DNV·GL

BIGMED

Drivers in rapid genetic diagnostics for rare diseases in infants

Oslo University Hospital, Dept. of Medical Genetics

Report No.: 2019-1187, Rev. 1
Document No.:
Date: 25.09.19

Project name:	BigMed
Report title:	Drivers in rapid genetic diagnostics for rare diseases in infants
Customer:	Oslo University Hospital, Dept. of Medical Genetics, Kirkeveien 166, Oslo
Date of issue:	25.09.19
Project No.:	10078801
Organisation unit:	GTR Precision Medicine
Report No.:	2019-1187, Rev. 1
Document No.:	

Applicable contract(s) governing the provision of this Report:

Objective: To investigate and explore the factors that contribute to the time to diagnosis for infants with rare diseases in a neonatal intensive care setting, with the goal of informing Oslo University Hospital design choices as it begins to offer similar services.

Prepared by:	Verified by:	Approved by:
Courtney Nadeau Sr. Researcher	Sharmin Alagaratnam Sr. Researcher	Bobbie Ray-Sannerud Program Director

DNV GL Distribution:

☒ Unrestricted distribution (internal and external)
☐ Unrestricted distribution within DNV GL Group
☐ Unrestricted distribution within DNV GL contracting party
☐ No distribution (confidential)

Keywords:
Next generation sequencing, clinical decision support, rare disease, phenotype, neonatal intensive care unit, *in vitro* medical device, molecular diagnostics

Rev. No.	Date	Reason for Issue	Prepared by	Verified by	Approved by
0	2019-09-25	First issue	CN	AS	BRS

Table of contents

1 RARE DISEASE DIAGNOSIS IN THE HEALTH SYSTEM

While individually rare, inherited and *de novo* genetic disorders collectively affect >5% of newborns (Baird, et al. 1988), representing a substantial challenge for the health system and significant costs to society (Baldovino, et al. 2016). Due to their rarity, diagnosing and treting these conditions is difficult. While some of these diseases are only detected later in life, many are diagnosed in infants after admission to a neo-natal ICU (NICU) after presenting with atypical or otherwise inexplicable symptoms. Most of these conditions have direct genetic causes or a substantial genetic component, so once admitted patients are generally subjected to a series of targeted genetic and metabolic tests. Negative results are followed by additional rounds of testing, each of which is subsequently less likely to identify a clinically actionable result.

These successive rounds of testing and re-hypothesizing require substantial clinical work and represent a significant cost to the health system, but more importantly consume critical weeks and months. Healthcare writing on the topic often refers to a 'diagnostic odyssey' (Black, Martineau and Manacorda 2015). and this is an appropriate term: in an EU-wide survey, 25% of patients had to wait between 5 and 30 years for a diagnosis, often after dealing with upwards of 5 to 10 physicians, and 40% of patients receive incorrect diagnoses, often leading to inappropriate medical or surgical treatment (EURORDIS 2009).

Clinical sequencing has the potential to drastically change the diagnostic odyssey of these children. By applying a broad, hypothesis-neutral first-line diagnostic such as whole genome sequencing, clinicians can quickly identify potential genetic causes. Depending on the sequencing approach and existing diagnostic landscape, exome or whole genome sequencing as a first-line diagnostic commonly leads to molecular diagnoses in 40-60% of infants, often within a period of weeks, rather than years (Farnaes, et al. 2018).

2 NEXT-GENERATION SEQUENCING IN A NEO-NATAL SETTING

When compared to other NGS applications, broad exome or whole-genome rare disease diagnostics offers unique challenges. Firstly, turn-around time (TaT) is critical as many of these diseases represent acute, life-threatening illnesses. The strict requirement for short TaT requires the optimization of work-flows, sequencing, and bioinformatics pipelines, but also represents a logistical and operational challenge. Hospital scheduling, sample transit, and wet-lab and bioinformatics logistics all have significant impacts on TaT. Secondly, WGS or exome diagnostics for rare diseases differ from many diagnostics in that the approach is hypothesis neutral: the exact genetic abnormality is not known prior to testing. As such, there can be significant uncertainty regarding test results, since full validation and verification (as is performed with large cancer cohorts, for example), is not possible. Instead, the interpretation and selection of appropriate variants to influence treatment decisions relies on the expertise and often iterative detective work performed by one or more clinical geneticists. The time and effort required for this interpretation is difficult to predict, and may vary from case to case.

3 IMPLEMENTING NICU-SEQ

The Department for Medical Genetics at OUS (AMG) is in the process of putting a production NICU sequencing pipeline in place. There are a wide variety of design choices that impact the speed, accuracy, and effectiveness of this testing program. The goal of this work, funded through the Norwegian BigMed project, was to provide insight into how certain design choices could impact TaT, specifically. AMG routinely conducts similar diagnostics in a production setting, however median TaT for routine analysis is around 8 weeks and for prioritized samples (medical urgency) around 4 weeks (OUS AMG). In contrast, paediatric hospitals such as the Rady Children's hospital in San Diego or Children's Mercy in Kansas City (Miller, et al. 2015) have demonstrated 26-48 h TaT, and in regular production see median TaT of around 7 days (range 3-12). More recent implementations relying heavily on automation have achieved regular TaT of under 24 hours, demonstrating that rapid WGS is achievable in a clinical setting (Clark, et al. 2019). Improving TaT leads to better outcomes for patients and lower overall healthcare costs, and the ultimate goal of this work should be to inform design decisions to reduce TaT for future clinical implementation.

A key objective of the work presented here was to identify the key factors for low TaT, and consider these within the broader context of quality, IT limitations, and pre-existing clinical workflows at AMG. The first part of this work was to map out fast NICU-seq pipelines from two leading children's hospitals in detail. Technical details, I/O requirements, technologies and software used, organizational considerations, and strategies for delivering fast clinical reports for infants were gathered and documented below.

