

Article

Not peer-reviewed version

---

# Analytical Validation of a Genomic Newborn Screening Workflow

---

[Kristine Hovhannesyan](#)<sup>\*,†</sup>, [Laura Helou](#)<sup>†</sup>, [Benoit Charlotiaux](#), Valerie Jacquemin, Flavia Piazzon, Myriam Mni, Charlotte Flohimont, [Corinne Fasquelle](#), [Davood Mashhadizadeh](#), [Tamara Dangouloff](#), Vincent Bours, [Laurent Servais](#), [Leonor Palmeira](#), [Francois Boemer](#)

Posted Date: 23 July 2025

doi: 10.20944/preprints202507.1861.v1

Keywords: newborn screening; genomic; BabyDetect; dried blood spot; next generation sequencing; analytical validation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

*Article*

# Analytical Validation of a Genomic Newborn Screening Workflow

Kristine Hovhannesian <sup>1,2,\*†</sup>, Laura Helou <sup>2,3,†</sup>, Benoit Charlotiaux <sup>2,3</sup>, Valerie Jacquemin <sup>1,2</sup>, Flavia Piazzon <sup>4</sup>, Myriam Mni <sup>1,2</sup>, Charlotte Flohimont <sup>1,2</sup>, Corinne Fasquelle <sup>2</sup>, Davood Mashhadizadeh <sup>4</sup>, Tamara Dangouloff <sup>4</sup>, Vincent Bours <sup>2</sup>, Laurent Servais <sup>5</sup>, Leonor Palmeira <sup>2,3</sup> and François Boemer <sup>1,2</sup>

<sup>1</sup> Biochemical Genetics Lab, Department of Human Genetics, CHU of Liege, University of Liege, Avenue de l'Hôpital 1, 4000, Liege, Belgium

<sup>2</sup> Department of Human Genetics, CHU of Liege, University of Liege, Avenue de l'Hôpital 1, 4000, Liege, Belgium

<sup>3</sup> Bioinformatics Unit, Department of Human Genetics, CHU of Liege, Avenue de l'Hôpital 1, 4000, Liege, Belgium

<sup>4</sup> Division of Child Neurology, Reference Center for Neuromuscular Diseases, Department of Pediatrics, CHU of Liege, University of Liege, Bd du Douzième de Ligne 1, 4000, Liege, Belgium

<sup>5</sup> MDUK Oxford Neuromuscular Centre & NIHR Oxford Biomedical Research Centre, University of Oxford, John Radcliffe, Headington, Oxford, OX3 9DU Oxford, UK

\* Correspondence: khovhannesian@chuliege.be

† These authors contributed equally to this work.

## Abstract

Newborn Screening (NBS) has evolved significantly since its inception, yet many treatable rare diseases remain unscreened due to technical limitations. The BabyDetect study used a targeted next-generation sequencing (tNGS) panel to expand NBS to treatable conditions not covered by conventional screening. We present here the analytical validation of this workflow, assessing sensitivity, precision, and reproducibility using dried blood spots from newborns. We implemented strict quality control thresholds for sequencing, coverage, and contamination, ensuring high reliability. Longitudinal monitoring confirmed consistent performance across more than 5,900 samples. Automation of DNA extraction improved scalability, and a panel redesign enhanced the coverage and selection of targeted regions. By focusing on known pathogenic/likely pathogenic variants, we minimized false positives and maintained clinical actionability. Our findings demonstrate that tNGS-based NBS is feasible, accurate, and scalable, addressing critical gaps in current screening programs.

**Keywords:** newborn screening; genomic; BabyDetect; dried blood spot; next generation sequencing; analytical validation

## 1. Introduction

Newborn Screening (NBS) has come a long way since its first introduction to the public health system back in the 1960's [1]. The advent of new technologies has paved the path for progressive inclusion of more metabolic and endocrine disorders, resulting in saving thousands of children from severe disability and/or early death [2]. The inclusion of a disease in NBS is driven by the criteria of Wilson and Jungner, that include the existence of an efficient treatment. Moreover, early screening and diagnosis of diseases are of primary importance to obtain the best effect of such treatments and to accelerate new drug development. However, today, a set of rare diseases that completely match the criteria of Wilson and Jungner are not screened at birth. Among those diseases are congenital myasthenic syndromes [3], neurotransmitters diseases [4], pyridoxine and pyridoxal-dependent

epilepsy [5], Brown-Vialetto-Von Laere syndrome [6] or Wilson disease [7]. Even though there is substantial pre-clinical and clinical evidence that early treatment is much more effective than late treatment, and that pre-symptomatically treated patients have a much better prognosis than patients treated after the onset of symptoms, spinal muscular atrophy (SMA) is only recently being screened at birth in Europe. The case of severe combined immunodeficiency (SCID) constitutes another perfect example [8].

NBS is far to be implemented under the standard neonatal care in large countries worldwide.

The main reasons for the absence of NBS for eligible diseases are technical. All NBSs today are implemented as metabolic-based and, for some exceptions - DNA-based. As a matter of consequence, diseases without a well identified biomarker or metabolite, measurable in excess or in default, are generally not screened.

Another issue is the time needed in several countries to introduce new diseases one by one in the official programs. Spinal muscular atrophy constitutes an excellent example. Indeed, this condition that constitutes today the paradigm of a “must-be-screened” is detected at birth since 2018 in Belgium [9] or Germany [10], but not yet in the UK [11].

Pilot initiatives in genomic newborn sequencing have been blossoming worldwide over the past few years [12–18]. We have initiated a 3-year research project, BabyDetect (ClinicalTrials.gov: NCT05687474; [www.babydetect.com](http://www.babydetect.com)), using targeted panel sequencing (tNGS). In this paper, we describe the workflow of BabyDetect together with the analytical validation. We highlight the importance of defining and implementing the quality parameters for longitudinal monitoring of the routine genomic NBS workflow. We demonstrate the potential and challenges of panel-based screening to be implemented as NBS program with the capacity to screen more diseases.

## 2. Materials and Methods

### *Samples*

Parental informed decision was taken before collection of newborns' samples on dedicated filter paper cards (LaCAR MDx) designed for BabyDetect project [14,19]. The project was approved by the Institutional Review Board (CHU Liege ethics committee (n° 2021/239)) and complied with the WMA Declaration of Helsinki.

### *Design of Validation Plates*

Two microtiter-plates of 88 samples were designed for method validation. Plates contained DNA extracted from newborns' dried blood spots (DBS) and the DNA extracted from whole blood. The outline of validation plates is presented in supplementary materials - Table S1. The following samples were used for validation purposes:

8 positive newborn samples (NBPOS-1 to NBPOS-8, 3-plex of each), with pathogenic (P) or likely pathogenic (LP) variants in PAH, ACADM, MMUT, G6PD, CFTR, DDC genes, confirmed with screening and diagnostic methods

8 negative newborn samples (NBNEG-1 to NBNEG-8, 3-plex of each)

4 negative adult samples (ADNEG-1 to ADNEG-4, 4-plex of each). These are adult whole blood samples with no reported conditions.

Same 4 negative adult samples spotted on DBS (DBS-ADNEG-1 to DBS-ADNEG-4, 4-plex each)

HG002-NA24385 Genome in a Bottle (GIAB) reference DNA (Coriell Institute, USA, <https://www.nist.gov/programs-projects/genome-bottle>), 8-plex

### *DNA Extraction*

DNA was extracted from DBS and whole blood using QIAamp DNA Investigator Kit (Qiagen, USA) according to manufacturer's instructions with modifications for plate-based extraction. Intact double-stranded DNA was quantified by Qubit fluorometer (ThermoFisher scientific). DNA quality

and size were checked on agarose gel and Agilent fragment analyzer (Figure S1). Initial validation samples and one-year screening samples were manually extracted.

To ensure scalability of the population-based NBS and to improve turnaround time (TAT), automated extraction was implemented using the QIA Symphony SP instrument with the QIA Symphony DNA Investigator Kit (Qiagen). Both methods were validated and compared.

#### *tNGS Panel Design and Sequencing*

Eligible diseases were selected by doing systematic literature review and having discussions with local medical experts. Several criteria were considered before selecting genes of interest: treatable conditions, early onset of severe disease, significant benefit of treatment, specified standard of care [14,19]. Firstly, a custom target-panel covering 359 genes for 126 diseases was designed (panel-v1) and, it was later expanded to reach 405 genes for 165 diseases (panel-v2) [14]. Twist Bioscience technology was used for library preparation and the high-performing probes were selected for target enrichment.

For panel-v1, probes were designed to capture all coding regions of genes of interest, including 3'UTRs and 5'UTRs. For 20 complex genes, introns were also targeted. Altogether, targets of interest (TOI) in panel-v1 encompass 1.6M bases of genomic DNA.

To improve the quality and performance of BabyDetect panel, we designed a second version of the panel (panel-v2). Original design of panel-v1 was modified to focus on only the coding regions and intron-exon boundaries (~50 base pairs from the intronic borders) of selected genes. Deep intronic variants, promoter and UTRs, and homopolymeric regions were not targeted. With panel-v2 approximately 1.5 Mb are targeted for capture and sequencing.

