

Article

# Occurrence of L1M Elements in Chromosomal Rearrangements Associated to Chronic Myeloid Leukemia (CML): Insights from Patient-Specific Breakpoints Characterization

Alberto L'Abbate <sup>1,†</sup> , Vittoria Moretti <sup>2,†</sup> , Ester Pungolino <sup>3</sup>, Giovanni Micheloni <sup>2,\*</sup> , Roberto Valli <sup>2</sup> , Annalisa Frattini <sup>4</sup> , Matteo Barcella <sup>5</sup>, Francesco Acquati <sup>6,7</sup> , Rolland A Reinbold <sup>8</sup>, Lucy Costantino <sup>9</sup>, Fulvio Ferrara <sup>9</sup> , Alessandra Trojani <sup>3</sup> , Mario Ventura <sup>10,\*</sup> , Giovanni Porta <sup>2,‡</sup> and Roberto Cairoli <sup>3,‡</sup> 

- <sup>1</sup> Institute of Biomembranes, Bioenergetics, and Molecular Biotechnologies, National Research Council (IBIOM-CNR), 70125 Bari, Italy; a.labbate@ibiom.cnr.it
- <sup>2</sup> Genomic Medicine Research Center, Department of Medicine and Surgery, University of Insubria, Via JH Dunant 5 Varese, 21100 Varese, Italy; vmoretti@uninsubria.it (V.M.); roberto.valli@uninsubria.it (R.V.); giovanni.porta@uninsubria.it (G.P.)
- <sup>3</sup> Division of Hematology, ASST Grande Ospedale Metropolitano Niguarda, 20162 Milano, Italy; estermaria.pungolino@ospedaleniguarda.it (E.P.); alessandra.trojani@ospedaleniguarda.it (A.T.); roberto.cairoli@ospedaleniguarda.it (R.C.)
- <sup>4</sup> Genetics and Biomedical Research Institute, National Research Council (IRGB-CNR), 20090 Milano, Italy; annalisa.frattini@irgb.cnr.it
- <sup>5</sup> Department of Health Science, University degli Studi of Milan, Via Rudini 8, 20142 Milan, Italy; matteobarcella84@gmail.com
- <sup>6</sup> Department of Biotechnology and Life Science, University of Insubria, Via JH Dunant 3, 21100 Varese, Italy; francesco.acquati@uninsubria.it
- <sup>7</sup> Genomic Medicine Research Center, Department of Biotechnology and Life Science, University of Insubria, Via JH Dunant 3, 21100 Varese, Italy
- <sup>8</sup> Institute of Biomedical Technologies, National Research Council of Italy, 20054 Segrate, Milano, Italy; rollandreinbold@me.com
- <sup>9</sup> Department of Molecular Genetics, Centro Diagnostico Italiano, 20147 Milano, Italy; lucy.costantino@cdi.it (L.C.); fulvio.ferrara@cdi.it (F.F.)
- <sup>10</sup> Department of Biology, University of Bari 'Aldo Moro', Via Edoardo Orabona 4, 70124 Bari, Italy
- \* Correspondence: gmicheloni@uninsubria.it (G.M.); mario.ventura@uniba.it (M.V.); Tel.: +39-0332397106 (G.M.); +39-0805443583 (M.V.)
- † These authors contributed equally to the work.
- ‡ These authors also contributed equally to the work.



**Citation:** L'Abbate, A.; Moretti, V.; Pungolino, E.; Micheloni, G.; Valli, R.; Frattini, A.; Barcella, M.; Acquati, F.; Reinbold, R.A.; Costantino, L.; et al. Occurrence of L1M Elements in Chromosomal Rearrangements Associated to Chronic Myeloid Leukemia (CML): Insights from Patient-Specific Breakpoints Characterization. *Genes* **2023**, *14*, 1351. <https://doi.org/10.3390/genes14071351>

Academic Editor: Andrey Sudarikov

Received: 1 June 2023  
Revised: 19 June 2023  
Accepted: 22 June 2023  
Published: 27 June 2023



**Copyright:** © 2023 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/>).

**Abstract:** Chronic myeloid leukemia (CML) is a rare myeloproliferative disorder caused by the reciprocal translocation t(9;22)(q34;q11) in hematopoietic stem cells (HSCs). This chromosomal translocation results in the formation of an extra-short chromosome 22, called a Philadelphia chromosome (Ph), containing the *BCR-ABL1* fusion gene responsible for the expression of a constitutively active tyrosine kinase that causes uncontrolled growth and replication of leukemic cells. Mechanisms behind the formation of this chromosomal rearrangement are not well known, even if, as observed in tumors, repetitive DNA may be involved as core elements in chromosomal rearrangements. We have participated in the explorative investigations of the PhilosoPhi34 study to evaluate residual Ph+ cells in patients with negative FISH analysis on CD34+/lin- cells with gDNA qPCR. Using targeted next-generation deep sequencing strategies, we analyzed the genomic region around the t(9;22) translocations of 82 CML patients and one CML cell line and assessed the relevance of interspersed repeat elements at breakpoints (BP). We found a statistically higher presence of LINE elements, in particular belonging to the subfamily L1M, in BP cluster regions of both chromosome 22 and 9 compared to the whole human genome. These data suggest that L1M elements could be potential drivers of t(9;22) translocation leading to the generation of the *BCR-ABL1* chimeric gene and the expression of the active *BCR-ABL1*-controlled tyrosine kinase chimeric protein responsible for CML.