4 EXAMPLE NICU-SEQ PROCESS (MILLER, ET AL. 2015)

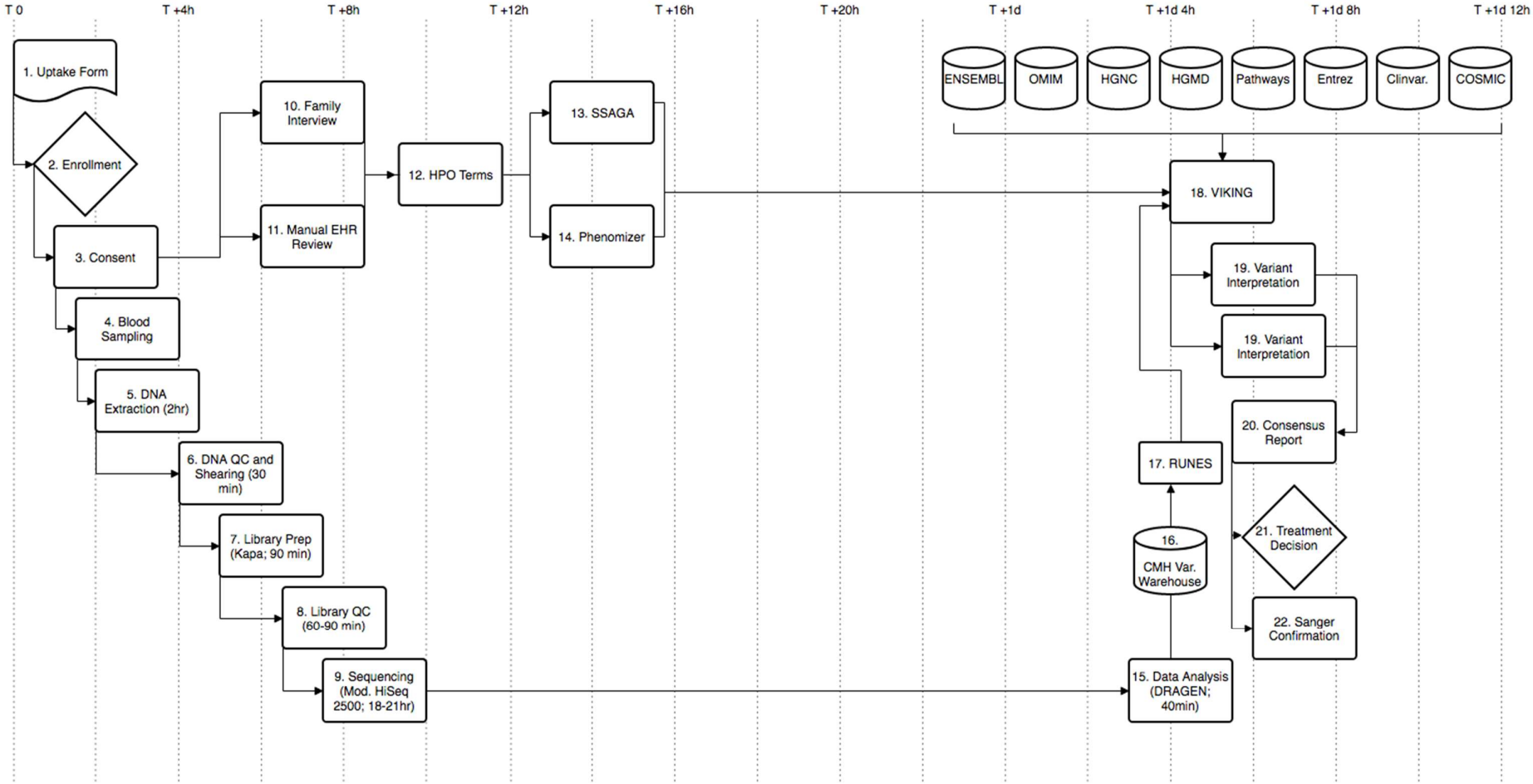

Figure 1: Process map of NICU-Seq pipeline from Miller et al. (2015). Steps corresponding to the processes listed below are placed on a time scale beginning with patient presentation.

This process reflects the previous world-record holding paediatric WGS pipeline, in use at the Rady Children's hospital in San Diego, Children's Mercy in Kansas City and others (Miller, et al. 2015). In brief, patient enrolment, library preparation, and the start of the sequencing run is completed on the first day. While this is ongoing, in parallel family interviews and an interrogation of the patient's journal generate a set of HPO terms. These terms are used as input for two tools which correlate these with databases of genetic diseases to identify lists of candidate diagnoses/genes. Sequencing is performed on two flow cells of a modified HiSeq 2500, which has been optimized to require 18-21 hours instead of the standard 25,5. During the second day, run data is streamed into a standalone server, which executes read alignment, duplicate removal, and variant calling in approximately 40 minutes. Variants are *in silico* annotated with custom software, and clinical geneticists examine the annotated variants in conjunction with clinical data in a custom explorer/mark-up portal. Points below refer to the steps in process map (Figure 1).

1. Standardized uptake form is completed by referring physician. Intended to collect primary symptoms, a brief list of past diagnostics and previous attempts, family history, etc. Rather than being evaluated by a panel in a regular multidisciplinary meeting or NICU board, once completed this is immediately evaluated by experts.

2. The enrollment decision is based on whether likely diagnosis could be detectable by NGS. There is no clinical screening for parents prior to uptake.

3. No specific notes on informed consent protocols are given.

4. No specific notes on collection protocols are given. Samples are taken from trios where possible, otherwise from single parents or only the proband. Blood samples for the proband are kept within maximum daily phlebotomy limits. Note that even for small infants it should be possible to obtain the required volume, but that physicians will need to prioritize the importance of NGS vs. other diagnostics in some cases.

5. DNA extraction is automated on an MSMI Chemagen (PE) with standard kits and a 24-well head. The entire instrument is enclosed in a biosafety cabinet. DNA isolation and purification is bead-based. Input for the procedure is 1,8 ml blood per sample. Isolation takes 2 h and delivers 40 µg of DNA per ml blood. According to PE, the DNA blood kit runs in only 45 min, so the 2 h here likely includes set-up time, blood transport and sample intake, etc.

6. DNA QC protocols are not detailed, however the total duration of DNA isolation, QC, and shearing is 2:30. Shearing is performed with a Covaris S2 using 2 µg of input DNA.

7. PCR-free library prep kits are used in all instances, and follow a standard end-repair, A-tailing, and ligation protocol. All purifications are SPRI-based, no information on whether single or double cutoffs are used. For the ultra-fast workflow, Kapa Hyper Prep kits are used (90 minutes), otherwise TruSeq DNA.

8. Library QC takes 60-90 minutes. No details are given, so it is not known of only fluorometric methods or also qPCR standards are used. Language in the papers indicates clustering needed optimization, potentially indicating libraries are quantified with fluorometer only (ie. no fragment analysis) and rely on consistent fragment sizes, concentration, and adapter load to cluster well.

9. Sequencing with a HiSeq 2500 in rapid mode with 2x100nt PE kit. Using one or usually two flow cells (300 or 600mio PE reads). Officially a 27 h run time, this can be done in 25:30 in practice. There is a second option for a custom, ultra-rapid run mode that reduces this 25:30 to 18-21 hr. This modification requires changing the SBS cycle, temp ramp, and microfluidics. After optimization, quality and cluster density does not appear significantly different from the commercially supported protocol. Runs typically generate 1,15 bn 101 nt reads and yield 37x coverage after quality filtering and alignment.

10. While library prep and sequencing is ongoing, physicians interview the family and review the patient's EHR to generate a list of phenotypic features mapped to HPO terms.

11. See 10.

12. Between 2 and 20 HPO terms are typically gathered from a patient.

13. While sequencing is ongoing, Phenomizer and/or SSAGA are used to identify candidate genes of interest based on the HPO terms and a list of 6000 genetic diseases (differential diagnosis). Phenomizer calculates a similarity score between the HPO terms entered and HPO terms associated with the diseases. P values from Phenomizer are used to sort potential diseases, and diseases without known causative genes are removed. Inheritance pattern, if known, is also used to filter results, and lists are reduced to 100-250 disease entries, if necessary, based on manual inspection or re-evaluation of mandatory/optional HPO terms.

14. Clinical Presentation: SSAGA takes HPO terms and provides differential diagnoses based on OMIM, Orphanet, and DECIPHER disease entries that include at least one feature. SSAGA is essentially a correlation tool, and maps features of 591 well-established recessive genetic diseases with paediatric phenotypes. Includes 227 clinical terms in 9 categories. Diagnoses generated by SSAGA can be stack-ranked based on the number of matching terms to prioritize.

15. Data analysis for the rapid workflow is performed using an Edico DRAGEN processor operating on a separate server running 2x Xeon CPUs and the DRAGEN FPGA connected to 32 GB of DDR3 RAM. Computationally intensive functions are performed on the FPGA, other functions run on the multi-threaded Xeon CPUs. Storage is supplied by SSDs in RAID 0. .bcl can be converted to .fastq by standard Illumina software, but for additional speed .bcl can be streamed directly into the FPGA and .fastq files can be generated there. After alignment, duplicates are flagged, and the DRAGEN variant caller generates a .vcf file. The entire process from .fastq or .bcl to .vcf is executed as a single workflow within the FPGA. Previous pipeline with Casava and GATK took 15 hrs, now 40 mins. Around 5,2 mio variants are called per sample, of which around 5 k are potentially relevant clinically. Primary data is not maintained, but .fastq (104 GB), .bam (71 GB), .vcf (1,2 GB), and an annotated variant file (825 MB) are put in permanent storage.