Study samples were sequenced using Illumina technologies. Three validation runs were realized: two independent runs with 2x100 bp reads on NovaSeq 6000 (20040719, NovaSeq 6000 SP Reagent Kit v1.5, 200 cycles) and one run with 2x75 bp on a High Output NextSeq 500/550 systems (20024907, NextSeq 500/550 High Output Kit v2.5, 150 Cycles). Both sequencing platforms were analyzed with defined criteria.

#### *Bioinformatic Analysis*

The raw sequencing results were aligned to the reference genome GRCh37/hg19 using a homemade pipeline (Humanomics v3.15, <https://gitlab.uliege.be/bif-chu/humanomics> [20]) that utilizes published algorithms: BWA-MEM v0.7.17 for mapping the reads, elPrep for filtering the reads and removing duplicates, HaplotypeCaller for detecting variants, and GenotypeGVCFs for producing variant calling format (VCF) files. VCF files are used for the filtering and interpretation of variants. Humanomics allows identification of single-nucleotide polymorphisms (SNPs) and short insertions and deletions (indels, from 1-15bp) located within exons or at the intron-exon boundary (~50 base pairs of flanking regions). The pipeline does not call copy number variations (CNVs), large deletions, mosaicism, or other structural aberrations (e.g., translocations, inversions).

#### *Sensitivity and Precision of Sequencing*

To estimate sensitivity and precision, we used GIAB sample. This sample was sequenced and analyzed within the same workflow as the DNA extracted from DBS. The generated VCFs were subsequently compared to publicly available data provided by the GA4GH consortium [21]. These reference data consist of a VCF encompassing all variants confidently identified in this sample, referred to as GIAB gold standard, and a BED file containing all regions fully characterized for this sample. These regions, known as High Confidence Regions (HCR), cover approximately 85% of the genome (hg19) and serve as the standard reference regions. Real Time Genomics (RTG) tools v3.12 (vcfeval) was used for VCF comparisons. Vcfeval from RTG Tools serves as a tool for evaluating the accuracy of genetic variant calls. It compares two sets of variants: one extracted from the VCF being tested and another from a designated set file – the GIAB gold standard. The analysis was restricted



to TOI, looking only at variants (i) present in the HCR and (ii) at positions covered with at least 30 reads. True positive (TP) represent variants found both in GIAB gold standard and in our VCF, false negative (FN) represent variants from GIAB gold standard that are missed in our VCF, and false positive (FP) are variants present in our VCF but not in GIAB gold standard. Sensitivity is defined as the fraction of GIAB gold standard detected, hence is equal to  $TP/(TP+FN)$ . Precision is defined as the fraction of variants from our VCF present in GIAB gold standard, hence is equal to  $TP/(TP+FP)$ . These two metrics provide a quantitative measure of the reliability of variant calls.

### *Reproducibility of the Results*

The pairwise comparison of samples from various groups was performed using RTG tools using one VCF as test sample and the other as “truth” set. We assessed concordance with the formula:  $C / (A + B + C)$ , where C corresponds to variants found in both VCF1 and VCF2 (hence classified as TP by VcfEval), A corresponds to variants exclusively inferred in VCF1 (hence classified as FP by VcfEval), and B corresponds to variants inferred only in VCF2 (hence classified as FN by VcfEval).

The analysis was limited to the TOI and HCR regions and filtered to retain only those with a minimum depth of 30 reads ( $DP \geq 30$ ).

### *Definition and Selection of Quality Metrics*

To ensure the reliability and accuracy of our results, we check quality control (QC) parameters that span various stages of sample processing, from library preparation, sequencing to variant calling. Among the numerous QC parameters provided by the Picard tools suite [22], the ten have been chosen for longitudinal monitoring:

#### *Sequencing*

Q30\_pct: The percentage of bases with a quality score of 30 or higher.

PF\_BASES: Total number of bases.

TOT\_Duplicates\_pct: The percentage of duplicated reads (PCR or optical duplicates).

#### *Library and capture*

TARGET\_BASES\_30X\_pct: Percentage of targeted bases that reached a minimum effective coverage of 30X.

MEAN\_TARGET\_COVERAGE: Average effective coverage of sequencing across the TOI.

SELECTED\_BASES\_pct: Percentage of aligned bases located on/near the captured targeted regions.

MEAN\_INSERT\_SIZE: Average insert size.

FOLD\_80\_BASE\_PENALTY: Resequencing rate required for 80% of targeted bases to achieve the average coverage of target.

#### *Alignment*

Reads\_aligned\_pct: Percentage of reads aligned to the reference genome.

#### *Variant Inference*

SNP\_REFERENCE\_BIAS: Average fraction of reads where the reference allele is observed at sites with a heterozygous single nucleotide polymorphism (SNP).

### *Threshold Setting*

For every sample processed, QC values generated by Picard tools are systematically stored in a centralized database. While the complete set of metrics is extensive, only a curated subset is required to validate the quality of an analysis under routine conditions. A sample is flagged as potentially problematic when, for each retained metric, its inferred value is outside a defined quality threshold.

To define QC thresholds, 34 independent runs were considered for panel-v1 and 43 runs for panel-v2. Duplicate and triplicate samples were excluded to avoid bias. For each metric, distribution

plots were created to visualize the Interquartile Range (IQR) and determine cutoff values, with most parameters adopting a 1.5xIQR threshold to identify potentially problematic samples.

For the longitudinal follow-up, we have adopted ChronQC as an integral tool for monitoring the quality of our data over time [23]. Leveraging ChronQC's capabilities, we can efficiently track changes and trends in our datasets across multiple time points, ensuring the reliability and consistency of our longitudinal analyses. All quality metrics are monitored for each sequencing run. We have defined three very important metrics (VIP): Q30\_pct, TARGET\_BASES\_30X\_pct and SNP\_REFERENCE\_BIAS. In the case of failure of one of these VIP metrics, the whole workflow for that sample is repeated.

### *Variant Interpretation Pipeline*

After the secondary analysis was completed and the sequencing quality was checked, the VCFs were used for variant interpretation of asymptomatic neonates, where no phenotypic data is available for the variant selection process. Using Alissa Interpreter (Agilent) we have developed and validated the BabyDetect Variant Interpretation Pipeline. The pipeline includes a) the decision tree that has criteria for filtering variants with no human phenotype ontology, b) the information of genes in the panel [14].

In BabyDetect, we filter and report class 4 and 5 variants according to ACMG guidelines [24] from our knowledge-based database, as well as from ClinVar. Variants which are not present in ClinVar were not reported. Filtered P/LP variants were assessed manually using additional platforms of Franklin (<https://franklin.genoox.com>) and VarSome [25]. Variants with conflicting assertions of pathogenicity between these three platforms (ClinVar, Franklin and VarSome) were not considered for reporting.

## **3. Results**

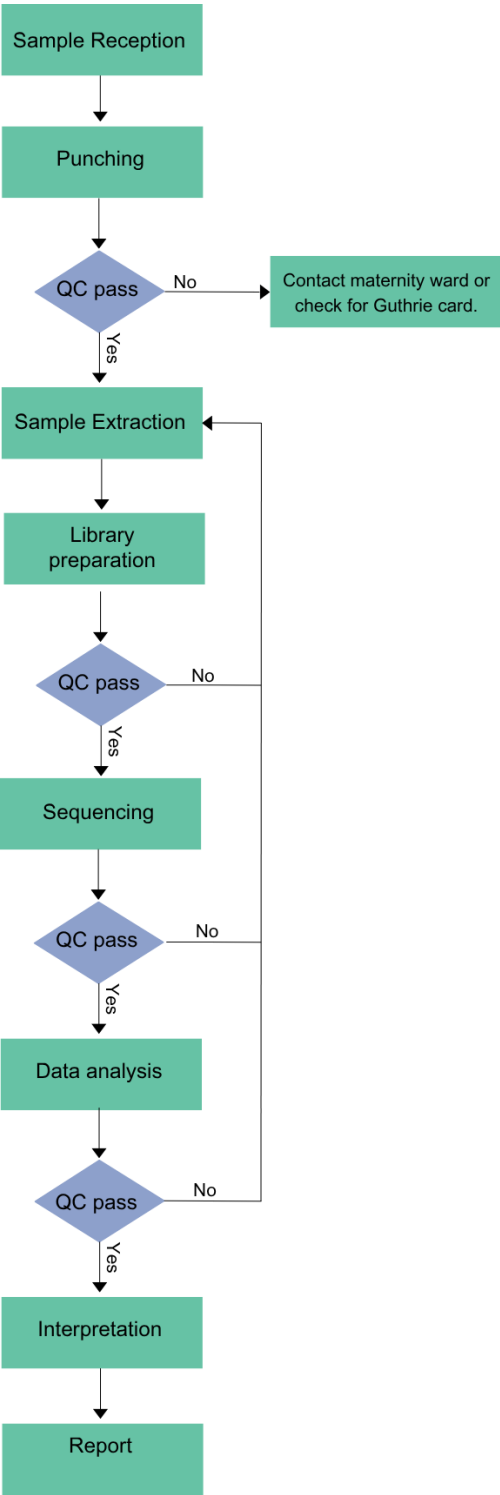
### *Workflow*

The BabyDetect analysis was developed starting in 2020 and is now accredited under ISO-15189:2022 certification and IVDR labelled. Its analytical workflow is defined from sample punching at the laboratory facility up until the clinical reporting (Figure 1). The mean laboratory processing time for workflow is from 8 to 10 days.

