

Article

# Validation of an NGS Panel Designed for Detection of Actionable Mutations in Tumors Common in Latin America

Mauricio Salvo <sup>1</sup>, Evelin González-Feliú <sup>1</sup>, Jessica Toro <sup>1</sup>, Iván Gallegos <sup>1,2</sup>, Ignacio Maureira <sup>1,3</sup>, Nicolás Miranda-González <sup>1</sup>, Olga Barajas <sup>1,4,5</sup>, Eva Bustamante <sup>5</sup>, Mónica Ahumada <sup>1,4</sup>, Alicia Colombo <sup>1,2</sup>, Ricardo Armisén <sup>6</sup>, Camilo Villamán <sup>1</sup>, Carolina Ibañez <sup>7</sup>, María Loreto Bravo <sup>7</sup>, Verónica Sanhueza <sup>8</sup>, M. Loreto Spencer <sup>9</sup>, Gonzalo de Toro <sup>10</sup>, Erik Morales <sup>11,12</sup>, Carolina Bizama <sup>13</sup>, Patricia García <sup>13</sup>, Ana María Carrasco <sup>5</sup>, Lorena Gutiérrez <sup>14</sup>, Justo Lorenzo Bermejo <sup>15</sup>, Ricardo A. Verdugo <sup>1,16,\*</sup> and Katherine Marcelain <sup>1,\*</sup>

**Citation:** Salvo, M.; González-Feliú, E.; Toro, J.; Gallegos, I.; Maureira, I.; Miranda-González, N.; Barajas, O.; Bustamante, E.; Ahumada, M.; Colombo, A.; et al. Validation of an NGS Panel Designed for Detection of Actionable Mutations in Tumors Common in Latin America. *J. Pers. Med.* **2021**, *11*, 899. <https://doi.org/10.3390/jpm11090899>

Academic Editor: Maria Graciela Castro

Received: 7 July 2021

Accepted: 3 September 2021

Published: 8 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- <sup>1</sup> Department of Basic and Clinical Oncology, Faculty of Medicine, Universidad de Chile, Santiago 8330015, Chile; mauricio.salvo@roche.com (M.S.); evelin@uchile.cl (E.G.-F.); jessicatoro@med.uchile.cl (J.T.); igallegos@hcuch.cl (I.G.); ignaciomaureira@uchile.cl (I.M.); miranda528491@gmail.com (N.M.-G.); olbeba@gmail.com (O.B.); mahumada@hcuch.cl (M.A.); acolombo@hcuch.cl (A.C.); c.villaman@gmail.com (C.V.)
  - <sup>2</sup> Department of Pathology, Hospital Clínico de la Universidad de Chile, Santiago 8380456, Chile
  - <sup>3</sup> Department of Medical Technology, Faculty of Medicine, Universidad de Chile, Santiago 8330015, Chile
  - <sup>4</sup> Department of Internal Medicine, Hospital Clínico Universidad de Chile, Santiago 8380456, Chile
  - <sup>5</sup> Fundación Arturo López Pérez, Santiago 7500921, Chile; bustamantee@falp.org (E.B.); dracarrasqui@gmail.com (A.M.C.)
  - <sup>6</sup> Center for Genetics and Genomics, Instituto de Ciencias e Innovación en Medicina, Facultad de Medicina Clínica Alemana Universidad del Desarrollo, Santiago 8320000, Chile; rarmisen@udd.cl
  - <sup>7</sup> Department of Hematology & Oncology, Faculty of Medicine, Pontificia Universidad Católica de Chile (PUC), Santiago 3580000, Chile; doctoraibanez@gmail.com (C.I.); marialoretobravo@gmail.com (M.L.B.)
  - <sup>8</sup> Department of Pathology, Hospital Padre Hurtado, Santiago 8710022, Chile; verosanhueza@yahoo.com
  - <sup>9</sup> Department of Pathology, Hospital Clínico Regional Guillermo Grant Benavente, Concepción 4070038, Chile; loreto Spencer@gmail.com
  - <sup>10</sup> School of Medical Technology, Universidad Austral de Chile at Puerto Montt, Puerto Montt 5110566, Chile; gdetoro@gmail.com
  - <sup>11</sup> Department of Pathology, Hospital Regional de Talca, Talca 3460000, Chile; emoralesm@hospitaldetalca.cl
  - <sup>12</sup> Department of Preclinical Sciences, Faculty of Medicine, Universidad Católica del Maule, Talca 3460000, Chile
  - <sup>13</sup> Department of Pathology, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago 3580000, Chile; carobizama@gmail.com (C.B.); pagarmu@gmail.com (P.G.)
  - <sup>14</sup> Department of Pathology, Hospital San Juan de Dios, Santiago 8320000, Chile; lorenagutierrezc@yahoo.es
  - <sup>15</sup> Institute of Medical Biometry, Universität Heidelberg, 69120 Heidelberg, Germany; lorenzo@imbi.uni-heidelberg.de
  - <sup>16</sup> Human Genetics Program, ICBM, Faculty of Medicine, Universidad de Chile, Santiago 8330015, Chile
- \* Correspondence: raverdugo@uchile.cl (R.A.V.); kmarcelain@uchile.cl (K.M.);  
Tel.: +56-22978-9527 (R.A.V.); +56-22978-9562 (K.M.)

**Abstract:** Next-generation sequencing (NGS) is progressively being used in clinical practice. However, several barriers preclude using this technology for precision oncology in most Latin American countries. To overcome some of these barriers, we have designed a 25-gene panel that contains predictive biomarkers for most current and near-future available therapies in Chile and Latin America. Library preparation was optimized to account for low DNA integrity observed in formalin-fixed paraffin-embedded tissue. The workflow includes an automated bioinformatic pipeline that accounts for the underrepresentation of Latin Americans in genome databases. The panel detected small insertions, deletions, and single nucleotide variants down to allelic frequencies of 0.05 with high sensitivity, specificity, and reproducibility. The workflow was validated in 272 clinical samples from several solid tumor types, including gallbladder (GBC). More than 50 biomarkers were detected in these samples, mainly in BRCA1/2, KRAS, and PIK3CA genes. In GBC, biomarkers for PARP, EGFR, PIK3CA, mTOR, and Hedgehog signaling inhibitors were found. Thus, this small NGS panel is an accurate and sensitive method that may constitute a more cost-efficient alternative

to multiple non-NGS assays and costly, large NGS panels. This kind of streamlined assay with automated bioinformatics analysis may facilitate the implementation of precision medicine in Latin America.

**Keywords:** NGS-panel; target therapies; predictive biomarkers; somatic variants; gallbladder cancer; Latin America

---

## 1. Introduction

In recent decades, molecular pathology has advanced substantially thanks to the exponential growth of genetic sequencing technology. The introduction of next-generation sequencing (NGS) opened the doors to high-throughput, multi-gene, massive data collection. This tool's ability to sequence more, faster, and at a reduced cost has made it attractive for many clinical research applications. In cancer, using this technology to interrogate solid tumor samples has propelled a massive characterization of genes involved in the disease [1,2]. This rise in "oncogenomics" has been accompanied by an increase in cancer drug approval and development [3,4]. Identifying tumor-specific genetic signatures and correlating them to treatment outcome has evolved into a strategy termed "precision medicine", a new diagnostic and treat process based on approved genomic biomarkers [4].

In Latin America and the Caribbean, 1.4 million new cancer cases were estimated to occur in 2018, while mortality rates varied among and within the region [5,6]. The most common types of cancer with the highest incidence are prostate (age standardized rate (ASR) 60.4), breast (ASR 56.8), colorectal (ASR 18.6), cervix uteri (ASR 15.2), lung (ASR 13.1), and stomach (ASR 9.5) cancers [1]. Overall, estimated age-standardized cancer incidence rates in Latin America are lower than those reported in North America and some European countries; however, the region exhibits higher mortality rates [7]. This paradox reflects the disparities in early diagnosis and treatment opportunities in the region.

In high- and medium-income countries, precision medicine is making its way into standard cancer treatment, improving survival and investigational drug trial success for many patients. A combination of factors prevents this helpful tool from becoming accessible to most of the world's population. In Latin America, approved and available gene-based cancer screening assays are often solutions designed to meet first world standards. These large panels are great diagnostic tools in regions of abundant therapy options, but for Latin America and other regions, they are not cost effective, leaving behind a need for more comprehensive regional solutions. Additionally, the absence of automated clinician-ready reporting for many of these approved panels creates another major cost and obstacle to their widespread implementation. As a result, NGS-based oncology panels do not appear to be cost-effective solutions for many governments and are not being implemented in health and insurance systems despite local sequencing capabilities. This scenario creates an urgent need for customized validated solutions and data interpretation in a clinical environment [8].