16. Variants are stored in an in-house DB, CMH variant warehouse. As of 2016 CMH contained 69,8 mio variants, 4584 samples, and 4,6 bn variant calls. A hadoop job updates allele frequency for all variants 5x daily, which is used to help prioritize variants.

17. RUNES is a variant annotation pipeline. RUNES applies a series of sequential steps that each add information to all variants in the dataset. Annotation tools include the ENSEMBL Variant Effect Predictor (VEP), a comparison with dbSNP, CMH splice impact evaluator and transcript context characterizer, and comparison with HGMD. At the end of annotation, variants are classified automatically to ACMG guidelines based on the accumulated evidence in the annotated .vcf. This relies heavily on status in the CMH variant db. Variants previously classified as pathogenic by clinicians in the CMH db are pathogenic. New variants that are likely pathogenic (ie. truncated) based on annotations are classified as likely pathogenic. Variants with >2% MAF in dbSNP are classified likely benign. Variants with >2% MAF in

dbSNP and annotated as benign in ClinVar are classified as benign. OMIM is screened monthly for genes to update the internal db with.

18. VIKING is a user-facing software used to interpret and filter sequencing results. Input to VIKING includes the annotated .vcf from RUNES and candidate genes and diagnoses from Phenomizer and/or SSAGA. Various static, pre-defined gene lists can also be used to filter, such as genes with OMIM records, genes associated with mitochondrial disorders, etc. Filtering can be performed through a web-based GUI, but filtering strategies can also be saved so that in future sessions a user can apply the same criteria (for example, filter out VUS, benign, or likely benign as annotated by RUNES, filter out anything without a SSAGA or Phenomizer predictor, filter out anything above a certain allele frequency in CMH warehouse etc.) in seconds. VIKING includes links to the CMH variant warehouse, OMIM, HGNC, HGMD, Entrez Gene, ENSEMBL, and pathways information, which assists in literature curation.

19. Variants are interpreted sequentially by two independent experts. In general, users only look at variants in silico classified as ACMG VUS, P, or LP, with MAF of <1%, 0,5%, or 0,01%, or at variants that are unique to the patient or the family. Manual .bam review in IGV is performed to exclude low coverage in the proband or parents as a source of potential artifacts. Real-time IGV links are included in VIKING. Users examine all monogenic inheritance patterns (de novo, mt, dominant, x-linked etc) for potential causative variants.

19.1 User story: VIKING filtering to ACMG cat. 3-5, <0,1% MAF, recessive inheritance, and in OMIM as monogenic disease-associated identified 16 variants in 8 genes in an affected infant. Of these, 2 variants in a single gene fit clinical features of patient and actual inheritance pattern (IGV vs. parents: low coverage in parents or other quality issues for most genes). Each variant heterozygous in one of the parents, leading to rapid diagnosis. In fast cases, variant annotation in VIKING and diagnosis can take 40 mins, for complex cases, this can take hours.

20. Reporting performed by ABMG-certified geneticists with experience in monogenic diseases. VIKING exports data and mark-up to a clinical report template.

21. If patient's phenotype differs from previous mutational reports for that disease, additional experts are contacted, and additional functional diagnostics are employed for confirmation.

22. Causative variant confirmation through bi-directional sanger sequencing. Sanger confirmation takes substantially longer than NGS: at CHM WGS takes 3-12 days (median 7), while sanger confirmation requires an additional 14 days. In some instances, physicians may choose to begin low-risk therapies prior to confirmation, depending on the severity of the condition and relative risks to the patient of either delaying treatment or commencing the incorrect treatment

5 EXAMPLE NICU-SEQ PROCESS (CLARK, ET AL. 2019)

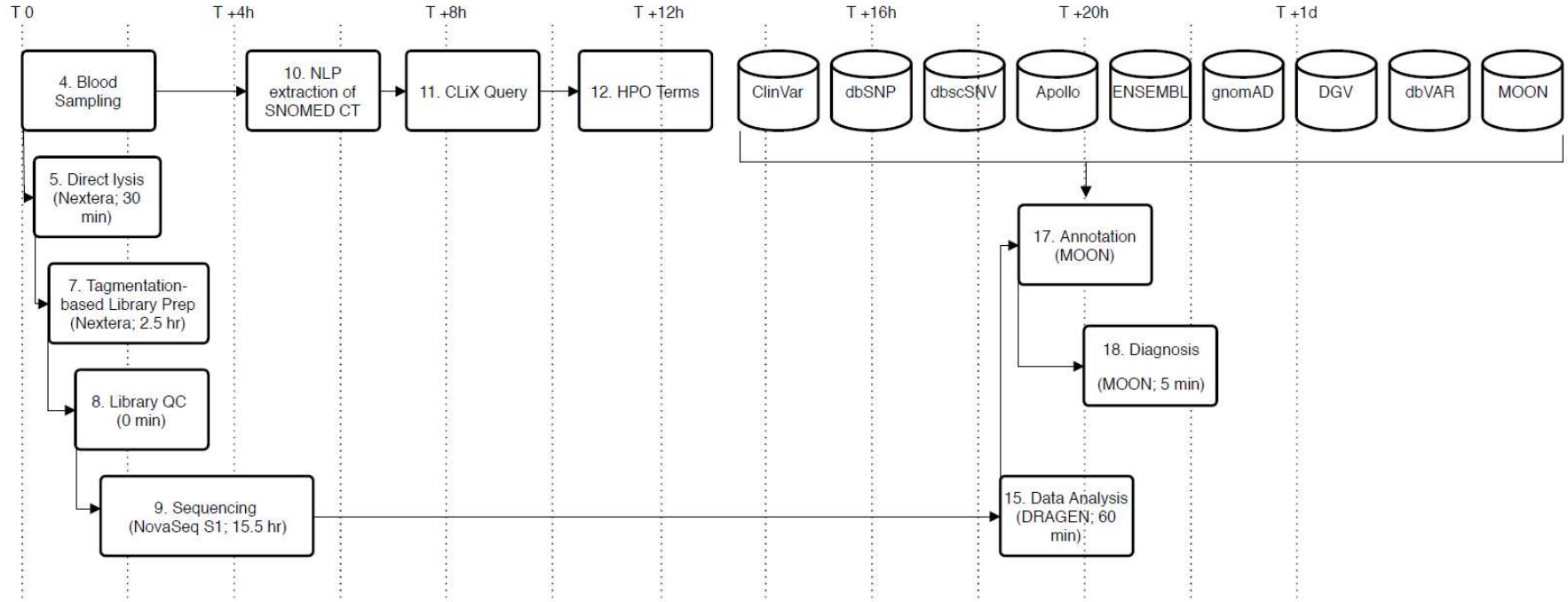

Figure 2: Process map of NICU-Seq pipeline from Clark et al. (2019). This pipeline is more limited in scope and does not include pre-sampling or treatment decisions as in Figure 1. Steps are labelled to correspond as best as possible with similar processes in Figure 1.