Samples for the study and for the analytical validation were collected, recorded and stored by trained personnel under conditions for clinical application. DBS cards are delivered to our dispatching unit and encoded with a unique identifier in information management system.

Following sample reception, the analytical workflow begins with the punching of DBS card. This step is followed by an initial QC evaluation; in the event of failure, due to insufficient material or poor sample integrity, the corrective action is initiated, including re-contacting the maternity ward or verifying the availability of an alternative card. If sample passes QC, it proceeds to DNA extraction, which is likewise subjected to QC before advancing to library preparation. To ensure better DNA yield, we run an overnight 30°C lysis of the DBS punches and proceed with Qiagen column-based extraction protocol (validation samples and one-year screening samples). Twist Bioscience technology is used for library preparation; the average size of prepared libraries with ligated adapters is 350bp, which allows sequencing of fragments by both NovaSeq and NextSeq systems. At each stage, QC is systematically applied. QC failure at any point triggers a loopback to the extraction phase, ensuring that only high-quality input progresses through the workflow. This looped design not only safeguards the fidelity of downstream analyses but also minimizes the risk of reporting inaccurate findings. The final steps involve variant interpretation and clinical reporting. When a positive result is identified, the workflow is repeated from the punching of DBS until variant interpretation. This allows us to confirm the positive screening result as well as to check for inversion of the sample.

The workflow described in this section was used for analytical validation, whereas the optimizations for DNA extraction as well as a panel redesign are presented in a dedicated section below.



**Figure 1. Laboratory operational workflow of BabyDetect analysis.** Green boxes represent working steps in the flow from top to bottom. Blue diamonds represent quality control steps which lead to decision points on how to proceed.

*Variant Interpretation Pipeline*

In BabyDetect, for conditions with autosomal-recessive inheritance, we report homozygote and compound heterozygote P and LP variants. Carriers of those variants are not reported in the scope of our project. In case of autosomal-dominant diseases, the presence of one P or LP variant is reported. For X-linked diseases, hemizygous identification of P or LP variants in males and homozygous or possible compound heterozygous identification of P or LP variants in females are reported [14]. The variants of uncertain significance (VUS) are not reported.

Our decision tree incorporates different filtering criteria including quality, allelic frequency, mutation type and operates to decipher variants of clinical interest. Briefly, after secondary analysis and sequencing quality verifications, VCF are uploaded in the interpretation pipeline. Variants are first filtered based on our gene panel. Then, variants are filtered by read depth and variant confidence by depth. Filtering based on variant allele and population frequency is then applied. Variants with population frequency <1% or any variants in frequent genes are checked in our managed variant list (MVL). Variants that match P and LP classification from MVL are selected for filtering based on X-linked genes in the panel and subsequently filtered by their zygosity. Further, variants are filtered on published P and LP ClinVar variants, and matching variants are again filtered based on X-linked genes and zygosity. All filtered homozygote, compound heterozygote and hemizygote variants are manually reviewed.

The filtering pipeline was designed with a semi-automatic approach; when no variants of interest are to review, the sample is automatically closed, whereas samples with P and/or LP variants are flagged for manual review. Variants are reviewed independently by three scientists and decisions on reporting are taken. Variants not registered in ClinVar or in our MVL are not reported. For variants with conflicting interpretations in ClinVar, Franklin, VarSome databases and relevant publications are checked. Consensus on reporting is taken if supporting evidence is found in three databases.

Validation Samples

Initial validation was performed on a GIAB reference (HG002-NA24385) and on 8 positive samples previously identified by an alternative method. Additional 8 newborns’ negative samples and 4 adult negative samples were used for comparative analysis of different parameters of the workflow. These adult whole blood samples with no reported conditions were also spotted as DBS and included in the validation process, to assess concordance between our DBS workflow with classical whole-blood workflow.

Sample selection, included in validation plate, are described in “Material and methods”. For initial validation (panel-v1) we used the first 8 positive samples (NBPOS-1 to NBPOS-8) indicated in table 1; these samples were selected by their availability at the conventional NBS laboratory, which were positive based on biochemical marker and/or identified P variants at diagnostic setting. 88 samples included in validation plates (Table S1) were sequenced twice in independent sequencing runs on NovaSeq 6000 and 48 samples were sequenced on NextSeq 500/550.

After validation, 2,600 newborn samples included in the project were processed with the validated protocol in 34 independent batches (96 or 48 samples/batch). The protocol with a redesigned panel (v2) was revalidated on a set of 8 additional positive samples (Table 1) and same GIAB. These 8 samples were sequenced on two independent sequencing runs. This second protocol was then used for the remainder of the study (3,319 samples in 43 independent batches).

Table 1. Positive samples used in analytical validation.

sample ID	disease	Gene	variant 1	variant 2	method of confirmation	Validation panel-v1
NBPOS-1	Phenylketonuria (PKU)	PAH	c.1066-11G>A	c.1315+1G>A	Conventional NBS and panel-sequencing	



NBPOS-2	Phenylketonuria (PKU)	PAH	c.1169A>G	c.898G>T	Conventional NBS and panel-sequencing	
NBPOS-3	Aromatic l-amino acid decarboxylase (AADCD)	DDC	c.823G>A	c.1037A>G	WGS and biochemical testing	
NBPOS-4	Cystic fibrosis (CF)	CFTR	c.1521_1523delCTT	c.1521_1523delCTT	Conventional NBS and phenotyping CFTR	
NBPOS-5	Medium-Chain-Acyl- CoA- Déshydrogénase (MCAD)	ACADM	c.948+2T>C	c.1045-2A>C	Sanger Sequencing	
NBPOS-6	Glucose-6-phosphate dehydrogenase deficiency (G6PD)	G6PD	c.466A>G	c.292G>A	Conventional NBS and panel-sequencing	
NBPOS-7	Medium-Chain-Acyl- CoA- Déshydrogénase (MCAD)	ACADM	c.1084A>G	c.1084A>G	Conventional NBS and phenotyping ACADM	
NBPOS-8	Cystic fibrosis (CF)	CFTR	c.3752G>A	c.3752G>A	Conventional NBS and phenotyping CFTR	
NBPOS-9	Hemophilia B	F9	c.1024A>G		panel-sequencing	
NBPOS-10	Short-chain acyl-CoA dehydrogenase (SCAD) deficiency	ACADS	c.1147C>T	c.596C>T	panel-sequencing	Validation panel-v2
NBPOS-11	Glucose-6-phosphate dehydrogenase (G6PD) deficiency	G6PD	c.466A>G	c.292G>A	panel-sequencing	
NBPOS-12	Cystic fibrosis	CFTR	c.1865G>A	c.1865G>A	panel-sequencing	
NBPOS-13	Cystic fibrosis	CFTR	c.1397C>G	c.3209G>A	panel-sequencing	
NBPOS-14	Wilson disease	ATP7B	c.3207C>A	c.1877G>C	panel-sequencing	
NBPOS-15	glucose-6-phosphate dehydrogenase (G6PD) deficiency	G6PD	c.1437G>C		panel-sequencing	
NBPOS-16	very long-chain acyl-CoA dehydrogenase (VLCAD) deficiency	ACADVL	c.325G>A	c.601_603delGAG	panel-sequencing	