In addition, an important caveat to interpreting Latin American cancer patient's genetic data is the under-representation of Latin American individuals in global resources characterizing the frequency of both germinal and cancer genome variants. Great cancer genomics efforts, such as TCGA and ICGC, are deprived of minorities (including subjects of Hispanic ethnicity [9]), limiting their capacity to describe somatic mutations with a prevalence below 10% and overcome the somatic background mutation frequency in specific ethnic groups [10]. For instance, the average Amerindian ancestry in cancer patients across all cohorts in TCGA is about 4% [9,11]. Additionally, the Latin American population is under-represented in germline variant repositories, which may induce a false categorization and overestimation of somatic variants [12–14]. Thus, an additional blood sample should accompany the tumor sample, increasing the sequencing costs.

To address these challenges, we designed, optimized, and validated a hybridization-based target enrichment workflow with multiple automated analyses capable of detecting variants in 25 genes; 23 of them with an association to a drug's response supported by the FDA or well-powered studies with consensus from experts in the field. Although this panel was designed to meet current and near-future Chilean precision oncology needs, we expect the panel and workflow to be relevant to other countries in the region. This workflow was locally validated using breast, colorectal, gastric, ovarian, pancreatic, and gallbladder tumor tissue samples and we reported its ability to detect single nucleotide variants (SNVs) and small insertions and deletions with high sensitivity and specificity. Additionally, high reproducibility was obtained for non-synonymous variants between and within runs. Finally, to address the shortage of health professionals trained in bioinformatics, the entire workflow, including quality control of sequencing data and calling for somatic variants, was automated and made available.

## 2. Materials and Methods

### 2.1. Panel Design

The panel targets hotspots, selected exons, or complete coding regions of 25 genes and includes predictive biomarkers in solid tumors. We refer to this panel, plus its associated workflow and analysis, as TumorSec™. For selecting targeted regions, biomarker genes classified with evidence 1, 2, 3a, 3b, R1, and R2 were selected for solid tumors in the OncoKB database ([www.oncokb.org](http://www.oncokb.org), accessed on 1 June 2021) [15]. Next, biomarker mutations with level of clinical evidence A, B, and C were selected in the Clinical Interpretation of Variants in Cancer, CiVic database (<https://civicdb.org/home>, accessed on 1 October 2018) [16] and the Variant Interpretation for Cancer Consortium (VICC) meta-knowledgebase [17]. Biomarkers were selected based on their level of evidence and incidence in the targeted tumor in Latin America. *TP53* and *ARID1A* complete coding regions were incorporated, as they contain prognosis and predictive chemotherapy biomarkers. The complete list of genes and drug associations is provided in Table 1.

Synthesis of the soluble, biotinylated probe library was performed on the NimbleGen cleavable array platform (SeqCap EZ Choice (RUO); Roche/NimbleGen, Basel, Switzerland). The probe design was optimized using the NimbleDesign software utility (NimbleGen, Roche, Basel, Switzerland).

**Table 1.** Genes included in the panel and their therapy association.

Gene	Drugs	Tumor Type	Evidence <sup>1</sup>
AKT1 *	AZD-5363	Breast cancer	B
		Ovarian cancer	
		Endometrial cancer	
ALK	Ceritinib	Non-small cell lung cancer	A
	Crizotinib		
	Alectinib		
	Brigatinib		
	Lorlatinib		
ARID1A *	Trastuzumab	Breast cancer	C
	ENMD-2076	Ovarian clear cell cancer	
	Bevacizumab	Renal cell carcinoma	
	Everolimus		
BRAF *	Encorafenib + Cetuximab	Colorectal cancer	A
	Vemurafenib	Melanoma	
	Dabrafenib	Non-small cell lung cancer	
	Trametinib + Dabrafenib	Anaplastic thyroid cancer	
	Cobimetinib + Vemurafenib	Hairy cell leukemia	
	Trametinib	Pilocytic astrocytoma	
	Encorafenib + Binimetinib	Ganglioglioma	

	Vemurafenib + Cobimetinib, Trametinib + Dabrafenib Vemurafenib + Cobimetinib Encorafenib + Panitumumab	Pleomorphic xanthoastrocytoma	
BRCA1 *	Olaparib Niraparib Rucaparib Talazoparib	Ovarian cancer Peritoneal serous carcinoma Breast cancer Prostate cancer Ovary/fallopian tube Pancreatic cancer	A
BRCA2 *	Olaparib Rucaparib Talazoparib	Ovarian cancer Peritoneal serous carcinoma Breast cancer Prostate cancer Ovary/fallopian tube	A
CDK4 *	Palbociclib Abemaciclib	Liposarcoma	B
EGFR	Erlotinib Afatinib Osimertinib Gefitinib Dacomitinib	Non-small cell lung cancer	A
ERBB2	Trastuzumab Fam-Trastuzumab deruxtecan-nxki Trastuzumab + Pembrolizumab Afatinib	Breast cancer Gastric adenocarcinoma Gastroesophageal junction adenocarcinoma Non-small cell lung cancer	A
ESR1	Anastrozole Fulvestrant Palbociclib	Breast cancer	B
IDH2	Enasidenib	Acute myeloid leukemia	A
KIT	Sunitinib Imatinib Regorafenib Sorafenib Ripretinib	Gastrointestinal stromal tumor Melanoma	A
KRAS *	Cetuximab Panitumumab Erlotinib Lapatinib Regorafenib Selumetinib Gefitinib Afatinib Icotinib Irinotecan	Colorectal cancer Non-small cell lung cancer	A
MET	Crizotinib Capmatinib Tepotinib	Non-small cell lung cancer	A
MTOR *	Everolimus Temsirolimus	Renal cell carcinoma Bladder Cancer	B
NRAS *	Cetuximab Panitumumab	Colorectal cancer	A
PDGFRA	Imatinib Sunitinib Regorafenib	Gastrointestinal stromal tumor	A

PI3KCA *	Buparlisib Serabelisib Alpelisib Copanlisib	Breast cancer	A
PTCH1 *	Vismodegib	Skin basal cell carcinoma Squamous cell carcinoma Medulloblastoma	A
PTEN *	Everolimus Pembrolizumab Cetuximab Sorafenib	Renal cell carcinoma Glioma Head and neck squamous cell carcinoma Colorectal cancer Hepatocellular carcinoma	B
ROS1	Crizotinib Alectinib Ceritinib	Non-small cell lung cancer	C
SMO	Vismodegib	Skin basal cell carcinoma	B
TP53 *	Prognosis	Various	A
TSC1 *	Everolimus	Giant cell astrocytoma Renal cell carcinoma Renal angiomyolipoma	A
TSC2 *	MTOR inhibitors	Giant cell astrocytoma Renal cell carcinoma Renal angiomyolipoma	A

<sup>1</sup> Level of evidence according to AMP/ASCO/CAP Level A: FDA-approved therapy included in professional guidelines; B: well-powered studies with consensus from experts in the field; C: FDA-approved therapies for different tumor types or investigational therapies [18]. The highest level of evidence is shown. Some biomarkers may have additional indications with lower levels of evidence for different cancer types or protocols. \* For these genes, all exons were targeted.

## 2.2. Sample Information

In total, 183 tumor tissue samples were sequenced for this study. In total, 19 were freshly frozen (FF): 13 colorectal and 6 breast; 164 were formalin-fixed paraffin-embedded (FFPE) blocks: 9 breast, 71 ovary, 1 gastric, 43 gallbladder, and 40 colorectal tumors. Additionally, DNA from 89 whole blood or buffy coat samples were sequenced: 7 from colorectal and 72 from breast cancer patients. Colorectal and gastric cancer samples were obtained from the “Biobanco de Tejidos y Fluidos de la Universidad de Chile.” To capture real world heterogeneity in sample quality, breast, ovary, and gallbladder FFPE tissue samples were collected from the pathology services from several sites along the country (Fundación Arturo López Pérez, Clínica Dávila, Clínica Indisa, Red UC Christus, Biobanco de Tejidos y Fluidos, Hospital Padre Hurtado, Hospital Regional de Concepción, Hospital Regional de Talca, Hospital de Puerto Montt, Hospital San Juan de Dios, Hospital Santiago Oriente Doctor Luis Tisné Brousse, Instituto Nacional del Cáncer, Hospital del Salvador, Hospital Regional de Coquimbo, Hospital Regional de Arica, Hospital Clínico San Borja Arriarán).

