



Article

# A Novel Tool for the Analysis and Detection of Copy Number Variants Associated with Haemoglobinopathies

Anna Minaïdou <sup>1</sup>, Stella Tamana <sup>1</sup> , Coralea Stephanou <sup>1</sup> , Maria Xenophontos <sup>1</sup> , Cornelis L. Harteveld <sup>2</sup>, Celeste Bento <sup>3</sup>, Marina Kleanthous <sup>1,†</sup> and Petros Kountouris <sup>1,\*,†</sup>

<sup>1</sup> Molecular Genetics Thalassemia Department, The Cyprus Institute of Neurology and Genetics, Nicosia 2371, Cyprus

<sup>2</sup> Human and Clinical Genetics Department, Leiden University Medical Center, P.O. Box 9600, 2333 ZC Leiden, The Netherlands

<sup>3</sup> Department of Haematology, Centro Hospitalar e Universitário de Coimbra, 3000-075 Coimbra, Portugal

\* Correspondence: petrosk@cing.ac.cy; Tel.: +357-22392623

† These authors contributed equally to the manuscript.

**Abstract:** Several types of haemoglobinopathies are caused by copy number variants (CNVs). While diagnosis is often based on haematological and biochemical parameters, a definitive diagnosis requires molecular DNA analysis. In some cases, the molecular characterisation of large deletions/duplications is challenging and inconclusive and often requires the use of specific diagnostic procedures, such as multiplex ligation-dependent probe amplification (MLPA). Herein, we collected and comprehensively analysed all known CNVs associated with haemoglobinopathies. The dataset of 291 CNVs was retrieved from the IthaGenes database and was further manually annotated to specify genomic locations, breakpoints and MLPA probes relevant for each CNV. We developed IthaCNVs, a publicly available and easy-to-use online tool that can facilitate the diagnosis of rare and diagnostically challenging haemoglobinopathy cases attributed to CNVs. Importantly, it facilitates the filtering of available entries based on the type of breakpoint information, on specific chromosomal and locus positions, on MLPA probes, and on affected gene(s). IthaCNVs brings together manually curated information about CNV genomic locations, functional effects, and information that can facilitate CNV characterisation through MLPA. It can help laboratory staff and clinicians confirm suspected diagnosis of CNVs based on molecular DNA screening and analysis.

**Keywords:** haemoglobinopathies; copy number variants (CNVs); MLPA



**Citation:** Minaïdou, A.; Tamana, S.; Stephanou, C.; Xenophontos, M.; Harteveld, C.L.; Bento, C.; Kleanthous, M.; Kountouris, P. A Novel Tool for the Analysis and Detection of Copy Number Variants Associated with Haemoglobinopathies. *Int. J. Mol. Sci.* **2022**, *23*, 15920. <https://doi.org/10.3390/ijms232415920>

Academic Editor: Richard Jäger

Received: 17 November 2022

Accepted: 8 December 2022

Published: 14 December 2022

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



**Copyright:** © 2022 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/>).

## 1. Introduction

Haemoglobinopathies are the most common monogenic diseases, with millions of carriers and patients worldwide [1]. They are caused by pathogenic variants in the globin gene clusters, which either reduce or abolish the synthesis of globin chains leading to thalassaemia syndromes, or alter the haemoglobin structure leading to structural haemoglobinopathies [2,3]. Over 2300 disease-causing variants have been reported in the IthaGenes database [4], with their varying severity affecting the clinical manifestations and survival of haemoglobinopathy patients [5].

While most known variants are substitutions and short indels, 291 copy number variants (CNVs) associated with haemoglobinopathies have been reported to date, according to the IthaGenes database. In addition to haemoglobinopathies, the presence of CNVs has been associated with genomic variability and many other genetic disorders. CNVs, mainly large deletions and duplications, within the  $\beta$ -globin gene cluster are relatively rare, causing approximately 10% of all  $\beta$ -thalassaemia cases [6]. CNVs leading to  $\beta$ -haemoglobinopathies involve *HBB* and/or the upstream locus control region  $\beta$ LCR [7]. In contrast to  $\beta$ -haemoglobinopathies, approximately 80–90% of the  $\alpha$ -thalassaemia cases are triggered by large genomic deletions/duplications due to meiotic unequal crossovers

between misaligned and highly homologous sequence repeats of the  $\alpha$ -globin locus [8]. CNVs associated with  $\alpha$ -thalassaemia phenotypes affect the  $\alpha$ -globin genes and/or the major regulatory element HS-40 of the  $\alpha$ -globin locus on chromosome 16 [6,8].

Although asymptomatic carriers and mildly affected patients of the extensive deletional forms of haemoglobinopathies are mainly identified with routine laboratory testing based on haematological and biochemical parameters, their definitive diagnosis requires molecular DNA analysis. Nevertheless, the molecular detection and characterisation of CNVs are challenging and often inconclusive using conventional methods, such as gap-PCR, reverse dot blot (RDB), fluorescent in situ hybridization (FISH) analysis, and Southern blot analysis [6,8]. Alternatively, multiplex ligation-dependent probe amplification (MLPA), in combination with other specific PCR-based methods, has been widely and effectively employed to identify and characterise known, rare, and novel CNVs associated with haemoglobinopathies [6]. The MLPA method can detect large or medium size deletions/duplications, down to a few hundred bases or a single exon. However, data about breakpoints are neither widely available nor easily accessible for all cases and, because of the limitations of the MLPA analysis, the boundaries of several large CNVs are still unclear. The exact genomic breakpoints of a deletion/duplication, following MLPA analysis, require further investigation for complete characterisation, and this is a challenging procedure due to the repetitive regions and highlights of the  $\alpha$ - and  $\beta$ -globin loci [9]. However, MLPA is an important tool for grouping chromosomal rearrangements and producing a picture of the known or novel CNVs according to the size and affected genes, including regulatory elements. Accurate identification and structural analysis of CNVs are necessary for the clinical diagnosis of thalassaemia syndromes [10]. Available information on the exact location and effect of previously reported CNVs is often fragmented in the scientific literature or incomplete in different databases, thus further impeding the characterisation and molecular diagnosis of such challenging cases. Consequently, it is often unclear whether several independently reported CNVs are identical or, in fact, novel, while their characterisation through MLPA is challenging due to the lack of knowledge about the exact breakpoints being affected.

Herein, we compile and analyse all the available evidence related to CNVs associated with haemoglobinopathies and we develop an open web tool to facilitate the identification and diagnosis of haemoglobinopathy-specific CNVs. IthaCNVs brings together the available MLPA information and manually curated information available on the ITHANET portal, while providing a user-friendly interface for filtering CNVs based on specific MLPA probes, breakpoints, and affected globin genes and/or regulatory elements.

## 2. Results

Figure 1 illustrates the main functionalities of IthaCNVs, which organises CNV information and facilitates CNV filtering in the following four different ways: (a) according to type of breakpoint information (i.e., the exact location or a range of coordinates harbouring the breakpoint), (b) according to specific locus or chromosomal positions, (c) in relation to MLPA probes, and (d) according to affected gene(s). Initial filtering based on one of four options produces a table with all relevant CNV entries accompanied with further filters acting either on the coordinates or the MLPA probes of the 5' and 3' breakpoints, or the affected genes. By expanding the filtering results row, the individual locus and chromosomal breakpoint positions can be displayed. Filtering using the MLPA probes is always based on the chromosomal location from 5' to 3' breakpoints, which is the same order that is shown in the ratio chart for the MLPA analysis.

Filter table by: Location

Filter Range by:  $\alpha$ -locus

Query IthaCNVs for  $-(\alpha)20.5$  for example and explore the type of data it has to offer filtering by locus location (using ranging position filtering).

**5' Breakpoint**

**3' Breakpoint**

**Exact position filtering:** use the same breakpoint position in From and To.

From: e.g. 70068

From: e.g. 70068

To: e.g. 70068

To: e.g. 70068

☐ Out of locus
 ☐ Out of locus

Locations of the breakpoints and MLPA probes are based on GRCh38.p13.  
 The SALSA MLPA Probemix P102-D1 HBB and P140-C1 HBA kit were used.  
 A tool for converting genome coordinates and genome annotation files between assemblies can be found here.  
 The individual locus and chromosome breakpoint positions can be viewed by expanding each filtering results row.

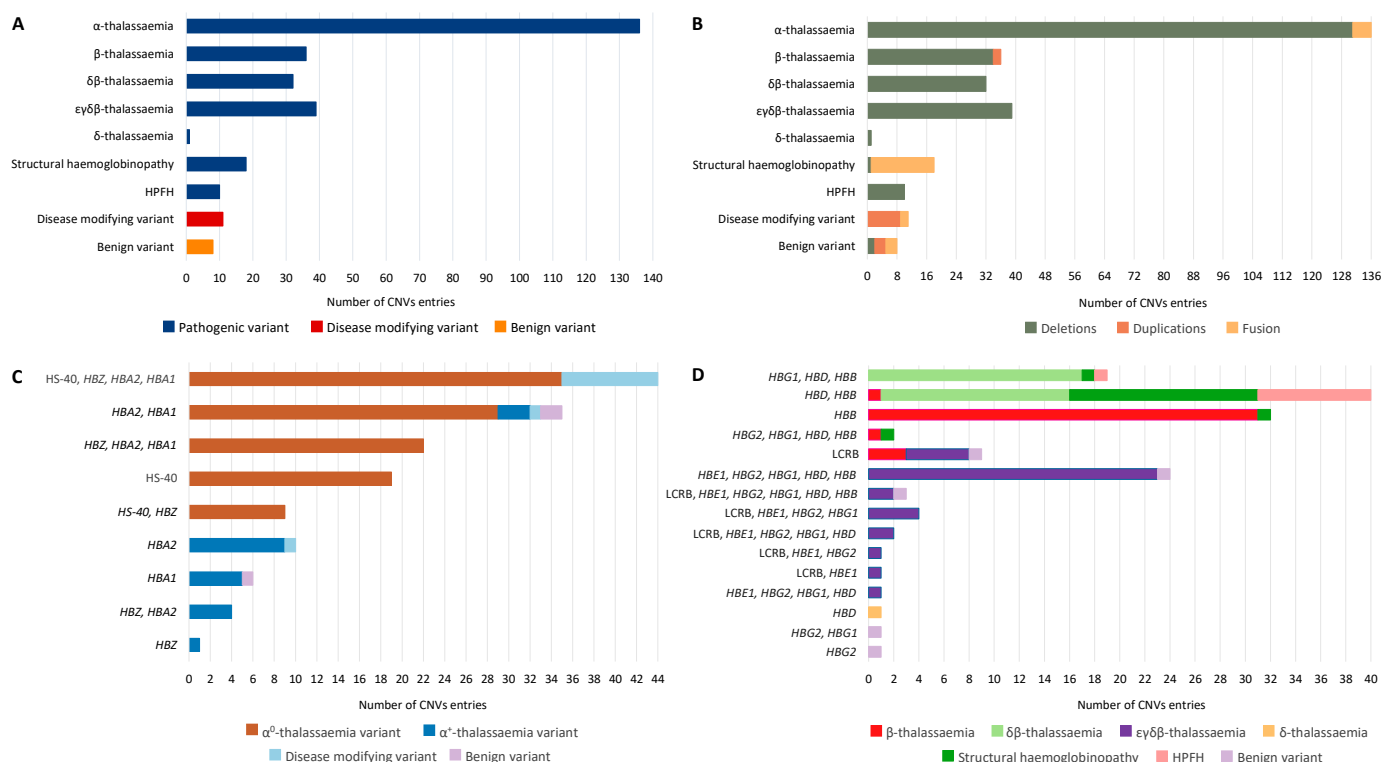
Showcase example for normal and affected probes

Print
 Search:

IthaID	Common name	Genes	Type	MLPA Probes				Type
				Last normal	1 <sup>st</sup> affected	Last affected	1 <sup>st</sup> normal	
300	$\alpha$ 3.7 (type I)	$\alpha$ 3.7 hybrid, HBA1, HBA2	Exact	HBA1&2-1 (214nt)	HBA2-intr.2 (160nt)	HBA1-intr.2 (165nt)	HBA1-intr.2 (250nt)	Exact
<b>HGVS name:</b> NG_000006.1:g.34247_38050del <b>Functionality:</b> Causative <b>Chromosome:</b> 16 <b>Locus:</b> $\alpha$ -locus			<b>Locus 5' BP</b> <b>From:</b> 34247 <b>To:</b> 38050		<b>Locus 3' BP</b> <b>From:</b> 38050 <b>To:</b> 38050			
			<b>Chromosome 5' BP</b> <b>From:</b> 173384 <b>To:</b> 173384		<b>Chromosome 3' BP</b> <b>From:</b> 177187 <b>To:</b> 177187			
312	--MED I	HBA1, HBA2	Range	HBZP1 region-down (292nt)	HBM region-up (184nt)	HBA1-down (310nt)	HBQ1-3 (400nt)	Range
<b>HGVS name:</b> NG_000006.1:g.(23641_23662)_(41183_41203)del <b>Functionality:</b> Causative <b>Chromosome:</b> 16 <b>Locus:</b> $\alpha$ -locus			<b>Locus 5' BP</b> <b>From:</b> 23641 <b>To:</b> 23662		<b>Locus 3' BP</b> <b>From:</b> 41183 <b>To:</b> 41203			
			<b>Chromosome 5' BP</b> <b>From:</b> 162778 <b>To:</b> 162799		<b>Chromosome 3' BP</b> <b>From:</b> 180320 <b>To:</b> 180340			
316	$-(\alpha)5.2$	HBA1, HBA2	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
<b>HGVS name:</b> N/A <b>Functionality:</b> Causative			<b>Locus 5' BP</b>		<b>Locus 3' BP</b>			
Information unclear								

Figure 1. Home page of the IthaCNVs tool. Filtering is based on the CNV location.

Figure 2A illustrates the distribution of CNVs by according to their associated disease groups/phenotypes, and variant functionality. The dataset comprises 291 CNVs, including 272 pathogenic (93.47%), 11 disease-modifying (3.78%) and 8 benign variants (2.75%). Of the pathogenic CNVs, 136 (50%) are associated with  $\alpha$ -thalassaemia, 39 with  $\epsilon\gamma\delta\beta$ -thalassaemia (14.34%), 36 with  $\beta$ -thalassaemia (13.24%) and 32 with  $\delta\beta$ -thalassaemia (11.76%). In rare cases, pathogenic CNVs are associated with a structural haemoglobinopathy (6.62%, 18/272), hereditary persistence of foetal haemoglobin—HPFH (3.68%, 10/272) and  $\delta$ -thalassaemia (0.37%, 1/272). Notably, one CNV (Hb Jambol; IthaID: 209) can be categorised into two haemoglobinopathy subgroups as the resulting hyperunstable Hb is also characterised by the thalassaemia phenotype [11].



**Figure 2.** Descriptive analysis of the CNV dataset. (A) Haemoglobinopathy groups/phenotypes and variant functionality, (B) Haemoglobinopathy groups/phenotypes and variant type, (C) Affected genes associated with  $\alpha$ -globin locus variants, (D) Affected genes associated with  $\beta$ -globin locus variants.

Figure 2B shows the distribution of CNVs based on the variant type and associated disease groups/phenotypes. The CNV dataset contains mainly large deletions (85.91%, 250/291) associated predominantly with thalassaemia phenotypes and HPFH. Of the pathogenic variants, two duplication CNVs are associated with  $\beta$ -thalassaemia (CD 62–83 (+65 bp), IthaID: 2190; >29.5 kb duplication, IthaID: 3869), while five CNVs are associated with  $\alpha$ -thalassaemia caused by a fusion of *HBA1-HBA2*. Structural haemoglobinopathies are generally associated with the fusion of genes resulting from either a deletion or duplication of a large region within the  $\beta$ -globin locus. Fusion between the different  $\alpha$ - or  $\beta$ -globin genes is usually created by an unequal crossing over from the misaligned globin genes, resulting in structurally abnormal haemoglobins [12]. Disease-modifying variants are caused by the duplication and fusion of genes only affecting the  $\alpha$ -globin locus (Figure 2C), while benign variants are related with both  $\alpha$ - and  $\beta$ -globin loci (Figure 2D).

Importantly, the exact breakpoints on either one or both ends of 26.46% (77/291) of CNVs remain unclear. Because CNVs were either identified by RFLP analysis and/or only recently analysed by MLPA, the precise breakpoints were not captured. Therefore, the 3' and 5' breakpoints of these CNVs are characterised as a range, since they have not

been sequenced. For these CNVs, GAP-PCR and Sanger sequencing can be performed for the exact specification of the breakpoints. In some cases, the 3' or the 5' end of the CNVs is not defined because of the limitation of the technique used. A common problem in characterising CNVs affecting the globin gene loci is that one or both breakpoints fall out of the range of MLPA probes. In cases where no additional techniques were used, the exact breakpoints beyond the probe locations remain unclear. When breakpoints extend beyond the MLPA probes, the size and approximate breakpoint localisation can be determined by performing an array CGH analysis [13].

Furthermore, and despite rigorous research, 9.27% (27/291) of all the CNVs presented in this study have unknown breakpoints due to limited information. Since some of the CNVs have only been reported in a single case study or have been identified in the context of other diseases, there is limited information available to clarify their breakpoints. The information provided reflects only the name of the deleted genes or the approximated size of the CNV. In addition, information about CNVs reported only once and around 30–40 years ago is often limited and inconclusive regarding the exact location.

Figure 2C and Supplementary Figure S1 focus on CNVs affecting the  $\alpha$ -globin locus. A schematic representation of all CNVs along with the corresponding locations of  $\alpha$ -globin cluster on chromosome 16, is presented in Supplementary Figure S1. Supplementary Figure S1A presents the CNVs restricted within the  $\alpha$ -globin locus (31.91%) and the specifically affected genes. As illustrated in Supplementary Figure S1, the majority of CNVs associated with  $\alpha$ -globin locus variants (68.09%) are very large CNVs extending beyond the  $\alpha$ -globin locus leading to  $\alpha^0$ -thalassaemia. Among the pathogenic  $\alpha$ -globin locus variants, 83.82% (114/136) are related to large deletions affecting both *HBA1*, *HBA2* and/or HS-40 leading to  $\alpha^0$ -thalassaemia, with the remaining 16.18% (22/136) causing defects in either *HBA1* or *HBA2* and resulting in  $\alpha^+$ -thalassaemia (Figure 2C)—except for four deletions ( $-\alpha 3.7$  (type I), IthaID: 300;  $-\alpha 3.7$  (type II), IthaID: 2230;  $-\alpha 3.7$  (type III), IthaID: 2231; and  $\alpha 12$ , IthaID: 2555) which, despite extending within *HBA1* and *HBA2* (Supplementary Figure S1A), create a single functional fusion-gene leading to  $\alpha^+$ -thalassaemia. Despite affecting *HBA2*, *HBA1* and/or HS-40, and part of the telomeric region of chromosome 16, as illustrated in Supplementary Figure S1B, the disease-modifying variants ( $\alpha$ -globin locus duplication variants) are caused by large duplications or triplications of  $\alpha$ -globin genes and do not usually lead to pathogenesis, unless they are combined with  $\beta$ -thalassaemia alleles. Benign variants are created by deletions or fusions that do not alter the functionality of *HBA1* and *HBA2*.

Figure 2D and Supplementary Figure S2 focus on CNVs affecting the  $\beta$ -globin locus. A related schematic representation of all CNVs, along with the corresponding locations of  $\beta$ -globin cluster on chromosome 11, is presented in Supplementary Figure S2. There is no significant difference in the number of CNVs restricted within the  $\beta$ -globin locus (56.20%, Supplementary Figure S2A) and larger CNVs extending beyond the  $\beta$ -globin locus (43.80%, Supplementary Figure S2B). According to Figure 2D, pathogenic  $\beta$ -globin locus CNVs are divided into the following six groups:  $\delta\beta$ -thalassaemia (23.53%, 32/136),  $\beta$ -thalassaemia (26.47%, 36/136),  $\epsilon\gamma\delta\beta$ -thalassaemia (28.68%, 39/136),  $\delta$ -thalassaemia (0.74%, 1/136), structural haemoglobinopathies (13.24%, 18/136) and HPFH (7.35%, 10/136). CNVs associated with  $\epsilon\gamma\delta\beta$ -thalassaemia represent the largest deletions that disturb the entire  $\beta$ -globin locus (59%, 23/39) including *HBB*, *HBD*, *HBG1*, *HBG2*, *HBE1* and  $\beta$ LCR. Notably, the locus control region  $\beta$ LCR is affected in almost all CNVs (92.31%, 36/39) associated with  $\epsilon\gamma\delta\beta$ -thalassaemia, demonstrating the important role of  $\beta$ LCR in the gene expression of the  $\beta$ -globin locus. CNVs associated with the  $\delta\beta$ -,  $\beta$ -, and  $\delta$ -thalassaemia are smaller deletions that affect both *HBB* and *HBD*, or only *HBB* or *HBD*, respectively. Exceptions include five CNVs that cause  $\beta$ -thalassaemia. In the first three cases (1992 bp deletion, IthaID: 2479; Caribbean, IthaID: 2566; 12.4 kb Mediterranean deletion, IthaID: 3935), the deletions affect a limited region of  $\beta$ LCR, as illustrated in Supplementary Figure S2A. The deletion involves only one or two (HS3 or/and HS4) of the five DNase I hypersensitivity (HS) regions of  $\beta$ LCR, which only impairs the expression of *HBB*, leaving the expression



of the remaining genes of the  $\beta$ -globin locus intact [14]. Moreover, another CNV (223 kb deletion; IthaID: 3394) does not cause  $\delta\beta$ -thalassaemia, despite affecting both *HBB* and *HBD*, because it extends from the 3'UTR of the *HBD* without affecting its expression [15]. The last CNV (>29.5 kb duplication; IthaID: 3869) is a duplication including *HBG2*, *HBG1*, *HBD*, and part of *HBB*. Therefore, only *HBB* is disturbed, resulting in  $\beta$ -thalassaemia. