



Review

# What Does This Mutation Mean? The Tools and Pitfalls of Variant Interpretation in Lymphoid Malignancies

Yann Guillermin <sup>1</sup>, Jonathan Lopez <sup>2,3,4</sup>, Kaddour Chabane <sup>5</sup> , Sandrine Hayette <sup>3,5</sup>,  
Claire Bardel <sup>6,7</sup>, Gilles Salles <sup>3,4,8</sup>, Pierre Sujobert <sup>3,4,5</sup> and Sarah Huet <sup>3,5,9,\*</sup>

<sup>1</sup> Centre Léon Bérard, Service d'Hématologie Clinique, 69008 Lyon, France; yannguillermin@orange.fr

<sup>2</sup> Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, Laboratoire de Biochimie et Biologie moléculaire, 69495 Pierre-Bénite CEDEX, France; jonathan.lopez@chu-lyon.fr

<sup>3</sup> Cancer Research Center of Lyon, INSERM U1052 UMR CNRS 5286, Equipe labellisée LIGUE Contre le Cancer, 69008 Lyon, France; sandrine.hayette@chu-lyon.fr (H.S.); gilles.salles@chu-lyon.fr (S.G.); pierre.sujobert@chu-lyon.fr (S.P.)

<sup>4</sup> Université de Lyon, Université Lyon 1, Faculté de Médecine et de Maïeutique Lyon Sud Charles Mérieux, 69921 Oullins CEDEX, France

<sup>5</sup> Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, Laboratoire d'Hématologie, 69495 Pierre-Bénite CEDEX, France; kaddour.chabane@chu-lyon.fr

<sup>6</sup> Hospices Civils de Lyon, Service de Biostatistique—bioinformatique et plateforme de séquençage haut débit NGS-CHU Lyon, 69677 Bron CEDEX, France; claire.bardel@chu-lyon.fr

<sup>7</sup> Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, équipe biostatistique-santé, F-69424 Lyon, France

<sup>8</sup> Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, Service d'Hématologie Clinique, 69495 Pierre-Bénite CEDEX, France

<sup>9</sup> Université de Lyon, Université Lyon 1, Faculté de Pharmacie Rockefeller, 69373 Lyon CEDEX, France

\* Correspondence: sarah.huet@chu-lyon.fr; Tel.: +33-478-863-639-57

Received: 15 March 2018; Accepted: 14 April 2018; Published: 20 April 2018



**Abstract:** High throughput sequencing (HTS) is increasingly important in determining cancer diagnoses, with subsequent prognostic and therapeutic implications. The biology of cancer is becoming increasingly deciphered and it is clear that therapy needs to be individually tailored. Whilst translational research plays an important role in lymphoid malignancies, few guidelines exist to guide biologists and routine laboratories through this constantly evolving field. In this article, we review the challenges of interpreting HTS in lymphoid malignancies and provide a toolkit to interpret single nucleotide variants obtained from HTS. We define the pre-analytical issues such as sequencing DNA obtained from formalin-fixed and paraffin-embedded tissue (FFPE), the acquisition of germline DNA, or the bioinformatic pitfalls, the analytical issues encountered and how to manage them. We describe the main constitutional and cancer databases, their characteristics and limitations, with an emphasis on variant interpretation in lymphoid malignancies. Finally, we discuss the challenges of predictions that one can make using in silico or in vitro modelling, pharmacogenomic screening, and the limits of those prediction tools. This description of the current status in genomic interpretation highlights the need for new large databases and international collaboration in the lymphoma field.

**Keywords:** next-generation sequencing; lymphoid malignancies; variant interpretation

## 1. Introduction

Hematological malignancies, and in particular lymphoproliferative neoplasms, are characterized by considerable heterogeneity. To ensure accurate ontological classification, one needs to integrate phenotypic information from morphology and immunophenotyping as well as genetic aspects such as cytogenetics or molecular biology. More recently, the technological breakthrough of high-throughput sequencing (HTS) has provided valuable information, which is increasingly useful in the diagnostic workflow of lymphoproliferative neoplasms. For example, mutational analysis of a panel of genes can help in the establishment of a diagnosis due to the specificity of particular mutations for a given entity (for example, the *BRAF* p.V600E mutation in hairy cell leukemia). Moreover, there is increasing evidence for a prognostic role of mutational analysis (for example, the m7-FLIPI score, which increases the value of the Follicular Lymphoma International Prognostic Index (FLIPI) clinical scoring system in follicular lymphoma (FL)) [1]. Furthermore, lymphoid malignancies have entered the personalized medicine era, where treatment is based on the mutational status of a gene. For example, the EZH2 inhibitor tazemetostat appears to have a very high activity in EZH2-mutated follicular lymphoma and patients with a tumor carrying this mutation may be preferentially treated with this new agent [2].

There are many challenges that need to be overcome to ensure the optimal use of HTS in the diagnosis of lymphoid malignancies. First, tumors are genetically heterogeneous, so the analysis of a given biopsy is not necessarily representative of the whole tumor burden. Second, the DNA can be damaged by the fixation process in formalin-fixed paraffin-embedded (FFPE) specimens, which induces artifacts in the DNA sequence that need to be distinguished from real mutations. Third, the sequencing strategy should be adapted to the clinical needs; choosing an optimal panel is a compromise between clinical, economical, and practical considerations to ensure that the clinician will obtain reliable and relevant mutational data within a reasonable time. Last but not least, the bioinformatic analysis strategy is critical; most importantly, the strategy of variant calling and filtering can introduce great variability that needs to be assessed and controlled [3]. Most laboratories perform tumor-only sequencing (i.e., without a paired germline sample), therefore it can be difficult to assess if a variant was somatically acquired during oncogenesis, and to determine its phenotypic consequences. In this review, we will describe the existing databases and strategies that can help to deal with these issues.

## 2. Is It a Somatic Variant?

The first challenges faced when dealing with HTS data from tumor tissues are to distinguish true sequence variations from technical artifacts, and then to distinguish whether a variant is somatically acquired by the tumor (or even potentially by non-tumor cells) or represents a germline variant that may or may not be implicated in tumorigenesis. To help deal with this problem, the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) have published practical guidelines for the interpretation and reporting of sequence variants [4], which have been since updated twice [3,5].

### 2.1. Pre-Analytical and Bioinformatic Issues

#### 2.1.1. Obtaining Germline DNA

Ideally, one would compare tumor samples to germline DNA (e.g., from blood, saliva, hair, skin, or nails). This is sometimes difficult in routine practice due to the technical difficulties in obtaining a sufficient amount of DNA from these sources (nails, hair), invasiveness (skin), or the possible contamination by cells from the hematological lineages. Moreover, this strategy represents an additional expense.

#### 2.1.2. Technical Considerations: FFPE Tissues

FFPE tissue is the most common form of tissue that is received for molecular testing in the diagnostic laboratory. The fixation process causes various damages to DNA [6], making it

challenging to interpret the variants identified by massively parallel sequencing. Formaldehyde is highly reactive with DNA bases and proteins, generating crosslinks that affect both the isolation of DNA and the amount of amplifiable DNA. Formalin fixation also causes DNA strand to break, leading to extensive fragmentation of DNA and low amounts of template amplifiable by PCR. In highly fragmented DNA samples, detecting true mutations is challenging because of the stochastic variation in allelic representation. The risk of false negatives is particularly increased if tumor purity is low [6]. Moreover, uracyl and thymine, which result from the deamination of cytosine and 5-methylcytosine by formaldehyde, are present in disproportionate levels in FFPE samples when compared to matched-frozen tissues [7]. Amplification by PCR generates C:G > T:A false-positive variants by incorporating an adenine opposite the uracil lesions. This is particularly important in the low (<10%) allele frequency range and inversely correlated with coverage, because of stochastic enrichment in the low copy number context [8]. Amplicon-based sequencing strategies are more prone to these false positive calls than capture-based approaches as they do not retain the information on the number of initiating templates of sequence reads. Accuracy and sensitivity can be improved by workflows that allow sequencing of sense and antisense strands independently (artifacts will be present in only 1 of the 2 DNA strands) [9]. Molecular tagging of the DNA templates by unique molecular identifiers (UMI) is another powerful way to reduce these FFPE artifacts (true mutations are present in all daughter molecules) [10]. Finally, variants (and in particular those of low allele frequency) should be confirmed by replicating the experiment, sequencing the antisense DNA strand, or validated by another approach. Minimizing these artifacts arising from formalin fixation is crucial to accurately detect actionable mutations from the FFPE samples. For all these reasons, if available, fresh tissue is the preferred source of DNA for molecular testing.