## 2.3. Control Samples

Three reference standard DNA samples from Horizon Discovery (Cambridge, United Kingdom) were used as positive controls for variant calling: HD200 (FFPE somatic), HD793, and HD794 (germline BRCA1/2 variants).

## 2.4. DNA Extraction, Quantification, and Sample Quality Control

DNA from frozen tissues was extracted using the QIAamp DNA Mini Kit (Qiagen, Maryland, USA). FFPE tissue DNA was extracted using GeneJet FFPE DNA Purification Kit and RecoverAll™ Total Nucleic Acid Isolation (Invitrogen, Thermo Fisher Scientific,

Carlsbad, USA), following the manufacturer's instructions, with overnight lysis instead of the suggested 1–2 h for FFPE tissue. Germline DNA was purified from whole blood samples or buffy coat using the Wizard® Genomic DNA Purification Kit (Promega, Madison, USA), according to manufacturer's protocol.

Purified DNA was quantified using the Qubit™ dsDNA HS Assay and Quant-IT™ Picogreen® dsDNA Reagent Kit (Invitrogen, Thermo Fisher Scientific, Carlsbad, USA). The purity of DNA was assessed by measuring the 260/280 nm absorbance ratio. For FFPE samples, fragment length and degradation were assessed using the HS Genomic DNA Analysis Kit (DNF-488) in a Fragment Analyzer (Agilent, formerly Advanced Analytical). DNA ranged from >1000 bp to 200 bp. Samples with <200 bp are not recommended for processing with the TumorSec workflow.

### 2.5. Library Preparation

Then, 100–150 ng of DNA (blood and frozen tissues) and 200 ng of DNA (FFPE) were used as input for sequencing library preparation. NGS libraries were prepared using KAPA HyperPlus Library Preparation Kit (Kapa Biosystems, Cape Town, South Africa). A double size selection was performed in libraries prepared with DNA from frozen tissue and blood and a single size selection for DNA from FFPE. Libraries were quantified using the Qubit™ dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific, Carlsbad, USA) and Quant-IT™ Picogreen® dsDNA Reagent Kit (Invitrogen). The quality of the amplified library was checked by measuring the 260/280 absorbance ratio and fragment's length, using the HS NGS Analysis Kit (DNF-474) in a Fragment Analyzer (Agilent, formerly Advanced Analytical).

### 2.6. Target Enrichment

Prepared DNA libraries (1200 ng total mass) were captured by hybridization probes (Roche NimbleGen SeqCap EZ. Roche, Pleasanton, USA). The number of samples used for pre-capture multiplexing was based on sample type: six were pooled for FFPE and blood samples, while fresh frozen tumor samples were pooled in reactions of four. Captured libraries were assessed for concentration and size distribution to determine molarity.

### 2.7. Sequencing Run Set-Up

Libraries were diluted to a concentration of 4 nmol/L and processed for sequencing, according to the manufacturer's instructions (Illumina, San Diego, CA, USA). The final captured library concentration for sequencing was 9.4–9.5 pM. Libraries were sequenced in an Illumina® MiSeq System using paired-end, 300 cycles (MiSeq Reagent Kits v2, Illumina® Illumina, San Diego, CA, USA).

### 2.8. Bioinformatic Analyses

The bioinformatic pipeline is summarized in Supplementary Figure S1. Filtering of reads and base correction were performed with the fastp v0.19.11 tool. The filtered reads align with the reference genome GRCh37/hg19 using Burrows–Wheeler Alignment mem (BWA mem v0.7.12). MarkDuplicates tool of Picard v2.20.2-8 was applied to identify duplicates. To reduce the number of mismatches to the reference genome, the reads were realigned with RealignerTargetCreator and IndelRealigner from GATK v3.8 [19]. Finally, the quality scores were re-calibrated with the combination of GATK's BaseRecalibrator and PrintReads tools [19].

The SomaticSeq v.3.3.0 program was used to call the variants in single-mode using only tumor sequence data [20]. This tool maximizes the sensitivity by combining the result of five next-generation variant SNV callers—Mutect2 [21], VarScan2 [22], VarDict [23], LoFreq [24], and Strelka [25]—adding Scalpel for indels. The reported SNVs were identified by at least three out of five SNV callers, and the reported indels by at least three out

of the six callers. The consensus variants obtained by SomatiSeq were annotated using the Cancer Genome Interpreter (<https://www.cancergenomeinterpreter.org/>, accessed on 1 June 2021) [26] and ANNOVAR [27] using RefGene, GnomAD v2.1.1 (genome and exome), ESP6500, ExAC v0.3, 1000 Genomes phase 3, CADD v1.3, dbSNP v150, COSMIC v92, and CLINVAR.

### 2.9. Variant Filtering and Sequence Quality Reporting

Variants with allele frequencies greater than 0.5 and with an altered allele depth  $\geq 12$  reads were selected. These thresholds were established as the limit of detection (LOD) for the NGS TumorSec panel following the recommendations of the Association for Molecular Pathology and College of American Pathologists [28]. Polymorphisms were eliminated, discarding alleles reported in 1000 Genomes, ESP6500, GnomAD, or ExAC [27] with a frequency greater than 0.01. Filtering was extended to include all under-represented populations that had information in GnomAD and ExAC. Additionally, a dataset containing genetic variants in Chilean individuals was used for further filtering.

The bioinformatics pipeline was executed automatically, creating pdf reports that allowed an easy view of quality metrics per sample. For this purpose, the programs FastQC v0.11.8, Qualimap v2.2.2a, Mosdepth v0.2.5, and MultiQC v1.8 were executed between the pre-processing of the bioinformatics workflow. The main metrics are the number of initial raw reads, the percentage of filtered reads, the duplication rate, the number of reads on target regions, the average depth on-target regions, the uniformity percentage, and the ratio of on-target regions with a minimum coverage of  $100\times$  to  $500\times$ . For variant calling and annotation, the coverage threshold for FFPE and FF was set at  $300\times$  in at least 80% of target regions.

The bioinformatic pipeline and tutorial can be found in the GitHub repository called Pipeline-TumorSec (<https://github.com/u-genoma/Pipeline-TumorSec>, accessed on 1 June 2021).

### 2.10. Germline Variant Calling

Data were pre-processed following the pipeline shown in Supplementary Figure S1. Variant calls were made using the GATK HaplotypeCaller tool. A minimum confidence threshold of 30 was set for variant calling. Additionally, a coverage threshold was set at  $200\times$  in at least 80% of target regions. Finally, a variant calling hard-filter for SNPs and indels was applied separately following GATK recommendations [29].

## 3. Results

### 3.1. Panel Design and Sequencing Metrics

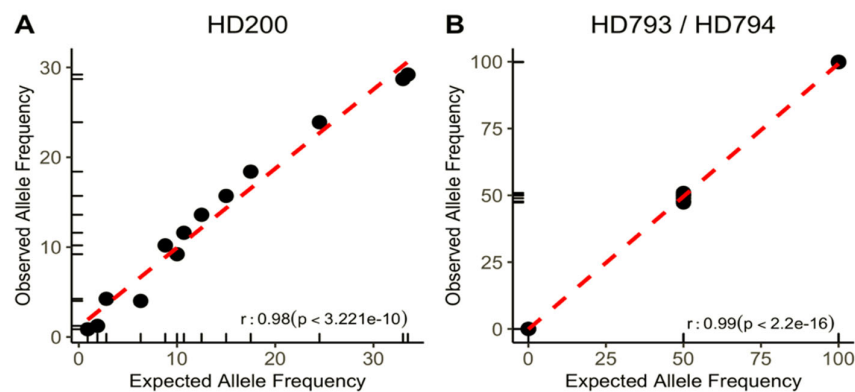
A total of 25 genes were included in this target enrichment panel, covering 98 kb of sequence length. Design details broken down by gene are shown in Table 1. In total, 79% (15/19) of fresh frozen, 71% (116/164) of FFPE, and 89% (79/89) of blood samples processed passed the sequencing quality threshold, capturing a minimum of 80% of target regions at a depth of  $300\times$ . A summary of relevant sequencing metrics for all 210 passed samples is shown in Supplementary Table S1. FFPE samples showed a high percentage of duplicates and off-target reads. Uniformity was  $>90\%$  for all sample types and  $>90\%$  of targeted regions had  $\geq 300\times$  coverage (Supplementary Figure S2).