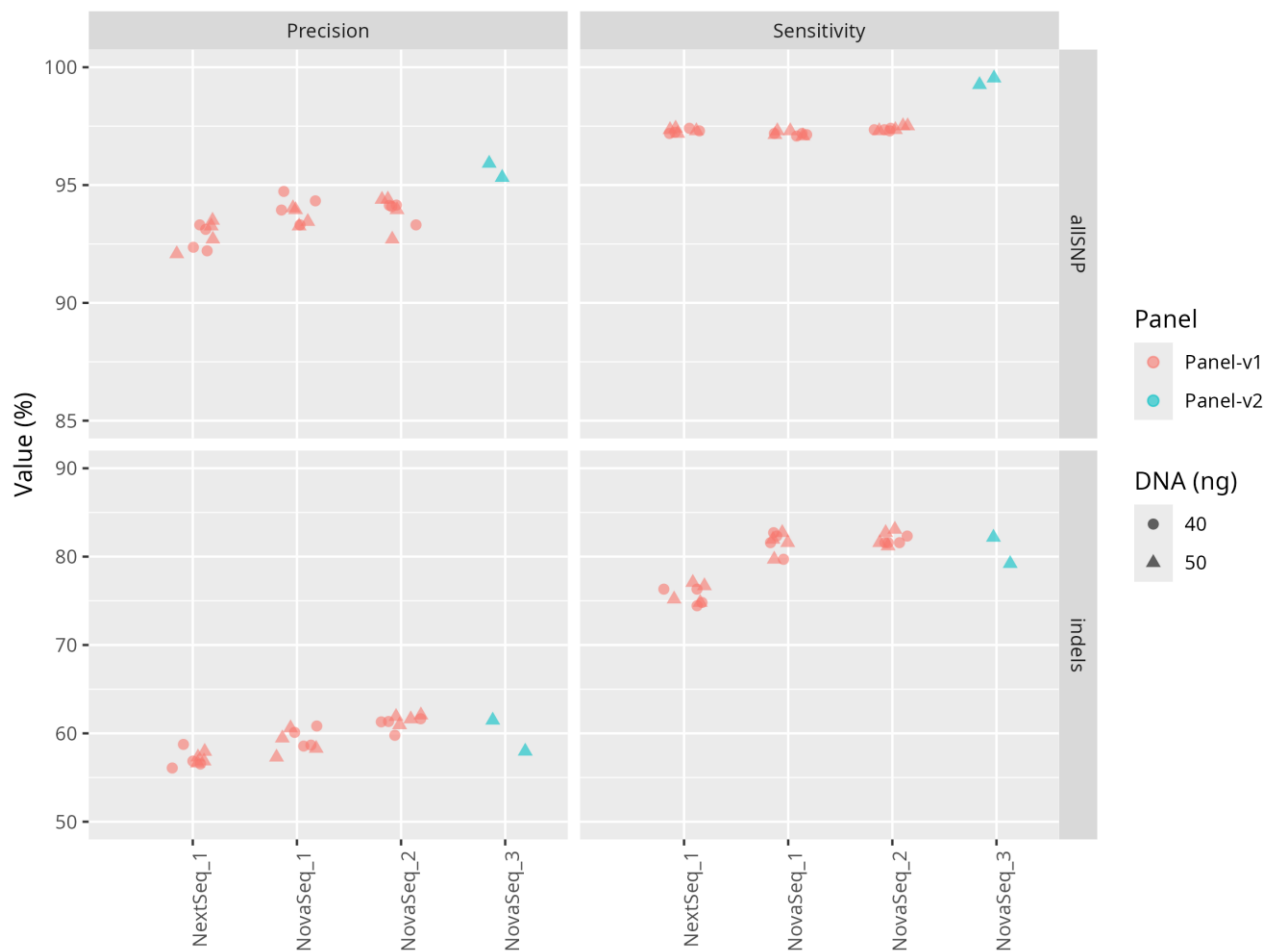
Performance of the Analysis

Sensitivity, Precision

Our results indicate that the initial validated workflow exhibits high sensitivity for SNP detection, with values exceeding 97% and a precision of 94% (panel-v1), whereas for the detection of indels, sensitivity is 81% and precision is estimated at 59%.

Our findings underscore the method’s high sensitivity and precision for SNP detection together with its reliable sensitivity but moderate precision for detecting small indels. Our obtained results were in accordance with the number of accredited NGS analysis of the diagnostic unit at university hospital of Liege (CHU Liege, oral communications at laboratory meetings and validation protocols), both for panel-based sequencing and whole-exome sequencing (WES).

The redesign of our target panel (v2) and its improved performance are described in the section “Optimizations” below.



**Figure 2. Performance metrics for SNPs and indels across sequencing panels and DNA inputs.** Sensitivity and precision are shown for SNPs and indels across different sequencing panels and DNA input quantities. Colors indicate the sequencing panel (red = Panel-v1, blue = Panel-v2), and shapes represent input DNA quantity in ng. Facets separate metrics (columns) and variant types (rows). Y-axis scales are adapted per variant type.

BabyDetect Variant Interpretation Pipeline was validated on positive and negative samples (n=216 sample). In the initial analytical validation, eight positive samples were analyzed in triplicate in two NovaSeq runs (48 samples) and, in duplicates, in a NextSeq run (16 samples). All disease-causing variants, identified by alternative methods, were correctly selected by our variant decision tree which leads to a 100% sensitivity for P variants of our workflow. Negative samples included in the validation plates also were interpreted with the pipeline and no P variants were retained after interpretation (Novaseq runs n= 112 samples, Nextseq run n=24 samples) as expected.

The 8 positive samples used in revalidation (panel-v2) of the workflow were sequenced in two runs (n=16 samples) and interpreted with the interpretation pipeline; same P variants were identified in both runs with 100% sensitivity.

***Intra-run and inter-run concordance***

We have used the concordance as an estimation for the similarity between two replicates of a sample within the workflow. Concordance is defined as the number of identical variants between both replicates divided by the sum of unique variants found in both replicates.

For GIAB, the estimated intra-run mean concordance between replicates of the same batch was 93%. The concordance analysis was also done for positive controls, and control extracted from whole blood and DBS (intra-run mean concordance - 90%).

Inter-run concordance between two replicates passed in two independent batches was used to estimate their similarity. Our results show an inter-run concordance close to 90% for SNPs for GIAB

and positive controls. The majority of discordant variants (92%) were FP with low allele frequencies, and are not relevant for clinical interpretation, as they are systematically excluded by our variant filtering pipeline.

Overall, these results indicate a high reproducibility of our workflow.

### **Quality control**

To ensure the reliability and accuracy of our sequencing data, we established a set of QC thresholds. Samples that do not meet these thresholds are flagged for re-analysis, ensuring data integrity and consistency. The quality metric thresholds for both panel-v1 and panel-v2 used in routine are presented in Figure S2.

#### ***Evaluating sequencing quality***

To ensure the reliability of sequencing data, we routinely monitor key quality metrics that reflect both yield and accuracy. We used core indicators PF\_BASES and Q30\_pct.

A minimum threshold of 85% (panel-v1) is applied for Q30\_pct as recommended by the instrument manufacturer. Based on revalidated panel-v2 and the rest of the sequenced runs, the minimum threshold was raised to 90% to further increase the stringency of our QC (Figure S2). All samples in our cohort exceeded these values, confirming the high quality of our sequencing runs.

For PF\_BASES, we established a lower limit based on inter-sample variability observed across multiple validation runs. Using the IQR as a robust measure of spread, we defined the threshold as  $1.5 \times \text{IQR}$  below the cohort median — an approach used to flag statistical outliers.

Samples below this limit are flagged for review. If sample failed Q30\_pct it is excluded from further analysis and is re-prepared from the newly punched DBS. This strategy helps ensure that only high-quality data is retained for downstream interpretation.

#### ***Evaluating target selection quality***

To monitor library preparation, we used TARGET\_BASES\_30X\_pct and MEAN\_TARGET\_COVERAGE metrics.

Initially, the TARGET\_BASES\_30X\_pct metric has a lower threshold set at 93%. Samples falling below this threshold exhibit insufficient coverage and are systematically reviewed and the flow is repeated for that sample (Figure 1). When MEAN\_TARGET\_COVERAGE also falls outside the acceptable range, this strongly indicates inadequate sequencing depth, warranting review. The MEAN\_TARGET\_COVERAGE threshold is arbitrarily set at 100x, ensuring robust depth across the targeted regions.

To evaluate the capture efficiency in library preparation, we follow the QC metrics of SELECTED\_BASES\_pct. For initial panel -v1, the minimum threshold of this parameter was set to 40% (Figure S2). The redesign of the panel (v2) allowed us to improve the on-target capture (min. threshold 78%), and the details are described in the dedicated section below.

#### ***Evaluating inversions and contaminations***

In our routine workflow, to control plate or sample inversion during manual preparation, we have introduced two risk mitigations:

Every 96-well plate includes in position H12 a control, containing P variants in two genes (SERPINA1: c.1096G>A and ALDOB: c.448G>C). This sample serves as “internal control sample”.