### 2.1.3. Bioinformatic Aspects

There are different ways to generate a list of variants from raw HTS data: first, most sequencer manufacturers and some reagent manufacturers offer their own analysis pipelines. Second, several stand-alone or cloud-based commercial software packages are available. Third, numerous open source bioinformatic tools are shared by the scientific community. A commonly used variant caller in laboratories performing constitutional HTS is the Genome Analysis Toolkit (GATK) [11], an open source suite developed at the Broad Institute. However, variant calling of somatic samples is more complex as true mutations can often be found with a variant allele frequency (VAF) of less than 10% (see below). Consequently, laboratories performing somatic HTS use various analysis pipelines without a clear gold standard. In this context, one strategy to increase true positive results is to combine outputs from two or more variant callers (e.g., Mutect2 of the GATK suite, VarScan2, VarDict, Freebayes...) [12–15] as a combination of different callers will perform better than each variant caller taken separately [16–18].

Irrespective of the bioinformatic pipeline used, each variant is characterized by its VAF, which is the proportion of alternate sequences at a given position. Many confounding factors can interfere with its interpretation. In an ideal sample containing a pure tumor with only one clone, the VAF would be either 100% or 50% if the variant were homozygous or heterozygous, respectively. However, samples also contain normal cells, hence lowering the observed VAFs. Moreover, tumor cell heterogeneity will also induce a high range of mutation VAFs. This challenges the accurate interpretation of variants as true variants with a low VAF might be hard to distinguish from technical noise. Different strategies aiming to distinguish low-frequency variants from sequencing background noise have been proposed [19–22].

Another important confounding factor to consider is the copy number variation (CNV). Indeed, when a mutated gene is present in more than one copy (or when a deletion occurs on the wild-type allele), it will artificially increase the VAF leading to interpretation errors. For example, in a sample containing 66% of tumor cells that all harbor a loss of a particular genomic region, a mutation located on the remaining allele would be detected in approximately 50% of the reads (66% of

cells with one single mutated allele and 34% of cells with two wild-type alleles,  $66/(66 + 34 \times 2) = 50\%$ ). Such a VAF of 50% is reminiscent of a germline polymorphism and can lead to a misinterpretation. Conversely, a germline polymorphism, initially present in one out of two alleles in each cell, would then be detected in 25% of the reads if the alternate allele were deleted in 66% of tumor cells (34 alleles with the variant for 34 + 66 wild-type alleles,  $34/134 = 25\%$ ). Such a VAF of 25% might wrongly lead to consider a rare, germline variant as a somatic mutation. This information is not available with all sequencing techniques, therefore remains hidden in most studies. Yet, copy number alterations impact 10% of the human genome [23], and its role in cancer is probably underappreciated.

The downside of decreasing the minimal calling VAFs is that it will also increase the calling of sequencing artifacts, which need to be filtered out using different strategies:

- Some variant callers such as FreeBayes or VarScan can be run in a multi-sample mode, allowing sequencing data to be obtained for all samples regarding a particular allele, if found to be mutated in one sample. From this, the minimal/mean/median frequency of reads supporting the altered sequence among all samples can be computed and inform about the background signal at this position; a VAF/median frequency ratio can be calculated, and a filtering threshold applied to select more probable true somatic mutations.
- The distribution of reference allele and alternate allele between forward and reverse sequencing strand should be similar, i.e., the ratio (reference forward reads)/(reference reverse reads) should be comparable to the ratio (alternate forward reads)/(alternate reverse reads). An Allele Strand Ratio (ASR) can be calculated, (reference F/R reads)/(alternate F/R reads); for real mutations ASR should be close to 1, conversely an ASR very distant from 1 suggests possible artifact variation.
- The occurrence of a variant in samples of the same run should be calculated, a too high recurrence would point out an artifact.
- >Over time, a local database of recurrent artifacts can be built to help remove known false positive calls.

The open source Integrative Genomics Viewer (IGV) allows visualization of the sequencing reads (bam files) [24]. Loading the data of all samples from the same run allows the evaluation of the background signal. PCR artifacts are suggested by variation in the same reads (same start, same stop) and not by others. Visualizing mutations in IGV (or a similar viewer) is essential before inclusion in a clinical report.

## 2.2. Databases for Germline and Somatic Variants

After the variant has been identified as true (i.e., not artifactual), the question of whether it is a germline polymorphism, or an acquired somatic mutation could be difficult to answer, particularly given that most labs do not perform the sequencing of germline DNA in parallel to that of the tumor. The first step to help filter out germline polymorphisms is an assessment of the VAF. Then, the phenotypic evaluation of the analyzed samples represents important information as the level of tumor purity is crucial to analyze the meaning of variants. Beyond the characterization of the sample, the interrogation of large genomic databases is required. However, one needs to understand how these databases are produced: where does the DNA come from (healthy and/or sick population, ethnic and geographical origin), how it was analyzed (quality metrics), and what kind of information is released. Various types of databases are relevant for the interpretation of somatic sequence variants including constitutional population-based databases to exclude polymorphisms and cancer/mutation-specific databases. In this article, we will focus on the databases that are most relevant and helpful when analyzing a sequence variant in routine practice in oncology.

### 2.2.1. Constitutional Databases

These catalogues of germline variants intend to provide a comprehensive list of variants and their frequency in the general population. In a cancer analysis setting, they are useful to exclude some

polymorphisms that could be interpreted as somatic (Tables 1 and 2). Nevertheless, population-based databases should be interpreted with caution. One metric often described is the minor allele frequency (MAF), which represents the frequency at which the second most common allele occurs in a given population. As the MAF varies with ethnicity, the interpretation of a variant should include this information. Importantly, most of the databases are skewed towards an overrepresentation of the Caucasian population. Moreover, some well-established oncogenic mutations are described in these databases, which is not surprising given the high prevalence of detectable oncogenic mutations in healthy people [25]. To stress this point, we will use as an example the MYD88 p.L265P mutation, initially described as a polymorphism. This problem indeed raises epistemological questions about the causal role of gene mutations in oncogenesis, which are neither necessary [26,27] nor sufficient [28]. There is no consensus on the MAF threshold to be used to filter out potential polymorphisms, but most clinical studies use a 1% cut-off to filter-out such inherited polymorphisms.

**Table 1.** Catalogues of germline variants databases.

Database	Cell of Origin	Healthy/Non Cancer Disease	Data	Number of Exome/Genome	URL
ExAC	Germline	both	Exome	60,706	<a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a>
gnomAD	Germline	both	Exome/Genome	136,632 *	<a href="http://gnomad.broadinstitute.org/about">http://gnomad.broadinstitute.org/about</a>
1000 Genomes	Germline	Healthy	Exome/Genome	2504	<a href="http://www.internationalgenome.org/home">http://www.internationalgenome.org/home</a>
dbSNP	Germline	both	Exome/Genome	NA	<a href="https://www.ncbi.nlm.nih.gov/gate2.inist.fr/projects/SNP/">https://www.ncbi.nlm.nih.gov.gate2.inist.fr/projects/SNP/</a>
ESP	Germline	both	Exome	6503	<a href="http://evs.gs.washington.edu/EVS/">http://evs.gs.washington.edu/EVS/</a>

\* 123,136 exomes and 15,496 genomes; NA: No data available.

**Table 2.** Ethnic representation in each database of germline variants, and reported frequencies for the MYD88 p.L265P mutation.

Database	% African /African American	% Latino /Mixed Americans	% East Asian	% Finnish	% Non-Finnish European	% South Asian	% Ashkenazi	% Other	MYD88 p.L265P Allele Frequency	Reference
ExAC	8.57	9.53	7.13	5.45	54.97	13.6	NA	0.75	0.01%	[29]
gnomAD	8.80	12.60	6.91	9.44	46.38	11.26	3.72	2.37	0.0036%	[29]
1000 Genomes	26.4	13.86	19.53	3.95	16.13	20.13	NA	0	0.02%	[30,31]
dbSNP	NA	NA	NA	NA	NA	NA	NA	NA	*	[32]
ESP	NA	NA	NA	NA	NA	NA	NA	NA	Not present	[33]

\* dbSNP reports VAF from different studies; NA: No data available.

