Targeted Second-Tier Confirmatory Sequencing NBS Pipeline

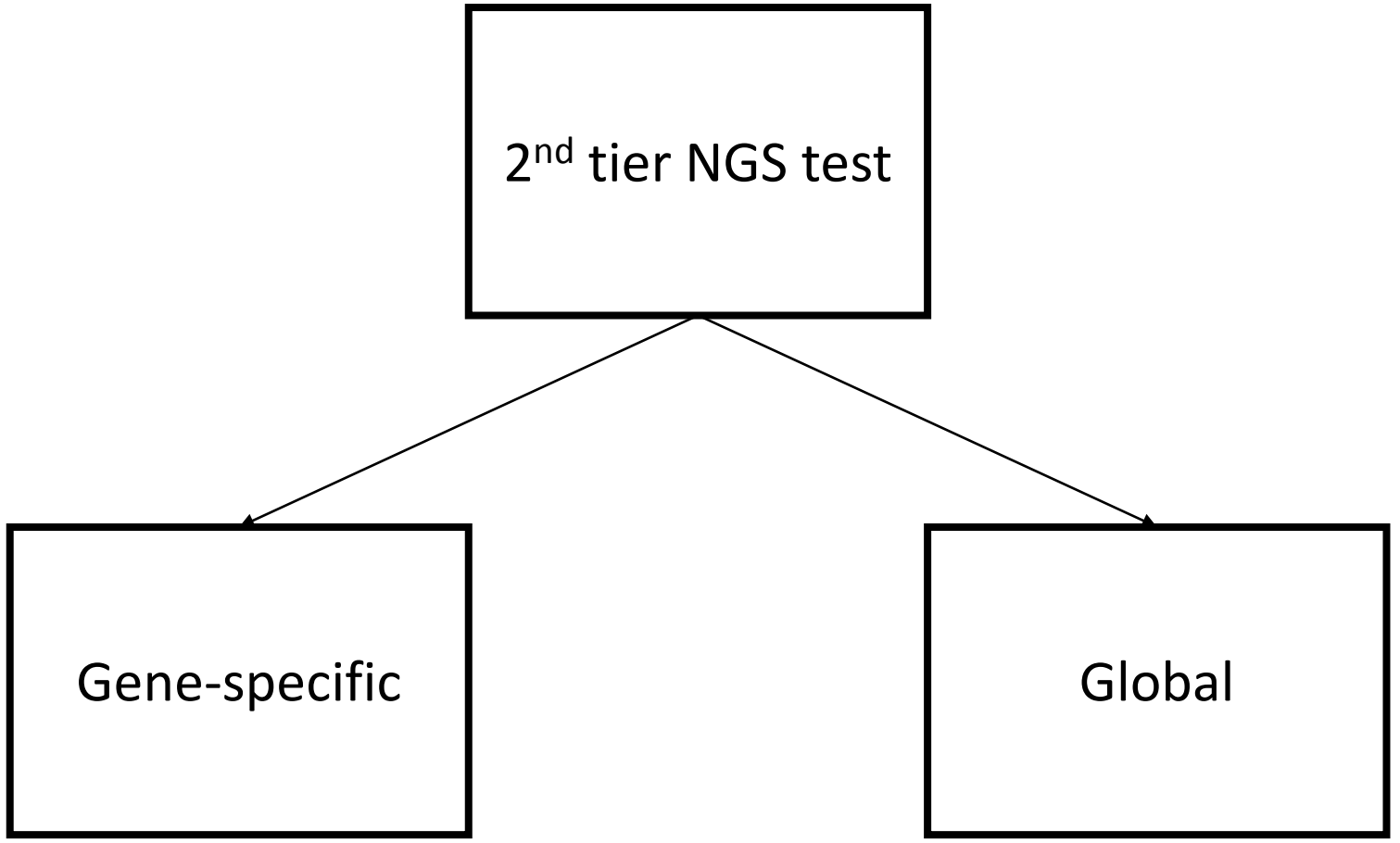
Nicole Ruiz-Schultz

Utah Department of Health Newborn Screening Program

Outline

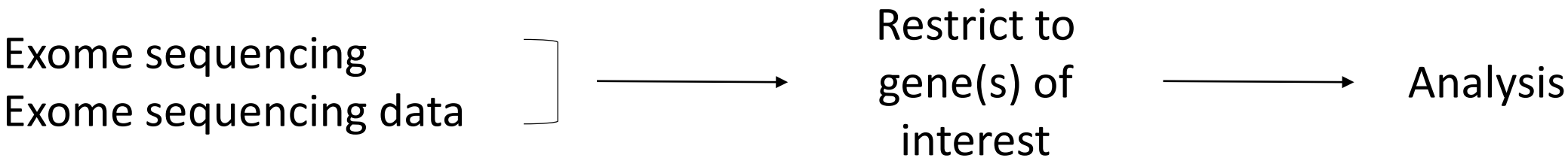
- Next Generation Sequencing (NGS) overview
- NGS pipelines in NBS
- Utah NBS targeted sequencing pipeline
- Variant databases
- Pipeline validation
- Distribution of pipeline
- Significance
- Upcoming work

What is our goal?