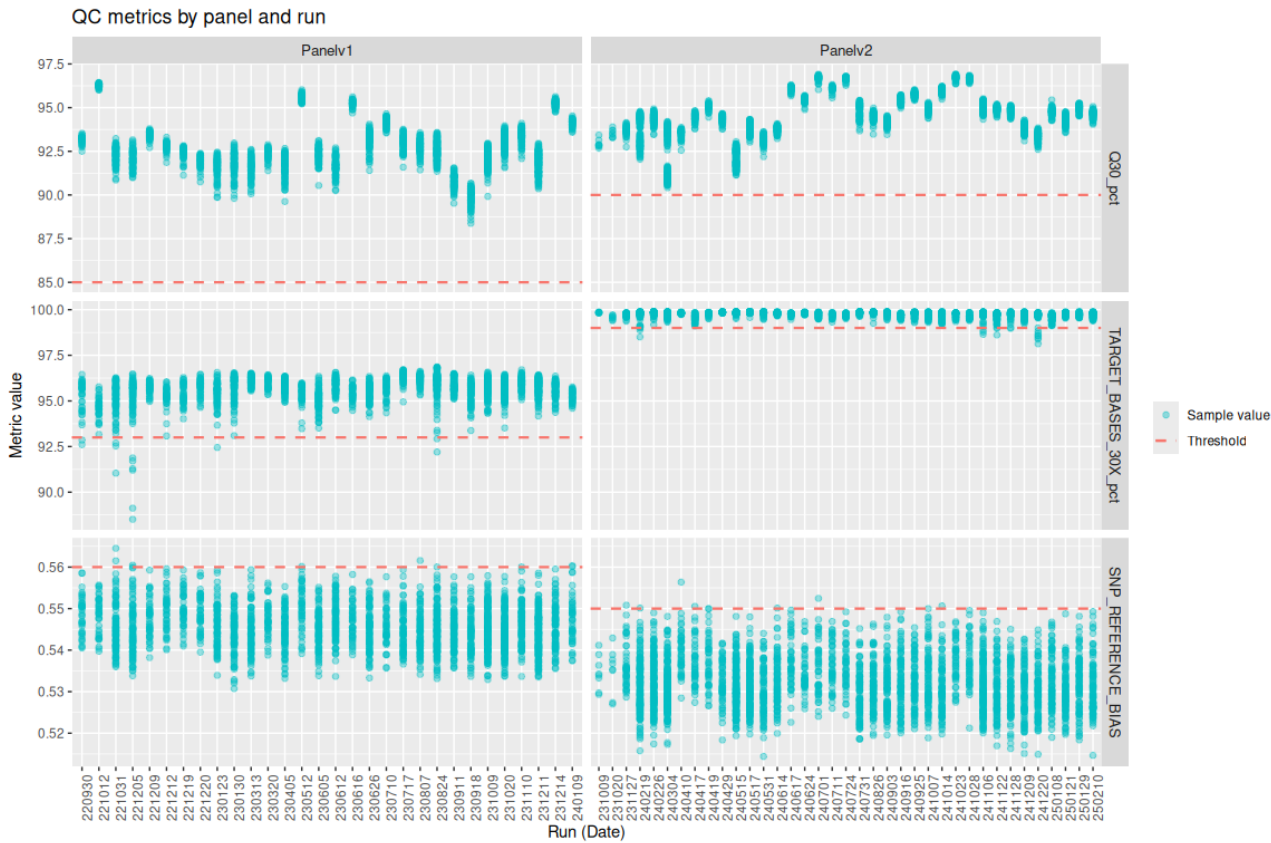
Each positive case is completely reanalyzed from the original DBS to confirm the initial sequencing finding.

Additionally, we use the SNP\_REFERENCE\_BIAS to detect potential contamination. The SNP\_REFERENCE\_BIAS upper threshold is fixed at 0.56, corresponding to  $1.5 \times \text{IQR}$  above the median. Values exceeding threshold suggest contamination and are reviewed and re-prepared.

“Internal control sample” is monitored after each variant interpretation of the batch to ensure that P variants are identified in every sequencing run with sufficient read and allelic depth, call and mapping quality, and true genotype. For each monitored parameter the Levey Jennings chart is generated, and the values are plotted after each batch interpretation. This monitoring allows us to confidently state that for that sample both P variants are being identified with high confidence and uniform (Chart S1).

*Longitudinal monitoring*

We use ChronQC (Figure 3) to automate the monitoring and evaluation of sequencing metrics over time, ensuring that the data consistently adheres to predefined quality standards throughout the study. We monitor all the metrics outlined in the ChronQC report. The samples failing the defined thresholds for VIP metrics are being repeated from the start of the laboratory workflow.



**Figure 3. Sequencing quality metrics across runs for Panel v1 and Panel v2.** Quality control metrics are shown for sequencing runs performed using two capture panels: Panel v1 (left) and Panel v2 (right). Each subplot displays one of three VIP metrics (top to bottom): the proportion of bases with a quality score  $\geq 30$  (Q30\_pct), the percentage of target bases covered at  $\geq 30\times$  (TARGET\_BASES\_30X\_pct), and the SNP reference bias (SNP\_REFERENCE\_BIAS). Each dot (blue) corresponds to an individual sample within a run, while horizontal dashed lines (pink) indicate the predefined quality threshold for each metric and panel. On the x-axis, runs are labeled by their run date; on the y-axis, metric values are shown.

*Decision criteria for sample quality*

Across the first year of the study using panel-v1 (2600 samples), 95.6% (n=2486) of samples passed all QC criteria without requiring repeating the workflow. The remaining 4.4% (n = 114) failed at least one metric: 103 samples failed one metric, 10 samples failed two metrics, and 1 sample failed three. The most frequent failure was SELECTED\_BASES\_pct (n=82). By systematically integrating these thresholds into our workflow, we enhanced both sequencing efficiency and downstream data accuracy.

*Robustness to workflow variations*

We have assessed whether variations in the standard conditions of the protocol for library preparation have any influence on results quality.

*Initial DNA quantity*

Initial DNA quantity can vary due to DNA extraction and manual manipulation. To evaluate the robustness of our workflow to such DNA quantity variations, we estimated sensitivity and precision for the detection of known GIAB variants using DNA quantities of 40 ng and 50 ng for

library preparation as recommended by Twist Bioscience. Our results indicate that both DNA quantities lead to a similar sensitivity for SNP detection (97.15% - 40 ng and 97.21% - 50 ng) and precision (94.08% - 40 ng and 93.67% - 50 ng). Sensitivity and precision for indels are also similar between both DNA quantities (sensitivity at 81.58% - 40 ng and 81.49% - 50 ng, precision at 59.54% - 40 ng and 58.92% - 50 ng) (Figure 2).

Further, to facilitate the manual workflow we didn't quantify the amount of DNA to be taken for library preparation but took 40  $\mu$ l of extracted DNA for fragmentation. Of the note, extracted DNA concentration varied from 50 to 400 ng. We have tested the robustness of the analysis by varying the DNA input 40 ng vs. 50 ng and 40  $\mu$ l of extract vs. 40 ng. Results show that the concordance observed here does not differ from the inter-run concordance (90%). The amount of DNA used for library preparation has little impact on concordance. This demonstrates consistency and shows that our method is robust to variations of initial DNA quantity.

#### ***Initial material variation: whole blood or dried blood spot***

Sensitivity and precision were evaluated on GIAB but our routine flow for neonatal screening uses DNA extracted from DBS [26]. Therefore, we assessed concordance between DNA extracted from whole blood and the same DNA extracted from DBS. Results show that the concordance observed here does not differ from the concordance observed in inter-run concordance data (90%) and that results obtained from DBS are no different from those of whole blood.

#### ***Sequencing instrument***

Depending on the number of samples to screen, we have used both NovaSeq (96 samples) and NextSeq (48 samples) sequencers. To evaluate whether sequencing instruments influence the quality of obtained results, we calculated concordance, precision and sensitivity for the GIAB between NextSeq and NovaSeq, for DNA quantities of 40 and 50 ng. The sequencing runs on NovaSeq are referenced as NovaSeq\_1 and NovaSeq\_2, the NextSeq run is referenced as NextSeq\_1.

The results show that the concordance between NovaSeq and NextSeq fluctuates approximately between 86 and 88% for the control samples, whereas for GIAB sample was lower, between 83 and 85%, with no difference between DNA input 40 and 50ng.

#### ***Optimizations***

##### ***DNA extraction automation***

To ensure scalability and to improve TAT, the automated QIAAsymphony DNA extraction was validated and implemented in our workflow. It has since replaced the manual extraction of our routine workflow, and 3,319 samples have been screened.

For both extractions, the amounts of isolated DNA fluctuated between 50 to 400 ng, with the size superior to 20 kb with manual column-based (Figure S1-b) and superior to 40 kb with beads-based QIAAsymphony (Figure S1-a) extractions. Concentrations obtained for selected DNA-plates are available in Figure S3. ANOVA test was conducted, and it showed a p-value = 0.188 ( $p < 0.05$ ), indicating that there is no significant difference between DNA concentrations extracted by both methods. Additionally, after running extracted samples with the entire analytical flow, pairwise concordance (per sample) was calculated between samples extracted manually and with QIAAsymphony. Samples were sequenced in different runs and results are presented in Table S2. The mean concordance between analyzed samples is 91.6. The comparisons of both methods allowed us to integrate, with confidence, the QIAAsymphony workstation in our further flow.