- The Single Nucleotide Polymorphism Database (dbSNP) of nucleotide sequence variation (SNV) from the National Cancer Bioinformatics Institute (NCBI):** This public domain catalogue, started in 1998 [32] and first released in 2000 by the NCBI as a part of the PubMed website includes only variants from non-tumor samples. The definition of small variations comprises SNPs, small (<50 nucleotides) insertions or deletions (InDels), and retroposable element insertions and microsatellite repeat variations. The catalogue is based on voluntary contributions as any public laboratory and private organizations can submit data that, after review, will be implemented as “first class data”. The “second-class data” available are computed from the original submitted data, automatically gathered from PubMed during the dbSNP build cycle. In 2004, its false positive rate was estimated at 15–17% by Mitchell et al. [34] mainly due to its conception. It accepts submissions from many sources that are difficult to verify regarding uncritical bioinformatic alignments of highly similar but distinct DNA sequences, or PCRs with primers that cannot discriminate between similar but distinct DNA sequences. In February 2017, the last build regarding Homo Sapiens was released, build 150 comprising 325,658,303 variants of which

135,967,291 were validated (as first-class data). In this database, MYD88 L265P is known as rs387907272, and was considered as a SNP until very recently. It has now been curated as a single nucleotide variant with a pathogenic clinical significance and a variant frequency <0.01%. This well-known pathogenic mutation highlights the difficulty to filter out polymorphisms due to false positives in such databases that are constantly evolving.

- **1000 Genomes Project:** The goal of the 1000 Genomes Project was to identify genetic variants with frequencies of at least 1% in the populations studied. It ran between 2008 and 2015 and sequenced 2504 samples from 26 populations [30,31]. Whilst the samples for the 1000 Genomes Project had no associated medical or phenotype data, all participants had to declare themselves to be healthy with self-reported ethnicity and gender. MYD88 L265P is described with an overall allele frequency of 0.02%, from one non-Finnish European population. In this population, the MAF was still below a threshold that most studies would consider as rare (<0.01%) [35]. It is of note that the SNPs and short Indels of the 1000 Genomes Project are included in the dbSNP, making it redundant to consult both databases.
- **The Exome Aggregation Consortium (ExAC):** This database, compiled by the Broad Institute, tends to aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects. All of these projects provided their raw sequencing data (generated using various technologies), which were then reprocessed and variant called through one unique pipeline to increase consistency. Notably, the germline information obtained by The Cancer Genome Atlas (see below) is available in the ExAC database. Given the nature of the projects aggregated (such as “Inflammatory Bowel Disease”, “Jackson Heart Study”, or “Schizophrenia Trios from Taiwan”), not all of the patients sequenced were healthy. In ExAC, 60,706 unrelated individuals have been sequenced [29], where people with severe pediatric diseases and their first-degree relatives have been removed. This database is not intended to be further extended, and a new project called “The Genome Aggregation Database” (gnomAD; see below) includes all the data contained in the ExAC database. In this ExAC database, MYD88 p.L265P is present at a frequency of 0.01% in the general population and is considered as too common to plausibly cause disease.
- **The Genome Aggregation Database (gnomAD):** This database aims to aggregate the data from genome and exome studies into one database and is mainly driven by the Broad Institute. It uses the data from the ExAC database and from a consortium of more than 100 investigators and uses the same process as that used for ExAC (same pipeline and variant calling to re-process all data). It therefore contains data from the ExAC, 1000 Genomes Project, and the Cancer Genome Atlas (TCGA; see below) among others and spans 123,136 exomes and 15,496 genomes from unrelated individuals. In the gnomAD, MYD88 p.L265P is described with an allele frequency 0.0036% and count of 9.
- **National Heart, Lung and Blood Institute Grand Opportunity (NHLBI GO) Exome Sequencing Project (ESP):** This is an on-going project bringing together US investigators aiming to discover new genes and mechanisms leading to heart disorders, lung disorders, or benign blood disorders. No cancer sequencing data are included in this database comprising 6503 samples in the ESP6500SI-V2 release. The variant calling and analysis of the data are centralized. A subset of the data was published in 2012 [33] and is therefore present in the dbSNP.

### 2.2.2. Cancer Databases

There are no databases specific to lymphoid malignancies, and even if hematological malignancies display many genetic alterations, no dedicated project has been undertaken (Table 3). The TCGA, although less relevant for lymphoid malignancies when compared to other cancers, will be discussed given its importance.

**Table 3.** Catalogues of cancer databases.

Database	Cell of Origin	Data	Number of Exome/Genome	Link	Reference
TCGA	Somatic	Exome /Genome	11,077	<a href="https://tcga-data.nci.nih.gov/docs/publications/tcga/">https://tcga-data.nci.nih.gov/docs/publications/tcga/</a>	[36]
ICGC	Somatic	Exome /Genome	17,000	<a href="http://icgc.org/">http://icgc.org/</a>	[37,38]
COSMIC	Somatic	Exome /Genome	32,000 genomes + 25,000 peer reviewed papers (genomes and/or exomes)	<a href="http://cancer.sanger.ac.uk/cosmic">http://cancer.sanger.ac.uk/cosmic</a>	[39,40]

- **The Cancer Genome Atlas (TCGA):** This project, that ended in 2017, was a collaboration between the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI) [36]. More than 11,000 patients suffering from a total 33 types of cancer (in hematology, only acute myeloblastic leukaemia and diffuse large B cell lymphoma were targeted) were included. Cancer tissue and matched normal tissues were collected as well as clinical data, then sequenced, and registered.
- **The International Cancer Genome Consortium (ICGC):** The ICGC is a confederation of international working groups that aims to describe driver somatic mutations in more than 50 types or subtypes of cancers. Most working groups are required to sequence at least 500 samples by Whole Genome Shotgun analyses, with exceptions made for rare or very homogenous types of cancer. As this technique is not yet available everywhere for such large-scale projects, interim goals are accepted such as sequencing only the region of interest, analysis of low genome coverage of paired-end reads for rearrangements, or genotyping arrays. In addition to building this catalogue of somatic mutations, analyses of DNA methylation and RNA expression are planned. Processing the samples must be conducted according to the ICGC guidelines to ensure similar quality in the projects. Lymphoid malignancies are well represented with six different dedicated projects [37,38].
- **The Catalogue of Somatic Mutation in Cancer (COSMIC):** As is the case for dbSNP, two different types of data are present in the COSMIC. The first-class data is expert-curated, with manual input data after comprehensive review of selected genes after its submission by a group or laboratory. These genes are those presented in Census genes, a dynamic catalogue of genes that have been recognized as implicated in cancer that was initially published in Nature Reviews Cancer [40]. The second type is genome-wide screening data, uploaded from publications or imported from other databases such as the TCGA or ICGC. The uploaded data from publications imply that some false positives are included in this database through the lack of germline sample controls, simple laboratory errors, or poor-quality samples such as FFPE ones. Nevertheless, this catalogue is the most comprehensive resource for information on somatic mutations in human cancer and aims at providing somatic mutation frequencies [39].

Even if a somatic mutation is detected in a patient, its relevance in oncogenesis still needs to be determined separately. ClinVar [41] is a public archive of evidence-based data on the relationship between a variant and a phenotype. This NCBI project is still on-going with continuous submissions of germline or somatic variants. Other helpful tools include stand-alone software packages and web-based content such as the NCI Genomic Data Commons or cbiportal [42] that compiles different types of data from various projects in an attempt to facilitate the interpretation of detected variants. The latter started as a project from the Memorial Sloan Kettering Cancer Center aiming to sequence 10,000 genomes with a new hybridization capture-based HTS panel [43] and now contains 170 cancer studies from various consortia.