### 3.2. Panel Performance

The panel's performance was calculated using the reference HD200 (Horizon Discovery) standard FFPE sample containing characterized mutations in the following genes: *BRAF*, *KIT*, *EGFR*, *KRAS*, *NRAS*, *PIK3CA*, *ARID1A*, and *BRCA2*. As observed in Figure 1A, the assay captured all 13 positive variants. A 0.98 coefficient of correlation (r-squared)

was extrapolated with a  $p$ -value of  $3.221 \times 10^{-10}$  between expected variant allele frequencies (VAF) from the positive control and those reported by the assay. VAFs ranged from 24.5% to as low as 1%, showing the assay's high analytical sensitivity.

Additionally, the panel's performance for detecting *BRCA1/2* germline mutations, which are predictive biomarkers for PARP inhibitor therapy in breast, ovarian, and prostate cancer, was tested (Table 1). Thus, references DNA HD793 and HD794 (Horizon Discovery), which contain known germline variants in *BRCA1* and *BRCA2* at VAF of 50 and 100%, were sequenced. Figure 1B shows that 11 out of 11 reported variants were detected at the expected VAF. The correlation coefficient between expected and reported VAF is 0.99 with a  $p$ -value of  $2.2 \times 10^{-16}$ . Importantly, no mutations were detected in the 15 positions reported as "no-mutated" (true negatives), showing the assay's high specificity.

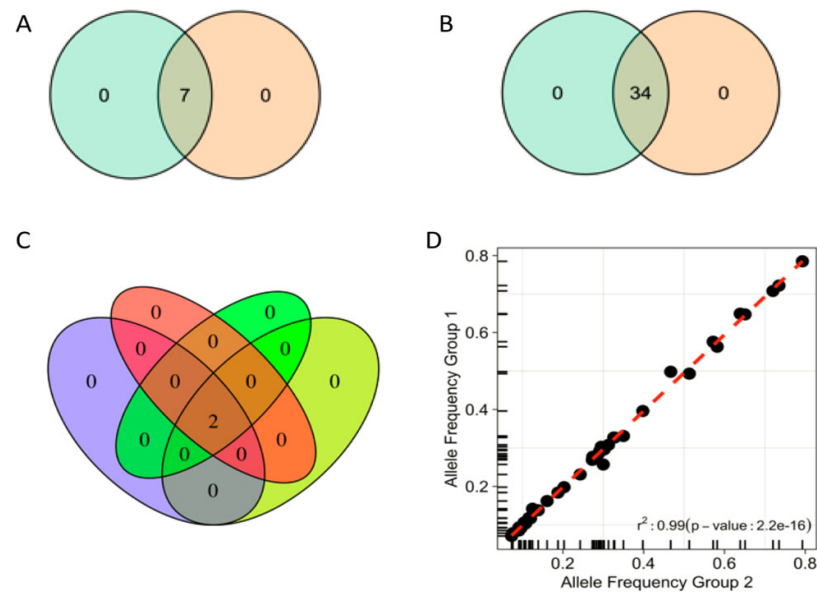


**Figure 1.** Performance and analytical sensitivity of the panel. Correlation between expected allelic frequencies for reported variants in commercial standard control samples and those observed by the assay: (A) FFPE Horizon Discovery sample (HD 200); (B) germline DNA samples with *BRCA1/2* mutations (HD 793 and HD 794).

For reproducibility assessment, three FFPE samples from different tumors (colorectal, ovary, and gallbladder) were used to prepare two separate libraries each. All samples passed the sequencing metrics threshold with  $\geq 300\times$  coverage in 99.6% of target regions and 89% uniformity. A 100% concordance among non-synonymous variants detected in the different libraries was observed (Figure 2A).

Inter-runs repeatability was assessed using four FFPE samples (three ovaries and one HD control). Different libraries were sequenced in different runs. Reproducibility of sequencing metrics (94% of target regions with  $\geq 300\times$  coverage and  $\geq 87\%$  uniformity) and concordance of detected variants were also observed (Figure 2B). One ovarian FFPE sample was assessed in three different library preparations and three separate sequencing runs (Figure 2C). A high correlation ( $r = 0.99$ ) was observed between VAFs called in all the different settings (Figure 2D).



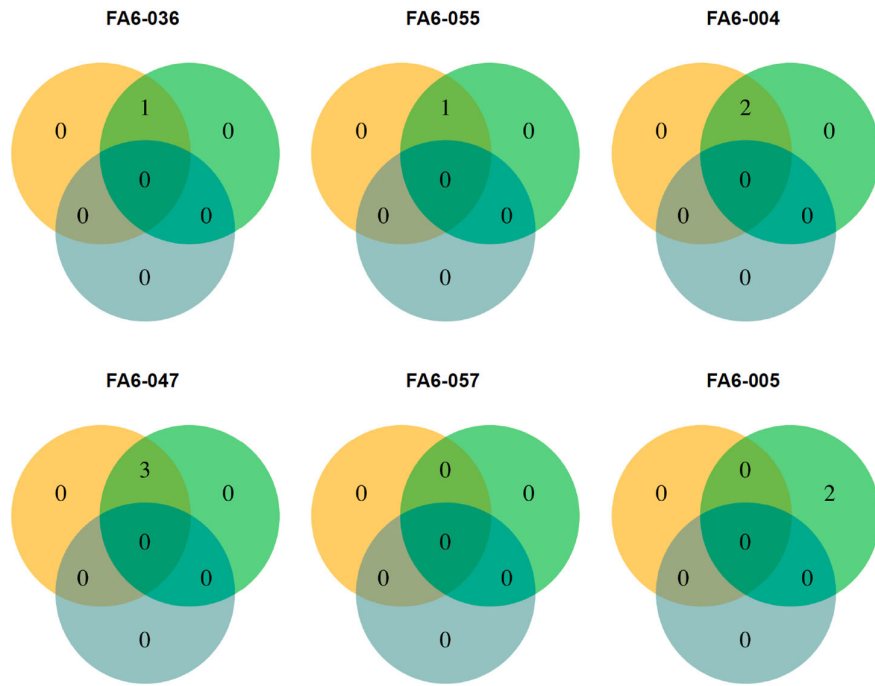


**Figure 2.** Reproducibility and repeatability of the assay: (A) within sequencing run reproducibility based on three patient samples, each from a different tumor type (colon, ovaries, gallbladder). Venn diagram displays variants observed in two different library preparations for each of the three samples. (B) Repeatability between sequencing runs was assessed using the same libraries from three different FFPE cancer and control (HD200) samples in two separate sequencing runs. Venn diagram displays variants observed between two sequencing runs. (C) One ovarian cancer sample was processed using four different combinations (colors) of library preparation and sequencing runs. (D) Correlation between allele frequencies of variants obtained in repeatability and reproducibility tests (displayed in (A,B)).

### 3.3. Comparison between FFPE, Fresh Frozen, and Blood gDNA

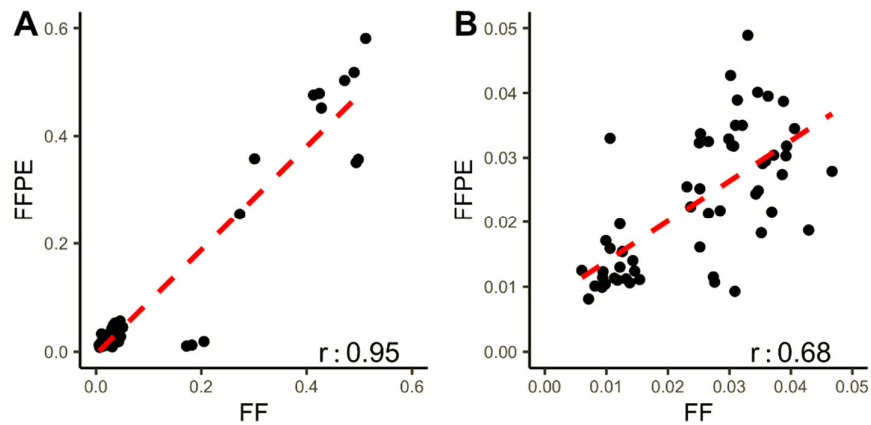
To assess whether the protocol and bioinformatic workflow for detecting somatic mutations discard FFPE-induced artifacts and germline variants, DNA from FFPE, fresh frozen (FF) tumor, and buffy coat samples from six ductal breast carcinoma subjects was sequenced.