##### ***Target of interest: improving performance and clinical impact***

After the first year of this study, we re-evaluated the gene list of our targeted panel, removed and added some genes. The re-evaluation of gene list meant a redesign of the panel. This opportunity of redesigning has allowed us to improve our general performance by focusing on the two following aspects. First, we removed some genes because they couldn't meet our quality control criteria due to technical limitations (homologous regions, large rearrangements). Second, we reduced the number of targeted regions (namely by removing large intronic regions) to reduce the number of off-target reads. The initially validated panel of BabyDetect contained 359 genes and the redesigned panel –

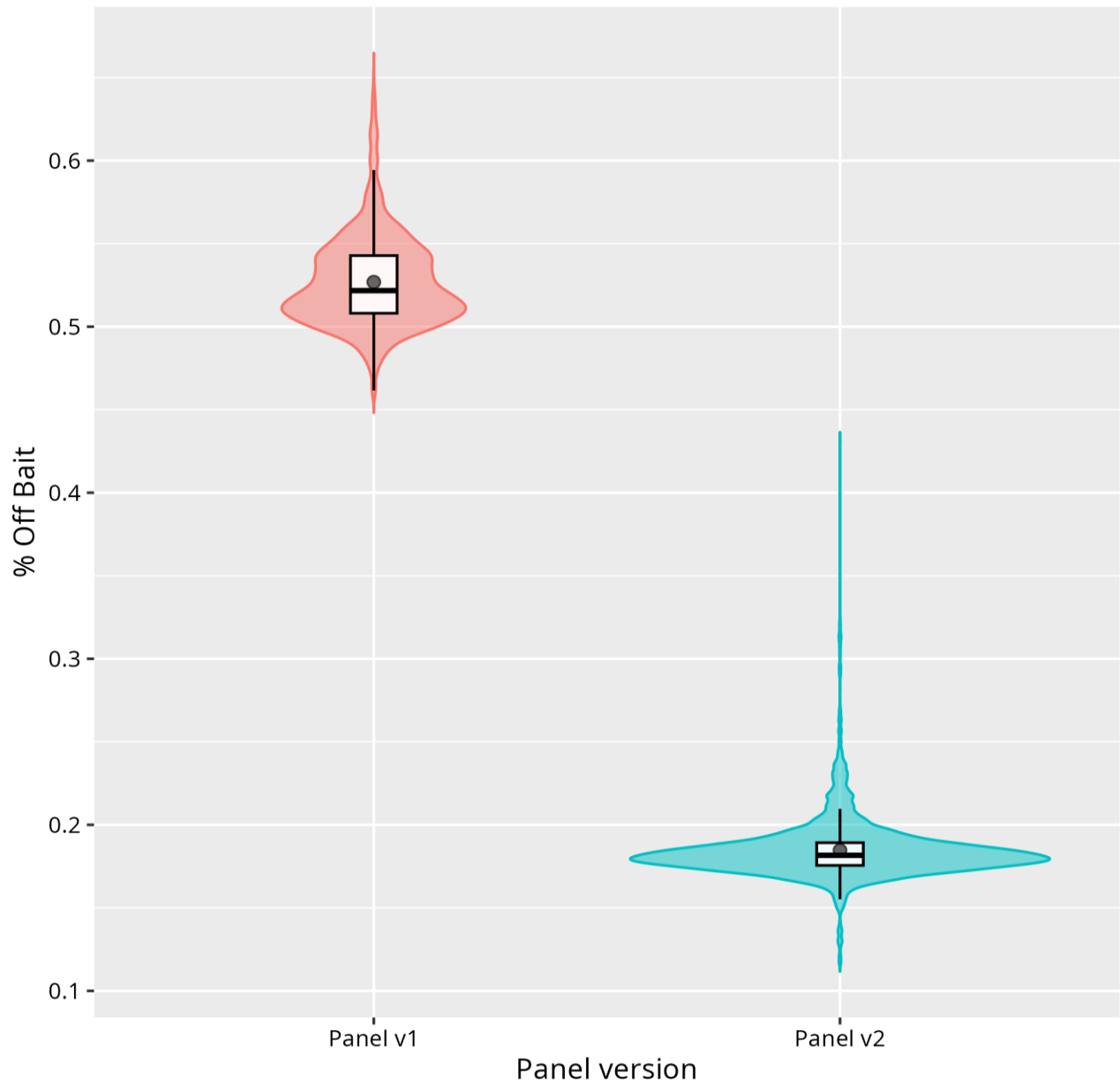


405 genes. Precisely, 15 genes were removed from the initial panel and 61 were added to the redesigned panel.

Beyond its broader screening scope, the redesigned panel demonstrated improved technical performance. A key indicator of this improvement is the significant reduction in off-target sequencing: the average percentage of off-target reads decreased from 52.7% with panel-v1 to 18.5% with panel-v2, representing a ~30% relative reduction (Figure 4). This improvement in on-target efficiency is critical (SELECTED\_BASES\_pct), as off-target reads can obscure variant interpretation and reduce overall data quality. By minimizing off-target capture, panel-v2 not only improves hybridization specificity but also enhances the reliability of downstream variant calling.

In line with these technical optimizations, we also observed a marked improvement in variant detection performance with panel-v2. While the initial workflow using panel-v1 already demonstrated high sensitivity for SNP detection, >97%, and a precision of 94%, the redesigned workflow with panel-v2 yielded even better results (Figure 2). Across all replicates, SNP sensitivity reached 99%, with precision ranging from 95% to 97%, reflecting more accurate identification of true-positive variants. Although indel calling remained more variable, slight improvements were also observed in panel-v2, with sensitivity ranging from 79% to 83% and precision from 56% to 64%.

These improvements at the panel design level complement the performance gains observed during the revalidation of our workflow and contribute to a more robust and clinically impactful screening tool.



**Figure 4. Distribution of off-target sequencing reads (% off bait) for exome capture panels v1 and v2.** Violin plots represent the full distribution of % off bait across samples for each panel. Box plots indicate the median and interquartile range, and black dots denote the mean. Panel v2 (blue, right) shows consistently lower off-target rates compared to panel v1 (red, left), reflecting improved target enrichment performance.

**4. Discussion**

In the BabyDetect project, we currently employ a tNGS panel encompassing 405 genes associated with 165 diseases (panel-v2). Our approach delivers high sensitivity and consistent detection of SNPs across all replicates, with robust precision metrics. As expected, detection of small insertions and deletions exhibits greater variability, particularly in regions of low sequence complexity, capture inefficiency, or homopolymers—challenges intrinsic to enrichment-based capture techniques. Indeed, indel calling remains technically more complex than SNP calling, as enrichment protocols introduce elevated indel error rates, especially in A/T-rich homopolymer stretches.

Improving capture efficiency can enhance indel sensitivity in tNGS and WES. Yet, whole-genome sequencing (WGS) offers a clear technical advantage by enabling more uniform coverage of the genome, with estimates suggesting that 60x is needed to recover 95% of indels [27]. WGS further extends these capabilities by including both coding and non-coding regions. This comprehensive

view enables the detection of structural variants, deep intronic and regulatory mutations, repeat expansions, and CNVs with higher sensitivity and resolution than targeted or exome-based approaches. Such breadth is particularly advantageous to increase diagnostic yield in neonatal screening programs, especially in disorders characterized by heterogeneous genetic architectures [28,29].

The transition from our initial panel (v1) to panel v2 was driven by the need to increase diagnostic yield by including newly validated disease genes and optimizing capture performance. Each modification of the panel, whether the inclusion of additional regions or the removal of poorly performing ones, requires extensive revalidation, which is resource-intensive and time-consuming. These repeated validation cycles can create bottlenecks in implementation and, in some cases, lead to sample backlogs. Retrospectively, the use of WES or WGS could have accelerated this development phase by enabling broader exploratory variant detection, reducing the need for iterative panel redesigns. WES enables the analysis of nearly all protein-coding regions of the genome, which harbor the majority of known disease-causing variants [30]. This allows for the detection of a wider spectrum of pathogenic single nucleotide variants (SNVs), small indels, and, with appropriate bioinformatic tools, some larger CNVs [31]. Nevertheless, the high per-sample cost and limited throughput of WES and WGS remain significant barriers to its use in population-scale screening programs, also considering the limitations associated with sequencing large number of samples on a single flowcell [32–34].

To date, over 6,500 samples have been processed using BabyDetect panels, representing approximately 6 % of Belgian newborns in a year. Extrapolating WES or WGS costs to this national scale underscores the current economic impracticality of these broader approaches for routine screening. Furthermore, beyond cost, these genome-wide approaches also introduce significant challenges. First, they generate a vast amount of data, increasing the burden of computational analysis, storage, and interpretation, especially in a time-sensitive context such as neonatal care [28]. Second, they yield a higher proportion of VUS, which complicates clinical decision-making and can lead to ethical dilemmas regarding disclosure and follow-up [24]. Moreover, the incidental identification of secondary findings, which are not related to the primary indication for testing, raises additional ethical and logistical concerns regarding consent, counseling, and long-term follow-up [35]. In contrast, targeted panels allow for streamlined analysis, reduced turnaround time, and alignment with established clinical actionability frameworks, making them more immediately implementable in public health settings.

Therefore, while WES and WGS hold promises for future implementation, especially in cases where a broader genetic investigation is warranted, current evidence supports the continued use of targeted panels as a pragmatic and scalable solution for population-wide neonatal screening. Strategic integration of CNV detection into panel-based methods, through algorithmic enhancements or supplemental assays, may further improve diagnostic performance without compromising operational feasibility.

Variant filtering and interpretation remain challenging for genomic NBS, especially in the case of neonatal screening in a considered healthy population where no phenotypic data is accounted. Our developed filtering strategy and conservative approach in reporting only known pathogenic and likely pathogenic variants allowed us to identify 1.8% of positive cases, with 0.8% not identified by conventional screening [14]. Filtering of pathogenic and likely pathogenic variants based on only known databases (ClinVar, Franklin, VarSome) and our own MVL may increase the number of clinical false negatives (1 false negative out of 71 positives [14] and 2 false negatives out of 114 positives – BabyDetect unpublished data). Alternatively, reporting VUS still remains challenging and is a subject of debate. In the context of newborn screening, overloading the variant interpretation pipeline with VUS variants and reviewing them manually can cause increase of screening TAT, whereas reporting of VUS can cause anxiety among screened population. Our developed filtering pipeline allows us to keep VUS in separate location in the database while not reviewing and not

reporting them. The VUS datasets could be used for future re-classification in case of available proof of pathogenicity and/or new scientific projects.