**Keywords:** LINE; L1M; chronic myeloid leukemia; rearrangements; BCR-ABL1

## 1. Introduction

Genomic instability is a hallmark of most types of cancer and leads to structural variants (SVs) including translocations, inversions, deletions, and duplications [1]. Genomic changes, ranging from simple translocations to complex karyotypes that arise from tangled phenomena such as chromothripsis, reshape cancer genomes and create de novo fusion genes with oncogenic properties [2]. Genomic rearrangements originate from DNA double strand breaks (DSBs) and misrepair through non-homologous end joining (NHEJ), which requires little or no homology, or homologous recombination (HR) [3].

Repeated DNA sequences are implicated as promoting elements of some of these recurrent chromosomal rearrangements [4]. The high density of repetitive DNA in some BP regions suggests that these sequences may provide “hot spots” for recombination and mediate translocation processes, increasing the likelihood of chromosomal rearrangement phenomena [4]. Four classes of repeated and interspersed sequences have been identified in the mammalian genome: DNA transposons, and LINE (long interspersed nuclear element), SINE (short interspersed nuclear element), and LTR (long terminal repeat) retrotransposons. The last three classes are the main represented categories of retrotransposons in mammals and constitute about 20%, 13%, and 8% of the human genome, respectively [5]. These elements, previously described as “junk DNA”, are now considered key components for the evolution and plasticity of the genome [5]. Their involvement in genetic dysfunctions is now evident, as they can trigger chromosomal rearrangements and cause mutations, resulting in the development of different pathologies including cancer [5]. The LINE’s most common family is L1, while Alu sequences are the most represented SINE elements. Both these elements are classified into subfamilies based on their sequence variants and evolution: L1 can be divided into L1M (mammalian-specific, oldest), L1P (primate-specific, intermediate), and L1H (human-specific, youngest) subfamilies; Alu elements are classified as *AluJ* (oldest), *AluS* (intermediate), and *AluY* (youngest) [6]. The recombination between Alu elements has been reported to be responsible for the partial duplication of the *MLL* gene in acute myeloid leukemia as well as in the generation of reciprocal translocations in tumors [7].

With the development of next generation sequencing (NGS) technologies (whole genome sequencing, exome, and targeted panels), a deeper understanding of genetic mechanisms involved in aberrant genetic alterations has become possible [8]. Limited efforts have been made to clarify molecular phenomena underlying the formation of such non-random genomic rearrangements, and little is known regarding the mechanisms responsible for a variety of translocations [9]. The precise knowledge of sequences surrounding BPs is fundamental to exploring genetic causes in patients with balanced translocations or inversion.

Chronic myeloid leukemia is a clonal myeloproliferative disorder characterized by the reciprocal translocation t(9;22)(q34;q11), involving the proto-oncogene *ABL1* (Abelson leukemia kinase proto-oncogene) on chromosome 9 and the *BCR* (breakpoint cluster region) on chromosome 22. The translocation generates the characteristic Ph chromosome and *BCR-ABL1* fusion gene [10], a constitutively active tyrosine kinase that promotes proliferation and survival of leukemic cells through the activation of downstream pathways such as RAS, RAF, JUN kinase, MYC, and STAT [11].

Three well-defined *BCR* BP regions in chromosome 22 have been characterized: the most frequent, named the Major BP cluster region (M-bcr), occurs between exons 12 and 16. CML patients show mostly M-bcr BPs resulting in a b2a2 or b3a2 transcript, containing *BCR* exon 13 (b2) or 14 (b3) and *ABL1* exon 2 (a2) [12]. In a small subset of patients, the region between exons 19 and 20 (micro-bcr,  $\mu$ -bcr) or the region distal to exon 1 (minor-bcr, m-bcr) in the *BCR* are involved. The BP region in *ABL1* includes a more variable and extended region of about 200 kb, from 10 kb upstream of the 5' of the gene to exon 2 [12]. These BP regions are associated with the production of p210, p230, and p190 *BCR-ABL* fusion protein variants, respectively.

Previously, in a cohort of 27 CML patients, we identified the presence of Alu at the *BCR-ABL1* BP junctions, underlying the evidence that repeated sequences may facilitate the pairing process and the resulting chromosomal translocation [9].

As participants in the exploratory investigations of the PhilosoPhi34 study that investigates the efficacy of nilotinib 300 mg BID in depleting bone marrow (BM) leukemic stem cells (CD34+/linPh+) in newly diagnosed chronic-phase (CP) CML patients at specific time points of treatment [13], we have analyzed a cohort of 82 CML patients with the goal to identify the patient-specific rearrangement of the *BCR/ABL1* in order to evaluate the minimal residual disease (MRD) in the Ph<sup>+</sup> CML patients after 6 or 12 months of nilotinib treatment. The aim of this study was to evaluate, through a gDNA-qPCR assay designed on patient-specific BP sequences, the number of Ph<sup>+</sup> residual cells [9,14–17].

We identified patients' BPs with the target enrichment approach and next generation sequencing (NGS) [14–17]. Moreover, we have included the KCL22 CML cell line as a positive control in our analyses. The analysis showed that *BCR-ABL1* junctions were clustered in 9 kbs on chromosome 22, mostly including the M-bcr BP cluster region, while on chromosome 9 they were clustered in a wider region of 154 kbs. Thus, we investigated the causes behind the generation of the Ph chromosome by analyzing sequences flanking each single breakpoint.

The analysis of repetitive elements in these regions showed a high presence of LINE elements belonging to the subfamily L1M. The comparison of L1M elements' frequency at the BP regions compared with their distribution throughout the entire human genome highlighted that L1M elements were much more recurrent at BP regions than in the whole genome. Therefore, we hypothesized, for the first time, the involvement of L1M in the translocation t(9;22)(q34;q11) and in the consequent creation of the Ph chromosome and *BCR-ABL1* fusion gene responsible for CML.

## 2. Materials and Methods

For any specific details of the analysis we performed, look at Appendix A.

### 2.1. Patient Samples

The PhilosoPhi34 study, which included 15 centers in Italy, collected bone marrow (BM) samples from 87 consecutive patients with CML on behalf of the Rete Ematologica Lombarda (REL). The PhilosoPhi34 study enrolled newly diagnosed Ph<sup>+</sup> CML patients in the chronic phase (CP-CML), aged  $\geq 18$  years, either male or female. BM samples were obtained and analyzed in accordance with the declaration of Helsinki, after written consent. We analyzed 82 CML patients of the PhilosoPhi34 study.

### 2.2. Selection of BM CD34+/lin- Cells

Mononuclear cells (MNCs) from bone marrow (BM) blood samples of the CML patients were isolated and BM CD34+/lin- cells were selected using a Diamond CD34 Isolation kit and an autoMACS Pro separator (Miltenyi Biotec, Bologna, Italy) according to the manufacturer's instructions (Miltenyi Biotec). Method details were described in "<http://dx.doi.org/10.17504/protocols.io.yncfvaw> (assessed on 19 July 2019)" and in our previous study [18].

### 2.3. CML Cell Line

An aliquot of the leukemic cell line KCL22 was kindly provided by Papa Giovanni XXIII Hospital (Bergamo, Italy).

### 2.4. Breakpoint Coordinates

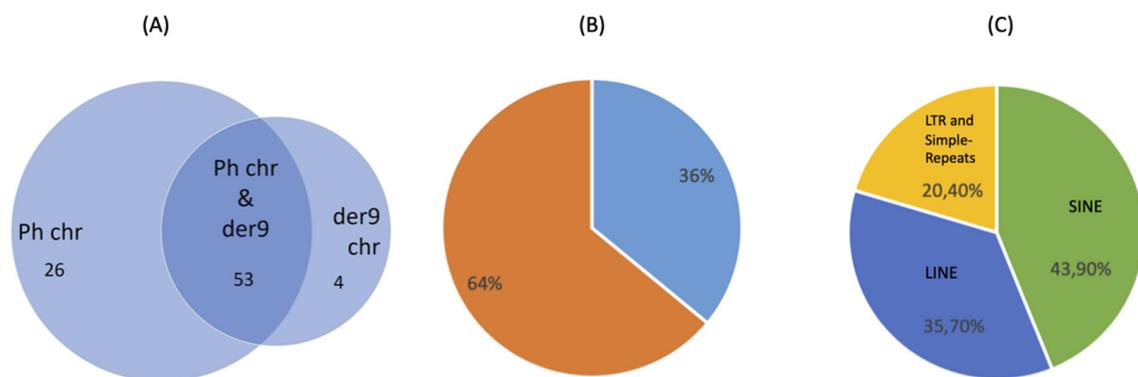
*BCR-ABL1* BPs were characterized through target enrichment and NGS on DNA from patients' CD34+/lin- cells and the cell line, using a SureSelectQXT Target Enrichment custom panel (Agilent Technologies SpA, Milano (MI), Italy) and paired-end sequencing technologies (HiSeq2500, Illumina, IGA technology services s.r.l., Udine (UD), Italy). Raw

reads were aligned to the human chromosome 9 and 22 sequences (GRCh38/hg38) using a BWA-MEM algorithm with default parameters. PCR duplicates were removed by means of the Picard MarkDuplicates tool version:2.23.9, "<http://broadinstitute.github.io/picard/>" (accessed on 1 May 2022)". Structural variations (SVs) identification and BPs refinement were performed using DELLY software (version 0.8.7). All the breakpoints' coordinates were submitted to GenBank (BankIt2713948: OR137812–OR137913).

### 3. Results

#### NGS and DELLY Analysis

Target enrichment and NGS characterization of 82 CML patients and one cell line showed peculiar BP DNA coordinates. DELLY analysis and visual inspection with the Integrative Genomics Viewer (IGV) identified a total number of 51556 SVs, among which 3248 were translocation events, in the 83 samples. We identified both Ph and reciprocal der9 chromosomes in 53 samples (106 translocations), whereas Ph alone was identified in 26 samples and der9 in four (Figure 1), resulting in 136 total translocation rearrangements and 272 BPs in total.



**Figure 1.** (A) Venn diagram shows the number of samples containing the Ph chromosome and/or the der9 chromosome. The overlapping area corresponds to the samples for which both chromosomes have been identified. (B) Graph represents the percentage of BPs mapped in repetitive elements (36%) and in non-repetitive elements (64%). (C) Graph indicates the distribution of repetitive elements in the BP cluster regions on chr22 and chr9.

DELLY reported 102 of 136 BP junction consensus sequences: 86 were characterized by micro-homologous regions (ranging from one to twenty nucleotides) and two BPs showed non-template insertion sequences of one base at the joined ends of the rearrangements (88%). The remaining 14 junction sequences were blunt (12%) (Tables 1 and S1).

Based on the BP coordinates, all samples showed a loss of genomic material at the BP regions either for chr9 or chr22, except for one case.

BPs on chromosome 22 spanned into a genomic region of almost 4.65 Mb (chr22:23289283–27941711). With the exception of one BP located 4.62 Mb distally to *BCR* (belonging to a der9 chromosome) and one BP mapping in the region between M-bcr and  $\mu$ -bcr, all BPs clustered within the M-bcr boundaries, with most of them mapping in the introns 13, 14, and 15 of the *BCR* gene (NM\_021574.3), and only seven BPs in exons (one in exon 13, three in exon 14, and three in exon 15). No BPs were identified inside the m-bcr and  $\mu$ -bcr. Almost all the BPs on chromosome 9 (130 of 136) mapped inside of intron 1 of the *ABL1* gene (NM\_007313.3) in a region of 137.7 Kbs (chr9:130715146–130852871). The remaining six BPs were located proximally to the *ABL1* gene (chr9:130662791–130710869), with one of them in 9p11.2 (chr9:41102675). With the exception of two, all the BPs were embedded into the “*ABL1*” enriched region.

**Table 1. Summary information associated with the characterized translocations.** The table reports the total number of translocations identified in the cohort and the number of consensus junction sequences identified with DELLY software. In detail, we have reported the percentage of consensus junction sequences characterized by micro-homologous regions or blunt ends at junction sites, as well as the distribution of BPs on chr9 and chr22, the percentage of those mapping in repetitive elements, and the type of repetitive elements in which they map. Finally, enriched L1M LINE elements in *ABL1* and M-bcr regions and their corresponding fold change values are indicated.

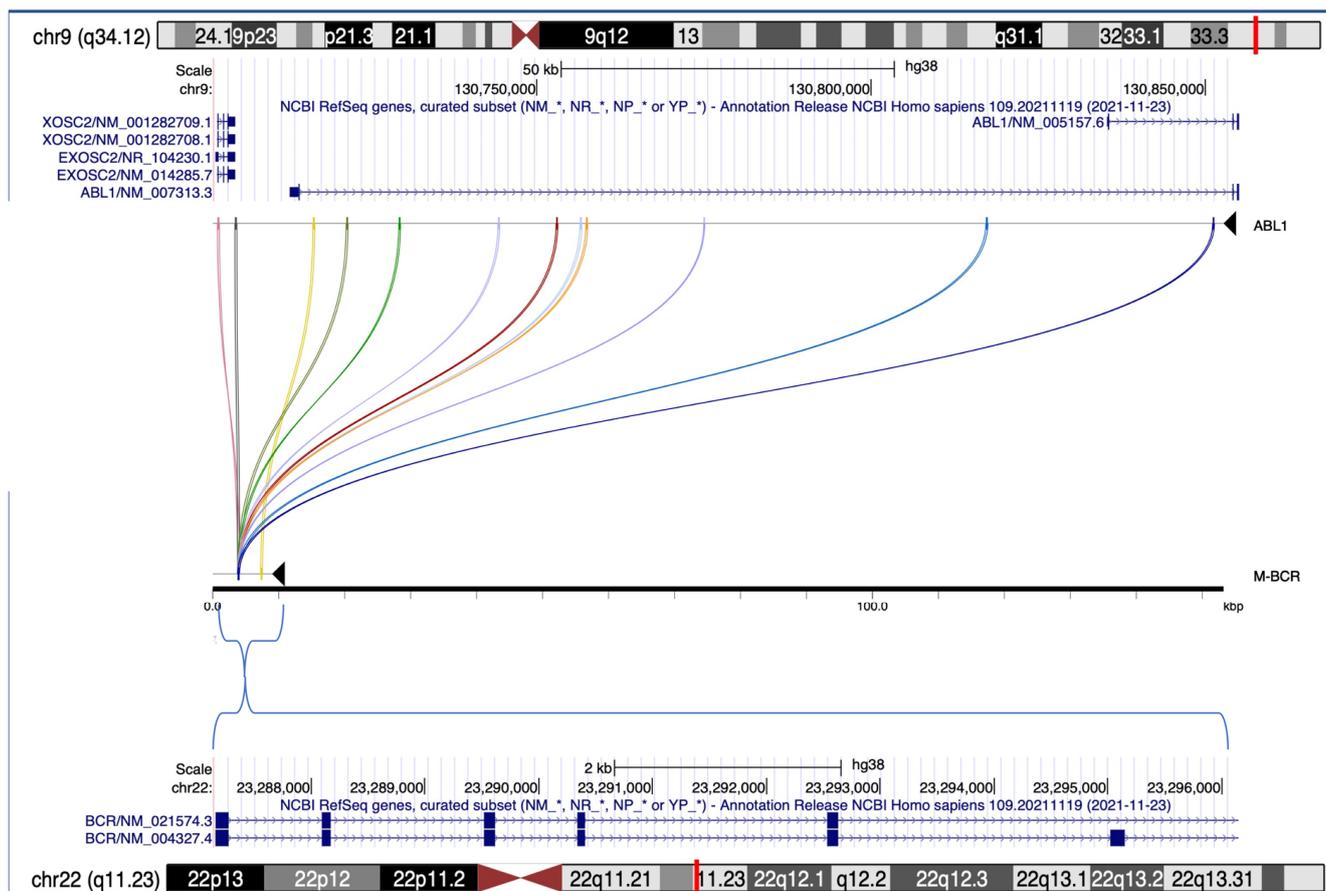
<b>Number of translocations</b>		136		
<b>Number of BP consensus sequences</b>		102		
		88% micro-homologies/non-template insertion		12% blunt
<b>BPs chr22</b>	M-bcr 134	m-bcr -	$\mu$ -bcr -	outside 2
<b>BPs chr9</b>	ABL1 region 128			outside 8
<b>36% BPs in repetitive elements</b>	SINE			43.9%
	LINE			35.7%
	other			20.4%
		FC <i>ABL1</i> /RI_154 kb_mean	FC M-bcr/RI_9 kb_mean	
<b>L1M LINE enrichment (FC<sup>a</sup> &gt; 2)</b>	L1ME1	9.5	10.5	
	L1MCc	12.4	-	
	L1MEc	10.6	-	
	L1MEf	8.5	-	
	L1Med	6.6	-	
	L1MC5	6.4	-	
	L1ME4a	4.7	-	
	L1MB4	4.5	-	
	L1M5	2.6	-	
	L1MC4	2.4	-	
	L1MC1	-	50.9	

Additionally, 36% of the total BPs (98/272) mapped into a repetitive element (Figure 1). The most represented classes were SINE (43.9%) and LINE (35.7%), followed by DNA transposons, LTR, and simple repeats (20.4% in total) (Figure 1).

Due to the BPs clustering in the M-bcr and *ABL1* regions, we evaluated the genomic content of these intervals of ~9 kbs and ~154 kbs, respectively. Both regions were characterized by the predominant presence of LINES and SINES. In detail, based on the UCSC's RepeatMasker annotation, the *ABL1* region contains 101 LINE elements and 75 of them are L1M\* type (Figure 2). Likewise, in the small M-bcr region, five LINES are annotated and three of them belong to two different L1M\* subfamilies (Figure 2).

The screening of the random intervals (RIs, see Materials and Methods) confirmed an enrichment of LINES as well as SINES in the regions of BP clustering in our samples. We paid more attention to the LINE elements due to the highest level of enrichment (Table 1). In detail, considering a FC value greater than two, the *ABL1* region was characterized by the enrichment of 13 different LINES: HAL1ME, L1MCc, L1MEc, L1ME1, L1MEf, L1Med, L1MC5, L1P4, L1ME4a, L1MB4, HAL1, L1M5, and L1MC4. With the exception of HAL1ME, L1P4, and HAL1, all elements are included in the L1M subfamily. Similarly, in M-bcr, we revealed an enrichment of three LINE elements; two of them are members of the L1M subfamily (L1MC1 and L1ME), and one of the L2b family. L1M copy number values in the samples' M-bcr and *ABL1* intervals resulted significantly higher than the corresponding L1M element mean values in RIs (Figure S1).





**Figure 3.** Pairwise alignment of the breakpoint cluster regions *ABL1* and *M-bcr*. On the top and the bottom are reported the chromosome 9 and 22 ideograms with the corresponding NCBI RefSeq genes content, respectively. In the middle, pairwise alignment shows the connection of the genomic blocks with a sequence similarity in *ABL1* (black line on the top) and *M-bcr* (black line on the bottom) by means of colored arches.

#### 4. Discussion

Transposons represent around 54% of the mammalian genome, of which almost 21% is represented by long interspersed nuclear elements (631,64 Mb) [19]. Most of these elements are silent and inactive, but 10–15% of them play an active role in the regulation of gene expression. In this study, we explored the role of transposable elements in the rearrangements of *BCR* and *ABL1*, investigating if a topological feature or a specific sequence close to these genes could promote their fusion. We characterized patient-specific BP sequences at chr9 and chr22 and we found that almost all 82 CML samples had the *M-bcr* (mostly in introns 13, 14, and 15). The analysis of regions spanning the BPs of ~9 kbs and ~154 kbs for chr22 and chr9, respectively, showed a predominant presence of SINEs and LINES. Indications that Alu elements may have a role in the generation of *BCR-ABL* have already been reported [20]. With a fine characterization of the breakpoints mapping, we narrowed the BPs regions involved in the t(9;22) in CML, and we assessed a high presence of LINE1 elements belonging to the L1M subfamily: 10 different L1M elements in the *ABL1* cluster region (L1MCc, L1MEc, L1ME1, L1MEf, L1Med, L1MC5, L1ME4a, L1MB4, L1M5, and L1MC4) and two L1M elements in the *M-bcr* cluster region (L1MC1 and L1ME). A previous study on a big cohort of samples was not able to identify consensus sequences around breakpoints but they analyzed only a small region across breakpoints [21], while in our study we analyzed a bigger region spanning all the breakpoints of our cohort.

According to the literature, a LINE1 insertion can affect the genome, epigenome, and the whole transcriptome of cells. Studies revealed that LINE1s increase gene mobilization

leading to the formation of chimeric genes generating new transcripts that could rise to chimeric proteins. An *in silico* study found 988 genes with LINE1 insertions that could generate chimeric transcripts, of which twenty have been associated with cancer [22].

The involvement of LINE1s in both gene deletion and chromosomal translocations was demonstrated by Rodriguez-Martinez and colleagues in an esophageal adenocarcinoma and head-and-neck and colorectal cancers. Authors reported that the aberrant L1 integrations can delete large regions of chromosomes, leading to the removal of tumor-suppressor genes and inducing complex translocations and large-scale duplications [23].

Repetitive elements are known to be heavily methylated in normal somatic tissues, but their methylation status is to a lesser extent in malignant tissues, driving the global genomic hypomethylation [24]. The cancer genome is frequently characterized by promoter hypomethylation of specific genes with an overall decrease at the level of five-methylcytosine. This hypomethylation affecting repeat sequences and transposable elements results in chromosomal instability and mutation events [25]. Previous studies demonstrated that hypomethylation of LINE1 promoters and other transposable elements appear at early stages of CML development. Roman-Gomez et al. hypothesized that destabilization of repetitive sequences (i.e., L1 hypomethylation) could be one of such mechanisms employed by *BCR-ABL* to generate genomic instability in the malignant cell, suggesting that repetitive DNA hypomethylation is closely associated with CML progression [26]. In addition, hypomethylation is also associated with open chromatin compartments, which turn out to be easily accessible. Engreitz et al. reported that translocations whose partners lie in the open chromatin regions are more significantly proximal than translocations with one or both partners in the closed compartments. By means of Hi-C experiments using a karyotypically normal lymphoblastoid cell line (GM06990), they demonstrated a significant contact frequency between the *BCR* and *ABL* loci [27]. Considering all these data, we could speculate that the higher density of hypomethylated L1M in these regions might be responsible for their nuclear proximity and open chromatin structure, thus allowing more frequent chromosomal rearrangements.

## 5. Conclusions

Our data highlighted the high distribution of L1M in *BCR* and *ABL1* gene regions, suggesting that L1M elements could be potential drivers of the t(9;22) translocation leading to the generation of the *BCR-ABL1* gene and the consequent expression of the *BCR-ABL1* active tyrosine kinase chimeric protein. Although it is not possible to confirm that L1M elements work as main actors in mediating the t(9;22) rearrangement, we can argue for a synergic role of theirs with the Alu elements, which can trigger the translocation rearrangement in association with the tendency of *BCR* and *ABL* loci to be spatially proximal [28].

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/genes14071351/s1>: Figure S1: L1M occurrence in random intervals; Table S1: List of breakpoints and associated features.

**Author Contributions:** Conceptualization: M.V., G.P. and R.C.; methodology: R.V., A.F. and M.V.; validation, M.B., L.C., F.F. and R.A.R.; formal analysis, A.L., R.V. and M.V.; investigation, R.V. and G.M.; resources, F.A., L.C. and F.F.; data curation, V.M., A.L. and M.B.; writing—original draft preparation, V.M., G.M. and A.L.; writing—review and editing, A.F., G.P., E.P. and R.C.; visualization, F.A. and A.T.; supervision, R.V. and E.P.; project administration, M.V. and G.P.; funding acquisition, A.T., E.P. and R.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee at ASST Grande Ospedale Metropolitano Niguarda (Milan, Italy) and the following local Ethics Committees of the participants' centers (Lombardia, Italy): EC ASST Spedali Civili Brescia, EC Desio Hospital, EC IRCCS Policlinico San Matteo (Pavia), EC Valduce Hospital (Como), EC Monza Brianza, EC IRCCS Ca' Granda Ospedale Maggiore Policlinico (Milan), EC San Raffaele Scientific Institute (Milan), EC Sacco Hospital (Milan), EC IRCCS

Istituto Nazionale dei Tumori (Milan), EC Valle Olona Ospedale di Circolo (Busto Arsizio), EC ASST Valle Olona Sant'Antonio Abate (Gallarate), EC Hospital of Cremona, and EC ASST Lecco (EudraCT: 2012-005062-34, approved 14 December 2012).

**Informed Consent Statement:** BM samples were obtained and analyzed in accordance with the Declaration of Helsinki, after written consent was obtained from patients.

**Data Availability Statement:** WGS data are available at the NCBI Sequence Read Archive "<https://www.ncbi.nlm.nih.gov/sra> (accessed on 29 June 2022)" under accession number PRJNA853897.

**Acknowledgments:** The authors would like to thank Novartis Pharma SpA Milan, Italy, for providing partial financial support, and the participants of the PhilosoPhi34 study.

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

## Appendix A.

### Appendix A.1. Patient Material

All protocols were approved by the Ethics Committee at ASST Grande Ospedale Metropolitano Niguarda (Milan, Italy) and the local Ethics Committees of the participants' centers (Lombardia, Italy) [29,30]. In this study, we enrolled 82 chronic phase (CP)-CML patients. Patients received first-line therapy with 300 mg BID Nilotinib (Tasigna<sup>®</sup>, Novartis Farma spA, Origgio (VA), Italy). Genomic DNA was extracted from peripheral blood (PB) or bone marrow (BM) samples, collected at the diagnosis using a DNeasy Blood & Tissue Kit (Qiagen s.r.l., Milan (MI), Italy) following the manufacturer's instructions. The extracted nucleic acids were quantified with the NanoDrop ND1000 spectrophotometer.

### Appendix A.2. CML Cell Line

The cell culture showed suspension cells arranged in small clusters; cells contained the t(9;22) translocation, leading to the *BCR-ABL1* e13-a2 (b2a2) fusion gene (American Type Culture Collection<sup>®</sup>, Manassas, Virginia, USA). The KCL22 cell line was sub-cultured in RPMI 1640 culture medium (Euroclone SpA, Pero (MI), Italy) supplemented with 20% heat-inactivated fetal bovine serum (FBS) (Euroclone SpA, Pero (MI), Italy), 1% L-glutamine (Euroclone SpA, Pero (MI), Italy), and 1% penicillin/streptomycin (Euroclone SpA, Pero (MI), Italy). Genomic DNA was extracted from cells using a DNeasy Blood & Tissue Kit (Qiagen s.r.l., Milan (MI), Italy) following the manufacturer's instructions. The extracted nucleic acids were quantified with the NanoDrop ND1000 spectrophotometer.

### Appendix A.3. Breakpoint Coordinates

Genomic DNA was enzymatically fragmented, and hybridization was performed with RNA probes spanning a 250 kb region on chromosomes 22 and 9 generally involved in translocation. The DNA fragments were captured with biotin-conjugated probes and selected with streptavidin-bound magnetic beads. The bead-probe-DNA complex was captured with a magnet, resulting in enrichment in *BCR-ABL1* BP regions. After beads removal and RNA digestion, libraries were obtained from enriched DNA fragments, sequenced, and BPs localization were identified at the single nucleotide level (GRCh38/hg38).

### Appendix A.4. NGS Data Analysis

Raw reads were aligned to the human chromosome 9 and 22 sequences (GRCh38/hg38) using the BWA-MEM algorithm with default parameters [31]. PCR duplicates were removed by means of the Picard MarkDuplicates tool version:2.23.9, "<http://broadinstitute.github.io/picard/> (accessed on 1 May 2022)" Structural variations (SVs) identification and BPs refinement were performed using DELLY software (version 0.8.7) [32]. In addition, all the identified SVs were checked by visual inspection with the Integrative Genomics Viewer (IGV), which we also used in samples for which DELLY failed in calling SVs [33]. We filtered out all SVs mapping outside of the enriched regions, and we focused the analysis on those involved in the formation of the Ph chromosomes and derivative chromosome 9.