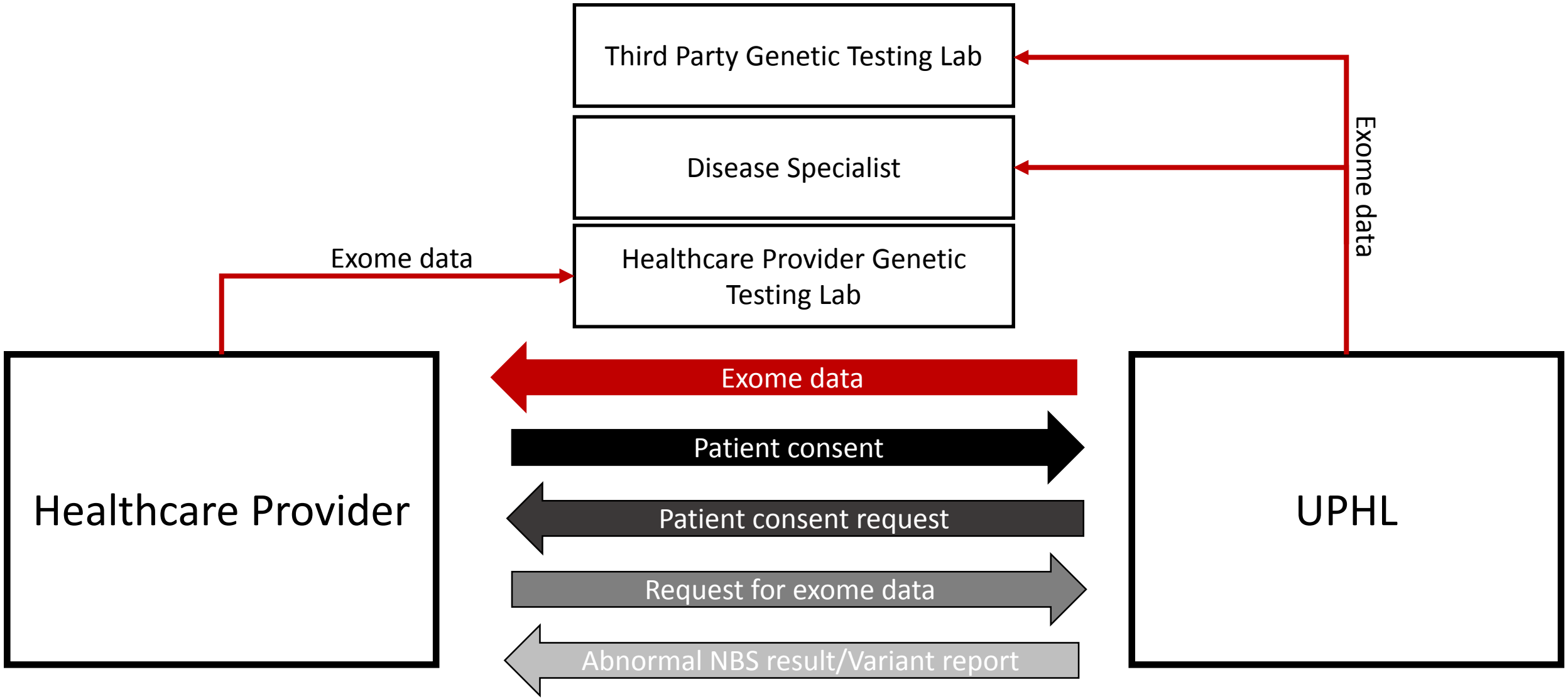


Global 2nd tier NGS test

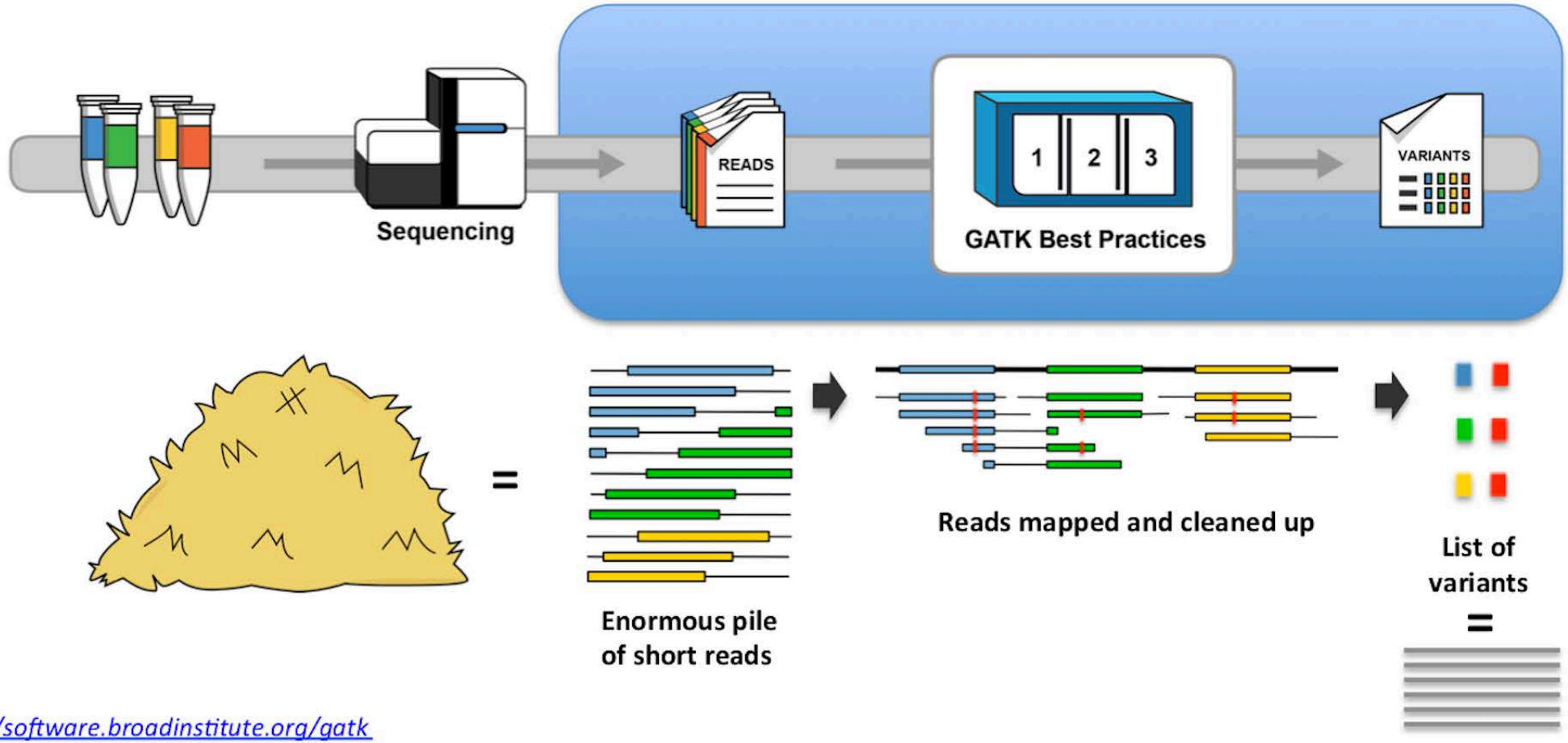
- Our approach: Global method but also gene-specific



Benefit of exome sequencing approach



Overview of NGS pipeline



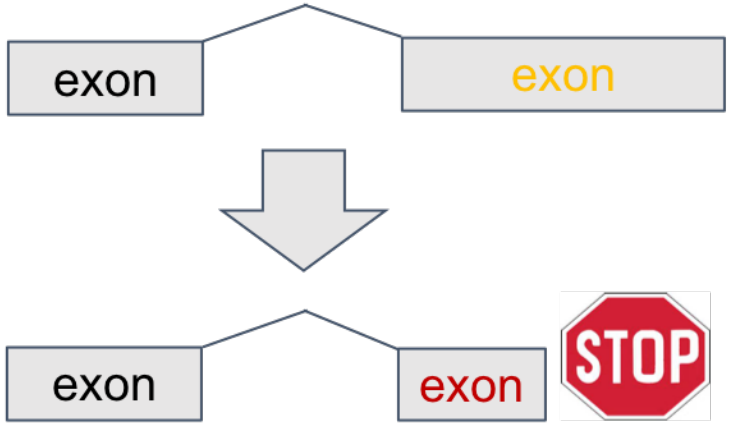
https://software.broadinstitute.org/gatk

What are we looking for in NGS analysis?