This process reflects an updated pipeline in use at the Rady Children's hospital in San Diego. Median TaT for clinical samples is 20:10h, but this notably excludes pre-test (consent, sampling), and interpretation or reporting activities. This pipeline differs significantly from the 2015 record. Firstly, it takes advantage of the new NovaSeq platform. While this does not result in significantly lower TaT (due to the HiSeq modifications of the previous workflow), it does provide significantly higher read depth using stock sequencing protocols. The molecular biology pipeline is wholly different, and uses a tagmentation-based direct whole-blood method and library amplification rather than a PCR-free, shearing and dsDNA-ligation protocol. Clinician phenotyping has been replaced with an automated, NLP algorithm, and a second AI filters and prioritizes variants and provides geneticists with a short-list of diagnoses. While the automated production of phenotypic terms from the EHR likely has no affect on TaT, the automated diagnosis has the potential to drastically impact TaT providing it is accurate and trustworthy. Overall, most of the time savings of this pipeline over the 2019 version arise from excluding upstream sampling and interpretation from the process, by using a direct-blood tagmentation-based library prep, by optimizing most processes to take less time than manufacturers indicate, and by automating the diagnosis with machine-learning. Points below refer to steps in the above process map (Figure 2), and are harmonized with steps in Figure 1. For example, point 8 discusses library prep quality control protocols in both pipelines (Figure 1 and Figure 2).

5. To increase TaT, a direct library preparation from whole blood or punch cards was used. From whole blood, 10 µL of EDTA blood was used with the optional blood lysis protocol from the Nextera DNA Flex Library prep protocol. Briefly, whole blood is lysed in a supplied lysis buffer and proteinase K and purified with an on-bead protocol. Estimate processing time is 40 min.

7. In contrast to the 2015 workflow, a tagmentation-based library prep was used, followed by 5 cycles of PCR, as is usually necessary with enzymatic methods. According to Illumina, the whole procedure including post-library amplification requires 3.5 h for small numbers (<16) of samples. The authors report that library prep takes 2:45 h.

8. Library QC was done with a pico-green assay only, and as in the previous workflow no fragment analysis was conducted. The use of direct lysed blood as an input material, without DNA quantification, pre-library prep fragment QC, and with only a fluorometric (no qPCR or fragment analysis) QC step pre-sequencing increases the probability of sequencing low-quality libraries, but skipping these steps does increase TaT. Estimated time for library QC is 30-60 min (however may be 60-90 min if the same protocol was used in 2015).

9. Libraries were not pooled. Simple samples were sequenced with a NovaSeq 6000 and S1 flowcell with a 2x101nt PE kit. According to the manufacturer, runs take 19 h and typically deliver 266-333 Gb pass-filter data, which corresponds to 80-100x coverage before removing duplicates and quality filtering. According to the authors, their S1 runs take 15:30 h and typically deliver 404-537 Gb of data, which is significantly more than This represents significantly more data per sample than many other pipelines produce. For duo or trio analysis, it appears that each sample was ran separately on a NovaSeq, either increasing TaT due to sequential analysis or necessitating the concurrent availability of 2-3 instruments.

10. While in 2015 a manual process was used to directly developed HPO terms from the EHR, an automated method for developing HPO terms was developed here. Clinical records for the patient were exported from an Epic EHR into a proprietary .json format and loaded into CLiX ENRICH (Clinithink). This data was processed by an NLP package, which coded entries into post-coordinated SNOMED CT expressions. The medical records of 16 children were used to train CLiX ENRICH.

11. CLiX query library: the CLiX software is built around SNOMED CT expressions, while HPO phenotypes are needed for diagnosis. The lab developed a custom library of CLiX queries using a semi-automated process. HPO terms were passed through the CLiX encoding engine, which generated CLix post-coordinated SNOMED CT expressions (i.e. queries) for each recognized HPO term. Where no exact matches were found, expressions were generated manually, and where no matches or incorrect matches were found, entries were added to the SNOMED CT terminology files in Clinthink to ensure appropriate matches between SNOMED CT and HPO were available. The query set used for diagnosis covered 60% (7706 of 12786) HPO terms.

12. After NLP generated a set of SNOMED CT terms from the unstructured EHR data, the library of CLiX queries were applied. Exact matches or matches to a SNOMED CT parent/child in the SNOMED CT terms derived from the patient EHR record via NLP returned the corresponding HPO term mapped out in the CLiX query library. This results in a list of HPO terms for each patient. NB: Versus the 2015 pipeline, while this process for developing HPO terms may be more standardized and certainly requires less clinician input, it is unlikely to influence TaT as this can occur concurrently with sequencing.

15. Primary data analysis including alignment and variant calling were performed with an Illumina DRAGEN processor (v2.5.1) following the same protocol as in the 2015 paper. Due to the increased amount of data per sample, it's likely that processing time is higher than in the 2015 paper: between 90-110 minutes depending on throughput and assuming no hardware or software bottlenecks are introduced. The authors report that alignment and variant calling takes a median of 1 h for 150 Gb of sequence, suggesting that they do not use most of the data generated by the sequencer.

17. Variant annotation and interpretation was automated with MOON (Diploid software). Variants were annotated with ClinVar, dbSNP, dbscSNV, Apollo, Ensembl, GnomAD, HPO (see 12.), DGV, dbVAR, and MOON.

18. MOON uses a black-box machine learning approach to generate a list of provisional diagnoses using various methods including NLP, Bayesian models, decision trees, and neural networks. No information on how this process is conducted or how the software is trained is available. A complex series of filters was applied by MOON, but what these are is not specifically indicated in most cases, but do rely on various annotations. This software outputs a ranked list of candidate variants with associated diseases. Presumably a geneticist or other professional evaluates these diagnoses, the quality of data underlying them, and considers confirmatory testing, however these are not considered part of the 20:10 h pipeline in this study.

6 DRIVERS OF TURNAROUND TIME

After reviewing this process map, several findings were identified and potential areas for further investigation were ranked by OUS and DNV GL. Key drivers in the pipeline are the sequencing process (which can take from under one to several days depending on the sequencer and reagent choice), bioinformatics (in particular read alignment and variant calling), and interpretation (which depends heavily on infrastructure available to clinical geneticists and up-stream annotation efforts). This is supported by the in-depth mapping of the CMKC pipeline as well as experience from OUS in rare disease diagnostics.

6.1 Sequencing

There are a limited number of design decisions available with regards to sequencer and reagent choice. While the example pipeline involved modifying an Illumina HiSeq, this approach requires significant technical skills, may make troubleshooting difficult, and makes upgrades to chemistry or software difficult if not impossible. Outside of 'hacking' the sequencer, discussions turned to sequencer and reagent choice. One of the clinical requirements for this pipeline is that duos or trios can be sequenced concurrently, since this reduces the number of false positive variants significantly and can aid in interpretation. This could be either done by running samples on separate sequencers (which increases cost significantly and introduces scheduling, capacity, and reproducibility issues), or by running a higher-throughput instrument and multiplexing samples.

1. AMG has access to most appropriate sequencing platforms.
2. For 30-35x coverage, many centers aim for around 100Gb raw data per sample.
3. Possible sequencing options are listed below. Note that any choice will require tradeoffs between run time, throughput, and cost (not included as a factor here, since institutional pricing varies significantly).

4. If the aim is to support trios with a short TaT, the NovaSeq running short reads with S1 (2x100 for 19h or 2x150 for 25h) or S2 (2x50 for 16h) flow cells is likely the most appropriate platform. Note that cost has not been considered here, nor have downstream considerations due to differences in read length (ie. greater utility of split-reads alignment with longer chemistries for rearrangements).