Variants detected among all sample triads for each of the six subjects are outlined in Figure 3. Nine variants in the FF sample set were reported. Seven of these variants were also reported in the matching FFPE samples. It is worth noting that the two variants detected in an FF sample (FA6-005) were found in the FFPE sample but at frequencies <5% (the LOD established for the assay).



**Figure 3.** Mutations found in FFPE and fresh frozen tumor tissue, and blood from same individuals. Venn diagram outlining non-synonymous variants reported for six different patients diagnosed with ductal breast carcinoma among three different sample types for each: FFPE tumor (orange), fresh frozen tumor (green), and buffy coat (light blue).

To further explore FF and FFPE samples’ concordance, the allele frequencies of both synonymous and non-synonymous variants detected in both sample types were plotted (Figure 4). Variants with AF < 5% display a low r-value (0.68,  $p$ -value of  $2.479 \times 10^{-9}$ ). However, when all variants (73) are analyzed, correlation increases ( $r = 0.95$ ,  $p$ -value of  $< 2.2 \times 10^{-16}$ ). Importantly, no germline variants and no variants exclusive for FFPE samples were detected using the pipeline for somatic mutations.



**Figure 4.** Correlation between allele frequencies in synonymous and non-synonymous variants found in paired FFPE and fresh frozen samples: (A) correlation between variant allele frequencies (VAF) in all variants found in FFPE and fresh frozen samples (FF). (B) Correlation between frequencies of variants found at VAF below LOD (0.05).

### 3.4. Validation of the Assay in Clinical Samples

To validate the assay and analysis capabilities in “real world” samples, 183 tumor biopsies from different clinical sites were processed. In total, 131 out of the 183 were successfully sequenced (116 FFPE and 15 FF): breast (14), ovary (69), gastric (1), gallbladder (31), and colorectal (16).

All variants with allelic frequency > 0.01 reported in at least one of the following four germline population variant’s databases (PVDs): GnomAD, ESP6500, ExAC, and 1000 Genomes, were eliminated. However, given that the Latin American population is not well represented in these repositories, somatic mutations were initially overestimated due to their absence or low (<1%) VAF. Thus, the classification of remaining variants was performed using the algorithm depicted in Supplementary Figure S2. This algorithm was designed based on recommendations from Sukhai et al. (2019) [30] and using annotations in COSMIC, dbSNP, CLINVAR, and PVDs databases. Filtering was extended to include all under-represented populations that have information in GnomAD and ExAC. Additionally, a local genetic germline database from Chilean individuals was used. The resulting variants were classified as germline, somatic, putative germline, putative somatic, putative germline novel, or putative somatic novel.

A total of 256 protein-affecting variants were found in the 131 samples. Among these, 197 non-synonymous somatic and putative known and novel somatic variants were identified in 111 out of the 131 samples (85%) (Table 2); moreover, 144 were unique variants. Figure 5 shows a breakdown of somatic mutations by gene and tumor type. Non-synonymous variants are shown according to mutation type (missense and nonsense mutations, frameshift causing deletions and insertions, and variants positioned in splice sites). Overall, TP53, BRCA2, PIK3CA, ARID1A, KRAS, TSC2, PTEN, and BRCA1 were the most frequently mutated genes. TP53 and PIK3CA were the most prevalent mutations in ovary cancer; TP53, BRCA2, and KRAS in colorectal cancer; and TP53 and PIK3CA in breast cancer. In 31 samples of GBC, we found 36 mutations in TP53 (16), KRAS (4), ARID1A (3), ERBB2 (2), TSC2 (2), PIK3CA (2), PTCH1 (2), TSC1 (2), and BRCA1, PTEN, and BRAF (1 each).

**Table 2.** Classification of non-synonymous variants found in 131 quality-passed tumor samples. The number of total and unique variants by classification is shown.

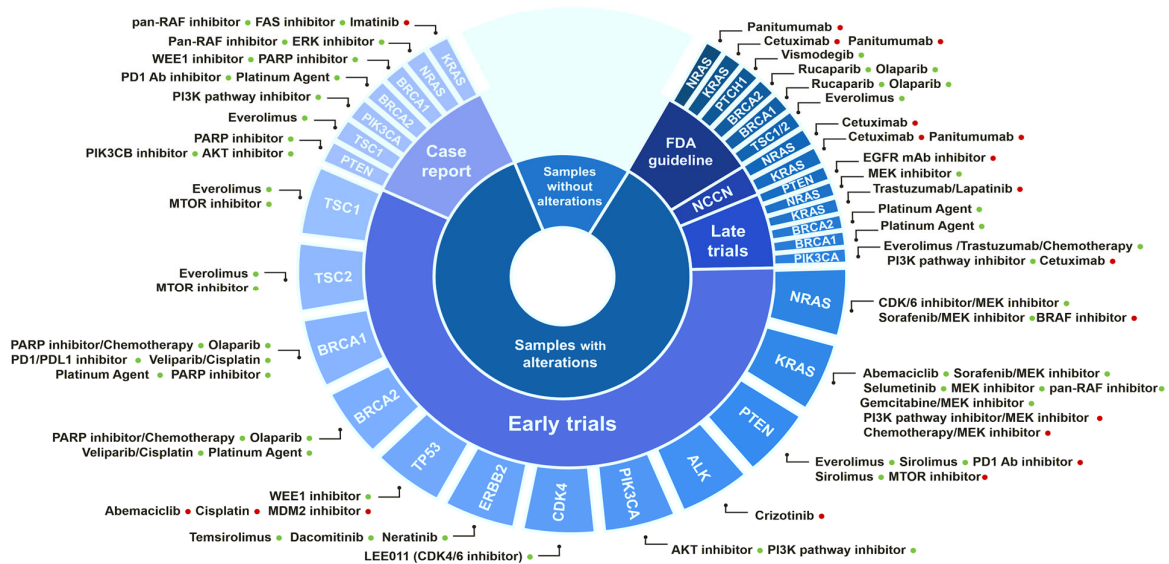
Classification of Variants	Total Variants	Unique Variants
Germline	55	26
Putative Novel Germline	4	3
Somatic	125	86
Putative Somatic	13	13
Putative Novel Somatic	59	45
<b>Total</b>	<b>256</b>	<b>173</b>



**Figure 5.** Non-synonymous somatic variants detected across all tumor samples; 197 non-synonymous somatic variants were detected in 111 out of 131 gallbladder, colorectal, breast, and ovary cancer samples.

### 3.5. Identification of Biomarkers for Targeted Therapies

In total, 137 (69.5%) out of the 197 identified somatic variants are described as a biomarker for drug response, which is supported by different levels of evidence: FDA guidelines (44) and clinical guidelines (9), late trials (37), early trials (119), case report (42), and (113) pre-clinical data. The affected gene and target drug associations with supporting evidence from “case reports” to “FDA guidelines” are depicted in Figure 6, which outlines: (1) the fraction of samples with reported genetic alterations; (2) the level of existing evidence for the biomarker; (3) the gene affected; and (4) the drug association (resistant or responsive). Table 3 contains a detailed description of the biomarker mutations supported by FDA and NCCN clinical guidelines found in this study in all samples.



**Figure 6.** Predictive biomarkers for therapy response identified in the analyzed clinical tumor samples. The biomarker gene, the associated drugs, and the level of evidence supporting this association in solid tumors are depicted. Positive

(drug-responsive) and negative (drug-resistant) associations are indicated by green and red dots, respectively. Information on these interactions was obtained from “The Cancer Genome Interpreter” [26].

**Table 3.** Biomarker mutations supported by FDA (*n* = 44) and NCCN guidelines (*n* = 9), found in the analyzed clinical samples. The associated drug and the mutation’s predictive effect are indicated.