Reference sequence …TCAGACATATACCAA…

Newborn sequence …TCAGACATATAGCAA…

Single Nucleotide Variation (SNV)



Effect of SNV: Gain of stop codon which would prematurely truncate protein and potentially inhibit proper protein function

NGS analysis pipeline – High-level overview

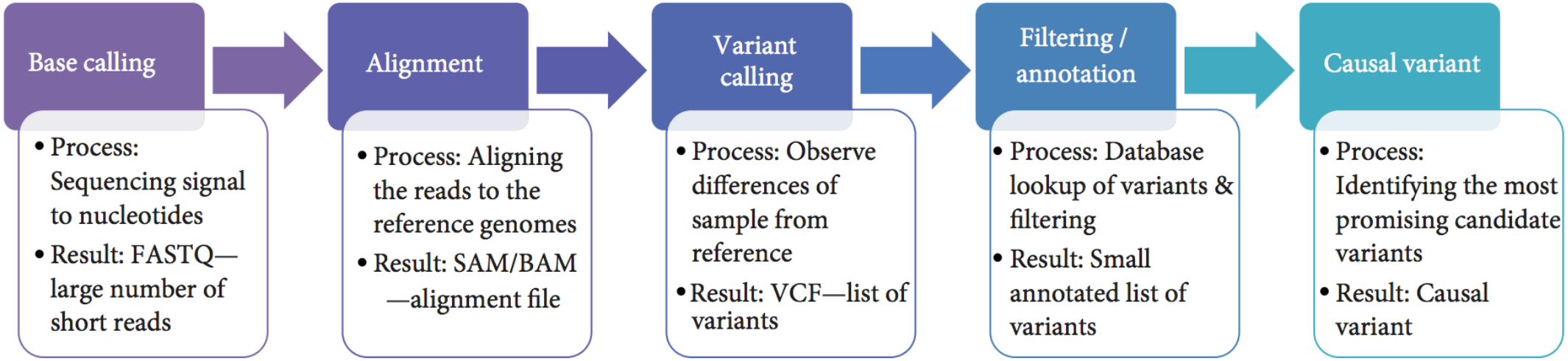


FIGURE 1: Next-generation sequencing bioinformatics workflow.

Dolled-Filhart MP, Lee M, Ou-Yang C-W, Haraksingh RR, Lin JC-H. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. ScientificWorldJournal. 2013 Jan;2013:730210.

Publicly available NGS pipelines

- Galaxy docker exome sequencing pipeline https://github.com/bgruening/docker-galaxy-exome-seq
- NGS-Pipe https://github.com/cbg-ethz/NGS-pipe
- bcbio https://github.com/bcbio/bcbio-nextgen
- ngs-easy https://github.com/KHP-Informatics/ngseasy
- Utah Genome Project (UGP) variant pipeline http://weatherby.genetics.utah.edu/UGP/wiki/index.php/UGP_Variant_Pipeline_1.4.0

Exome capture method

- Illumina Nextera DNA Exome kit
- Sequencing platform: Illumina HiSeq
- Potential issues to troubleshoot
  - Extract enough DNA from dried blood spot
  - Adequate coverage of target genes

Utah NBS sequencing pipeline

- Targeted second-tier sequencing for confirmatory testing
  - Whole-exome sequencing
  - A priori restriction to disease-specific genes
- Why are we choosing this method?
  - Cheaper to sequence entire exome versus using sequencing panels for each disorder (economies of scale)

When will the NGS pipeline run?

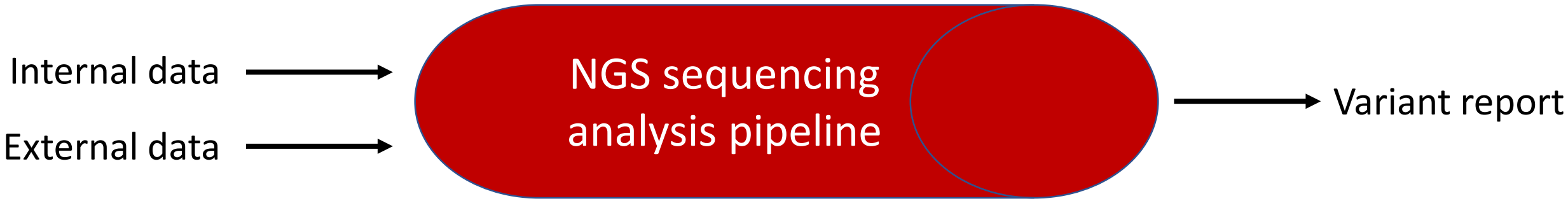
- Cystic Fibrosis
  - 1st screen – Abnormal IRT
  - 2nd screen – Abnormal IRT
  - NGS analysis of CFTR gene