### 2.3. Assessing the Functional Consequences and Clinical Impact of the Variants

#### 2.3.1. In Silico Modelling

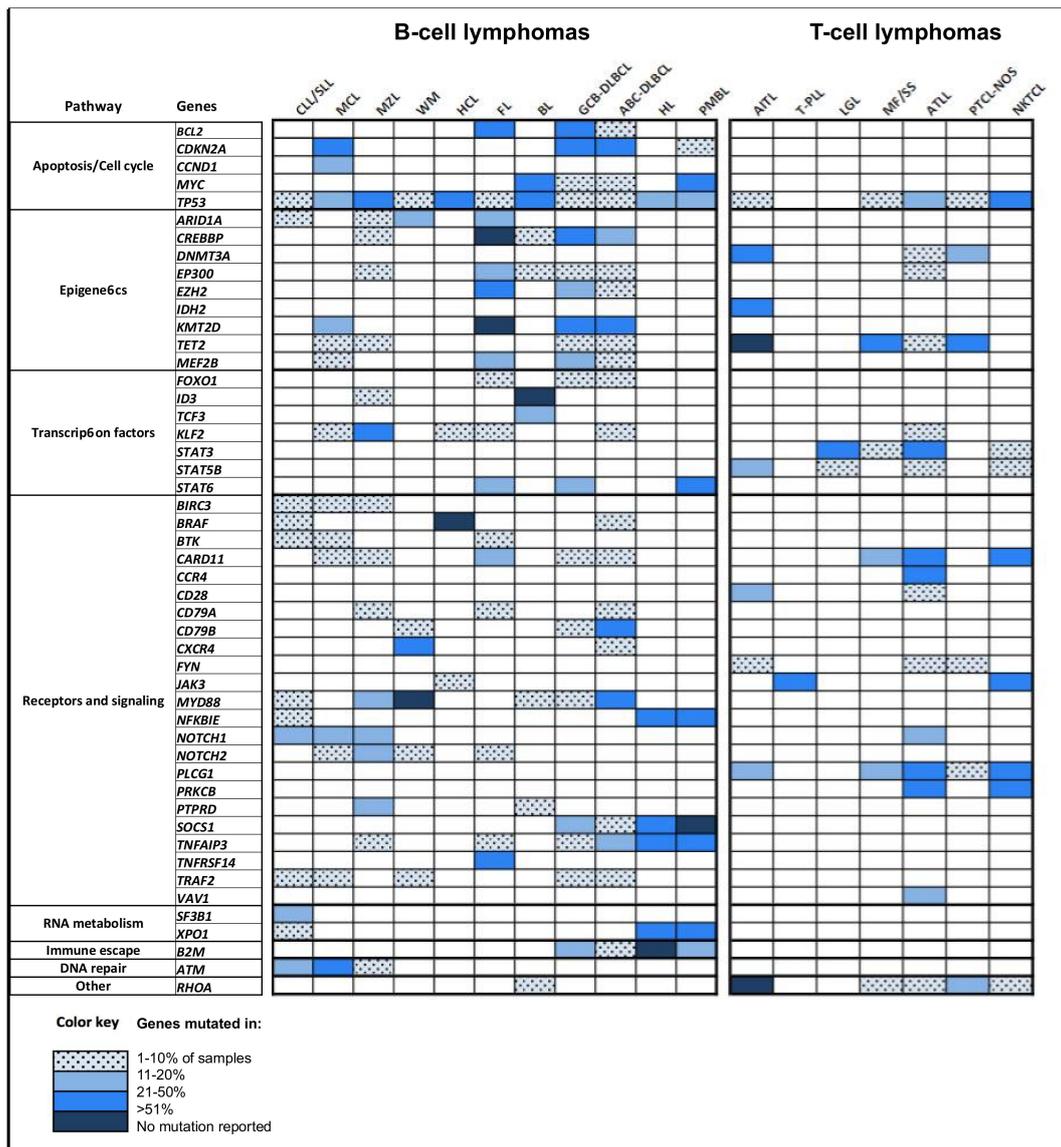
A few hotspots have been described and characterized in lymphoid malignancies (Table 4), showing direct clinical utility in refining diagnosis and/or directing treatment decisions [2,44–48]. The genetic landscape of most lymphomas has also been extensively studied and is now well characterized (Figure 1; for a complete review on recurrent genomic alterations, see [49,50]). Nevertheless, the functional consequences or potential clinical impact of most of the SNVs identified with HTS are currently poorly (if at all) described. In particular, tumor suppressor genes harbor a widely distributed pattern of mutations, making the interpretation of missense mutations more complex.

**Table 4.** Mutation hotspots in lymphoid neoplasms.

Gene	Hotspot Mutation	Lymphoid Neoplasms (Frequency)	Commentary
<i>BRAF</i>	V600E	HCL (>90%), MM (5%)	targeted therapy available
<i>EZH2</i>	Y646, A692 *	FL (30%), DLBCL (10%)	targeted therapy available; * Amino-acid numbering based on transcript NM_004456.4 (sometimes reported as Y641 and A687 with NM_001203247.1)
<i>IDH2</i>	R172	AITL (40%)	targeted therapy available
<i>K/N/H-RAS</i>	G12, G13, Q61	MM (40%), DLBCL (10%)	targeted therapy available
<i>MYD88</i>	L265P **	LPL (95%), MGUS (50%), DLBCL (10%), CLL (5%), PCNSL (50%), EMZL/MALT (5%), NMZL (5%)	** Amino-acid numbering based on transcript NM_002468.4 (sometimes reported as L273P with NM_001172567.1)
<i>RHOA</i>	G17V	AITL (60%), PTCL-NOS (20%)	-
<i>SF3B1</i>	K700E, K666	CLL (15%)	Prognostic impact in CLL
<i>XPO1</i>	E571	PMBL (25%), cHL (25%), CLL (5%)	targeted therapy available

Only recurrent mutations observed with a frequency >10% are presented. Abbreviations: HCL: hairy cell leukemia; MM: multiple myeloma; FL: follicular lymphoma; DLBCL: diffuse large B-cell lymphoma; AITL: angio-immunoblastic T-cell lymphoma; LPL: lymphoplasmacytic lymphoma; MGUS: monoclonal gammopathy of undetermined significance; CLL: chronic lymphocytic leukemia; PCNSL: primary central nervous system lymphoma; EMZL/MALT: extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue; NMZL: nodal marginal zone lymphoma; PTCL-NOS: peripheral T-cell lymphoma, not otherwise specified; PMBL: primary mediastinal B-cell lymphoma; cHL: classical Hodgkin lymphoma. \* Amino-acid numbering based on transcript NM\_004456.4 (sometimes reported as Y641 and A687 with NM\_001203247.1). \*\* Amino-acid numbering based on transcript NM\_002468.4 (sometimes reported as L273P with NM\_001172567.1).

Several prediction tools have been developed to determine the likelihood for a variant to be deleterious (that is, to adversely impact protein function), based on sequence conservation between species, evolutionary relationship in protein family, and putative impact on the protein structure or stability [51]. A non-exhaustive list of the most widely used bioinformatic tools is presented in Table 5. These approaches, although useful, have intrinsic limitations in predicting the impact of variants on protein function. A study comparing the SIFT, PolyPhen2, and MutationAssessor algorithms in lung adenocarcinoma found a concordance of 64% between these methods, that disagreed on multiple variants in clinically relevant genes [52]. The dbNSFP aggregates functional predictions and annotations of all potential non-synonymous and splicing-site SNVs in the human genome [53,54]. The latest version (v3.5) compiles prediction scores from 20 prediction algorithms, six conservation scores, related additional annotations, and allele frequencies observed in different databases including the 1000 Genomes Project, ExAC, gnomAD, and the NHLBI GO ESP. This allows the annotation of variants and comparison of all these prediction scores in one step.



**Figure 1.** Mutation frequencies in different lymphoma entities. Abbreviations: CLL: chronic lymphocytic leukemia; SLL: small lymphocytic lymphoma; MCL: mantle cell lymphoma; MZL: marginal zone lymphoma; WM: Waldenström’s macroglobulinemia; HCL: hairy cell leukemia; FL: follicular lymphoma; BL: Burkitt lymphoma; GCB-DLBCL: germinal-center B-cell-like diffuse large B-cell lymphoma; ABC-DLBCL: activated B-cell-like diffuse large B-cell lymphoma; HL: Hodgkin lymphoma; PMBL: primary mediastinal B-cell lymphoma. AITL: angioimmunoblastic T-cell lymphoma; T-PLL: T-cell prolymphocytic leukemia; LGL: large granular lymphocytic leukemia; MF: mycosis fungoides, SS: Sézary syndrome; ATLL: adult T-cell leukemia/lymphoma; PTCL-NOS: peripheral T-cell lymphoma not otherwise specified; NKTCL: extranodal NK/T-cell lymphoma, nasal-type.

**Table 5.** Bioinformatic resources for prediction of variant functional impact.

Resource	URL	References
SIFT Sorting Intolerant From Tolerant	<a href="http://sift.jcvi.org">http://sift.jcvi.org</a>	[55,56]
PROVEAN Protein Variation Effect Analyzer	<a href="http://provean.jcvi.org/index.php">http://provean.jcvi.org/index.php</a>	[57]
PolyPhen-2 Polymorphism Phenotyping v2	<a href="http://genetics.bwh.harvard.edu/pph2">http://genetics.bwh.harvard.edu/pph2</a>	[58]
MutationAssessor	<a href="http://mutationassessor.org">http://mutationassessor.org</a>	[59]
MutationTaster	<a href="http://www.mutationtaster.org/">http://www.mutationtaster.org/</a>	[60]
CAROL * Combined Annotation scoRing toOL	<a href="http://www.sanger.ac.uk/science/tools/carol">http://www.sanger.ac.uk/science/tools/carol</a>	[61]
Align GCGD **	<a href="http://agvgd.hci.utah.edu/">http://agvgd.hci.utah.edu/</a>	[62]
dbNSFP v3.0 *** database for Nonsynonymous SNPs' Functional Predictions	<a href="https://sites.google.com/site/jpopgen/dbNSFP">https://sites.google.com/site/jpopgen/dbNSFP</a>	[54]

\* Combines SIFT and PolyPhen-2; \*\* Cancer-specific database where users can either supply their own protein multiple sequence alignments or select from the library of alignments (currently available for ATM, BRCA1, BRCA2, CHEK2, and TP53); \*\*\* compiles prediction scores from 20 prediction algorithms (SIFT, Polyphen2-HDIV, Polyphen2-HVAR, LRT, MutationTaster2, MutationAssessor, FATHMM, MetaSVM, MetaLR, CADD, VEST3, PROVEAN, FATHMM-MKL coding, fitCons, DANN, GenoCanyon, Eigen coding, Eigen-PC, M-CAP, REVEL, MutPred) and 6 conservation scores (PhyloP  $\times$  2, phastCons  $\times$  2, GERP++ and SiPhy).