Platform	Flowcell	Reagents	Output	Run Time
NextSeq 550	High-Output Flowcell	2x150bp	100-120Gb	29h
HiSeq 3000/4000	Single Flowcell	1x50	105-125	1d
		2x75	325-375	~3d
		2x150	650-750	3.5d
HiSeq X	Single Flowcell	2x150	800-900	3d
NovaSeq 6000	SP	2x50	65-80	13h
		2x150	200-250	25h
		2x250	325-400	38h
	S1	2x50	134-167	13h
		2x100	266-333	19h
		2x150	400-500	25h
	S2	2x50	333-417	16h
		2x100	667-833	25h
		2x150	1000-1250	36h
	S4	2x100	1600-2000	36h
		2x150	2400-3000	44h
Output: Single sample only, duo possible, suitable for duos+trios, excess capacity Time: >1.5 days Costs not included in this table: Service-level costs in US from one institute are 7,4kUSD for S1 PE100, 12k for S2 PE50.				

An additional factor influencing flow-cell and reagent choice is process scheduling: there is no benefit to TaT from choosing a faster sequencing protocol if that introduces a wait time prior to the next step. In the CMKC example pipeline, library prep was completed in the afternoon of day 1 of testing, and sequencing ran from the evening until the middle of day 2. An automated set of scripts handled data transfer and initiated the bioinformatics pipeline, but by scheduling this in the middle of the day automation would not be a strict requirement. This combined with automated bioinformatics processing allowed interpretation to be carried out in the late afternoon of day 2.

5. For the OUS pipeline, assuming sequencing can also begin by 17:00 on day 1 of testing, sequencing on the NovaSeq could end by 09:00 (S2 2x50), 12:00 (S1 2x100), or 18:00 (S1 2x150).
6. If manual steps or human oversight are required to initiate and monitor the bioinformatics pipeline, shorter sequencing times may be preferred.
7. If the entire bioinformatics workflow can be completed within several hours, interpretation may be possible in the afternoon of day 2. If the bioinformatics pipeline takes longer than several hours, it may make sense to choose a longer sequencing run, initiate the bioinformatics in the afternoon of day 2, and begin interpretation on day 3.

6.2 Primary Data Analysis

One of the major drivers in bioinformatics pipeline run time is the time required for aligning reads, removing duplicates, and calling haplotypes (.fastq to .bam or .cram). The second most computationally expensive task is calling variants (.bam or .cram to .vcf). Together, these two steps can easily comprise over 1d in processing time. For fast TaT in a NICU setting, providers are essentially locked into using the Illumina DRAGEN processor, a specialized FPGA, hardware, and software platform for fast read alignment and variant calling. Other than optimizing scheduling, I/O to a local or AWS DRAGEN node, or similar tasks, there are few design decisions that could be made here. Note that several hospitals have found that scripting and automating transfer between tools within the bioinformatics pipeline improve TaT, partially due to less hands-on time, but more importantly through the ability to execute processes during off-hours.

6.3 Interpretation

Perhaps the greatest opportunities for improving TaT come from optimizing the annotation, interpretation, and reporting process. There is great variability in pipelines, which often comprise mixes of open-source tools, commercial software, and manual expert analysis to varying degrees. Furthermore, interpretation time can vary significantly from case to case, depending on the inherent difficulty, how much automated (provisional) diagnosis can be relied on, and how well geneticists are supported by interpretation portals and clinical decision support software. In practice, while pipelines analysed often claim TaT of around a day, in routine clinical use mean TaT often stretches to weeks, often due to difficult interpretation.

While a number of discussions surrounding what the appropriate databases, annotations, and user-facing platforms for variant interpretation are beyond the scope of this work, the question of how to best integrate phenotypic data into the pipeline is important, and it was decided to focus future efforts on this topic.

7 OVERVIEW OF RELEVANT HEALTH ONTOLOGIES

An overview of rapid NICU-seq and adult rare-disease pipelines and various studies show the utility of phenotypic information in promoting both speed and diagnostic yield. Phenotypic data in a rare disease setting is most often used in one of two ways: either to determine the contents of a molecular test (a so-called panel or filtering approach), or to guide the interpretation of a broad WGS or exome approach through prioritizing or otherwise ranking variants within genes related to disorders that affect the patient phenotype. Regardless of which approach is taken (filtering or prioritizing), labs and medical device manufacturers need to examine carefully how and what data they collect, how this is used in practice, and how this integrates with the IT of the wider health system. We conducted a short survey of medical ontologies here, meant as a high-level overview. Ontologies were chosen based on their use by common bioinformatics tools or health record systems, their capacity to describe human disease or phenotypic information, and their relevance in the modern health system.

a. Unified Medical Language System:

i. UMLS (https://www.nlm.nih.gov/research/umls/) is a metathesaurus that links terms and codes from several ontologies and groups concepts in a semantic network. UMLS unifies hundreds of application-specific vocabularies (https://www.nlm.nih.gov/research/umls/sourcereleasedocs/index.html), including ICD-10, LOINC, MeSH, HPO, and SNOMED CT. A mapping of SNOMED CT to ICD-10 codes is maintained by the NLM, however direct mappings of other ontologies including HPO are not available.

b. SNOMED CT:

i. Owned by the International Health Terminology Standards Organization (IHTSDO) since 2007, this standard encompasses around 350 000 medical concepts in a structured ontology. At its core, SNOMED consists of clinical concepts, which consist of a description (including a fully-specified name and one or more synonyms), Unique identified, and relationship(s) with other concepts. The department of e-Helse has recently completed an evaluation project of medical ontologies and has begun steps to develop a Norwegian version of SNOMED CT (https://ehelse.no/nyheter/ny-utgave-av-helsespraket-snomed-ct-er-lansert-pa-norsk).

c. ICD-10:

i. Maintained by the WHO, ICD is an ontology of medical diagnoses put in place in many national health systems (sometimes as-is, and other times modified, such as ICD-10-CA in Canada). ICD-10 consists of around 68 000 diagnoses, symptoms, and additional information that can be used to describe a particular diagnosis. While ICD-10 is limited to diagnoses (some of which are very specific), SNOMED concepts may be diagnoses, symptoms, or other data. Both ICD-10 and SNOMED are supported in many EHR systems, and both are compatible with HL7 FHIR. Information on Norwegian implementation here (https://ehelse.no/standarder-kodeverk-og-referansekatalog/helsefaglige-kodeverk/kodeverket-icd-10-og-icd-11)

e. LOINC:

i. LOINC is a research-grant funded project that provides a set of codes (https://loinc.org/panels/) and terms for types of health observations, documents, and measurements, including codes for particular lab tests. NGS tests would be included under Laboratory Order-Mol Pathology Panels, for example. LOINC is compatible with HL7, and works in conjunction with SNOMED CT concepts.

ii. There is a Phenotypes domain panel, but this is meant to capture protocols for determining exposure or other rough phenotypic testing (ie. the skeletal panel includes only a few protocols for testing for specific bone disorders, no way to include an inborn phenotype). The ontology is designed to allow users to describe particular tests within a hierarchy of test types, not phenotypic patient data *per se*.

f. HPO:

i. Academic project (core of the Monarch NIH grant) develops an ontology of >13 000 phenotypes in a structured hierarchy along with ORPHA disease associations, ORPHA and OMIM gene associations, and LOINC associations.

ii. Most commonly used ontology for NGS testing for rare disease.

g. HL7:

i. Within HL7, phenotypes can be included as observations (http://hl7.org/fhir/observation.html). Observation category is likely exam, which is a catchall for general physical findings and other direct observations made by clinicians. These can be added as short text, in which case clinicians could upload free text describing the phenotype or HPO terms directly. In either case this would not be stored as structured data, so the integration of phenotupic data within HL7 is weak at best.

ii. There is an active HL7 clinical genomics working group that has proposed a domain information model for genotype, phenotype, and interactions https://www.hl7.org/documentcenter/public_temp_E18464AA-1C23-BA17-0CCD810949D743D2/wg/clingenomics/docs/HL7%20Clinical%20Genomics%20Overview%20-%20May%202015%20-%20CG%20DIM%20-%20Shabo.pdf

iii. More information here, not clear that this is in progress or implemented. Based around a genotype-phenotype association data model: http://www.hl7.org/documentcenter/public_temp_E18464AA-1C23-BA17-0CCD810949D743D2/wg/clingenomics/2018%2004%2003%20-%20V2%20LRI%20-%20Ch.%205%20CG%20and%20Code%20System%20Tables.pdf

A recent publication from the Norwegian directorate for e-health covers similar topics here: https://ehelse.no/standarder-kodeverk-og-referansekatalog/helsefaglige-kodeverk

8 SELECTED TOOLS USING PHENOTYPIC DATA

We conducted a non-exhaustive survey of tools used by labs in rare disease diagnostics for collecting, curating, and automating parts of their analysis based on phenotypic data. A subset of bioinformatic tools which are used in clinical sequencing pipelines at other institutions are presented below.