- Hemoglobin disorders
  - 1st screen – Abnormal HPLC
  - 2nd screen – Abnormal HPLC
  - NGS analysis of HBA1, HBA2 genes

- Pompe Disease
  - 1st screen – Low/Absent enzyme activity
  - 2nd screen – Low/Absent enzyme activity
  - NGS analysis of GAA gene

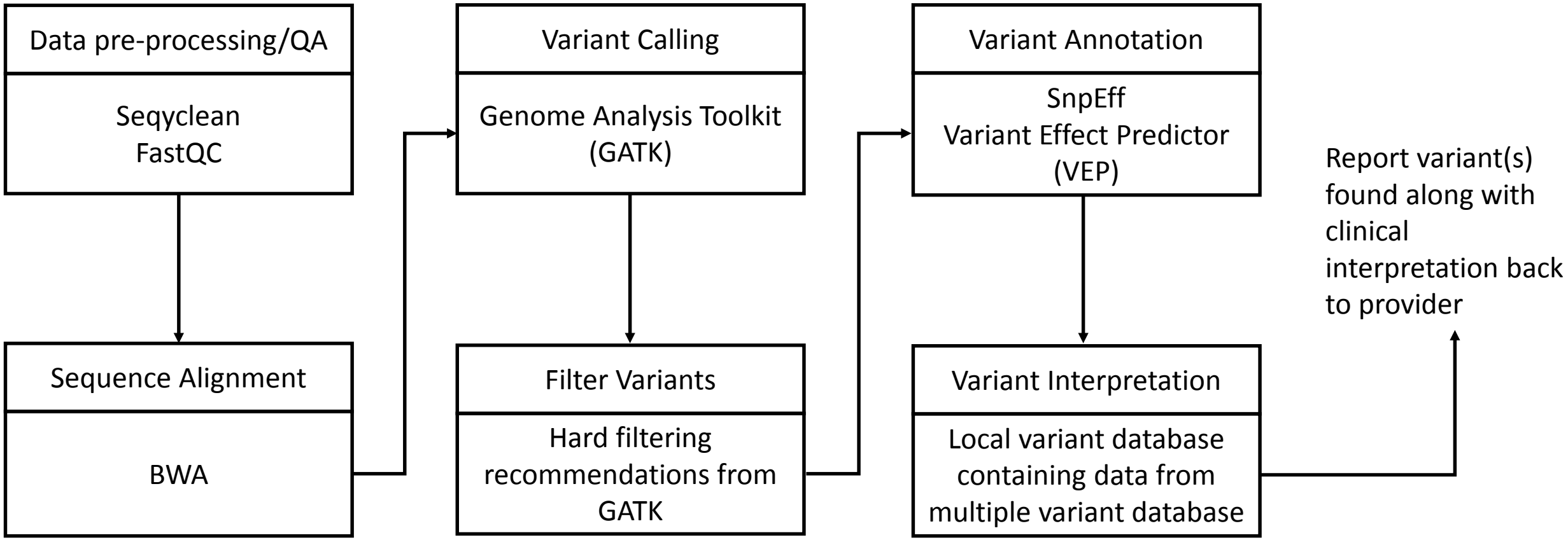
- MPS I
  - 1st screen – Low enzyme activity
  - 2nd screen – Low enzyme activity
  - NGS analysis of IDUA gene

Sequencing analysis pipeline



NGS sequencing analysis pipeline

Based on GATK Best Practices guidelines



Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. NIH Public Access; 2013;43(1110):11.10.1-33.