To define an appropriate balance between false negatives and false positives in BabyDetect, we assessed the trade-off between variant review workload and TAT, especially critical for early-onset treatable conditions. A higher number of variants flagged for manual review, especially VUS, would drastically increase the TAT, potentially delaying clinical intervention during the neonatal window where treatment timing is essential. Conversely, overly stringent filtering reduces the number of variants to review but increases the risk of false negatives. We therefore adopted a filtering strategy that minimizes manual burden without compromising sensitivity for known actionable conditions. This compromise was refined through retrospective analysis of early cohorts and benchmarking against conventional screening results, allowing us to calibrate filters to optimize both diagnostic yield and operational feasibility. Manual review is strictly limited to variants meeting defined pathogenicity criteria, and VUS are stored for future re-analysis, preserving both clinical rigor and workflow efficiency.

The main challenge for variant interpretation associated with asymptomatic newborns undergoing newborn screening is the absence of symptoms that can be predictive for any suspected disease. In BabyDetect, we implemented a structured variant filtering tree, not considering the Human Phenotype Ontology, which provides a standardized vocabulary of phenotypic abnormalities encountered in human disease. Our standardized framework integrates automated filtering, pathogenicity scoring based on ACMG/AMP guidelines, and manual curation steps. Each node of the decision tree was carefully designed to incorporate both objective evidence (e.g., variant frequency, gene-disease association, and expert-driven judgment). This architecture ensures reproducibility while preserving clinical oversight, especially for conflicting variants near the classification boundary between VUS and likely pathogenic.

There is another challenge associated with variant interpretation pipelines, such as the use of artificial intelligence (AI) based tools. Not underestimating the benefits associated with implementation and use of AI in variant filtering (exp. Franklin platform, <https://franklin.genoox.com>), it is important to have sufficient knowledge to correctly consider and analyze, evidence-based, the classification provided by AI tools for each filtered variant. In addition, as AI tools become more widely adopted in genomics, the field of explainable AI (XAI) has emerged as essential, particularly in clinical contexts where understanding the rationale behind a prediction is as important as the prediction itself. In variant interpretation, XAI approaches can enhance transparency by highlighting which features contributed most to a classification—such as functional annotations, conservation scores, or population data—thus allowing human experts to validate or challenge the output. This is especially relevant in NBS, where decisions must be timely, ethically sound, and clinically actionable. Recent efforts have underscored the importance of interpretable models in healthcare applications [36] and in variant classification workflows [37].

For laboratories aiming to implement newborn genomic screening, the integration of explainable AI into decision-support systems offers a promising path forward—but must be coupled with domain expertise, rigorous validation, and transparent governance structures. Defining thresholds for actionability, clarifying roles of AI outputs, and preserving the option for expert override are critical safeguards to ensure that automation enhances, rather than replaces, clinical judgment.

In the context of population-level screening, where clinical sensitivity and specificity are paramount, the bioinformatics pipeline must not only be accurate but also stable, auditable, and reproducible across time and environments [38]. Our current bioinformatics pipeline is under strict revision control with in silico revalidation of every major release, and execution is run on a dedicated High-Performance Cluster with containerized software tools under revision control as well. Our pipeline is implemented using a series of Bash scripts. While this legacy approach has been carefully maintained and regularly updated, it lacks features such as integrated workflow management, native scalability, and environment portability. As part of our continuous improvement strategy, and in line with our practices for other genomic workflows, we are planning to migrate this pipeline to a

workflow manager framework. This transition will improve modularity, reproducibility, containerization, and long-term maintainability, particularly critical in the context of clinical deployment at scale. We thus advocate for a modular, containerized architecture built with workflow managers such as Nextflow [39] or Snakemake [40], which allow explicit version control of all steps, tools, and reference files. Containerization (e.g., via Singularity [41,42] or Docker [43]) ensures computational reproducibility by isolating the runtime environment and avoiding hidden system-level dependencies. Each step in the pipeline—from base calling and demultiplexing to alignment, variant calling, annotation, and report generation—should be fully logged, traceable, and testable.

Our experience in deploying in-house pipelines under ISO 15189 accreditation has underscored the importance of rigorous development practices and traceability. All tools and custom scripts are versioned using Git, and changes are tracked through pull requests and structured code review. Every production release is associated with a changelog, and all parameter settings and reference files used during clinical runs are archived in immutable configurations. The use of containerization, primarily via Singularity, which is compatible with HPC environments, has proven essential for ensuring identical results across environments, a prerequisite for re-accreditation and external audits.

To support continuous integration and deployment, we have implemented automated test suites covering both unit-level validation of individual pipeline modules and integration-level checks using synthetic and well-characterized reference datasets. This approach allows us to detect regressions early, validate infrastructure updates, and minimize downtime during production transitions. Promoting pipeline changes from development to clinical production is governed by a formal approval workflow, which includes performance benchmarking, documentation updates, and review by a multidisciplinary team.

Importantly, we propose:

Immutable configuration tracking, where every analysis run is associated with a fixed pipeline version, tool versions, and parameters.

Automated unit and integration testing for all pipeline components, ensuring that updates or infrastructure changes do not introduce regressions.

Reference dataset benchmarking to regularly evaluate the pipeline against synthetic or known truth sets (e.g., Genome in a Bottle, synthetic mixtures), thereby safeguarding analytical performance.

Clear separation between development and production environments, with a formal promotion workflow when pipeline changes are validated and ready for deployment.

Data provenance mechanisms (e.g., checksums, sample lineage tracking) to ensure that outputs can be backtracked to raw data and initial parameters.

Furthermore, harmonization with external clinical guidelines should be embedded where applicable, particularly at the variant filtration and prioritization stages.

## 5. Conclusions

The BabyDetect project demonstrates that tNGS-based newborn screening is a reliable and scalable approach to detect treatable rare diseases not covered by conventional methods. Our validated workflow achieves high sensitivity and precision, with robust QC ensuring reproducibility across thousands of samples. Automation and panel optimization have improved efficiency and coverage. By focusing on clinically actionable variants, we balance diagnostic yield with manageable interpretation workload.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

**Author Contributions:** K.H. contributed to conceptualization, writing, methodology development, investigation, data curation, analysis, resources, software, validation, review and editing of the paper. L.H. contributed to conceptualization, writing, software development (consolidated the bioinformatics pipeline and sequencing results analysis), investigation, analysis, validation, review and editing of the paper. B.C. contributed to software development (consolidated the bioinformatics pipeline and sequencing results analysis) and the



writing, review and editing of the paper. V.J. contributed to data curation, formal analysis, investigation, validation, the writing, review and editing of the paper. F.P. contributed to methodology, validation (gene panel curation), data curation, and review and editing of the paper. M.M. contributed to formal analysis and resources. Ch.F. contributed to formal analysis and resources. C.F. contributed to methodology, investigation, analysis, validation, the review and editing of the paper. D.M. contributed to funding acquisition and project administration. T.D. contributed to funding acquisition, methodology, project administration, resources, and the writing, review and editing of the paper. V.B. contributed to resources and review and editing of the paper. L.S. contributed to conceptualization, funding acquisition, investigation, project administration, resources, supervision, and the writing, review and editing of the paper. L.P. contributed to conceptualization, writing, software development (consolidated the bioinformatics pipeline and sequencing results analysis), investigation, analysis, validation, supervision, review and editing of the paper. F.B. contributed to conceptualization, writing, funding acquisition, investigation, analysis, validation, data curation, project administration, resources, supervision, and the writing, review and editing of the paper. K.H. and L.H. contributed equally to this work. All authors approved the paper and agreed to be accountable for the accuracy of the reported findings.

**Acknowledgements:** This work was supported by private and public grants (Sanofi (SGZ-2019-12843), Orchard Therapeutics, Takeda, Léon Frédéricq Foundation) and by a donation from LaCAR MDx. V.B. was funded by a Région Wallonne grant (WALGEMED project 1710180). The funders of this study had no role in study design, data collection, data analysis, data interpretation or writing of the report. We thank BabyDetect Expert Panel : Serpil Alkan <sup>§</sup>, Christophe Barrea <sup>§</sup>, Paulina Bartoszek <sup>h</sup>, Emeline Bequet <sup>§</sup>, Hedwige Boboli <sup>§</sup>, Romain Bruninx <sup>§</sup>, Laure Collard <sup>§</sup>, Aurore Daron <sup>§</sup>, François-Guillaume Debray <sup>e</sup>, Adeline Jacquinet <sup>e</sup>, Julie Harvengt <sup>e</sup>, Saskia Bulk <sup>e</sup>, Marie-Françoise Dresse <sup>§</sup>, Jean-Marie Dubru <sup>§</sup>, Iulia Ebetiuc <sup>§</sup>, Benoit Florkin <sup>§</sup>, Caroline Jacquemart <sup>§</sup>, Céline Kempeneers <sup>§</sup>, Nadège Hennuy <sup>§</sup>, Marie-Christine Lebrethon <sup>§</sup>, Marie Leonard <sup>§</sup>, Patricia Leroy <sup>§</sup>, Angélique Lhomme <sup>§</sup>, Jacques Lombet <sup>§</sup>, Sabine Michotte <sup>§</sup>, Ariane Milet <sup>§</sup>, Anne-Simone Parent <sup>§</sup>, Vincent Rigo <sup>§</sup>, Marie-Christine Seghaye <sup>§</sup>, Sandrine Vaessen <sup>§</sup> g. CHR Citadelle, Liege, Belgium ; h. Cliniques Universitaires Saint-Luc, Brussels, Belgium.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of CHU Liege (n° 2021/239).