1. Phenomizer: (Koehler, et al. 2009) http://compbio.charite.de/phenomizer/
    a. Open source, RUO software for differential diagnosis.
    b. Clinicians select a set of HPO terms (complete ontology is available, along with ancestors/descendants in tree). HPO terms are described as observed or mandatory (if a specific HPO term is mandatory, any syndromes not associated with that HPO term will be disregarded).
    c. A statistical model based on semantic similarity is used to generate scores and p values, effectively creating a ranked list of syndromes most-to-least likely to be associated with the HPO terms chosen (differential diagnosis). This can be manually refined.
2. Exomiser (Smedley, et al. 2015) and Jannovar (Jaeger, et al. 2014):
    a. Exomiser and Jannovar are java-based tools for annotating .vcf files. Jannovar annotates variants with transcript definitions, predicted pathogenicity from dbSNP, and population-level variant frequency from several external databases. Exomiser is used to filter and prioritises potentially causative variants.
    b. Both tools are open-source and RUO. It appears that the work has been archived by the Sanger Institute and the developer has moved to a different institution, but continues to make semi-regular updates (https://github.com/exomiser/Exomiser/graphs/contributors).
    c. Inputs include several public databases, .vcf, and HPO terms.
3. PhenoTips (Girdea, et al. 2013):http://www.gene42.com, https://phenotips.org/
    a. Limited EHR for collecting, storing, and analysing phenotypic data (with/without NGS). Supports pedigrees, allows users to browse HPO, query OMIM with one or more HPO terms, and capture phenotypic data through UI.
    b. Different licensing options available. Local implementation possible.
    c. Front-end for building pedigrees, suggests genes for a panel, etc.
    d. If decision is made to use this in practice, several questions:
        i. Who does phenotyping? Do they have access to the tool?
        ii. In practice, using PhenoTips involves implementing a second, NGS-specific EHR within the OUS ecosystem, and should not be viewed as a small commitment. How is this maintained in the future, and how will it be integrated with existing EHR? Specifically, how are HPO terms and ICD codes from PhenoTips integrated into the patient record?
        iii. How to validate the software? Is PhenoTips considered a medical device under MDR? Seems like most functionalities are simple matching and the display of external databases, with the exception of predictive search. Language clearly indicates PhenoTips is for the diagnosis of human disease. Regardless of how

PhenoTips is classified, a second topic is due diligence and QA for the databases used by PhenoTips (or any other tools).

e. Growth charts are based on WHO growth charts for Canada, and do not provide accurate information for other regions. This could be misleading for clinicians using the tool.

f. When entering phenotypic information, PhenoTips provides a real-time assessment of the information content of the terms provided via the Monarch phenotype specificity meter. This tool is not developed for diagnostic purposes in humans, and may give clinicians a false sense of security, suggesting that their entries are sufficient or of high quality. Furthermore, feedback from this tool may prompt clinicians to add additional phenotypes, re-phrase entries to similar terms that provide a better score, or otherwise enter inaccurate or superfluous data.

g. After entering phenotypic information, PhenoTips provides a set of suggested genes which can be used to prioritize or filter genes for analysis. This tool indicates which of the HPO terms included in the analysis lead to inclusion for each gene. While this does provide some level of transparency, it is not clear how this module determines a given gene is included. Several examples with test HPO terms included genes with no clear link to the phenotypes in question when viewed in OMIM. PhenoTips also provides links to Ensembl and RefSeq entries for each gene on the suggested list.

    i. OMIM entries are submitted when mutations are published in the scientific literature and are reviewed by staff at Johns Hopkins, paid for through crowdfunding. OMIM is free for research use only, but requires licenses for commercial use and is not intended for the diagnosis of disease. OMIM entries do include the names of contributors and dates of submissions and edits, but not which content was edited. OMIM entries contain substantial links to related scientific literature, so while thorough do take substantial time to fully evaluate.

4. Patient Archive (http://patientarchive.org/):

    a. Garvan institute's online portal for sharing and storing clinical cases (including HPO terms). Based on HPO and Bio-LarK CR concept recognition (Groza, et al. 2015).

    b. RUO, proprietary.

5. Moon (http://www.diploid.com/moon):

    a. Commercial software developed by Diploid, a small Belgian startup.

    b. .vcf from exome or WGS, age of onset, gender, and HPO terms are entered into web interface. Black-box AI suggests a causative variant. Clinicians can review, or choose to look at additional variants, and add these to a report. Automatically generates a discussion section about that variant, and clinicians manually add a classification and references. Positioned as a competitor to Exomiser. Unclear how the model is built or what sorts of data are used to train or test algorithm. No detailed performance data available.

    c. RUO. Language surrounding diagnostics, possibilities to run unsupervised (and also re-run periodically on previous results) and issue automated reports most likely make this a standalone medical device under MDR.

9 ONTOLOGIES IN PRACTICE

To inform design choices in the OUS NICU-seq pipeline regarding the gathering and use of phenotypic data in rare disease diagnostics, we conducted a survey of practices amongst other NICU-seq pipelines in use by both hospitals and commercial diagnostics providers.

1. **Blueprint Genetics**: A commercial integrated diagnostics provider. Relies on free text with submissions, interpretation done by in-house geneticists.
2. **Children's Mercy Kansas City**: Manual review of EHR and interviews with family to gather HPO terms in SAGA software (Thiffault, et al. 2019).
3. **Rady San Diego**: CLiX software to generate SNOMED CT codes from EHR free text, custom SNOMED CT to HPO mapping for a limited set of SNOMED CT concepts that relate to HPO terms, then use tools developed for HPO for variant prioritization (Clark, et al. 2019).
    a. Fastest performance previously was 26h, 37h to diagnosis in a clinical setting, but mean time to diagnosis was ***16d***. Here median TaT 20:10h.
        i. Old pipeline: Manual order, TruSeq DNA manual library prep, HiSeq 2500 rapid run, DRAGEN v1, HPO from EHR via manual review. HPO and vcf loaded into Opal (Fabric). Manual transfer of some files (Miller, et al. 2015).
        ii. New pipeline (Clark, et al. 2019): ePortal order from EHR. Manual or automated Nextera, NovaSeq S1, DRAGEN v2, NLP for HPO terms from EHR. Automated comparison of HPO and vcf (differential diagnosis) through custom platform and display highest-ranked answers. Scripts to transfer files. HPO and vcf transferred to Moon (Diploid, belgium) for automated diagnosis (direct competitor to exomiser, runtime ~1h).
    b. Their ICD and DRG codes are sparse and not specific enough, this is the need for NLP.
    c. Use CLiX enrich NLP to generate SNOMED CT. Manual mapping of SNOMED to HPO for many terms.
    d. Very small training/test sets for NLP vs. other AI applications.
    e. Moon giving too many FP, using InterVar (Li and Wang 2017) to filter out variants post-annotation. InterVar is an open source, RUO package which automatically annotates variants with classifications according to ACMG criteria. This pipeline only retains variants classified as P and LP.
    f. Scheduling library prep in PM, sequencing and data analysis O/N, and reporting following AM.
4. **Centogene**: Ordering physician checks off phenotypes from a pre-defined list (not given as HPO terms). This is a slightly different context, as physicians are also required to pre-determine the disease to test for (provided from a 93-page list of diseases and their respective OMIM gene, broken down into several categories). Seems they are using a panel-first approach, in contrast to many programs.
5. **Sick Kids Toronto**: PhenoTips (Girdea, et al. 2013) used by clinical geneticist to generate HPO terms for prioritization.

7. **Genomics England**: For rare disease program, allows clinicians to choose from a curated set of HPO terms, broken up by specialty (see rare disease data models: 352-page guidance document that presents selected HPO terms by suspected disorder group) https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/information-for-gmc-staff/rare-disease-documents/. There is also a guidelines doc on developing an HPO model for rare diseases. Like Centogene, these seem to be used in a panel-first approach.

    a. Disease models are broken up into 21 medical specializations.

    b. These guidelines are fairly subjective (ie. aim for 20-40 HPO terms per disease, avoid 'common' terms), but it seems that disease model development is transparent and open to criticism and additional layers of internal continuous improvement.

    c. Within the guidance documents, phenotypes are listed in HPO terms, but they also supply indications for testing and a list of clinical tests. Disorders are also ordered in a structured hierarchy based on clinical specialty. Two categories are available (ultra-rare disorders and genomic medicine service indications) in the event that clinicians have no idea of the type of disorder affecting the patient.

    d. Genomics England uses Open Clinica for submission and data capture and has a separate tool (labkey) for validating data. More on HPO data capture here (https://community.openclinica.com/sites/fileuploads/akaza/cms-community/marketing/OC15_Matser.pdf)

    e. Other sources of disease data models: NIH GARD has HPO terms tiered into commonness for each disorder (https://rarediseases.info.nih.gov/diseases/, under symptoms). Not all disorders have HPO terms, and for many disorders phenotypic information is more presented as text.

    f. OMIM contains extensive clinical data for each entry, however these are not coded as HPO. While these are less interesting from an automated differential diagnosis perspective, referencing OMIM and/or GARD as a part of the diagnostic process after candidate variants are presented seems beneficial.

    g. The HPO database directly links to Orpha and OMIM diseases, and presents a list of HPO terms divided into high-level categories for the disease in question. Many diseases have a long list of HPO terms (60-70), and these have no indication of prevalence, so the primary diagnostic power of this may be limited.

10 CHOOSING AN ONTOLOGY FOR PHENTOYPIC DATA

Given this background understanding of available ontologies and tools for incorporating phenotypic data into diagnostics, we examined how exactly phenotypic data in incorporated within each ontology and which tools and processes could be used in the OUS context.

The first question to address is what ontology to collect phenotypic data in. Primary choices here are SNOMED CT and HPO. Other ontologies may be useful, but primarily serve supporting purposes.

SNOMED CT is used more widely in healthcare, and alongside a commitment to develop Norwegian resources based around SNOMED CT is likely to tie into more medical services in the future. HPO, in contrast, is used more widely in the rare disease and translational research context. While

interoperability with other medical services may be low in the future, this does however mean that systems for differential diagnosis arising in the research setting are available, and that HPO data is more actionable in the context of rare disease.

There are dozens of tools for gathering and utilizing HPO data for the purpose of rare-disease diagnostics already, and notable key initiatives across the globe rely on HPO. Additionally, almost all pre-existing NICU-seq pipelines rely solely or significantly on phenotypic data in HPO format as part of their variant filtering or prioritization processes, providing a comprehensive pre-existing body of experience to build from. While SNOMED CT contains roughly 27x more entries than HPO, most phenotypes from HPO do not have equivalents in SNOMED CT (roughly 70%) or UMLS (roughly 66%) (Dhombres and Bodenreider 2016).In general, the deep phenotyping needed for variant prioritization is of limited use in a wider medical context, however is essential for rare disease diagnostics.

In light of the merging of research and clinical infrastructures in genomics and the greater prevalence of NGS-based diagnostics in healthcare, there has been public commitments from both SNOMED CT and GA4GH to develop comprehensive bi-directional linkages between SNOMED CT and HPO in the future.

We found two published HPO to SNOMED CT mappings as of May 2019. One manual mapping of HPO to SNOMED CT terms found only 30% complete equivalence, and partial (lexical or logical) mappings for an additional 62% of the HPO corpus (Dhombres and Bodenreider 2016).

A second set of partial mappings was developed and presented in a 2019 paper (Clark, et al. 2019). In brief, the authors used NLP software to automate the extraction and encoding of HPO terms from free text from EHR entries. To do this, they used a commercial software that analyses free text and generates SNOMED CT concepts. Starting from the HPO corpus, they developed queries for approximately 60% of HPO terms, that would search for appropriate SNOMED CT concepts. In the final implementation, software automatically scours the EHR and generates SNOMED CT concepts, which are searched with the HPO query library. If encoded data matches the query, the corresponding HPO term is added to the list for that patient. The list of automatically generated HPO terms can then be used for diagnosing rare diseases, again with an automated machine learning pipeline.

Given the predominance of HPO-based input tools and moves towards a comprehensive HPO to SNOMED CT mapping in the future, it seems that gathering, processing, and storing phenotypic data as HPO terms is likely to be a more productive path forward in the near term, with the understanding that should this deep phenotyping data be integrated into other health services in the future, mapping to SNOMED CT will be possible without changing the up-front processes used at OUS.

11 USING HPO IN THE CLINIC

In addition to verifying and validating the software and databases used for storing and using HPO terms for rare disease diagnosis, there are numerous other considerations for OUS before implementing a clinical pipeline incorporating phenotypic information.

Firstly, SOPs and process maps for the phenotyping, data collection, and use of HPO terms in diagnosis should be codified and compliant with the existing quality management system. As some of these processes are by necessity subjective, test plans for both the phenotyping procedure and the use of phenotypic data in variant prioritization should be developed, and tested out in a clinical investigation. Particularly as both the phenotyping and interpretation can potentially require a broad range of stakeholders, sufficient training materials and a clinical roll-out plan should be developed, with the goals of ensuring accurate and consistent phenotyping and use of phenotypic data.

Additional technical questions require decisions. Firstly, whether to allow clinicians to input free-form HPO terms, or whether these should be limited to specific disease-directed inputs. Free-form term entry may lead to an abundance of non-specific or irrelevant terms, decreasing diagnostic utility, and to decreased consistency between clinicians. Additionally, while some terms may be highly clinically relevant for certain types of patients, these may in fact have limited or negative clinical utility in other scenarios. A second technical question surrounds whether to use HPO terms to filter or to rank variants. While a strict filtering (ie. in silico panel) approach is more likely to remove a greater number of false positive variants and lead to faster TaT in cases with clear, causative, or previously identified variants, a ranking approach is likely to contribute to higher overall diagnostic yield.

While PhenoTips (Girdea, et al. 2013) is a candidate software for data capture, technical questions surround implementation and the downstream use of this data. Firstly, there is significant regulatory uncertainty surrounding PhenoTips and EU medical device regulations. The manufacturer's marketing language clearly indicates a diagnostic purpose under MDR: "PhenoTips® is a software tool for collecting and analyzing phenotypic information of patients with genetic disorders.", and some modules of PhenoTips do appear to function in clinical decision support. However the manufacturer does not appear to have an appropriate quality management system in place, nor is it clear who their authorised representative in the EU is, even though their software has been made available on the market.