We analyzed the genomic composition and structure of the regions enriched with M-bcr (chr22:23287159-23296156, hg38-based) and *ABL1* (chr9:130701881-130855096, hg38-based). In detail, we screened such regions for the presence of interspersed repeats and low complexity DNA sequences by intersecting them with the RepeatMasker track from the UCSC Genome Browser (<http://genome-euro.ucsc.edu/cgi-bin/hgTables>). In order to evaluate a potential enrichment of the repetitive elements enclosed in the region of the BPs clustering, we generated random intervals (RIs) of the same size as the M-bcr (~9 kbs) and *ABL1* (~154 kbs) regions, using bedtools' *random* (v2.29.0) [34]. Specifically, we intersected a total number of 170,000 and 10,000 RIs of 9 kbs and 154 kbs, respectively, covering almost half of the human genome, with the RepeatMasker track coordinates. We considered as enriched the repetitive elements in the intervals of interest with a fold change (FC) more than 2. FC was calculated as the ratio of copy number value of the element in the region of interest and the mean value of the same element in the RIs. Assuming the occurrence distribution of the repetitive elements in RIs is a Poisson distribution, we calculated the repetitive elements' mean occurrence '*l*' per intervals in the RIs and then the probability of having '*x*' occurrences (corresponding to the repeat element value of interest) in M-bcr and *ABL1* regions using the '*ppois*' R function. We considered as significant the elements in M-bcr and *ABL1* enriched regions with a *p*-value < 0.05.

#### Appendix A.5. Miropeats

To graphically assess the sequence similarity of the regions involved in the translocation events, we performed a pairwise alignment of the BP cluster regions M-bcr and *ABL1* using minimap2 aligner [35], with the following parameters: -x asm5 -B 2 -s 20 -p 0.01 --cs. The pairwise alignment format file was converted to a PostScript file and then annotated with NCBI RefSeq genes.

## References

1. Forbes, S.A.; Beare, D.; Gunasekaran, P.; Leung, K.; Bindal, N.; Boutselakis, H.; Ding, M.; Bamford, S.; Cole, C.; Ward, S.; et al. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **2015**, *43*, D805–D811. [[CrossRef](#)]
2. L'abbate, A.; Cunsolo, C.L.; Macrì, E.; Iuzzolino, P.; Mecucci, C.; Doglioni, C.; Coco, M.; Muscarella, L.A.; Salati, S.; Tagliafico, E.; et al. FOXP1 and TP63 involvement in the progression of myelodysplastic syndrome with 5q- and additional cytogenetic abnormalities. *BMC Cancer* **2014**, *14*, 396. [[CrossRef](#)] [[PubMed](#)]
3. Van Gent, D.C.; Hoeijmakers, J.H.; Kanaar, R. Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* **2001**, *2*, 196–206. [[CrossRef](#)] [[PubMed](#)]
4. Kolomietz, E.; Meyn, M.S.; Pandita, A.; Squire, J.A. The role of Alu repeats clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes Chromosomes Cancer* **2002**, *35*, 97–112. [[CrossRef](#)] [[PubMed](#)]
5. Chenais, B. Transposable elements in cancer and other human diseases. *Curr. Cancer Drug Targets* **2015**, *15*, 227–242. [[CrossRef](#)]
6. Price, A.L.; Eskin, E.; Pevzner, P.A. Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. *Genome Res.* **2004**, *14*, 2245–2252. [[CrossRef](#)]
7. Onno, M.; Nakamura, T.; Hillova, J.; Hill, M. Rearrangement of the human *trc* oncogene by homologous recombination between Alu repeats of nucleotide sequences from two different chromosomes. *Oncogene* **1992**, *7*, 2519–2523.
8. Peiris, M.N.; Li, F.; Donoghue, D.J. BCR: A promiscuous fusion partner in hematopoietic disorders. *Oncotarget* **2019**, *10*, 2738–2754. [[CrossRef](#)]
9. Mattarucchi, E.; Guerini, V.; Rambaldi, A.; Campiotti, L.; Venco, A.; Pasquali, F.; Lo Curto, F.; Porta, G. Microhomologies and Interspersed Repeat Elements at Genomic Breakpoints in Chronic Myeloid Leukemia. *Cancer* **2008**, *301*, 288–301. [[CrossRef](#)]
10. Baikie, A.G.; Court-BRWM; Buckton, K.E.; Harnden, D.G.; Jakobs, P.A.; Tough, I.M. A possible specific chromosome abnormality in human chronic myeloid leukaemia. *Nature* **1960**, *31*, 1165–1166. [[CrossRef](#)]
11. Carlesso, N.; Frank, D.A.; Griffin, J.D. Tyrosyl phosphorylation and DNA binding activity of signal transducers and activators of transcription (STAT) proteins in hematopoietic cell lines transformed by Bcr/Abl. *J. Exp. Med.* **1996**, *183*, 811–820. [[CrossRef](#)] [[PubMed](#)]
12. Melo, J.V. BCR-ABL gene variants. *Baillière's Clin. Haematol.* **1997**, *10*, 203–222. [[CrossRef](#)]
13. Pungolino, E.; D'adda, M.; De Canal, G.; Trojani, A.; Perego, A.; Elena, C.; Lunghi, F.; Turrini, M.; Borin, L.; Iurlo, A.; et al. Nilotinib-induced bone marrow CD34+/lin-Ph+ cells early clearance in newly diagnosed CP-Chronic Myeloid Leukemia: Final report of the PhilosoPhi34 study. *Eur. J. Haematol.* **2021**, *107*, 436–448. [[CrossRef](#)] [[PubMed](#)]

