to hundreds of thousands of euros depending on the model depending on the computer, as well as consumables and reagents. Expenses of experimental design. It's also essential to consider sample selection and sequencing library preparation. Furthermore, costs associated with the construction of sequencing pipelines and bioinformatical methods are often underestimated. The costs of data management, informatics equipment, and downstream data analysis, as well as the costs of improving such pipelines and performing downstream sequence analysis, are not factored in overall costs of NGS.

# methods:



**Figure1:** Summary of some widely used base callers software available for the Illumina platform.



**Figure2:** Summary of the main methods for calling structural variants (SV) and copy number variation (CNV) from next generation sequencing (NGS).

**Figure3-Table 1:** List with examples of widely used tools to perform an NGS functional filter.

Software	Short Description	Ref.
PhyloP Phylogenetic p-values	Based on a model of neutral evolution, the patterns of conservation (positive scores)/acceleration (negative scores) are analyzed for various annotation classes and clades of interest.	[146]
SIFT Sorting Intolerant from Tolerant	Predicts based on sequence homology, if an AA substitution will affect protein function and potentially alter the phenotype. Scores less than 0.05 indicating a variant as deleterious.	[112]
PolyPhen-2 Polymorphism Phenotyping v2	Predicts the functional impact of an AA replacement from its individual features using a naive Bayes classifier. Includes two tools HumDiv (designed to be applied in complex phenotypes) and HumVar (designed to diagnostic of Mendelian diseases). Higher scores (>0.85) predicts, more confidently, damaging variants.	[113]
CADD Combined Annotation Dependent Depletion	Integrates diverse genome annotations and scores all human SNV and Indel. It prioritizes functional, deleterious, and disease causal variants according to functional categories, effect sizes and genetic architectures. Scores above 10 should be applied as a cut-off for identifying pathogenic variants.	[114]
MutationTaster	Analyses evolutionary conservation, splice-site changes, loss of protein features and changes that might affect the amount of mRNA. Variants are classified, as polymorphism or disease-causing	[147]
Human Splice Finder	Predict the effects of mutations on splicing signals or to identify splicing motifs in any human sequence.	[133]
nsSNPAnalyzer	Extracts structural and evolutionary information from a query nsSNP and uses a machine learning method (Random Forest) to predict its phenotypic effect. Classifies the variant as neutral and disease.	[148]
TopoSNP Topographic mapping of SNP	Analyze SNP based on its geometric location and conservation information, produces an interactive visualization of disease and non-disease associated with each SNP.	[149]
Condel Consensus Deleteriousness	Condel integrates the output of different methods to predict the impact of nsSNP on protein function. The algorithm based on the weighted average of the normalized scores classifies the variants as neutral or deleterious.	[115]
ANNOVAR * Annotate Variation	Annotates the variants based on several parameters, such as identification whether SNPs or CNVs affect the protein (gene-based), identification of variants in specific genomic regions outside protein-coding regions (region-based) and identification of known variants documented in public and licensed database (filter-based)	[116]
VEP * Variant Effect Predictor	Determines the effect of multiple variants (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts and protein sequence, as well as regulatory regions.	[117]
snpEff *	Annotation and classification of SNV based on their effects on annotated genes, such as synonymous/nsSNP, start or stop codon gains or losses, their genomic locations, among others. Considered as a structural based tool for annotation.	[118]
SeattleSeq *	Provides annotation of SNVs and small indels, by providing to each the dbSNP rs IDs, gene names and accession numbers, variation functions, protein positions and AA changes, conservation scores, HapMap frequencies, PolyPhen predictions and clinical association.	[119]

# **analysis:**

**Primary-Analysis:** QualityControl(ReadFilteringandTrimming).

The identification and analysis of raw data (signal analysis), the targeting of the generation of legible sequencing reads (base calling), and the scoring of base quality are all part of the primary data analysis. This primary analysis usually produces a FASTQ file (Illumina) or an unmapped binary alignment map (uBAM) file type (Ion Torrent).

In the Ion Torrent platform, this task is basically performed in the Ion Torrent Suite Software [53].

**Secondary-Analysis:**

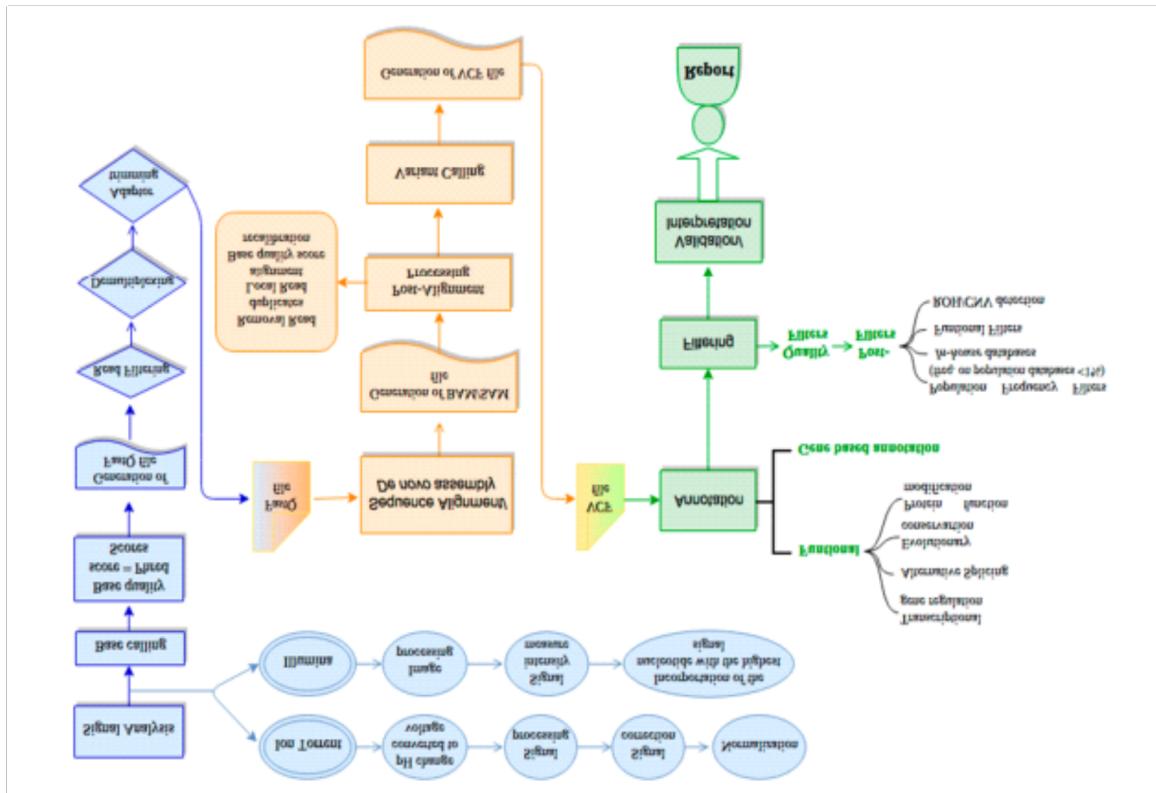
(SequenceAlignment,Post-AlignmentProcessing,VariantCalling).

Secondary analysis, which involves read alignment against the reference human genome (typically hg19 or hg38) and variant calling, is the next step in the NGS data analysis pipeline. Read alignment, which involves aligning sequenced fragments against a reference genome, or de novo assembly, which involves assembling a genome from scratch, are two options for mapping sequencing reads.

**Tertiary-Analysis:** (VariantAnnotation,VariantFiltering,PrioritizationandVisualization).

The third major phase in the NGS analysis pipeline tackles the critical issue of data interpretation, or determining the fundamental relation between variant data and the phenotype observed in a patient in the context of human clinical genetics.

Prioritization and data visualisation tools are used to begin the tertiary study. These analytical measures can be carried out in a variety of ways in a broad range of applications that should be updated on a regular basis to reflect recent scientific findings, necessitating ongoing support and development by the developers.



**figure4:** An description of the bioinformatics workflow for next-generation sequencing (NGS). Main (blue), secondary (orange), and tertiary (green) research are the three types of NGS bioinformatics. The identification and analysis of raw data is the first step in primary data analysis. The reads are then matched against the reference human genome (or constructed from scratch) in secondary research, and the calling is carried out. The tertiary analysis stage, which involves variant annotation, variant filtering, prioritisation, data visualisation, and reporting, is the final step. CNV stands for copy number variation.

## Results:

(1) Since the completion of the Human Genome Project, the cost of next-generation sequencing (NGS) has decreased at a dramatic rate, outpacing Moores Law. Through continuous innovation, Illumina has helped reduce the cost of NGS, enabling the \$1000 human genome.

As next-generation sequencing costs continue to decline, Illumina is leading the way in making NGS more affordable and accessible. We strive to help labs of all sizes access the potential of this powerful technology. With these resources, well guide you through key factors to consider when planning your NGS budget.

(2)The importance of next-generation sequencing technology not only reduced the cost, but also contributed to the rapid arrival of the new corona virus genome, which helped researchers and scientists speed up clinical trials to find the vaccine.

(3)NGS can be used to analyse DNA and RNA samples and is a popular tool in functional genomics. In contrast to microarray methods, NGS-based approaches have several advantages including:

a priori knowledge of the genome or genomic features is not required

it offers single-nucleotide resolution, making it possible to detect related genes (or

features), alternatively spliced transcripts, allelic gene variants and single nucleotide polymorphisms

higher dynamic range of signal

requires less DNA/RNA as input (nanograms of materials are sufficient)

higher reproducibility.

## Conclusion:

despite all of the progress made so far, genetics still has a long way to go before it can provide a definitive response to the diagnosis of all genetic diseases. In order to reduce error rates and improve data handling strategies, further changes in sequencing platforms and data handling strategies are needed. Scientists and clinicians will have to integrate knowledge from multiple -omics sources (such as genome, transcriptome, and proteome) to gain a better understanding of disease, especially complex and heterogeneous diseases (as well as epigenome). As a result, the NGS is increasingly evolving to deal with more than just the traditional genomic approach. One big obstacle, however, is dealing with and interpret all of the different layers of data. Large genomic and epigenomic data sets are being developed, and current computational methods may not be able to handle and extract their full potential. Above all, (bio)informaticians, scientists, and clinicians would need to collaborate in order to analyse data and create innovative methods for integrated systems level research. We believe that machine learning algorithms, such as neural networks and support vector machines, as well as emerging artificial intelligence technologies, will be critical in improving NGS platforms and software, which will aid scientists and clinicians in solving complex biological problems, improving clinical diagnostics and opening new avenues for novel therapy development.

## References:

1. Jackson, D.A.; Symons, R.H.; Berg, P. Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 1972, 69, 2904–2909. [CrossRef]
2. Sanger, F.; Coulson, A.R. A rapid method for determining sequences in DNA by

primed synthesis with DNA

polymerase. *J. Mol. Biol.* 1975, 94, 441–448. [CrossRef]

3. Maxam, A.M.; Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 1977, 74, 560–564.

[CrossRef]

4. Sanger, F.; Nicklen, S.; Coulson, A.R. Biochemistry DNA sequencing with chain-terminating inhibitors (DNA

polymerase/nucleotide sequences/bacteriophage 4X174). *Proc. Natl. Acad. Sci. USA* 1977, 74, 5463–5467.

[CrossRef] [PubMed]

5. Venter, J.C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.;

Evans, C.A.; Holt, R.A. The sequence of the human genome. *Science* 2001, 291, 1304–1351. [CrossRef]

[PubMed]

6. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Natl. Biotechnol.* 2008, 26, 1135–1145. [CrossRef]

[PubMed]

7. Schadt, E.E.; Turner, S.; Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet.* 2010,

19, R227–R240. [CrossRef] [PubMed]

8. Ameur, A.; Kloosterman, W.P.; Hestand, M.S. Single-Molecule Sequencing: Towards Clinical Applications.

*Trends Biotechnol.* 2019, 37, 72–85. [CrossRef] [PubMed]

9. Van Dijk, E.L.; Jaszczyzyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology.

*Trends Genet.* 2018, 34, 666–681. [CrossRef]

10. Aird, D.; Ross, M.G.; Chen, W.-S.; Danielsson, M.; Fennell, T.; Russ, C.; Jaffe, D.B.; Nusbaum, C.; Gnirke, A.

Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 2011, 12,

R18. [CrossRef]

11. Head, S.R.; Komori, H.K.; LaMere, S.A.; Whisenant, T.; Van Nieuwerburgh, F.; Salomon, D.R.; Ordoukhalian, P.

Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques* 2014, 56, 61.

[CrossRef] [PubMed]

12. Van Dijk, E.L.; Jaszczyzyn, Y.; Thermes, C. Library preparation methods for next-generation sequencing:

Tone down the bias. *Exp. Cell Res.* 2014, 322, 10–20. [CrossRef] [PubMed]

13. Knierim, E.; Lucke, B.; Schwarz, J.M.; Schuelke, M.; Seelow, D. Systematic Comparison of Three Methods for

Fragmentation of Long-Range PCR Products for Next Generation Sequencing. *PLoS ONE* 2011, 6, e28240.

[CrossRef] [PubMed]

14. Illumina. Nextera XT Library Prep: Tips and Troubleshooting. Available online: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/technologies/nextera-xttroubleshooting-technical-note.pdf> (accessed on 1 November 2019).

15. Ion Torrent. APPLICATION NOTE Ion PGM™ Small Genome Sequencing. Available online: <https://tools.thermofisher.com/content/sfs/brochures/Small-Genome-Ecoli-De-Novo-App-Note.pdf> (accessed on

13 December 2019).

16. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* 2015, 13, 278–289.

[CrossRef]

17. Sakharkar, M.K.; Chow, V.T.K.; Kangueane, P. Distributions of Exons and Introns in the Human Genome.

*Silico Biol.* 2004, 4, 387–393.

18. Samorodnitsky, E.; Jewell, B.M.; Hagopian, R.; Miya, J.; Wing, M.R.; Lyon, E.; Damodaran, S.; Bhatt, D.;

Reeser, J.W.; Datta, J.; et al. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for

Whole-Exome Sequencing. *Hum. Mutat.* 2015, 36, 903–914. [CrossRef]

19. Hung, S.S.; Meissner, B.; Chavez, E.A.; Ben-Neriah, S.; Ennishi, D.; Jones, M.R.; Shulha, H.P.; Chan, F.C.;  
Boyle, M.; Kridel, R.; et al. Assessment of Capture and Amplicon-Based Approaches for the Development of  
a Targeted Next-Generation Sequencing Pipeline to Personalize Lymphoma Management. *J. Mol. Diagn.*  
2018, 20, 203–214. [CrossRef]
20. Horn, S. Target Enrichment via DNA Hybridization Capture. In *Ancient DNA: Methods and Protocols*; Shapiro, B., Hofreiter, M., Eds.; Humana Press: Totowa, NJ, USA, 2012; pp. 177–188. [CrossRef] *J. Clin. Med.* 2020, 9, 132 24 of 30
21. Kanagawa, T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* 2003, 96,  
317–323. [CrossRef]
22. Sloan, D.B.; Broz, A.K.; Sharbrough, J.; Wu, Z. Detecting Rare Mutations and DNA Damage with  
Sequencing-Based Methods. *Trends Biotechnol.* 2018, 36, 729–740. [CrossRef]
23. Fu, Y.; Wu, P.-H.; Beane, T.; Zamore, P.D.; Weng, Z. Elimination of PCR duplicates in RNA-seq and small  
RNA-seq using unique molecular identifiers. *BMC Genom.* 2018, 19, 531. [CrossRef]
24. Hong, J.; Gresham, D. Incorporation of unique molecular identifiers in TruSeq adapters improves the accuracy  
of quantitative sequencing. *Biotechniques* 2017, 63, 221–226. [CrossRef] [PubMed]
25. Bentley, D.R.; Balasubramanian, S.; Swerdlow, H.P.; Smith, G.P.; Milton, J.; Brown, C.G.; Hall, K.P.; Evers, D.J.;  
Barnes, C.L.; Bignell, H.R.; et al. Accurate whole human genome sequencing using reversible terminator  
chemistry. *Nature* 2008, 456, 53–59. [CrossRef] [PubMed]
26. Fuller, C.W.; Middendorf, L.R.; Benner, S.A.; Church, G.M.; Harris, T.; Huang, X.; Jovanovich, S.B.; Nelson, J.R.;  
Schloss, J.A.; Schwartz, D.C.; et al. The challenges of sequencing by synthesis. *Nat. Biotechnol.* 2009, 27,  
1013–1023. [CrossRef] [PubMed]

27. Kircher, M.; Heyn, P.; Kelso, J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genom.* 2011, 12, 382. [CrossRef]
28. Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011, 475, 348. [CrossRef]
29. Merriman, B.; Ion Torrent R&D Team; Rothberg, J.M. Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis* 2012, 33, 3397–3417. [CrossRef]
30. Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* 2012, 13, 1. [CrossRef]
31. Thompson, J.F.; Steinmann, K.E. Single molecule sequencing with a HeliScope genetic analysis system. *Curr. Protoc. Mol. Biol.* 2010, 7, 10. [CrossRef]
32. Pushkarev, D.; Neff, N.F.; Quake, S.R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 2009, 27, 847–850. [CrossRef]
33. McCarthy, A. Third Generation DNA Sequencing: Pacific Biosciences Single Molecule Real Time Technology. *Chem. Biol.* 2010, 17, 675–676. [CrossRef]
34. Nakano, K.; Shiroma, A.; Shimoji, M.; Tamotsu, H.; Ashimine, N.; Ohki, S.; Shinzato, M.; Minami, M.; Nakanishi, T.; Teruya, K.; et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Hum. Cell* 2017, 30, 149–161. [CrossRef] [PubMed]
35. PacBio. SMRT Sequencing—Delivering Highly Accurate Long Reads to Drive Discovery in Life Science.

Available online:

<https://www.pacb.com/wp-content/uploads/SMRT-Sequencing-Brochure-Deliveringhighly-accurate-long-reads-to-drive-discovery-in-life-science.pdf> (accessed on 9 December 2019).

36. Merker, J.D.; Wenger, A.M.; Sneddon, T.; Grove, M.; Zappala, Z.; Fresard, L.; Waggett, D.; Utiramerur, S.;

Hou, Y.; Smith, K.S.; et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet. Med.* 2018, 20, 159–163. [CrossRef] [PubMed]

37. Technologies, O.N. Company History. Available online: <https://nanoporetech.com/about-us/history> (accessed on 13 December 2019).

38. Deamer, D.; Akeson, M.; Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* 2016, 34, 518.

[CrossRef] [PubMed]

39. Jain, M.; Fiddes, I.T.; Miga, K.H.; Olsen, H.E.; Paten, B.; Akeson, M. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* 2015, 12, 351–356. [CrossRef]

40. Nanopore, O. High-Throughput, Real-Time and On-Demand Sequencing for Your Lab. Available online:

<https://nanoporetech.com/sites/default/files/s3/literature/GridION-Brochure-14Mar2019.pdf> (accessed on 9

December 2019).

41. Goodwin, S.; Gurtowski, J.; Ethe-Sayers, S.; Deshpande, P.; Schatz, M.C.; McCombie, W.R. Oxford Nanopore

sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* 2015, 25,

1750–1756. [CrossRef]

42. De Coster, W.; De Rijk, P.; De Roeck, A.; De Pooter, T.; DHert, S.; Strazisar, M.; Sleegers, K.; Van Broeckhoven, C.

Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res.* 2019, 29, 1178–1187. [CrossRef]

43. Xu, L.; Seki, M. Recent advances in the detection of base modifications using the Nanopore sequencer. *J. Hum.*

Genet. 2020, 65, 25–33. [CrossRef]

44. Genomics, X. The Power of Massively Parallel Partitioning. Available online: [https://pages.10xgenomics.com/rs/446-PBO-704/images/10x\\_BR025\\_Chromium-Brochure\\_Letter\\_Digital.pdf](https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_BR025_Chromium-Brochure_Letter_Digital.pdf) (accessed on 13 December 2019).

45. Zheng, G.X.Y.; Lau, B.T.; Schnall-Levin, M.; Jarosz, M.; Bell, J.M.; Hindson, C.M.; Kyriazopoulou-Panagiotopoulou, S.; Masquelier, D.A.; Merrill, L.; Terry, J.M.; et al. Haplotyping germline

and cancer genomes with high-throughput linked-read sequencing. Nat. Biotechnol. 2016, 34, 303–311.

[CrossRef]

46. Zheng, G.X.Y.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; et al. Massively parallel digital transcriptional profiling of single cells. Nat. Commun.

2017, 8, 14049. [CrossRef]

47. Granja, J.M.; Klemm, S.; McGinnis, L.M.; Kathiria, A.S.; Mezger, A.; Corces, M.R.; Parks, B.; Gars, E.; Liedtke, M.; Zheng, G.X.Y.; et al. Single-cell multiomic analysis identifies regulatory programs in

mixed-phenotype acute leukemia. Nat. Biotechnol. 2019, 37, 1458–1465. [CrossRef]

48. Zeng, Y.; Liu, C.; Gong, Y.; Bai, Z.; Hou, S.; He, J.; Bian, Z.; Li, Z.; Ni, Y.; Yan, J.; et al. Single-Cell RNA

Sequencing Resolves Spatiotemporal Development of Pre-thymic Lymphoid Progenitors and Thymus

Organogenesis in Human Embryos. Immunity 2019, 51, 930–948. [CrossRef] [PubMed]

49. Laurentino, S.; Heckmann, L.; Di Persio, S.; Li, X.; Meyer zu Hörste, G.; Wistuba, J.; Cremers, J.-F.; Gromoll, J.; Kliesch, S.; Schlatt, S.; et al. High-resolution analysis of germ cells from men with sex chromosomal

aneuploidies reveals normal transcriptome but impaired imprinting. Clin. Epigenet. 2019, 11, 127. [CrossRef]

[PubMed]

50. Wang, X.; Xiong, X.; Cao, W.; Zhang, C.; Werren, J.H.; Wang, X. Genome Assembly of the A-Group Wolbachia

in *Nasonia oneida* Using Linked-Reads Technology. *Genome Biol. Evol.* 2019, 11, 3008–3013. [CrossRef]

[PubMed]

51. Delaneau, O.; Zagury, J.-F.; Robinson, M.R.; Marchini, J.L.; Dermitzakis, E.T. Accurate, scalable and integrative

haplotype estimation. *Nat. Commun.* 2019, 10, 5436. [CrossRef]

52. Nanopore, O. Terabases of Long-Read Sequence Data, Analysed in Real Time. Available online: <https://nanoporetech.com/sites/default/files/s3/literature/PromethION-Brochure-14Mar2019.pdf> (accessed on 9

December 2019).

53. Ion Torrent. Torrent Suite-Signal Processing and Base Calling Application Note Torrent Suite Software

Analysis Pipeline. Technical Note. Available online:  
[http://coolgenes.cahe.wsu.edu/ion-docs/Technical-Note--Analysis-Pipeline\\_6455567.html](http://coolgenes.cahe.wsu.edu/ion-docs/Technical-Note--Analysis-Pipeline_6455567.html) (accessed on 26 February 2019).

54. McKinnon, K.I. Convergence of the Nelder-Mead Simplex Method to a Nonstationary Point. *SIAM J. Optim.*

1998, 9, 148–158. [CrossRef]

55. Mehlhorn, K.; Sanders, P. Generic Approaches to Optimization. In *Algorithms and Data Structures: The Basic*

Toolbox; Springer Science & Business Media: Heidelberg, Germany, 2008. [CrossRef]

56. Illumina. Illumina Sequencing Technology: Technology Spotlight: Illumina® Sequencing. Available online:

[https://www.illumina.com/documents/products/techspotlights/techspotlight\\_sequencing.pdf](https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf) (accessed on 13

December 2019).

57. Erlich, Y.; Mitra, P.P.; delaBastide, M.; McCombie, W.R.; Hannon, G.J. Alta-Cyclic: A self-optimizing base

caller for next-generation sequencing. *Nat. Methods* 2008, 5, 679–682. [CrossRef]

58. Kao, W.-C.; Stevens, K.; Song, Y.S. BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing. *Genome Res.* 2009, 19, 1884–1895. [CrossRef]
59. Ledergerber, C.; Dessimoz, C. Base-calling for next-generation sequencing platforms. *Brief. Bioinform.* 2011, 12, 489–497. [CrossRef]
60. Cacho, A.; Smirnova, E.; Huzurbazar, S.; Cui, X. A Comparison of Base-calling Algorithms for Illumina Sequencing Technology. *Brief. Bioinform.* 2015, 17, 786–795. [CrossRef]
61. Zhang, S.; Wang, B.; Wan, L.; Li, L.M. Estimating Phred scores of Illumina base calls by logistic regression and sparse modeling. *BMC Bioinform.* 2017, 18, 335. [CrossRef] *J. Clin. Med.* 2020, 9, 132 of 30
62. Ewing, B.; Hillier, L.; Wendl, M.C.; Green, P. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome Res.* 1998, 8, 175–185. [CrossRef]
63. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* 2013, 11, 11.10.11–11.10.33. [CrossRef]
64. Patel, R.K.; Jain, M. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS ONE* 2012, 7, e30619. [CrossRef] [PubMed]
65. Zhou, Q.; Su, X.; Wang, A.; Xu, J.; Ning, K. QC-Chain: Fast and Holistic Quality Control Method for Next-Generation Sequencing Data. *PLoS ONE* 2013, 8, e60234. [CrossRef] [PubMed]
66. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 1 October 2019).
67. Kong, Y. Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing

- technologies. *Genomics* 2011, 98, 152–153. [[CrossRef](#)]
68. Renaud, G.; Stenzel, U.; Kelso, J. leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* 2014, 42, e141. [[CrossRef](#)]
69. Lindgreen, S. AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. *BMC Res. Notes* 2012, 5, 337. [[CrossRef](#)]
70. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30, 2114–2120. [[CrossRef](#)]
71. Del Fabbro, C.; Scalabrin, S.; Morgante, M.; Giorgi, F.M.; Binkley, G. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS ONE* 2013, 8, e85024. [[CrossRef](#)]
72. Gargis, A.S.; Kalman, L.; Lubin, I.M. Assay Validation. In Clinical Genomics; Kulkarni, S., Pfeifer, J., Eds.; Academic Press: Boston, MA, USA, 2015; pp. 363–376.
73. Flicek, P.; Birney, E. Sense from sequence reads: Methods for alignment and assembly. *Nat. Methods* 2009, 6, S6–S12. [[CrossRef](#)]
74. Ameur, A.; Che, H.; Martin, M.; Bunikis, I.; Dahlberg, J.; Höijer, I.; Häggqvist, S.; Vezzi, F.; Nordlund, J.; Olason, P.; et al. De Novo Assembly of Two Swedish Genomes Reveals Missing Segments from the Human GRCh38 Reference and Improves Variant Calling of Population-Scale Sequencing Data. *Genes* 2018, 9, 486. [[CrossRef](#)] [[PubMed](#)]
75. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
76. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance

genomics data visualization and exploration. *Brief. Bioinform.* 2012, 14, 178–192. [CrossRef] [PubMed]

77. Fonseca, N.A.; Rung, J.; Brazma, A.; Marioni, J.C. Tools for mapping high-throughput sequencing data.

*Bioinformatics* 2012, 28, 3169–3177. [CrossRef]

78. Miller, J.R.; Koren, S.; Sutton, G. Assembly algorithms for next-generation sequencing data. *Genomics* 2010,

95, 315–327. [CrossRef]

79. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*

2010, 26, 589–595. [CrossRef]

80. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA

sequences to the human genome. *Genome Biol.* 2009, 10, R25. [CrossRef]

81. Ruffalo, M.; LaFramboise, T.; Koyutürk, M. Comparative analysis of algorithms for next-generation

sequencing read alignment. *Bioinformatics* 2011, 27, 2790–2796. [CrossRef]

82. Homer, N. TMAP: The Torrent Mapping Program. Available online: <https://github.com/iantorrent/TMAP/blob/master/doc/tmap-book.pdf> (accessed on 1 October 2019).

83. Li, Z.; Chen, Y.; Mu, D.; Yuan, J.; Shi, Y.; Zhang, H.; Gan, J.; Li, N.; Hu, X.; Liu, B.; et al. Comparison of the

two major classes of assembly algorithms: Overlap–layout–consensus and de-bruijn-graph. *Brief. Funct. Genom.* 2012, 11, 25–37. [CrossRef]

84. Pop, M.; Phillippy, A.; Delcher, A.L.; Salzberg, S.L. Comparative genome assembly. *Brief. Bioinform.* 2004, 5, 237–248. [CrossRef]

85. Compeau, P.E.C.; Pevzner, P.A.; Tesler, G. How to apply de Bruijn graphs to genome assembly. *Nat. Biotechnol.*

2011, 29, 987–991. [CrossRef] [PubMed] *J. Clin. Med.* 2020, 9, 132 27 of 30

86. Sedlazeck, F.J.; Lee, H.; Darby, C.A.; Schatz, M.C. Piercing the dark matter: Bioinformatics of long-range

- sequencing and mapping. *Nat. Rev. Genet.* 2018, 19, 329–346. [CrossRef]
87. Tian, S.; Yan, H.; Kalmbach, M.; Slager, S.L. Impact of post-alignment processing in variant discovery from whole exome data. *BMC Bioinform.* 2016, 17, 403. [CrossRef] [PubMed]
88. Robinson, P.N.; Piro, R.M.; Jager, M. Postprocessing the Alignment. In Computational Exome and Genome Analysis, 1st ed.; Chapman and Hall/CRC: New York, NY, USA, 2017. [CrossRef]
89. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010, 20, 1297–1303. [CrossRef]
90. Nielsen, R.; Paul, J.S.; Albrechtsen, A.; Song, Y.S. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 2011, 12, 443. [CrossRef] [PubMed]
91. Libbrecht, M.W.; Noble, W.S. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 2015, 16, 321. [CrossRef]
92. Kuhn, M.; Stange, T.; Herold, S.; Thiede, C.; Roeder, I. Finding small somatic structural variants in exome sequencing data: A machine learning approach. *Comput. Stat.* 2018, 33, 1145–1158. [CrossRef]
93. Hwang, S.; Kim, E.; Lee, I.; Marcotte, E.M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci. Rep.* 2015, 5, 17875. [CrossRef]
94. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* 2011, 27, 2156–2158. [CrossRef]
95. The Variant Call Format (VCF) Version 4.2 Specification. Available online: <https://samtools.github.io/htsspecs/VCFv4.2.pdf> (accessed on 1 October 2019).
96. Feuk, L.; Carson, A.R.; Scherer, S.W. Structural variation in the human genome.

Nat. Rev. Genet. 2006, 7,

85–97. [CrossRef]

97. Stankiewicz, P.; Lupski, J.R. Structural Variation in the Human Genome and its Role in Disease. *Annu. Rev. Med.* 2010, 61, 437–455. [CrossRef] [PubMed]

98. Mitsuhashi, S.; Matsumoto, N. Long-read sequencing for rare human genetic diseases. *J. Hum. Genet.* 2020,

65, 11–19. [CrossRef] [PubMed]

99. Kraft, F.; Kurth, I. Long-read sequencing in human genetics. *Med. Genet.* 2019, 31, 198–204. [CrossRef]

100. Zhao, M.; Wang, Q.; Wang, Q.; Jia, P.; Zhao, Z. Computational tools for copy number variation (CNV)

101. Pirooznia, M.; Goes, F.S.; Zandi, P.P. Whole-genome CNV analysis: Advances in computational approaches.

*Front. Genet.* 2015, 6, 138. [CrossRef] [PubMed]

102. Korbel, J.O.; Urban, A.E.; Affourtit, J.P.; Godwin, B.; Grubert, F.; Simons, J.F.; Kim, P.M.; Palejev, D.;

Carriero, N.J.; Du, L.; et al. Paired-End Mapping Reveals Extensive Structural Variation in the Human

Genome. *Science* 2007, 318, 420–426. [CrossRef]

103. Chen, K.; Wallis, J.W.; McLellan, M.D.; Larson, D.E.; Kalicki, J.M.; Pohl, C.S.; McGrath, S.D.; Wendl, M.C.;

Zhang, Q.; Locke, D.P.; et al. BreakDancer: An algorithm for high-resolution mapping of genomic structural

variation. *Nat. Methods* 2009, 6, 677. [CrossRef]

104. Ye, K.; Guo, L.; Yang, X.; Lamijer, E.-W.; Raine, K.; Ning, Z. Split-Read Indel and Structural Variant Calling

Using PINDEL. In Copy Number Variants: Methods and Protocols; Bickhart, D.M., Ed.; Springer: New York, NY,

USA, 2018; pp. 95–105. [CrossRef]

105. Duncavage, E.J.; Abel, H.J.; Pfeifer, J.D.; Armstrong, J.R.; Becker, N.; Magrini, V.J. SLOPE: A quick and

accurate method for locating non-SNP structural variation from targeted

next-generation sequence data.

Bioinformatics 2010, 26, 2684–2688. [CrossRef]

106. Park, H.; Chun, S.-M.; Shim, J.; Oh, J.-H.; Cho, E.J.; Hwang, H.S.; Lee, J.-Y.; Kim, D.; Jang, S.J.; Nam, S.J.; et al.

Detection of chromosome structural variation by targeted next-generation sequencing and a deep learning

application. Sci. Rep. 2019, 9, 3644. [CrossRef]

107. Russnes, H.G.; Navin, N.; Hicks, J.; Borresen-Dale, A.-L. Insight into the heterogeneity of breast cancer

through next-generation sequencing. J. Clin. Investig. 2011, 121, 3810–3818. [CrossRef] J. Clin. Med. 2020, 9, 132 28 of 30

108. Magi, A.; Tattini, L.; Palombo, F.; Benelli, M.; Gialluisi, A.; Giusti, B.; Abbate, R.; Seri, M.; Gensini, G.F.;

Romeo, G.; et al. H3M2: Detection of runs of homozygosity from whole-exome sequencing data. Bioinformatics

2014, 30, 2852–2859. [CrossRef] [PubMed]

109. Zarrei, M.; MacDonald, J.R.; Merico, D.; Scherer, S.W. A copy number variation map of the human genome.

Nat. Rev. Genet. 2015, 16, 172–183. [CrossRef] [PubMed]

110. Ion Torrent. CNV Detection by Ion Semiconductor Sequencing. Available online: <https://assets.thermofisher.com/TFS-Assets/LSG/brochures/CNV-Detection-by-Ion.pdf> (accessed on 1 November 2019).

111. Scherer, S.W.; Lee, C.; Birney, E.; Altshuler, D.M.; Eichler, E.E.; Carter, N.P.; Hurles, M.E.; Feuk, L. Challenges

and standards in integrating surveys of structural variation. Nat. Genet. 2007, 39, S7–S15. [CrossRef]

[PubMed]

112. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res.

2003, 31, 3812–3814. [CrossRef] [PubMed]

113. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.;

- Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* 2010, 7, 248–249. [CrossRef] [PubMed]
114. Kircher, M.; Witten, D.M.; Jain, P.; ORoak, B.J.; Cooper, G.M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 2014, 46, 310–315. [CrossRef]
115. González-Pérez, A.; López-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score. *Condel. Am. J. Hum. Genet.* 2011, 88, 440–449. [CrossRef]
116. Yang, H.; Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* 2015, 10, 1556–1566. [CrossRef]
117. McLaren, W.; Pritchard, B.; Rios, D.; Chen, Y.; Flieck, P.; Cunningham, F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010, 26, 2069–2070. [CrossRef]
118. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 2012, 6, 80–92. [CrossRef]
119. Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009, 461, 272. [CrossRef] [PubMed]
120. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010, 38, e164. [CrossRef] [PubMed]
121. Keren, H.; Lev-Maor, G.; Ast, G. Alternative splicing and evolution: Diversification, exon definition and

- function. *Nat. Rev. Genet.* 2010, 11, 345–355. [CrossRef] [PubMed]
122. Pruitt, K.D.; Harrow, J.; Harte, R.A.; Wallin, C.; Diekhans, M.; Maglott, D.R.; Searle, S.; Farrell, C.M.; Loveland, J.E.; Ruef, B.J.; et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* 2009, 19, 1316–1323. [CrossRef]
123. McCarthy, D.J.; Humburg, P.; Kanapin, A.; Rivas, M.A.; Gaulton, K.; Cazier, J.-B.; Donnelly, P. Choice of transcripts and software has a large effect on variant annotation. *Genome Med.* 2014, 31, 26. [CrossRef]
124. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010, 467, 52–58. [CrossRef]
125. Stoneking, M.; Krause, J. Learning about human population history from ancient and modern genomes. *Nat. Rev. Genet.* 2011, 12, 603–614. [CrossRef]
126. Siva, N. 1000 Genomes project. *Nat. Biotechnol.* 2008, 26, 256. [CrossRef]
127. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016, 536, 285–291. [CrossRef]
128. Genomics, A. Practical Guidelines. Available online: <https://www.acmg.net/ACMG/Medical-GeneticsPractice-Resources/Practice-Guidelines.aspx> (accessed on 13 December 2019).
129. Harper, P.S. The European Society of Human Genetics: Beginnings, early history and development over its first 25 years. *Eur. J. Hum. Genet.* 2017, 2017, 1–8. [CrossRef]
130. Gilissen, C.; Hoischen, A.; Brunner, H.G.; Veltman, J.A. Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.* 2012, 20, 490–497. [CrossRef] [PubMed] *J. Clin. Med.* 2020, 9, 132 29 of 30
131. Sauna, Z.E.; Kimchi-Sarfaty, C. Understanding the contribution of synonymous

mutations to human disease.

Nat. Rev. Genet. 2011, 12, 683–691. [CrossRef] [PubMed]

132. Cartegni, L.; Chew, S.L.; Krainer, A.R. Listening to silence and understanding nonsense: Exonic mutations

that affect splicing. Nat. Rev. Genet. 2002, 3, 285–298. [CrossRef] [PubMed]

133. Desmet, F.-O.; Hamroun, D.; Lalande, M.; Collod-Béroud, G.; Claustres, M.; Béroud, C. Human Splicing

Finder: An online bioinformatics tool to predict splicing signals. Nucleic Acids Res.

2009, 37, e67. [CrossRef]

[PubMed]

134. Desvignes, J.-P.; Bartoli, M.; Miltgen, M.; Delague, V.; Salgado, D.; Krahn, M.; Béroud, C. VarAFT: A variant

annotation and filtration system for human next generation sequencing data. Nucleic Acids Res. 2018, 46,

W545–W553. [CrossRef]

135. MacArthur, D.G.; Balasubramanian, S.; Frankish, A.; Huang, N.; Morris, J.; Walter, K.; Jostins, L.; Habegger, L.;

Pickrell, J.K.; Montgomery, S.B.; et al. A Systematic Survey of Loss-of-Function Variants in Human

Protein-Coding Genes. Science 2012, 335, 823–828. [CrossRef]

136. Lelieveld, S.H.; Veltman, J.A.; Gilissen, C. Novel bioinformatic developments for exome sequencing. Hum.

Genet. 2016, 135, 603–614. [CrossRef]

137. Robinson, P.N.; Köhler, S.; Oellrich, A.; Sanger Mouse Genetics, P.; Wang, K.; Mungall, C.J.; Lewis, S.E.;

Washington, N.; Bauer, S.; Seelow, D.; et al. Improved exome prioritization of disease genes through

cross-species phenotype comparison. Genome Res. 2014, 24, 340–348. [CrossRef]

138. Eilbeck, K.; Quinlan, A.; Yandell, M. Settling the score: Variant prioritization and Mendelian disease. Nat. Rev.

Genet. 2017, 18, 599. [CrossRef]

139. Khurana, E.; Fu, Y.; Chen, J.; Gerstein, M. Interpretation of Genomic Variants Using a Unified Biological

Network Approach. PLoS Comp. Biol. 2013, 9, e1002886. [CrossRef]

140. Petrovski, S.; Wang, Q.; Heinzen, E.L.; Allen, A.S.; Goldstein, D.B. Genic Intolerance to Functional Variation

and the Interpretation of Personal Genomes. PLoS Genet. 2013, 9, e1003709. [CrossRef]

141. Zemojtel, T.; Köhler, S.; Mackenroth, L.; Jäger, M.; Hecht, J.; Krawitz, P.; Graul-Neumann, L.; Doelken, S.;

Ehmke, N.; Spielmann, M.; et al. Effective diagnosis of genetic disease by computational phenotype analysis

of the disease-associated genome. Sci. Transl. Med. 2014, 6, 252ra123. [CrossRef] [PubMed]

142. Singleton, M.V.; Guthery, S.L.; Voelkerding, K.V.; Chen, K.; Kennedy, B.; Margraf, R.L.; Durtschi, J.; Eilbeck, K.;

Reese, M.G.; Jorde, L.B.; et al. Phevor Combines Multiple Biomedical Ontologies for Accurate Identification

of Disease-Causing Alleles in Single Individuals and Small Nuclear Families. Am. J. Hum. Genet. 2014, 94,

599–610. [CrossRef] [PubMed]

143. Pereira, R.; Oliveira, M.E.; Santos, R.; Oliveira, E.; Barbosa, T.; Santos, T.; Gonçalves, P.; Ferraz, L.; Pinto, S.;

Barros, A.; et al. Characterization of CCDC103 expression profiles: Further insights in primary ciliary

dyskinesia and in human reproduction. J. Assist. Reprod. Genet. 2019, 36, 1683–1700. [CrossRef]

144. Pereira, R.; Barbosa, T.; Gales, L.; Oliveira, E.; Santos, R.; Oliveira, J.; Sousa, M. Clinical and Genetic Analysis

of Children with Kartagener Syndrome. Cells 2019, 8, 900. [CrossRef]

145. Stelzer, G.; Plaschkes, I.; Oz-Levi, D.; Alkelai, A.; Olander, T.; Zimmerman, S.; Twik, M.; Belinky, F.;

Fishilevich, S.; Nudel, R.; et al. VarElect: The phenotype-based variation prioritizer of the GeneCards Suite.

BMC Genom. 2016, 17, 444. [CrossRef]

146. Pollard, K.S.; Hubisz, M.J.; Rosenbloom, K.R.; Siepel, A. Detection of nonneutral substitution rates on

- mammalian phylogenies. *Genome Res.* 2010, 20, 110–121. [CrossRef]
147. Schwarz, J.M.; Rödelsperger, C.; Schuelke, M.; Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 2010, 7, 575–576. [CrossRef]
148. Bao, L.; Zhou, M.; Cui, Y. nsSNPAnalyzer: Identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* 2005, 33, W480–W482. [CrossRef]
149. Stitzel, N.O.; Binkowski, T.A.; Tseng, Y.Y.; Kasif, S.; Liang, J. topoSNP: A topographic database of non-synonymous single nucleotide polymorphisms with and without known disease association. *Nucleic Acids Res.* 2004, 32, 520–522. [CrossRef]
150. Esposito, A.; Colantuono, C.; Ruggieri, V.; Chiusano, M.L. Bioinformatics for agriculture in the Next-Generation sequencing era. *Chem. Biol. Technol. Agric.* 2016, 3, 9. [CrossRef]
151. Bamshad, M.J.; Ng, S.B.; Bigham, A.W.; Tabor, H.K.; Emond, M.J.; Nickerson, D.A.; Shendure, J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 2011, 12, 745–755. [CrossRef]
- [PubMed] *J. Clin. Med.* 2020, 9, 132 30 of 30
152. Ku, C.S.; Cooper, D.N.; Polychronakos, C.; Naidoo, N.; Wu, M.; Soong, R. Exome sequencing: Dual role as a discovery and diagnostic tool. *Ann. Neurol.* 2012, 71, 5–14. [CrossRef] [PubMed]
153. Sboner, A.; Mu, X.; Greenbaum, D.; Auerbach, R.K.; Gerstein, M.B. The real cost of sequencing: Higher than you think! *Genome Biol.* 2011, 12, 125. [CrossRef]
154. Mardis, E.R.; Mu, X.J.; Greenbaum, D.; Auerbach, R.K.; Gerstein, M.B.; Nisbett, J.; Guigo, R.; Dermitzakis, E.; Gilad, Y.; Pritchard, J.; et al. The \$1000 genome, the \$100,000 analysis? *Genome Med.* 2010, 2, 84. [CrossRef]
155. Moorthie, S.; Hall, A.; Wright, C.F. Informatics and clinical genome sequencing: Opening the black box. *Genet. Med.* 2013, 15, 165–171. [CrossRef]

156. Daber, R.; Sukhadia, S.; Morrisette, J.J.D. Understanding the limitations of next generation sequencing informatics, an approach to clinical pipeline validation using artificial data sets. *Cancer Genet.* 2013, 206, 441–448. [CrossRef]
157. Biesecker, L.G.; Green, R.C. Diagnostic clinical genome and exome sequencing. *N. Engl. J. Med.* 2014, 370, 2418–2425. [CrossRef]
158. Gagan, J.; Van Allen, E.M. Next-generation sequencing to guide cancer therapy. *Genome Med.* 2015, 7, 80. [CrossRef]
159. Oliveira, J.; Martins, M.; Pinto Leite, R.; Sousa, M.; Santos, R. The new neuromuscular disease related with defects in the ASC-1 complex: Report of a second case confirms ASCC1 involvement. *Clin. Genet.* 2017, 92, 434–439. [CrossRef]
160. Tebani, A.; Afonso, C.; Marret, S.; Bekri, S. Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations. *Int. J. Mol. Sci.* 2016, 17, 1555. [CrossRef] [PubMed]
161. Ohashi, H.; Hasegawa, M.; Wakimoto, K.; Miyamoto-Sato, E. Next-Generation Technologies for Multiomics Approaches Including Interactome Sequencing. *BioMed Res. Int.* 2015, 2015, 1–9. [CrossRef] [PubMed]
- © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).