**Informed Consent Statement:** Before any sampling was done, parental informed consent was obtained from all subjects involved in the study.

**Data availability and Publication Ethics:** In accordance with the informed consent agreements, the raw sequencing data can be stored for each patient for 10 years. Metadata files are retained with no time limit. The raw sequencing data and metadata files generated in the study cannot be made publicly available because of ethical and data protection constraints. Deidentified data that support the results reported in this article will be made available to suitably qualified researchers through any requests that comply with ethical standards to the corresponding author (K.H., khovhannesyan@chuliege.be). Data must be requested between 1 and 12 months after the paper has been published, and the proposed use of the data must be approved by an independent review committee identified for this purpose by mutual agreement. Requests will be forwarded by the corresponding author to the identified ethics review committee. Upon acceptance by that committee, deidentified data will be provided by the corresponding author to the applicants through a secure web platform within 2 months. The minimum dataset required to run our code and reproduce results is available via Zenodo at <https://doi.org/10.5281/zenodo.13935241> [ref 20].

**Conflicts of Interest:** F.B. has consulted for LaCAR MDx and has presented lectures for Novartis and Sanofi. T.D. has presented lectures for Biogen, Roche and Novartis. L.S. is a member of scientific advisory boards or has consulted for Biogen, Novartis, Roche, Illumina, Sanofi, Scholar Rock, LaCAR MDx and BioHaven. The other authors declare no competing interests.

## References

1. Andermann, A. Revisiting wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull. World Health Organ.* 86, 317–319 (2008).

2. Fernhoff, P. M. Newborn Screening for Genetic Disorder. *Pediatr. Clin. North Am.* Volume 56, Pages 505-513 (2009).
3. Engel, A. G. Congenital Myasthenic Syndromes. *Rosenb. Mol. Genet. Basis Neurol. Psychiatr. Dis. Fifth Ed. Acad. Press* Pages 1191-1208 (2015) doi:<https://doi.org/10.1016/B978-0-12-410529-4.00099-1>.
4. Brennenstuhl, H., Jung-Klawitter, S., Assmann, B. & Opladen, T. Inherited Disorders of Neurotransmitters: Classification and Practical Approaches for Diagnosis and Treatment. *Neuropediatrics* 50, 002–014 (2019).
5. Sharma, S. & Prasad, A. Inborn Errors of Metabolism and Epilepsy: Current Understanding, Diagnosis, and Treatment Approaches. *Int. J. Mol. Sci.* 18, 1384 (2017).
6. Amir, F. *et al.* The Clinical Journey of Patients with Riboflavin Transporter Deficiency Type 2. *J. Child Neurol.* 35, 283–290 (2020).
7. Roberts, E. A. & Schilsky, M. L. Current and Emerging Issues in Wilson's Disease. *N. Engl. J. Med.* 389, 922–938 (2023).
8. Van Der Burg, M., Mahlaoui, N., Gaspar, H. B. & Pai, S.-Y. Universal Newborn Screening for Severe Combined Immunodeficiency (SCID). *Front. Pediatr.* 7, 373 (2019).
9. Boemer, F. *et al.* Three years pilot of spinal muscular atrophy newborn screening turned into official program in Southern Belgium. *Sci. Rep.* 11, 19922 (2021).
10. Müller-Felber, W. *et al.* Newbornscreening SMA – From Pilot Project to Nationwide Screening in Germany. *J. Neuromuscul. Dis.* 10, 55–65 (2023).
11. Servais, L., Dangouloff, T., Muntoni, F., Scoto, M. & Baranello, G. Spinal muscular atrophy in the UK: the human toll of slow decisions. *The Lancet* 405, 619–620.
12. Ziegler, A. *et al.* Expanded Newborn Screening Using Genome Sequencing for Early Actionable Conditions. *JAMA* 333, 232 (2025).
13. Chen, T. *et al.* Genomic Sequencing as a First-Tier Screening Test and Outcomes of Newborn Screening. *JAMA Netw. Open* 6, e2331162 (2023).
14. Boemer, F. *et al.* Population-based, first-tier genomic newborn screening in the maternity ward. *Nat. Med.* (2025) doi:10.1038/s41591-024-03465-x.
15. The BabySeq Project Team *et al.* The BabySeq project: implementing genomic sequencing in newborns. *BMC Pediatr.* 18, 225 (2018).
16. Kingsmore, S. F. *et al.* Genome-based newborn screening for severe childhood genetic diseases has high positive predictive value and sensitivity in a NICU pilot trial. *Am. J. Hum. Genet.* 111, 2643–2667 (2024).
17. Jansen, M. E., Klein, A. W., Buitenhuis, E. C., Rodenburg, W. & Cornel, M. C. Expanded Neonatal Bloodspot Screening Programmes: An Evaluation Framework to Discuss New Conditions With Stakeholders. *Front. Pediatr.* 9, 635353 (2021).
18. Minten, T. *et al.* Data-driven consideration of genetic disorders for global genomic newborn screening programs. *Genet. Med.* 27, 101443 (2025).
19. Dangouloff, T. *et al.* Feasibility and Acceptability of a Newborn Screening Program Using Targeted Next-Generation Sequencing in One Maternity Hospital in Southern Belgium. *Children* 11, 926 (2024).
20. CHARLOTEAUX, B. *et al.* Humanomics. *GitLab* <https://doi.org/10.5281/ZENODO.13935241> (2024).
21. Wagner, J. *et al.* Benchmarking challenging small variants with linked and long reads. *Cell Genomics* 2, 100128 (2022).
22. Broad Institute. Picard toolkit. *Broad Inst. GitHub Repos.* (2019) doi:<https://broadinstitute.github.io/picard/>.
23. Tawari, N. R. *et al.* ChronQC: a quality control monitoring system for clinical next generation sequencing. *Bioinformatics* 34, 1799–1800 (2018).
24. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* 17, 405–424 (2015).
25. Kopanos, C. *et al.* VarSome: the human genomic variant search engine. *Bioinformatics* 35, 1978–1980 (2019).
26. Ding, Y. *et al.* Scalable, high quality, whole genome sequencing from archived, newborn, dried blood spots. *Npj Genomic Med.* 8, 5 (2023).
27. Fang, H. *et al.* Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med.* 6, 89 (2014).

28. Goldfeder, R. L. *et al.* Medical implications of technical accuracy in genome sequencing. *Genome Med.* 8, 24 (2016).
29. NIHR BioResource—Rare Disease *et al.* Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med.* 45, 627–636 (2019).
30. Taylor, J. C. *et al.* Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat. Genet.* 47, 717–726 (2015).
31. Royer-Bertrand, B. *et al.* CNV Detection from Exome Sequencing Data in Routine Diagnostics of Rare Genetic Disorders: Opportunities and Limitations. *Genes* 12, 1427 (2021).
32. Jegathisawaran, J., Tsiplova, K., Hayeems, R. & Ungar, W. J. Determining accurate costs for genomic sequencing technologies—a necessary prerequisite. *J. Community Genet.* 11, 235–238 (2020).
33. Schwarze, K., Buchanan, J., Taylor, J. C. & Wordsworth, S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet. Med.* 20, 1122–1130 (2018).
34. Nurchis, M. C. *et al.* Cost-Effectiveness of Whole-Genome vs Whole-Exome Sequencing Among Children With Suspected Genetic Disorders. *JAMA Netw. Open* 7, e2353514 (2024).
35. Green, R. C. *et al.* Correction: Corrigendum: ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* 19, 606 (2017).
36. Islam, M. R., Ahmed, M. U., Barua, S. & Begum, S. A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks. *Appl. Sci.* 12, 1353 (2022).
37. De Paoli, F. *et al.* Digenic variant interpretation with hypothesis-driven explainable AI. *NAR Genomics Bioinforma.* 7, lqaf029 (2025).
38. Wratten, L., Wilm, A. & Göke, J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nat. Methods* 18, 1161–1168 (2021).
39. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 35, 316–319 (2017).
40. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Research* 10, 33 (2021).
41. Kurtzer, G. M. *et al.* hpcng/singularity: Singularity 3.7.3. Zenodo <https://doi.org/10.5281/ZENODO.1310023> (2021).
42. Kurtzer, G. M., Sochat, V. & Bauer, M. W. Singularity: Scientific containers for mobility of compute. *PLOS ONE* 12, e0177459 (2017).
43. Merkel, D. Docker: lightweight Linux containers for consistent development and deployment. *Linux J.* 2014, 2 (2014).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.