14. Mattarucchi, E.; Spinelli, O.; Rambaldi, A.; Pasquali, F.; Curto, F.L.; Campiotti, L.; Porta, G. Molecular monitoring of residual disease in chronic myeloid leukemia by genomic DNA compared with conventional mRNA analysis. *J. Mol. Diagn.* **2009**, *11*, 482–487. [[CrossRef](#)]
15. Pagani, I.S.; Spinelli, O.; Mattarucchi, E.; Pirrone, C.; Pigni, D.; Amelotti, E.; Lilliu, S.; Boroni, C.; Intermesoli, T.; Giussani, U.; et al. Genomic quantitative real-time PCR proves residual disease positivity in more than 30% samples with negative mRNA-based qRT-PCR in Chronic Myeloid Leukemia. *Oncoscience* **2014**, *1*, 510–521. [[CrossRef](#)]
16. Porta, G.; Pagani, I.S.; Pirrone, C. gDNA Q-PCR for clinical monitoring of CML. *Cell Cycle* **2015**, *14*, 3659–3660. [[CrossRef](#)]
17. Rainero, A.; Angaroni, F.; D’Avila, F.; Conti, A.; Pirrone, C.; Micheloni, G.; Tararà, L.; Millefanti, G.; Maserati, E.; Valli, R.; et al. GDNA qPCR is statistically more reliable than mRNA analysis in detecting leukemic cells to monitor CML. *Cell Death Dis.* **2018**, *9*, 349. [[CrossRef](#)]
18. Trojani, A.; Pungolino, E.; Rossi, G.; D’Adda, M.; Lodola, M.; Camillo, B.D.; Perego, A.; Turrini, M.; Orlandi, E.; Borin, L.; et al. Wide-transcriptome analysis and cellularity of bone marrow CD34+/lin- cells of patients with chronic-phase chronic myeloid leukemia at diagnosis vs. 12 months of first-line nilotinib treatment. *Cancer Biomark.* **2017**, *21*, 41–53. [[CrossRef](#)] [[PubMed](#)]
19. Pradhan, R.K.; Ramakrishna, W. Transposons: Unexpected players in cancer. *Gene* **2022**, *808*, 145975. [[CrossRef](#)] [[PubMed](#)]
20. Jeffs, A.R.; Benjes, S.M.; Smith, T.L.; Sowerby, S.J.; Morris, C.M. The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. *Hum. Mol. Genet.* **1998**, *7*, 767–776. [[CrossRef](#)]
21. Ross, D.M.; O’Hely, M.; Bartley, P.A.; Dang, P.; Score, J.; Goyne, J.M.; Sobrinho-Simoes, M.; Cross, N.C.; Melo, J.V.; Speed, T.P.; et al. Distribution of genomic breakpoints in chronic myeloid leukemia: Analysis of 308 patients. *Leukemia* **2013**, *27*, 2105–2107. [[CrossRef](#)] [[PubMed](#)]
22. Criscione, S.W.; Theodosakis, N.; Micevic, G.; Cornish, T.C.; Burns, K.H.; Neretti, N.; Rodić, N. Genome-wide characterization of human L1 antisense promoter-driven transcripts. *BMC Genom.* **2016**, *17*, 463. [[CrossRef](#)] [[PubMed](#)]
23. Rodriguez-Martin, B.; Alvarez, E.G.; Baez-Ortega, A.; Zamora, J.; Supek, F.; Demeulemeester, J.; Santamarina, M.; Ju, Y.S.; Temes, J.; Garcia-Souto, D.; et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.* **2020**, *52*, 306–319. [[CrossRef](#)]
24. Ehrlich, M. DNA methylation in cancer: Too much, but also too little. *Oncogene* **2002**, *21*, 5400–5413. [[CrossRef](#)]
25. Eden, A.; Gaudet, F.; Waghmare, A.; Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **2003**, *300*, 455. [[CrossRef](#)] [[PubMed](#)]
26. Roman-Gomez, J.; Jimenez-Velasco, A.; Agirre, X.; Castillejo, J.A.; Navarro, G.; San Jose-Eneriz, E.; Garate, L.; Cordeu, L.; Cervantes, F.; Prosper, F.; et al. Repetitive DNA hypomethylation in the advanced phase of chronic myeloid leukemia. *Leuk. Res.* **2008**, *32*, 487–490. [[CrossRef](#)]
27. Engreitz, J.M.; Agarwala, V.; Mirny, L.A. Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS ONE* **2012**, *7*, e44196. [[CrossRef](#)]
28. Albano, F.; Anelli, L.; Zagaria, A.; Coccaro, N.; D’Addabbo, P.; Liso, V.; Rocchi, M.; Specchia, G. Genomic segmental duplications on the basis of the t(9;22) rearrangement in chronic myeloid leukemia. *Oncogene* **2010**, *29*, 2509–2516. [[CrossRef](#)]
29. Trojani, A.; Pungolino, E.; Dal Molin, A.; Lodola, M.; Rossi, G.; D’Adda, M.; Perego, A.; Elena, C.; Turrini, M.; Borin, L.; et al. Nilotinib interferes with cell cycle, ABC transporters and JAK-STAT signaling pathway in CD34+/lin- cells of patients with chronic phase chronic myeloid leukemia after 12 months of treatment. *PLoS ONE* **2019**, *14*, e0218444. [[CrossRef](#)]
30. Pungolino, E.; Rossi, G.; De Canal, G.; Trojani, A.; D’adda, M.; Perego, A.; Orlandi, E.M.; Lunghi, F.; Turrini, M.; Borin, L.; et al. Nilotinib induced bone marrow CD34+/lin-Ph+ cells early clearance in newly diagnosed CP-chronic myeloid leukemia. *Am. J. Hematol.* **2018**, *93*, E162–E164. [[CrossRef](#)]
31. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
32. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A.M.; Benes, V.; Korbel, J.O. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **2012**, *28*, i333–i339. [[CrossRef](#)] [[PubMed](#)]
33. 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](#)]
34. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)] [[PubMed](#)]
35. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]

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