Gene	Mutation	Drug	Effect
BRCA1	E1609 *, L702Wfs*5, N1745Tfs*20, Q1273 *, V370I		
BRCA2	A2603S, D1796Mfs*9, K3327Nfs*13, L1114V, splice_acceptor_variant, T2783Afs*13, T2790I, I1364M, L398P, D635G, R2034C	Rucaparib (PARP inhibitor) Olaparib (PARP inhibitor)	Responsive
KRAS	A146V, Q61H G12A, G12D, G12V, L19F, Q25*fs*1	Panitumumab (EGFR mAb inhibitor) Cetuximab (EGFR mAb inhibitor)	Resistant
NRAS	G12C, Q61R	Panitumumab (EGFR mAb inhibitor) Cetuximab (EGFR mAb inhibitor)	Resistant
PIK3CA	H1047R, E545A, E545K, E542K, R88Q, N345S, E579K	Alpelisib + Fulvestrant	Responsive
PTCH1	R441H, D717N, H1240R, P725S, V580A, T677A, N871D	Vismodegib (SHH inhibitor)	Responsive
TSC1	K375Sfs*30, L826Q, L827Q, T582S	Everolimus (MTOR inhibitor)	Responsive
TSC2	R1729C, S1530L, K533delK, A460T, A950T, D1084G, P1771L, S1096C, T154I		

\*: It is a mutation nomenclature.

In GBC samples, we found biomarker mutations in eight genes, with supporting evidence ranging from case reports to FDA and NCCN guidelines in different tumor types (Table 4). It is worth highlighting the presence of predictive biomarkers for drugs that are currently in use for treating different cancers, such as PARP, ERBB2, EGFR, PIK3CA, mTOR, and Hedgehog signaling inhibitors.

**Table 4.** Biomarker mutations found in GBC samples. The number of samples with the mutation, associated drug, supporting level of evidence, and tumor where the evidence was generated are indicated.

# of Samples	Gene	Mutation	Drugs	Evidence	Tumor Tested
1	BRCA1	V370I	Rucaparib (PARP inhibitor)	FDA guidelines	OV
			Olaparib (PARP inhibitor)		BRCA
			WEE1 inhibitor	Case report Early trials	BRCA
			Platinum agent (chemotherapy) Veliparib; Cisplatin (PARP inhibitor; chemotherapy)		OV OV
2	ERBB2	L755S	Dacomitinib (Pan ERBB inhibitor)	Early trials	NSCLC
			Neratinib (ERBB2 inhibitor)		CANCER, LUAD
			Temsirolimus (MTOR inhibitor)		
4	KRAS	G12A G12V Q61H G12D	Panitumumab (EGFR mAb inhibitor)	FDA guidelines	COREAD
			Cetuximab (EGFR mAb inhibitor)	FDA guidelines	LUAD
			Trastuzumab; Lapatinib (ERBB2 mAb inhibitor; ERBB2 inhibitor)	Late trials	PA
				Early trials	NSCLC, HC, BT, L
				Early trials	L

			Gemcitabine; MEK inhibitor (chemotherapy; MEK inhibitor)	Early trials	PA
			MEK inhibitor	Early trials	L
			Selumetinib (MEK inhibitor)	Case report	L
			PI3K pathway inhibitor; MEK inhibitor		GIST
			Abemaciclib (CDK4/6 inhibitor)		
			Imatinib (BCR-ABL inhibitor and KIT inhibitor)		
			PI3K pathway inhibitor		
1	PIK3CA	E545K	Everolimus; Trastuzumab; chemotherapy (MTOR inhibitor; ERBB2 mAb inhibitor; chemotherapy)	FDA guidelines Late trials	BRCA BRCA
			Cetuximab (EGFR mAb inhibitor)	Late trials	COREAD
			AKT inhibitor	Early trials	BRCA
			PI3K pathway inhibitor	Early trials	ED, OV, CESC
			PI3K pathway inhibitor	Case report	BLCA, HNSC, L
1	PTCH1	P725S	Vismodegib (SHH inhibitor)	FDA guidelines	BCC, MB
		E171 *			
		G244S			
		G266V			
		K321Ifs * 10			
		L257P			
		R280T			
		V173Gfs * 10	MDM2 inhibitor	Early trials	LIP
11	TP53	R213 *	Abemaciclib (CDK4/6 inhibitor)	Early trials	BRCA
		R248W	Cisplatin (chemotherapy)	Early trials	FGCT, MGCT
		R273C	WEE1 inhibitor	Early trials	OV
		W53 *			
		R273H			
		R248Q			
		Q192 *			
		C238F			
1	TSC1	L826Q	Everolimus (MTOR inhibitor)	FDA guidelines Early trials Case report	GCA, RA BLCA ST, S, R
2	TSC2	D1084G S1096C	Everolimus (MTOR inhibitor)	FDA guidelines	GCA, RA
1	ARID1A #	Splice acceptor variant	(EZH2 inhibitor)	Pre-clinical	OV
			(PD1 inhibitor)	Pre-clinical	OV
			(PARP inhibitor)	Pre-clinical	CANCER
			(ATR inhibitor)	Pre-clinical	CANCER

# For ARID1A, pre-clinical evidence is shown, \*: It is a mutation nomenclature.

#### 4. Discussion

As a result of its global adoption and implementation, the clinical utility of NGS in the field of oncology has a rapidly growing body of evidence. The ability to obtain massive amounts of genetic information from small amounts of tissue provides clear advantages for decision-making processes against cancer [2,4,31]. Nevertheless, a large portion of cancer patients around the world do not have this option readily available. This work at-

tempts to favor the implementation of NGS in the Latin American health system by showcasing a locally developed assay, accompanied by an open-source automated analysis focused on the target population's needs.

A critical consideration for implementing NGS in low resource settings is finding a workflow compatible with low-quality, highly degraded samples. Although fresh frozen tissue is the gold standard for molecular analyses, its use in clinical practice is impractical because of its high cost and technical difficulties related to its obtainment, processing, and storing. The sample storage infrastructure found in the developed world, with dedicated  $-80\text{ }^{\circ}\text{C}$  and  $-20\text{ }^{\circ}\text{C}$  freezers, is too often not in the budget for many Latin American diagnostic laboratories. FFPE tissue samples are much more cost effective, as they can be stored at room temperature. However, tumor biopsies in this region are often fixed with formalin with different protocols and laboratory environments, producing varied DNA damage during and after the formalin fixation process (e.g., fragmentation, degradation, crosslinking [32,33]).

DNA quality is affected by the type of formalin used for tissue fixation and the time since preservation [34], both of which vary highly in laboratories across Chile. A total of 48 FFPE samples failed to pass the DNA, library, or sequencing quality controls. Most of the failed samples were colorectal (33/40) and gallbladder (12/43) samples. Twenty-four failed colorectal samples had low on-target rates, suggesting issues with hybridization and/or capture, while the 12 gallbladder samples did not pass the library preparation quality metrics.

In general, FFPE samples showed a higher percentage of duplicates and off-target reads (Table S1). However, these characteristics do not affect sequencing results. FFPE samples have the highest mean on-target region's coverage compared to FF and BC. Removing duplicates is intended to reduce noise during the variant identification process and minimize false positives. These results suggest removing duplicates has little effect on this panel's performance at the sequencing depths of interest ( $\sim 300\times$ ). As sequencing technologies continue to advance, PCR duplicate removal will become less of an issue [35].

Somatic mutation analysis and fusion detection are critical in cancer research [1,15]. In this project, we focused on creating a cost-effective regional tool with clinical relevance. As a result of our approach, a tailored DNA based panel, we lack the ability to detect relevant RNA gene alterations such as fusions. Although side-by-side DNA and RNA analysis would provide a more complete understanding of gene alteration profiles, it can be costly and time consuming. Future solutions such as total nucleic acid library preparation followed by target enrichment could provide a more comprehensive analysis in one workflow. In the meantime, this panel detects point mutations in ALK, MET, and ROS1, which have been associated with resistance to TKI.

Currently, there is no community consensus about the most appropriate variant caller for somatic mutations [36]. For this reason, the bioinformatic pipeline for variant calling used in this work incorporated six variant callers capable of producing highly accurate somatic mutation calls for both SNVs and small indels. Somatic variant callers discard germline variants by interrogating the reference genomes and population databases, such as 1000 Genomes, where Latin American genetic variation is not well represented. Thus, we implemented a more accurate bioinformatic pipeline that allowed variants' classification as somatic, germline, and putative somatic/germline. This variant calling process highlights the extra layer of difficulty Latin American researchers and clinical laboratories face due to the absence of reference genomes representative of our population in the major databases. The overestimation of somatic variants is a problem when facing the tumor of a patient from any region or ancestry without a reference genome informative of the genetic variation in that specific population. This issue is critical for therapy determinants, such as tumor mutational load, which should be carefully interpreted in these patients [37].