Without suitable regulatory approvals, it may be possible for OUS to implement PhenoTips as a custom device, however the complete technical qualification and quality assurance for this process lies with OUS itself. Similarly, many open-source packages are available for downstream differential diagnosis, variant prioritization, or filtering based on HPO terms, and similar regulatory and safety considerations should be made when considering implementing these.

Long-term, assuming robust and standardized systems for gathering and storing HPO data are put in place, the possibility of moving these from clinical genomics-specific databases in AMG into general-use EHR should be considered. In such a case, whether to incorporate these as HPO, or to validate and use a SNOMED CT to HPO mapping or similar technology, and how to identify and mitigate risks for inappropriate use of this data in the broader health context will need to be addressed.

12 CONCLUSIONS

We conducted a systematic process mapping of rapid WGS pipelines for the diagnosis of rare disease in a NICU setting, and identified 3 main bottlenecks that determine turnaround time.

1. Firstly, while library preparation and upstream molecular biology does consume several hours, there is great variability in the time required to sequence samples. By choosing a reagent kit and sequencer that can sequence duos or trios in around one day, TaT can be reduced from half a week to less than 24 hours.

2. Secondly, upstream bioinformatics, consisting of .fastq QC, read alignment, and SNP calling, can take many hours with custom bioinformatics pipelines and manual jobs. By utilizing a DRAGEN FPGA (Illumina) and automating file transfers and scheduling steps with scripts, this part of the pipeline can be reduced from >16 h down to around one hour in a production setting.

3. Finally, the variant annotation, prioritization, and interpretation parts of the process can vary significantly based on the quality of available data and the difficulty of the diagnosis in question. In some straight-forward examples with trio sequencing and previously identified, clearly pathogenic variants, interpretation can take as little as a single hour, however in a production setting this part of the process can take many days. There are a wide variety of systems and tools available to geneticists performing this work, and these greatly impact the speed and accuracy of interpretation.

In particular, the use of phenotypic data to automate, prioritize, or otherwise aid geneticists in interpreting cases has the potential to drastically decrease TaT. We examined existing medical ontologies that capture phenotypic data, of which HPO is the most relevant for this work. We reviewed several tools that can aid in using phenotypic information for rare disease diagnosis, and examine how phenotypic data is used at several health institutions and integrated diagnostics providers. Overall, while the benefits of using phenotypic data for rare disease diagnosis are clear, there are wide discrepancies in how this is implemented: some sites rely on mostly manual protocols, while others have automated the extraction, annotation, and interpretation to various degrees.

For the primary purposes of NGS diagnostics, it will be more useful to collect data as HPO terms via a tool such as PhenoTips in the short term, and to abstract or translate to SNOMED CT terminology for use in the wider health system at a later date if desirable. Given the relatively recent addition of NGS-based diagnostics and rapidly evolving understanding of many genetic disorders, a prioritization or ranking approach, where HPO data is used to change the order of which variants are displayed to the geneticist, is a safer choice than an *in silico* panel or filtering approach, through which variants are excluded from analysis.

Several open questions remain about the clinical implementation. Firstly, decisions regarding tool choice, data flow, and roll-out plans need to be taken by OUS. Tools should be integrated into their existing quality management systems, and thorough technical qualification and an evaluation of the safety and risk profile of the proposed solution should be conducted.

13 REFERENCES

Baird, Patricia A., Terence W. Anderson, Howard B. Newcombe, and R. Brian Lowry. 1988. "Genetic Disorders in Children and Young Adults: A Population Study." *Am. J. Hum. Genet.* 42:677-693.

Baldovino, SImone, Antoni Montserrat Moliner, Domenica Taruscio, Erica Daina, and Dario Raccatello. 2016. "Rare Diseases in Europe: from a Wide to Local Perspective." *IMAJ* 18:359-363.

Black, Nick, Fred Martineau, and Tommaso Manacorda. 2015. *Diagnostic odyssey for rare diseases: exploration of potential indicators.* London: Policy Innovation Research Unit.

Clark, Michelle M., Amber Hildreth, Sergey Batalov, Yan Ding, Shimul Chowdhury, Kelly Watkins, Katarzyna Ellsworth, et al. 2019. "Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation." *Sci. Transl. Med.* 11, eaat6177.

Dhombres, Ferdinand, and Olivier Bodenreider. 2016. "Interoperability between phenotypes in research and healthcare terminologies-Investigating partial mappings between HPO and SNOMED CT." *J. Biomed. Semantics* 7:3.

EURORDIS. 2009. *The voice of 12,000 patients: experiences and expectations of rare disease patients on diagnosis and care in Europe.* Boulogne-Billancourt: EURORDIS.

Farnaes, Lauge, Amber Hildreth, Nathaly M. Sweeney, Michelle M. Clark, Shimul Chowdhury, Shareef Nahas, Julie A. Cakici, et al. 2018. "Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization." *npj Genomic Medicine* 3:10. doi:10.1038/s41525-018-0049-4.

Girdea, Marta, Sergiu Dumitriu, Marc Fiume, Sarah Bowdin, Kym M. Boycott, Sebastien Chenier, David Chitayat, et al. 2013. "PhenoTips: Patient Phenotyping Software for Clinical and Research Use." *Human Mutation* 34(8):1057-1065.

Groza, Tudor, Sebastian Koehler, Sandra Doelken, Nigel Collier, Anika Oellrich, Damian Smedley, Francisco M. Couto, Gareth Baynam, Andreas Zankl, and Peter N. Robinson. 2015. "Automatic concept recognition using the Human Phenotype Ontology reference and test suite corpora." *Database* bav500.

Jaeger, Marten, Kai Wang, Sebastian Bauer, Damian Smedley, Peter Krawitz, and Peter N. Robinson. 2014. "Jannovar: A Java Library for Exome Annotation." *Human Mutation* 35(5): 548-555.

Koehler, Sebastian, Marcel H. Schulz, Peter Krawitz, Sebastian Bauer, Sandra Doelken, Claus E. Ott, Christine Mundlos, Denise Horn, Stefan Mundlos, and Peter N. Robinson. 2009. "Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies." *Am. J. Hum. Genet.* 85(4): 454-464.

Li, Quan, and Kai Wang. 2017. "InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines." *Am. Jour. Hum. Genet.* (100): 267-280.

Miller, Niel A., Emily G. Farrow, Margaret Gibson, Laurel K. Willig, Greyson Twist, Byunggil Yoo, Tyler Marrs, et al. 2015. "A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases." *Genome Medicine* 7:100.

Smedley, Damian, Julius O. B. Jacobsen, Marten Jaeger, Sebastian Koehler, Manuel Holtgrewe, Max Schubach, Enrico Siragusa, et al. 2015. "Next-generation diagnostics and disase-gene discovery with the Exomiser." *Nat. Protoc.* 10(12): 2004-2015.

Thiffault, Isabelle, Emily Farrow, Lee Zellmer, Courtney Berrios, Neil Miller, Margaret Gibson, Raymond Caylor, et al. 2019. "Clinical genome sequencing in an unbiased pediatric cohort." *Genetics in Medicine* (21):303-310.

About DNV GL

Driven by our purpose of safeguarding life, property and the environment, DNV GL enables organizations to advance the safety and sustainability of their business. We provide classification and technical assurance along with software and independent expert advisory services to the maritime, oil & gas and energy industries. We also provide certification services to customers across a wide range of industries. Operating in more than 100 countries, our professionals are dedicated to helping our customers make the world safer, smarter and greener.