Another limitation of most databases is the lack of annotation and predicted functional impact of synonymous mutations, which change the sequence of a gene without altering the primary amino acid sequence of the encoded protein. As a consequence, they are largely ignored in clinical practice. However, accumulating experimental evidence has demonstrated that these so-called “silent” mutations actually contribute to human cancers by altering splicing, mRNA stability, protein folding, and translation [63,64]. This highlights the need to incorporate information about synonymous mutations in databases.

Deciphering whether a variant represents a driver in lymphomagenesis, and how to distinguish them from so-called “passenger” mutations is a fundamental challenge. A cancer driver mutation is defined as providing a selective growth advantage under given microenvironmental conditions, thus promoting clonal expansion. A passenger mutation provides modest (if any) advantage to the fitness of the cell, however, it might be selected in the expanding clone if it co-occurs with a driver mutation. As experimental evidence demonstrating that a mutation is responsible for a cellular phenotype that provides a selective advantage to the cell cannot be easily obtained, two approaches are considered as “surrogate” methods to distinguish driver from passenger mutations. First, the VAF has long been intuitively used as a surrogate for identifying potential drivers based on the assumption that early events, clonally represented with high VAF, likely represent alterations driving the cellular fitness and early tumor progression (although this cannot exclude a passenger mutation arising early in the progenitor cancer cells). Nevertheless, in routine practice, the VAFs are rarely corrected for copy-number status, which might greatly influence VAF. Second, several computational algorithms have been developed to address this question. Some of these in silico methods use the concept of “significantly mutated genes”, which calculate the mutation rate according to gene size and sequence context (silent mutations and non-coding mutations in the surrounding regions) to determine whether the observed mutation rate is higher than expected by chance in a given gene (e.g., MuSiC, MutSigCV) [65,66]. Other methods evaluate the expected functional impact (missense, loss-of-function, silent...) and distribution of the mutations along the gene, the entropy score for missense mutations (i.e., the degree of reoccurring mutations at a specific site within a gene), the nucleotide-level inter-species conservation, or the DNA and protein contexts (e.g., SNP density, modification of a functional domain, predicted secondary structure, change in hydrophobicity/polarity/charge...). Numerous algorithms have thus been proposed to predict the relevance of somatic mutations in cancer cells (CHASM, TUSON, Oncodrive FM, OncodriveCLUST) [67–70]. Nevertheless, a comparison

showed a huge variability in the genes identified as cancer drivers by the different prediction methods, thus questioning the frequency of false-positive calls when using such algorithms [71,72]. Moreover, some of these approaches are designed for whole genome or whole exome data and are not appropriate for gene panels that are currently used in clinical practice.

### 2.3.2. In Vitro Modelling

Beyond *in silico* predictions, the consequences of a given mutation can be tested *in vitro*. Whereas establishing an accurate model is time consuming and not compatible with the time constraints of the clinical setting, some public databases assessing the functional consequences of a large number of mutations can be very informative. However, one should keep in mind that *in vitro* assessment may not reproduce *in vivo* behavior; for example, despite all the models showing activation of the Mitogen-Activated Protein Kinase (MAPK) pathway by BRAF p.V600E mutations, colorectal cancers with this mutation are not sensitive to vemurafenib, because of the activation of a feedback loop leading to Epidermal Growth Factor Receptor (EGFR) activation [73].

- Large scale pharmacogenomic screening of cell lines: Two major initiatives have analyzed how the sensitivity of cell lines towards a large panel of chemical compounds is correlated with genomic features. Importantly, there is a good agreement between the results obtained with these two large scale projects [74].

The Genomics of Drug Sensitivity in Cancer project has described the sensitivity of 1001 cancer cell lines to 265 anticancer drugs, and compared the response rates with the analysis of DNA sequence, copy number anomalies, DNA methylation, and gene expression at the mRNA level [75]. The data are easily accessible via a website [76], allowing the evaluation of drug sensitivity when a given gene is mutated, amplified, or deleted. However, no distinction is made between the different mutations existing for a given gene (such as gain-of-function or loss-of-function mutations). This point is of importance when interpreting these data.

The Cancer Cell Line Encyclopedia project has produced similar data [77] on 242 cell lines exposed to 354 small molecules, which was later extended to 860 cell lines and 481 compounds in an updated version. The data are also available through a web interface [78], allowing the evaluation of how genomic or transcriptomic features (mutations, CNV analysis, gene expression) impact drug sensitivity. For example, the mTORC1 inhibitor sirolimus is predicted to be more active in the case of *PTEN* mutation or deletion, as expected, and less active in the case of *CD79B* mutation. Whether these data will translate into clinically meaningful information remains to be demonstrated. Caution is required when interpreting these data, which should be considered for generation of hypotheses and not to guide patient treatment.

- Large-scale phenotypic characterization of mutations: Recently, a team from the Broad Institute has succeeded in moving from high throughput genomic characterization of lung cancer to large-scale functional analysis of variants [52]. For the 194 most frequent mutations described in lung cancer, they measured how much the overexpression of the mutated gene impacted the transcriptome (using a reduced transcriptome called L1000) [79] when compared with the overexpression of the wild-type form; they demonstrated that only 69% of these variants had measurable functional consequences. However, to the best of our knowledge, no such database exists for lymphoma mutations, but this approach may not be feasible owing to the extreme heterogeneity between lymphoma subtypes.

### 2.3.3. Limits of Current Prediction Tools and Models to Predict the Clinical Impact

Despite efficient prediction tools and large-scale *in vitro* screening, none of these approaches can totally unravel the *in vivo* complexity of tumor biology. A first limit is that very small mutated sub-clones might not necessarily be identified as potential drivers, and their selection under treatment is not predictable by using only *in silico* approaches. In chronic lymphocytic leukemia, sub-clones

with *TP53* mutations have been shown to expand to dominant clones under the selective pressure of chemotherapy [80–83], and to greatly influence the response to chemotherapy and clinical outcome, irrespective of the VAF (however this concept has been challenged in a recently published study [84]). Second, the selection of small clones by a dysregulated tumor microenvironment during the course of the disease should probably reduce the importance that is given to the VAF or in silico approaches in mutational status interpretation at the time of diagnosis [80,85]. Adding another layer of complexity, the influence of a mutated clone on non-mutated neighboring cells probably represents a still unrecognized pathophysiological mechanism of drug resistance in some tumors bearing particular small mutated clones. Such a paracrine pro-survival signal was recently shown for the tyrosine-protein kinase BTK p.C481S mutation-bearing cells in Waldenström's Macroglobulinemia (WM) and diffuse large B-cell lymphoma of the activated B-cell subtype (DLBCL-ABC) [86]. Large-scale pharmacogenomics or functional screening of cell lines do not investigate the potential impact of the tumor microenvironment on mutated sub-clones, nor the interplay between tumor cells with different molecular features.

All of these findings confirm that limiting the tumor genetics to the dominant clone may hamper the accurate prediction of outcome and optimal therapeutic decisions. This may have direct consequences for the design of clinical trials when patient enrolment relies on mutational status, for example, whether a VAF threshold should be applied to enroll patients for targeted therapy, and whether a tumor with a small mutated sub-clone will respond similarly to one with a predominantly mutated population. Recent data suggest that the presence of small BTK p.C481S mutation-bearing clones may determine the response to BTK inhibitor ibrutinib [86].

The combined effect of co-occurring mutations might also greatly influence the biological and molecular properties of tumors, and the response to treatment. Two striking examples of such interaction were recently described in B-cell lymphomas: in WM, patients with MYD88<sup>L265P</sup> and CXCR4<sup>WT</sup> status showed the highest response rate to ibrutinib when compared to MYD88<sup>L265P</sup>CXCR4<sup>MUT</sup> and MYD88<sup>WT</sup>CXCR4<sup>WT</sup> patients [46]; in DLBCL, tumors with CD79A/B<sup>WT</sup>MYD88<sup>MUT</sup> did not respond to ibrutinib, whereas CD79A/B<sup>MUT</sup>MYD88<sup>MUT</sup>, CD79A/B<sup>WT</sup>MYD88<sup>WT</sup>, and CD79A/B<sup>MUT</sup>MYD88<sup>WT</sup> tumors did [44]. Nevertheless, in silico prediction of the impact of associated variants is highly challenging. To the best of our knowledge, this issue was addressed by at least one publication [87], but evidence regarding the accuracy of those predictions is lacking. Finally, the order in which the associated mutations were acquired in tumor-initiating cells has been shown to influence clinical features and the response to targeted therapy in myeloproliferative neoplasms [88,89], but such evidence has yet to be demonstrated in lymphoid neoplasms. Future studies incorporating single-cell genotyping might answer this question in lymphoma.

### 3. Conclusions

HTS allows refinement of molecular diagnosis in lymphoid malignancies, which has therapeutic and prognostic implications. For these reasons, it is becoming a method of choice for variant detection by many clinical laboratories. However, the difficulties for an accurate and reproducible report of variants between platforms and laboratories are far from trivial and span technical, computational, and biological challenges in data interpretation. For the assessment of the functional consequences of variants identified in lymphoid malignancies, specific databases, or the extension of existing ones are needed. Such a database dedicated to lymphoproliferative disorders could be built by collecting all known gene lesions published in peer-reviewed literature, and/or from user-submitted data, followed by manual curation on an ongoing basis. Given the expansion of the HTS studies, a large number of lymphoid variants might be quickly listed in a specific database that would be further implemented for both novel mutations and additional annotations. Ideally, the information needed would include the validation in germline controls (i.e., whether a mutation was proven to be somatic), functional consequences from in vitro experiments (if assessed), correlation with other

genomic, epigenomic or transcriptomic features of the tumors, phenotypic impact on drug sensitivity (if assessed), and curated information regarding the lymphoma subtype (such as detailed histology report) to allow appropriate interpretation.

**Acknowledgments:** No funds were received for this study. The authors are grateful to the Direction de la Recherche Clinique et de l'Innovation (DRCI) from Hospices Civils de Lyon for the English editing of the manuscript.

**Author Contributions:** Yann Guillermin, Pierre Sujobert and Sarah Huet conceived and designed the study. All authors contributed significantly to the literature search, data analysis and writing of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pastore, A.; Jurinovic, V.; Kridel, R.; Hoster, E.; Staiger, A.M.; Szczepanowski, M.; Pott, C.; Kopp, N.; Murakami, M.; Horn, H.; et al. Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: A retrospective analysis of a prospective clinical trial and validation in a population-based registry. *Lancet Oncol.* **2015**, *16*, 1111–1122. [CrossRef]
2. Morschhauser, F.; Salles, G.; McKay, P.; Tilly, H.; Schmitt, A.; Gerecitano, J.; Johnson, P.; Le Gouill, S.; Dickinson, M.J.; Fruchart, C.; et al. Interim Report from a Phase 2 Multicenter Study of Tazemetostat, an Ezh2 Inhibitor, in Patients with Relapsed or Refractory B-Cell Non-Hodgkin Lymphomas. *Hematol. Oncol.* **2017**, *35*, 24–25. [CrossRef]
3. Roy, S.; Coldren, C.; Karunamurthy, A.; Kip, N.S.; Klee, E.W.; Lincoln, S.E.; Leon, A.; Pullambhatla, M.; Temple-Smolkin, R.L.; Voelkerding, K.V.; et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J. Mol. Diagn.* **2018**, *20*, 4–27. [CrossRef] [PubMed]
4. 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. [CrossRef] [PubMed]
5. 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. [CrossRef] [PubMed]
6. Do, H.; Dobrovic, A. Sequence artifacts in DNA from formalin-fixed tissues: Causes and strategies for minimization. *Clin. Chem.* **2015**, *61*, 64–71. [CrossRef] [PubMed]
7. Oh, E.; Choi, Y.-L.; Kwon, M.J.; Kim, R.N.; Kim, Y.J.; Song, J.-Y.; Jung, K.S.; Shin, Y.K. Comparison of Accuracy of Whole-Exome Sequencing with Formalin-Fixed Paraffin-Embedded and Fresh Frozen Tissue Samples. *PLoS ONE* **2015**, *10*, e0144162. [CrossRef] [PubMed]
8. Wong, S.Q.; Li, J.; Tan, A.Y.-C.; Vedururu, R.; Pang, J.-M.B.; Do, H.; Ellul, J.; Doig, K.; Bell, A.; MacArthur, G.A.; et al. CANCER 2015 Cohort Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med. Genom.* **2014**, *7*, 23. [CrossRef] [PubMed]
9. Schmitt, M.W.; Kennedy, S.R.; Salk, J.J.; Fox, E.J.; Hiatt, J.B.; Loeb, L.A. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14508–14513. [CrossRef] [PubMed]
10. Kinde, I.; Wu, J.; Papadopoulos, N.; Kinzler, K.W.; Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9530–9535. [CrossRef] [PubMed]
11. Broad Institute. Genome Analysis Toolkit. Available online: <https://software.broadinstitute.org/gatk/> (accessed on 11 February 2018).
12. Cibulskis, K.; Lawrence, M.S.; Carter, S.L.; Sivachenko, A.; Jaffe, D.; Sougnez, C.; Gabriel, S.; Meyerson, M.; Lander, E.S.; Getz, G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **2013**, *31*, 213–219. [CrossRef] [PubMed]

13. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv*, **2012**. Available online: <https://arxiv.org/abs/1207.3907>(accessed on 11 February 2018).
14. 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. [[CrossRef](#)] [[PubMed](#)]
15. 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. [[CrossRef](#)] [[PubMed](#)]
16. Krøigård, A.B.; Thomassen, M.; Lænkholm, A.-V.; Kruse, T.A.; Larsen, M.J. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PLoS ONE* **2016**, *11*, e0151664. [[CrossRef](#)] [[PubMed](#)]
17. Wang, Q.; Jia, P.; Li, F.; Chen, H.; Ji, H.; Hucks, D.; Dahlman, K.B.; Pao, W.; Zhao, Z. Detecting somatic point mutations in cancer genome sequencing data: A comparison of mutation callers. *Genome Med.* **2013**, *5*, 91. [[CrossRef](#)] [[PubMed](#)]
18. Sandmann, S.; de Graaf, A.O.; Karimi, M.; van der Reijden, B.A.; Hellström-Lindberg, E.; Jansen, J.H.; Dugas, M. Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Sci. Rep.* **2017**, *7*, 43169. [[CrossRef](#)] [[PubMed](#)]
19. Kockan, C.; Hach, F.; Sarrafi, I.; Bell, R.H.; McConeghy, B.; Beja, K.; Haegert, A.; Wyatt, A.W.; Volik, S.V.; Chi, K.N.; et al. SiNVICT: Ultra-sensitive detection of single nucleotide variants and indels in circulating tumour DNA. *Bioinformatics* **2017**, *33*, 26–34. [[CrossRef](#)] [[PubMed](#)]
20. Hao, Y.; Xuei, X.; Li, L.; Nakshatri, H.; Edenberg, H.J.; Liu, Y. RareVar: A Framework for Detecting Low-Frequency Single-Nucleotide Variants. *J. Comput. Biol.* **2017**, *24*, 637–646. [[CrossRef](#)] [[PubMed](#)]
21. Fernandez-Cuesta, L.; Perdomo, S.; Avogbe, P.H.; Leblay, N.; Delhomme, T.M.; Gaborieau, V.; Abedi-Ardekani, B.; Chanudet, E.; Olivier, M.; Zaridze, D.; et al. Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer. *EBioMedicine* **2016**, *10*, 117–123. [[CrossRef](#)] [[PubMed](#)]
22. Muller, E.; Goardon, N.; Brault, B.; Rousselin, A.; Paimparay, G.; Legros, A.; Fouillet, R.; Bruet, O.; Tranchant, A.; Domin, F.; et al. OutLyzer: Software for extracting low-allele-frequency tumor mutations from sequencing background noise in clinical practice. *Oncotarget* **2016**, *7*, 79485–79493. [[CrossRef](#)] [[PubMed](#)]
23. Shlien, A.; Malkin, D. Copy number variations and cancer. *Genome Med.* **2009**, *1*, 62. [[CrossRef](#)] [[PubMed](#)]
24. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative Genomics Viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [[CrossRef](#)] [[PubMed](#)]
25. Martincorena, I.; Roshan, A.; Gerstung, M.; Ellis, P.; Van Loo, P.; McLaren, S.; Wedge, D.C.; Fullam, A.; Alexandrov, L.B.; Tubio, J.M.; et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **2015**, *348*, 880–886. [[CrossRef](#)] [[PubMed](#)]
26. Mack, S.C.; Witt, H.; Piro, R.M.; Gu, L.; Zuyderduyn, S.; Stütz, A.M.; Wang, X.; Gallo, M.; Garzia, L.; Zayne, K.; et al. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **2014**, *506*, 445–450. [[CrossRef](#)] [[PubMed](#)]
27. Parker, M.; Mohankumar, K.M.; Punchihewa, C.; Weinlich, R.; Dalton, J.D.; Li, Y.; Lee, R.; Tatevossian, R.G.; Phoenix, T.N.; Thiruvankatam, R.; et al. C11orf95-RELA fusions drive oncogenic NF- $\kappa$ B signalling in ependymoma. *Nature* **2014**, *506*, 451–455. [[CrossRef](#)] [[PubMed](#)]
28. Xie, M.; Lu, C.; Wang, J.; McLellan, M.D.; Johnson, K.J.; Wendl, M.C.; McMichael, J.F.; Schmidt, H.K.; Yellapantula, V.; Miller, C.A.; et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **2014**, *20*, 1472–1478. [[CrossRef](#)] [[PubMed](#)]
29. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285. [[CrossRef](#)] [[PubMed](#)]
30. Consortium, T. 1000 G. P. A global reference for human genetic variation. *Nature* **2015**, *526*, 68. [[CrossRef](#)]
31. Sudmant, P.H.; Rausch, T.; Gardner, E.J.; Handsaker, R.E.; Abyzov, A.; Huddleston, J.; Zhang, Y.; Ye, K.; Jun, G.; Fritz, M.H.-Y.; et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **2015**, *526*, 75. [[CrossRef](#)] [[PubMed](#)]
32. Sherry, S.T.; Ward, M.; Sirotkin, K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **1999**, *9*, 677–679. [[PubMed](#)]

33. Tennessen, J.A.; Bigam, A.W.; O'Connor, T.D.; Fu, W.; Kenny, E.E.; Gravel, S.; McGee, S.; Do, R.; Liu, X.; Jun, G.; et al. NHLBI Exome Sequencing Project Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **2012**, *337*, 64–69. [[CrossRef](#)] [[PubMed](#)]
34. Mitchell, A.A.; Zwick, M.E.; Chakravarti, A.; Cutler, D.J. Discrepancies in dbSNP confirmation rates and allele frequency distributions from varying genotyping error rates and patterns. *Bioinformatics* **2004**, *20*, 1022–1032. [[CrossRef](#)] [[PubMed](#)]
35. Panoutsopoulou, K.; Tachmazidou, I.; Zeggini, E. In search of low-frequency and rare variants affecting complex traits. *Hum. Mol. Genet.* **2013**, *22*, R16–R21. [[CrossRef](#)] [[PubMed](#)]
36. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **2015**, *19*, A68–A77. [[CrossRef](#)] [[PubMed](#)]
37. The ICGC MML-Seq Project. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat. Genet.* **2012**, *44*, 1316–1320. [[CrossRef](#)]
38. Ramsay, A.J.; Martínez-Trillos, A.; Jares, P.; Rodríguez, D.; Kwarciak, A.; Quesada, V. Next-generation sequencing reveals the secrets of the chronic lymphocytic leukemia genome. *Clin. Transl. Oncol.* **2013**, *15*, 3–8. [[CrossRef](#)] [[PubMed](#)]
39. Forbes, S.A.; Bhamra, G.; Bamford, S.; Dawson, E.; Kok, C.; Clements, J.; Menzies, A.; Teague, J.W.; Futreal, P.A.; Stratton, M.R. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.* 2008. [[CrossRef](#)]
40. Futreal, P.A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M.R. A census of human cancer genes. *Nat. Rev. Cancer* **2004**, *4*, 177. [[CrossRef](#)] [[PubMed](#)]
41. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; et al. ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868. [[CrossRef](#)] [[PubMed](#)]
42. MSKCC. cBioPortal for Cancer Genomics. Available online: <http://www.cbioportal.org/> (accessed on 11 February 2018).
43. Zehir, A.; Benayed, R.; Shah, R.H.; Syed, A.; Middha, S.; Kim, H.R.; Srinivasan, P.; Gao, J.; Chakravarty, D.; Devlin, S.M.; et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **2017**, *23*, 703–713. [[CrossRef](#)] [[PubMed](#)]
44. Wilson, W.H.; Young, R.M.; Schmitz, R.; Yang, Y.; Pittaluga, S.; Wright, G.; Lih, C.-J.; Williams, P.M.; Shaffer, A.L.; Gerecitano, J.; et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat. Med.* **2015**, *21*, 922–926. [[CrossRef](#)] [[PubMed](#)]
45. Tiacci, E.; Trifonov, V.; Schiavoni, G.; Holmes, A.; Kern, W.; Martelli, M.P.; Pucciarini, A.; Bigerna, B.; Pacini, R.; Wells, V.A.; et al. BRAF Mutations in Hairy-Cell Leukemia. *N. Engl. J. Med.* **2011**, *364*, 2305–2315. [[CrossRef](#)] [[PubMed](#)]
46. Treon, S.P.; Xu, L.; Yang, G.; Zhou, Y.; Liu, X.; Cao, Y.; Sheehy, P.; Manning, R.J.; Patterson, C.J.; Tripsas, C.; et al. MYD88 L265P Somatic Mutation in Waldenström's Macroglobulinemia. *N. Engl. J. Med.* **2012**, *367*, 826–833. [[CrossRef](#)] [[PubMed](#)]
47. Odejide, O.; Weigert, O.; Lane, A.A.; Toscano, D.; Lunning, M.A.; Kopp, N.; Kim, S.; van Bodegom, D.; Bolla, S.; Schatz, J.H.; et al. A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood* **2014**, *123*, 1293–1296. [[CrossRef](#)] [[PubMed](#)]
48. Palomero, T.; Couronné, L.; Khiabani, H.; Kim, M.-Y.; Ambesi-Impiombato, A.; Perez-Garcia, A.; Carpenter, Z.; Abate, F.; Allegretta, M.; Haydu, J.E.; et al. Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat. Genet.* **2014**, *46*, 166–170. [[CrossRef](#)] [[PubMed](#)]
49. Bogusz, A.M.; Bagg, A. Genetic aberrations in small B-cell lymphomas and leukemias: Molecular pathology, clinical relevance and therapeutic targets. *Leuk. Lymphoma* **2016**, *57*, 1991–2013. [[CrossRef](#)] [[PubMed](#)]
50. Rosenquist, R.; Rosenwald, A.; Du, M.-Q.; Gaidano, G.; Groenen, P.; Wotherspoon, A.; Ghia, P.; Gaulard, P.; Campo, E.; Stamatopoulos, K. European Research Initiative on CLL (ERIC) and the European Association for Haematopathology (EAHP) Clinical impact of recurrently mutated genes on lymphoma diagnostics: State-of-the-art and beyond. *Haematologica* **2016**, *101*, 1002–1009. [[CrossRef](#)] [[PubMed](#)]
51. Tsang, H.; Addepalli, K.; Davis, S.R. Resources for Interpreting Variants in Precision Genomic Oncology Applications. *Front. Oncol.* **2017**, *7*, 214. [[CrossRef](#)] [[PubMed](#)]