The addition of reference genomes representative of the population in combination with the genetic characterization of more cancers will help address questions involving the distribution of genetic alterations among different world populations.

The best approach for resolving the somatic vs. germline mutations issue is to include respective blood samples alongside biopsies. However, this raises the assay's cost per patient, which may delay the assay's implementation.

To achieve more accurate somatic variant calling, further efforts towards genetic characterization of the Latin American and other under-studied populations are needed. Building an inclusive tumor reference genome database will allow for the discovery of novel somatic mutations and non-explored correlations to the disease.

The assay was designed to meet the biomarker needs in countries with low participation in clinical trials and in which a limited number of drugs are available or currently in use in clinical practice. Among the 131 predictive biomarkers for therapies response detected by the assay, 52 are supported by evidence recognized by the FDA and NCCN clinical guidelines.

Although a small number of GBC samples was successfully sequenced, biomarker mutations in nine genes were identified. Importantly, these are predictive biomarkers for drugs that are currently in use for treating different cancers, such as PARP, ERBB2, EGFR, PIK3CA, mTOR, and Hedgehog signaling inhibitors. Since most of these drugs are available in Chile and LATAM, finding predictive biomarkers in GBC generates opportunities for specific and basket clinical trials including GBC patients from Chile and other regions in LATAM, where this disease has an unusually high prevalence.

**Supplementary Materials:** The following are available online at [www.mdpi.com/article/10.3390/jpm11090899/s1](http://www.mdpi.com/article/10.3390/jpm11090899/s1), Table S1: Sequencing metrics (median) for all samples that passed QC filters. Figure S1: General workflow of the bioinformatic pipeline for identifying somatic variants. Figure S2: Sequencing metrics and mean coverage of targeted genes according to sample type. Figure S3: Algorithm for variant's classification.

**Author Contributions:** Conceptualization, O.B., M.A., R.A.V. and K.M.; Formal analysis, E.G.-F., I.M., N.M.-G., C.V. and R.A.V.; Funding acquisition, M.S., C.B., J.L.B., R.A.V. and K.M.; Investigation, M.S., E.G.-F., J.T., I.M., N.M.-G., O.B., E.B., M.A., A.C., R.A., C.I., M.L.B., G.d.T., E.M., C.B., A.M.C. and K.M.; Methodology, M.S., E.G.-F., J.T., I.G., I.M., N.M.-G., C.V., P.G., R.A.V. and K.M.; Resources, I.G., E.B., M.A., A.C., C.I., M.L.B., V.S., M.L.S., G.d.T., E.M., C.B., P.G., A.M.C., L.G. and J.L.B.; Supervision, I.G., E.B., A.M.C., V.S., M.L.S., L.G., R.A.V. and K.M.; Writing—original draft, M.S., E.G.-F., J.T., R.A., R.A.V. and K.M.; Writing—review and editing, O.B., C.I., G.d.T. and J.L.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by ANID FONDEF IDEA, grant number IT16I10051, the European Union's Horizon 2020 research and innovation program (grant 825741); and ANID FONDECYT grant number 1171463. The HPC computational resources used in this work were funded by grants VID INFRAESTRUCTURA 0440/2018, VID U-Redes 704/2016 (University of Chile), FONDEF D10E1007 and D11I1029 (CONICYT), and FONDEQUIP EQM140157.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the "Comité de Ética Científico del Servicio Metropolitano de Salud Oriente", "Comité Ético Científico o de Investigación del Hospital Clínico de la Universidad de Chile" (approval number N°17-18), and the "Comité de Ética de Investigación en Seres Humanos de la Facultad de Medicina de la Universidad de Chile".

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The bioinformatic pipeline and tutorial can be found in the GitHub repository called Pipeline-TumorSec (<https://github.com/u-genoma/Pipeline-TumorSec>).

**Acknowledgments:** We thank Daniela Diez and Vania Montecinos for clinical coordination, and the personnel from "Biobanco de Tejidos y Fluidos de la Universidad de Chile" for their permanent support. We also thank all patients who altruistically donated their samples for this research. Probes and other consumables were partially provided by Roche Sequencing Solutions (Pleasanton, CA).



**Conflicts of Interest:** Mauricio Salvo is a Roche Sequencing Solutions employee. All other authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Shyr, D.; Liu, Q. Next Generation Sequencing in Cancer Research and Clinical Application. *Biol. Proced. Online* **2013**, *15*, 4, doi:10.1186/1480-9222-15-4.
- Wakai, T.; Prasoon, P.; Hirose, Y.; Shimada, Y.; Ichikawa, H.; Nagahashi, M. Next-Generation Sequencing-Based Clinical Sequencing: Toward Precision Medicine in Solid Tumors. *Int. J. Clin. Oncol.* **2019**, *24*, 115–122, doi:10.1007/s10147-018-1375-3.
- Martin-Liberal, J.; Hierro, C.; Ochoa de Olza, M.; Rodon, J. Immuno-Oncology: The Third Paradigm in Early Drug Development. *Target Oncol.* **2017**, *12*, 125–138, doi:10.1007/s11523-016-0471-4.
- Garralda, E.; Dienstmann, R.; Piris-Giménez, A.; Braña, I.; Rodon, J.; Tabernero, J. New Clinical Trial Designs in the Era of Precision Medicine. *Mol. Oncol.* **2019**, *13*, 549–557, doi:10.1002/1878-0261.12465.
- Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424, doi:10.3322/caac.21492.
- Carioli, G.; Bertuccio, P.; Malvezzi, M.; Rodriguez, T.; Levi, F.; Boffetta, P.; La Vecchia, C.; Negri, E. Cancer Mortality Predictions for 2019 in Latin America. *Int. J. Cancer* **2020**, *147*, 619–632, doi:10.1002/ijc.32749.
- Sierra, M.S.; Soerjomataram, I.; Antoni, S.; Laversanne, M.; Piñeros, M.; de Vries, E.; Forman, D. Cancer Patterns and Trends in Central and South America. *Cancer Epidemiol.* **2016**, *44* (Suppl. 1), S23–S42, doi:10.1016/j.canep.2016.07.013.
- Torres, Á.; Oliver, J.; Frecha, C.; Montealegre, A.L.; Quezada-Urbán, R.; Díaz-Velásquez, C.E.; Vaca-Paniagua, F.; Perdomo, S. Cancer Genomic Resources and Present Needs in the Latin American Region. *Public Health Genom.* **2017**, *20*, 194–201, doi:10.1159/000479291.
- Harismendy, O.; Kim, J.; Xu, X.; Ohno-Machado, L. Evaluating and Sharing Global Genetic Ancestry in Biomedical Datasets. *J. Am. Med. Inform. Assoc.* **2019**, *26*, 457–461, doi:10.1093/jamia/ocy194.
- Spratt, D.E.; Chan, T.; Waldron, L.; Speers, C.; Feng, F.Y.; Ogunwobi, O.O.; Osborne, J.R. Racial/Ethnic Disparities in Genomic Sequencing. *JAMA Oncol.* **2016**, *2*, 1070–1074, doi:10.1001/jamaoncol.2016.1854.
- Yuan, J.; Hu, Z.; Mahal, B.A.; Zhao, S.D.; Kensler, K.H.; Pi, J.; Hu, X.; Zhang, Y.; Wang, Y.; Jiang, J.; et al. Integrated Analysis of Genetic Ancestry and Genomic Alterations across Cancers. *Cancer Cell* **2018**, *34*, 549–560.e9, doi:10.1016/j.ccell.2018.08.019.
- Popejoy, A.B.; Fullerton, S.M. Genomics Is Failing on Diversity. *Nature* **2016**, *538*, 161–164, doi:10.1038/538161a.
- Bustamante, C.D.; Burchard, E.G.; De la Vega, F.M. Genomics for the World. *Nature* **2011**, *475*, 163–165, doi:10.1038/475163a.
- Sirugo, G.; Williams, S.M.; Tishkoff, S.A. The Missing Diversity in Human Genetic Studies. *Cell* **2019**, *177*, 26–31, doi:10.1016/j.cell.2019.02.048.
- Chakravarty, D.; Gao, J.; Phillips, S.M.; Kundra, R.; Zhang, H.; Wang, J.; Rudolph, J.E.; Yaeger, R.; Soumerai, T.; Nissan, M.H.; et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* **2017**, *2017*, 1–16, doi:10.1200/po.17.00011.
- Griffith, M.; Spies, N.C.; Krysiak, K.; McMichael, J.F.; Coffman, A.C.; Danos, A.M.; Ainscough, B.J.; Ramirez, C.A.; Rieke, D.T.; Kujan, L.; et al. CIViC Is a Community Knowledgebase for Expert Crowdsourcing the Clinical Interpretation of Variants in Cancer. *Nat. Genet.* **2017**, *49*, 170–174, doi:10.1038/ng.3774.
- Wagner, A.H.; Walsh, B.; Mayfield, G.; Tamborero, D.; Sonkin, D.; Krysiak, K.; Deu-Pons, J.; Duren, R.P.; Gao, J.; McMurry, J.; et al. A Harmonized Meta-Knowledgebase of Clinical Interpretations of Somatic Genomic Variants in Cancer. *Nat. Genet.* **2020**, *52*, 448–457, doi:10.1038/s41588-020-0603-8.
- Li, M.M.; Datto, M.; Duncavage, E.J.; Kulkarni, S.; Lindeman, N.I.; Roy, S.; Tsimberidou, A.M.; Vnencak-Jones, C.L.; Wolff, D.J.; Younes, A.; et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J. Mol. Diagn.* **2017**, *19*, 4–23, doi:10.1016/j.jmoldx.2016.10.002.
- McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303, doi:10.1101/gr.107524.110.
- Fang, L.T.; Afshar, P.T.; Chhibber, A.; Mohiyuddin, M.; Fan, Y.; Mu, J.C.; Gibeling, G.; Barr, S.; Asadi, N.B.; Gerstein, M.B.; et al. An Ensemble Approach to Accurately Detect Somatic Mutations Using SomaticSeq. *Genome Biol.* **2015**, *16*, 197, doi:10.1186/s13059-015-0758-2.
- Benjamin, D.; Sato, T.; Cibulskis, K.; Getz, G.; Stewart, C.; Lichtenstein, L. Calling Somatic SNVs and Indels with Mutect2. *BioRxiv Bioinform.* **2019**, doi:10.1101/861054.
- Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing. *Genome Res.* **2012**, *22*, 568–576, doi:10.1101/gr.129684.111.

23. Lai, Z.; Markovets, A.; Ahdesmaki, M.; Chapman, B.; Hofmann, O.; McEwen, R.; Johnson, J.; Dougherty, B.; Barrett, J.C.; Dry, J.R. VarDict: A Novel and Versatile Variant Caller for Next-Generation Sequencing in Cancer Research. *Nucleic Acids Res.* **2016**, *44*, e108, doi:10.1093/nar/gkw227.
24. Wilm, A.; Aw, P.P.K.; Bertrand, D.; Yeo, G.H.T.; Ong, S.H.; Wong, C.H.; Khor, C.C.; Petric, R.; Hibberd, M.L.; Nagarajan, N. LoFreq: A Sequence-Quality Aware, Ultra-Sensitive Variant Caller for Uncovering Cell-Population Heterogeneity from High-Throughput Sequencing Datasets. *Nucleic Acids Res.* **2012**, *40*, 11189–11201, doi:10.1093/nar/gks918.
25. Saunders, C.T.; Wong, W.S.W.; Swamy, S.; Becq, J.; Murray, L.J.; Cheetham, R.K. Strelka: Accurate Somatic Small-Variant Calling from Sequenced Tumor-Normal Sample Pairs. *Bioinformatics* **2012**, *28*, 1811–1817, doi:10.1093/bioinformatics/bts271.
26. Tamborero, D.; Rubio-Perez, C.; Deu-Pons, J.; Schroeder, M.P.; Vivancos, A.; Rovira, A.; Tusquets, I.; Albanell, J.; Rodon, J.; Taberero, J.; et al. Cancer Genome Interpreter Annotates the Biological and Clinical Relevance of Tumor Alterations. *Genome Med.* **2018**, *10*, 25, doi:10.1186/s13073-018-0531-8.
27. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* **2010**, *38*, e164, doi:10.1093/nar/gkq603.
28. Jennings, L.J.; Arcila, M.E.; Corless, C.; Kamel-Reid, S.; Lubin, I.M.; Pfeifer, J.; Temple-Smolkin, R.L.; Voelkerding, K.V.; Nikiforova, M.N. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J. Mol. Diagn.* **2017**, *19*, 341–365, doi:10.1016/j.jmoldx.2017.01.011.
29. GATK. (How to) Filter Variants Either with VQSR or by Hard-Filtering. Available online: <https://gatk.broadinstitute.org/hc/en-us/articles/360035531112--How-to-Filter-variants-either-with-VQSR-or-by-hard-filtering> (accessed on 6 July 2021).
30. Sukhai, M.A.; Misyura, M.; Thomas, M.; Garg, S.; Zhang, T.; Stickle, N.; Virtanen, C.; Bedard, P.L.; Siu, L.L.; Smets, T.; et al. Somatic Tumor Variant Filtration Strategies to Optimize Tumor-Only Molecular Profiling Using Targeted Next-Generation Sequencing Panels. *J. Mol. Diagn.* **2019**, *21*, 261–273, doi:10.1016/j.jmoldx.2018.09.008.
31. Tourneau, C.L.; Delord, J.-P.; Gonçalves, A.; Gavaille, C.; Dubot, C.; Isambert, N.; Campone, M.; Trédan, O.; Massiani, M.-A.; Mauborgne, C.; et al. Molecularly Targeted Therapy Based on Tumour Molecular Profiling versus Conventional Therapy for Advanced Cancer (SHIVA): A Multicentre, Open-Label, Proof-of-Concept, Randomised, Controlled Phase 2 Trial. *Lancet Oncol.* **2015**, *16*, 1324–1334, doi:10.1016/S1470-2045(15)00188-6.
32. Prentice, L.M.; Miller, R.R.; Knaggs, J.; Mazloomian, A.; Hernandez, R.A.; Franchini, P.; Parsa, K.; Tessier-Cloutier, B.; Lapuk, A.; Huntsman, D.; et al. Formalin Fixation Increases Deamination Mutation Signature but Should Not Lead to False Positive Mutations in Clinical Practice. *PLoS ONE* **2018**, *13*, e0196434, doi:10.1371/journal.pone.0196434.
33. Guyard, A.; Boyez, A.; Pujals, A.; Robe, C.; Tran Van Nhieu, J.; Allory, Y.; Moroch, J.; Georges, O.; Fournet, J.-C.; Zafrani, E.-S.; et al. DNA Degradation during Storage in Formalin-Fixed and Paraffin-Embedded Tissue Blocks. *Virchows Arch.* **2017**, *471*, 491–500, doi:10.1007/s00428-017-2213-0.
34. Nagahashi, M.; Shimada, Y.; Ichikawa, H.; Nakagawa, S.; Sato, N.; Kaneko, K.; Homma, K.; Kawasaki, T.; Kodama, K.; Lyle, S.; et al. Formalin-Fixed Paraffin-Embedded Sample Conditions for Deep Next Generation Sequencing. *J. Surg. Res.* **2017**, *220*, 125–132, doi:10.1016/j.jss.2017.06.077.
35. Ebbert, M.T.W.; Wadsworth, M.E.; Staley, L.A.; Hoyt, K.L.; Pickett, B.; Miller, J.; Duce, J. Alzheimer’s Disease Neuroimaging Initiative; Kauwe, J.S.K.; Ridge, P.G. Evaluating the Necessity of PCR Duplicate Removal from Next-Generation Sequencing Data and a Comparison of Approaches. *BMC Bioinform.* **2016**, *17* (Suppl. 7), 239, doi:10.1186/s12859-016-1097-3.
36. Barnell, E.K.; Ronning, P.; Campbell, K.M.; Krysiak, K.; Ainscough, B.J.; Sheta, L.M.; Pema, S.P.; Schmidt, A.D.; Richters, M.; Cotto, K.C.; et al. Standard Operating Procedure for Somatic Variant Refinement of Sequencing Data with Paired Tumor and Normal Samples. *Genet. Med.* **2019**, *21*, 972–981, doi:10.1038/s41436-018-0278-z.
37. Parikh, K.; Huether, R.; White, K.; Hoskinson, D.; Beaubier, N.; Dong, H.; Adjei, A.A.; Mansfield, A.S. Tumor Mutational Burden From Tumor-Only Sequencing Compared With Germline Subtraction From Paired Tumor and Normal Specimens. *JAMA Netw. Open* **2020**, *3*, e200202, doi:10.1001/jamanetworkopen.2020.0202