Consumer-driven reporting of variants

- Consult with clinicians as to how they would prefer variants to be reported
  - Primary Care Provider – Normal/Abnormal
  - Specialist
    - If condition is known
      - Variant
      - Variant classification
      - Example: Pathogenic GAA variant (p.Gly828_Asn882del)
    - If condition is unknown
      - All variant information
        - Text file format (TSV, CSV)
        - Variant Call Format (VCF)

Variant Reporting Formats

- TXT

```
Sample: A
Variants Detected: 1
Gene    HGVS_g  HGVS_c  HGVS_p  Classification
GAA     NC_000017.10:g.78078910delT     NM_000152.4:c.525delT   NP_000143.2:p.Glu176Argfs       pathogenic
```

- VCF

```
##fileformat=VCFv4.1
##fileDate=20121203
##phasing=none
##reference=file:///usr/local/db/homosapiens/b37/human_g1k_v37.fasta
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO    FORMAT  NA12878-NGv3-LAB1360-A
17      78078910        rs386834235     CT      C       6516.01 PASS    AB=0.82392;ABHom=0.878;ABP=277.33;AC=2;AF=1.00;AN=2;AO=248;BVAR;BaseQRankSum=4.243;CIG
AR=1X;DB;DP=246;DPRA=0;DS;Dels=0.00;EPP=7.24817;EPPR=8.37251;HRun=1;HWE=-0;HaplotypeScore=4.2298;LEN=1;MEANALT=5;MLEAC=2;MLEAF=1.00;MQ=45.81;MQ0=0;MQM=43.3548
;MQMR=23.5306;MQRankSum=5.543;NS=1;NUMALT=1;ODDS=110.983;OND=0.129;PAIRED=1;PAIREDR=1;QD=26.27;RO=49;RPP=64.7922;RPPR=92.7497;RUN=1;ReadPosRankSum=2.824;SAP=5
0.9578;SB=-2.514e+03;SRP=12.9813;TYPE=snp;VQSLOD=1.3662;XAI=0.000257697;XAM=0.0204421;XAS=0.0201844;XRI=0;XRM=0.00535366;XRS=0.00535366;culprit=QD;set=gatk-fr
eebayes;technology.illumina=1   GT:AD:DP:GQ:PL  1/1:30,216:246:99:6516,124,0
```

Ideas on NBS variant database

- Variant database for newborn screening
- Community solution
- Cost-free solution
- Community organization requirement

Building the local variant database

- Allows for easier, faster querying of variants
- Creating a MySQL database using medgen-mysql
- Populate using data from:
  - ClinVar
  - Human Gene Mutation Database (HGMD)
  - dbSNP
  - Exome Aggregation Consortium (ExAC)
  - Genome Aggregation Database (gnomAD)
  - Locus Specific Databases (LSDBs)
    - CFTR2
    - Pompe Disease Mutation Database

https://bitbucket.org/invitae/medgen-mysql

Potential problems with variant database curation

- Which databases should we use?
- Is the data current?
- Are there conflicting interpretations of the same variant in different databases? How can we resolve these conflicts?
  - ClinVar Miner

Henrie A, Hemphill SE, Ruiz-Schultz N, Cushman B, DiStefano MT, Azzariti D, et al. ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data. Hum Mutat. 2018 Aug;39(8):1051–60.

Re-evaluation of variants

- Variants of unknown significance (VUS)
- How often should our local database be updated?
  - Plan on updating as often as databases provide new releases
- What if the clinical interpretation of a variant changes?
  - Plan on re-evaluating patient variants every 6 months
  - Update LIMS to re-query VUSs and generate updated reports

Pipeline Validation

- Gene/Disease-specific validation
  - CFTR/CF
  - GAA/Pompe
  - ACADVL/VLCAD
- Validation of entire pipeline from sample to sequence
- Validation with third-party data

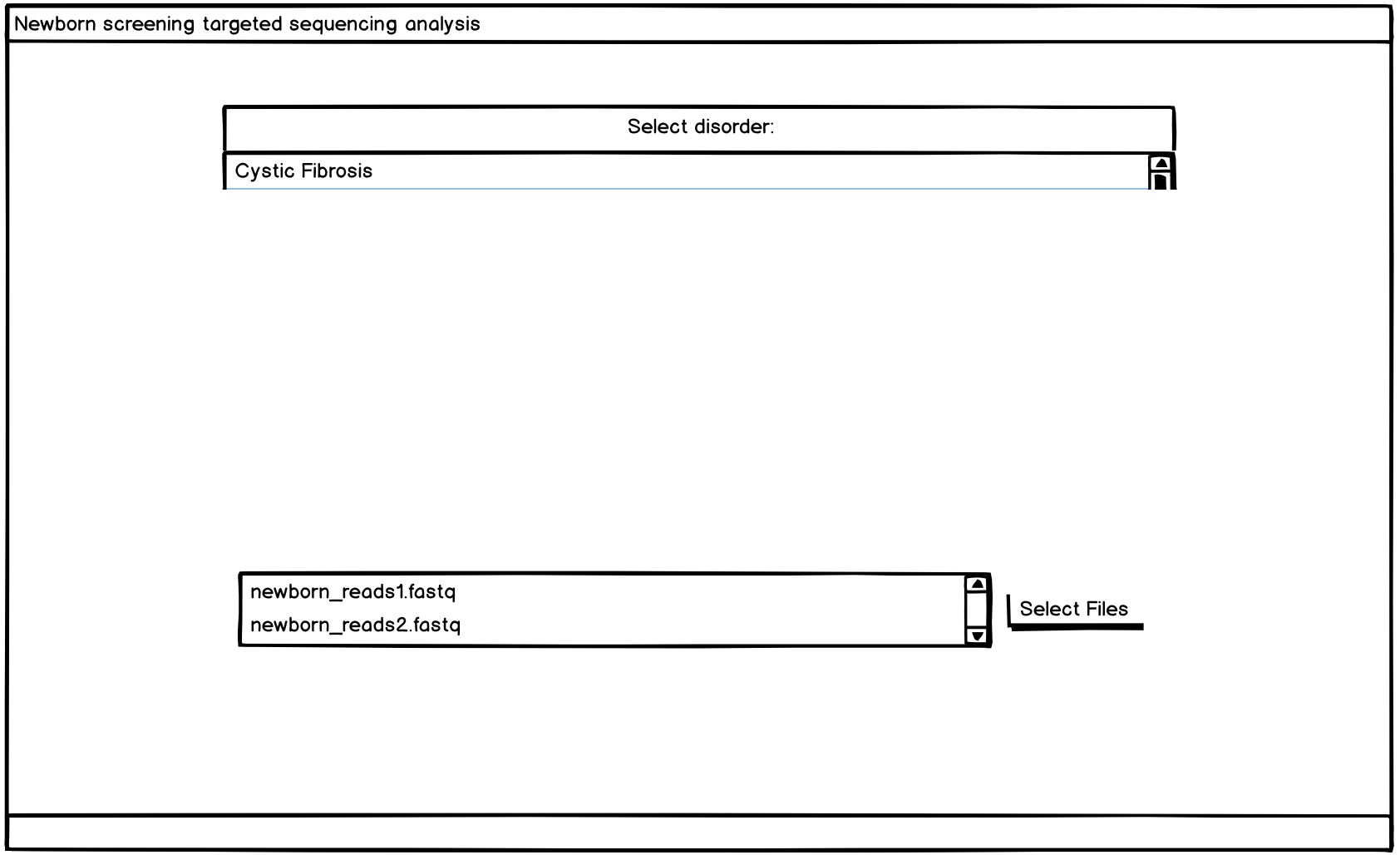
Distribution of pipeline

- Command-line version of pipeline

- Graphical User Interface (GUI) version of pipeline

https://hub.docker.com/r/tplcom/docker-presentation/
https://shiny.rstudio.com/

GUI development for gene-specific restriction

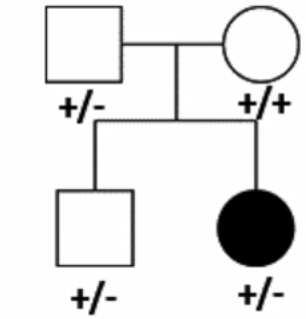


Why is this important? - Case Study

- Full-term apparently healthy baby girl, unremarkable family history; healthy older sibling
- First screen: absence of TREC, nearly undetectable
- Flow cytometry: T- NK+B+ SCID
- SCID targeted 19 gene panel: **heterozygous mutation on *FOXN1* c.1418dleC:p.P473fs**
- Bone Marrow transplant at 3.5 months
- At 2 years old, Flow cytometry still abnormal; transplant not successful

FOXN1

- Variant not found in ExAC database, not reported in the literature
- Reported *FOXN1* phenotype: congenital alopecia, nail dystrophy, absent thymus (requires thymic transplant)
- Normally autosomal recessive disorder (patient heterozygous)
- Incomplete penetrance reported: mild finger nail defects found in heterozygotes
- No nail or hair defects found in patient or family members, thymus present in patient
- Is Genetic cause determined in patient?
- Genomic sequencing of family: brother and father had same mutation as patient
- TREC retrospectively tested on brothers newborn screen specimen (State of birth was not screening at the time); TREC was 0



***FOXN1* p.P473fs**

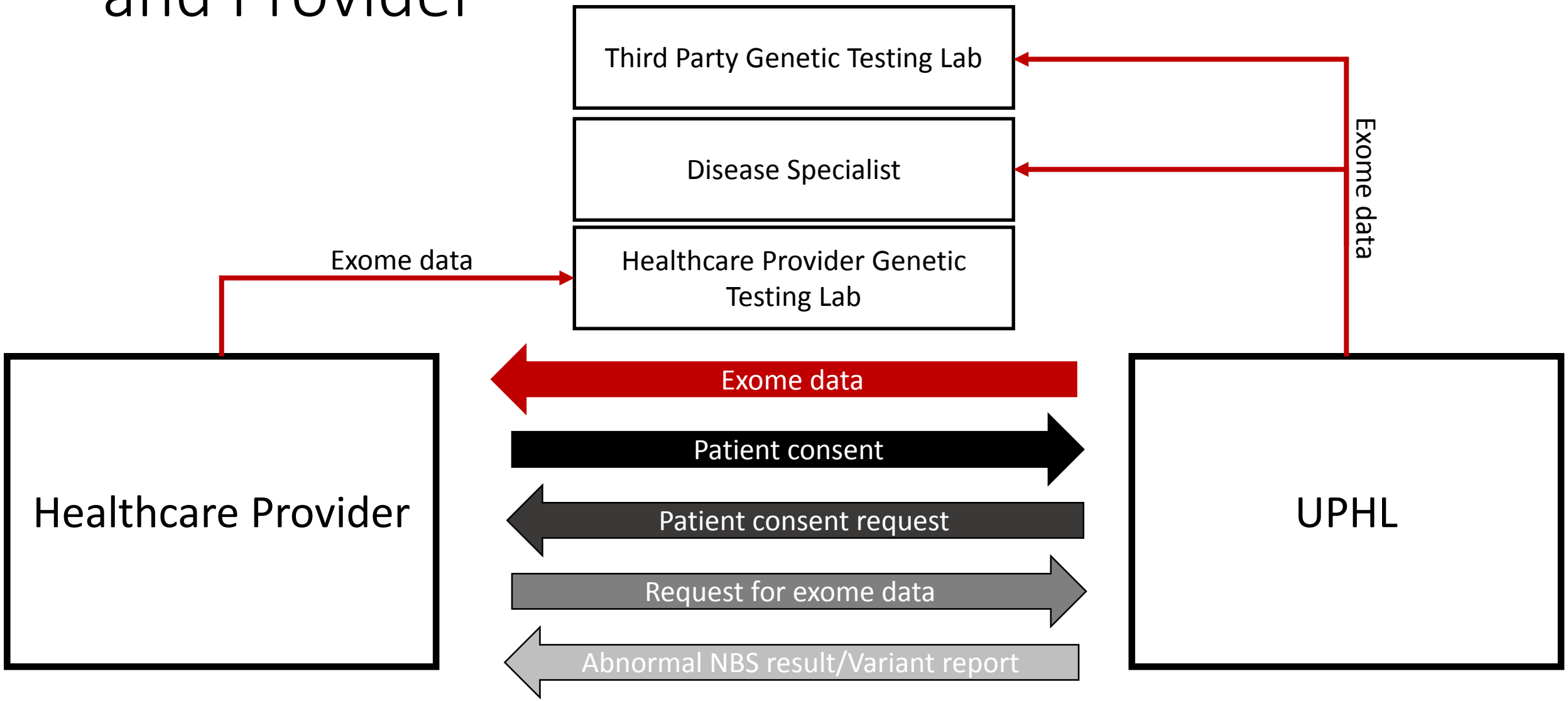
Case Study - Summary

- Heterozygous mutation in *FOXN1* likely cause of T-cell lymphopenia early in life that is asymptomatic
- Close follow-up recommended; treatment likely not required
- Patient had thymus transplant: transplant conditioning damages the thymus preventing normal T cell development

Sync for Genes phase 2 pilot site

- Sync for Genes Aim
  - Leverage Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) infrastructure for communicating information from clinical genomic labs in a format for universal use across medicine
- Utah Newborn Screening Program Aim
  - Enable electronic transfer of abnormal screening results to healthcare providers to reduce turnaround time for time-sensitive results to ultimately improve patient care

Sync for Genes: Data Transfer between UPHL and Provider



Request from the community

- Data
- Biological specimens
- Contribution towards development of shared repositories

Acknowledgements

Utah Department of Health Newborn Screening Program

- Andy Rohrwasser
- Kim Hart
- Mary Rindler
- David Jones

Utah Department of Health Infectious Disease

- Kelly Oakeson

University of Utah

- Karen Eilbeck (Biomedical Informatics)
- Mark Yandell (Human Genetics)
- Barry Moore (Human Genetics)
- Shawn Rynearson (Human Genetics)