52. Berger, A.H.; Brooks, A.N.; Wu, X.; Shrestha, Y.; Chouinard, C.; Piccioni, F.; Bagul, M.; Kamburov, A.; Imielinski, M.; Hogstrom, L.; et al. High-throughput Phenotyping of Lung Cancer Somatic Mutations. *Cancer Cell* **2016**, *30*, 214–228. [[CrossRef](#)] [[PubMed](#)]
53. Liu, X.; Jian, X.; Boerwinkle, E. dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **2011**, *32*, 894–899. [[CrossRef](#)] [[PubMed](#)]
54. Liu, X.; Wu, C.; Li, C.; Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* **2016**, *37*, 235–241. [[CrossRef](#)] [[PubMed](#)]
55. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [[CrossRef](#)] [[PubMed](#)]
56. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **2009**, *4*, 1073–1081. [[CrossRef](#)] [[PubMed](#)]
57. Choi, Y.; Sims, G.E.; Murphy, S.; Miller, J.R.; Chan, A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **2012**, *7*, e46688. [[CrossRef](#)] [[PubMed](#)]
58. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [[CrossRef](#)] [[PubMed](#)]
59. Reva, B.; Antipin, Y.; Sander, C. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* **2011**, *39*, e118. [[CrossRef](#)] [[PubMed](#)]
60. Schwarz, J.M.; Cooper, D.N.; Schuelke, M.; Seelow, D. MutationTaster2: Mutation prediction for the deep-sequencing age. *Nat. Methods* **2014**, *11*, 361–362. [[CrossRef](#)] [[PubMed](#)]
61. Lopes, M.C.; Joyce, C.; Ritchie, G.R.S.; John, S.L.; Cunningham, F.; Asimit, J.; Zeggini, E. A combined functional annotation score for non-synonymous variants. *Hum. Hered.* **2012**, *73*, 47–51. [[CrossRef](#)] [[PubMed](#)]
62. Mathe, E.; Olivier, M.; Kato, S.; Ishioka, C.; Hainaut, P.; Tavtigian, S.V. Computational approaches for predicting the biological effect of p53 missense mutations: A comparison of three sequence analysis based methods. *Nucleic Acids Res.* **2006**, *34*, 1317–1325. [[CrossRef](#)] [[PubMed](#)]
63. Supek, F.; Miñana, B.; Valcárcel, J.; Gabaldón, T.; Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **2014**, *156*, 1324–1335. [[CrossRef](#)] [[PubMed](#)]
64. Gotea, V.; Gartner, J.J.; Qutob, N.; Elnitski, L.; Samuels, Y. The functional relevance of somatic synonymous mutations in melanoma and other cancers. *Pigment Cell Melanoma Res.* **2015**, *28*, 673–684. [[CrossRef](#)] [[PubMed](#)]
65. Dees, N.D.; Zhang, Q.; Kandath, C.; Wendl, M.C.; Schierding, W.; Koboldt, D.C.; Mooney, T.B.; Callaway, M.B.; Dooling, D.; Mardis, E.R.; et al. MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* **2012**, *22*, 1589–1598. [[CrossRef](#)] [[PubMed](#)]
66. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **2013**, *499*, 214–218. [[CrossRef](#)] [[PubMed](#)]
67. Carter, H.; Chen, S.; Isik, L.; Tyekuceva, S.; Velculescu, V.E.; Kinzler, K.W.; Vogelstein, B.; Karchin, R. Cancer-specific high-throughput annotation of somatic mutations: Computational prediction of driver missense mutations. *Cancer Res.* **2009**, *69*, 6660–6667. [[CrossRef](#)] [[PubMed](#)]
68. Gonzalez-Perez, A.; Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **2012**, *40*, e169. [[CrossRef](#)] [[PubMed](#)]
69. Davoli, T.; Xu, A.W.; Mengwasser, K.E.; Sack, L.M.; Yoon, J.C.; Park, P.J.; Elledge, S.J. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **2013**, *155*, 948–962. [[CrossRef](#)] [[PubMed](#)]
70. Tamborero, D.; Gonzalez-Perez, A.; Lopez-Bigas, N. OncodriveCLUST: Exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* **2013**, *29*, 2238–2244. [[CrossRef](#)] [[PubMed](#)]
71. Tamborero, D.; Gonzalez-Perez, A.; Perez-Llamas, C.; Deu-Pons, J.; Kandath, C.; Reimand, J.; Lawrence, M.S.; Getz, G.; Bader, G.D.; Ding, L.; et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **2013**, *3*, 2650. [[CrossRef](#)] [[PubMed](#)]
72. Tokheim, C.J.; Papadopoulos, N.; Kinzler, K.W.; Vogelstein, B.; Karchin, R. Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14330–14335. [[CrossRef](#)] [[PubMed](#)]

73. Prahallad, A.; Sun, C.; Huang, S.; Di Nicolantonio, F.; Salazar, R.; Zecchin, D.; Beijersbergen, R.L.; Bardelli, A.; Bernards, R. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **2012**, *483*, 100–103. [CrossRef] [PubMed]
74. Cancer Cell Line Encyclopedia Consortium. Genomics of Drug Sensitivity in Cancer Consortium Pharmacogenomic agreement between two cancer cell line data sets. *Nature* **2015**, *528*, 84–87. [CrossRef]
75. Iorio, F.; Knijnenburg, T.A.; Vis, D.J.; Bignell, G.R.; Menden, M.P.; Schubert, M.; Aben, N.; Gonçalves, E.; Barthorpe, S.; Lightfoot, H.; et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **2016**, *166*, 740–754. [CrossRef] [PubMed]
76. Home Page—Cancerrxgene—Genomics of Drug Sensitivity in Cancer. Available online: <http://www.cancerrxgene.org/> (accessed on 11 February 2018).
77. Basu, A.; Bodycombe, N.E.; Cheah, J.H.; Price, E.V.; Liu, K.; Schaefer, G.I.; Ebright, R.Y.; Stewart, M.L.; Ito, D.; Wang, S.; et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **2013**, *154*, 1151–1161. [CrossRef] [PubMed]
78. Cancer Therapeutics Response Portal. Available online: <https://portals.broadinstitute.org/ctrp.v2.1/> (accessed on 11 February 2018).
79. Peck, D.; Crawford, E.D.; Ross, K.N.; Stegmaier, K.; Golub, T.R.; Lamb, J. A method for high-throughput gene expression signature analysis. *Genome Biol.* **2006**, *7*, R61. [CrossRef] [PubMed]
80. Rossi, D.; Khiabanian, H.; Spina, V.; Ciardullo, C.; Brusca, A.; Famà, R.; Rasi, S.; Monti, S.; Deambrogi, C.; De Paoli, L.; et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* **2014**, *123*, 2139–2147. [CrossRef] [PubMed]
81. Landau, D.A.; Carter, S.L.; Stojanov, P.; McKenna, A.; Stevenson, K.; Lawrence, M.S.; Sougnez, C.; Stewart, C.; Sivachenko, A.; Wang, L.; et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **2013**, *152*, 714–726. [CrossRef] [PubMed]
82. Malcikova, J.; Stano-Kozubik, K.; Tichy, B.; Kantorova, B.; Pavlova, S.; Tom, N.; Radova, L.; Smardova, J.; Pardy, F.; Doubek, M.; et al. Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. *Leukemia* **2015**, *29*, 877–885. [CrossRef] [PubMed]
83. Nadeu, F.; Delgado, J.; Royo, C.; Baumann, T.; Stankovic, T.; Pinyol, M.; Jares, P.; Navarro, A.; Martín-García, D.; Beà, S.; et al. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood* **2016**, *127*, 2122–2130. [CrossRef] [PubMed]
84. Blakemore, S.; Clifford, R.; Antoniou, P.; Parker, H.; Robbe, P.; Larrayoz, M.; Davis, Z. The Contribution of Gene Mutations to Long-Term Clinical Outcomes: Data from the Randomised UK LRF CLL4 Trial. *Am. Soc. Hematol.* **2017**, *130*, 259.
85. Cooper, J.N.; Young, N.S. Clonality in context: Hematopoietic clones in their marrow environment. *Blood* **2017**, *130*, 2363–2372. [CrossRef] [PubMed]
86. Chen, J.; Liu, X.; Munshi, M.; Xu, L.; Tsakmaklis, N.; Demos, M.; Kofides, A. BTK(Cys481Ser) Mutation Drives Ibrutinib Resistance through ERK1/2 Hyperactivation, and Can Confer a Protective Effect on Bystander Waldenström’s Macroglobulinemia and ABC DLBCL Cells through Paracrine Mediated Pro-Survival Signaling. *Am. Soc. Hematol.* **2017**, *130*, 803.
87. Liu, M.; Watson, L.T.; Zhang, L. Predicting the combined effect of multiple genetic variants. *Hum. Genom.* **2015**, *9*, 18. [CrossRef] [PubMed]
88. Ortmann, C.A.; Kent, D.G.; Nangalia, J.; Silber, Y.; Wedge, D.C.; Grinfeld, J.; Baxter, E.J.; Massie, C.E.; Papaemmanuil, E.; Menon, S.; et al. Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med.* **2015**, *372*, 601–612. [CrossRef] [PubMed]
89. Nangalia, J.; Nice, F.L.; Wedge, D.C.; Godfrey, A.L.; Grinfeld, J.; Thakker, C.; Massie, C.E.; Baxter, J.; Sewell, D.; Silber, Y.; et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica* **2015**, *100*, e438–e442. [CrossRef] [PubMed]

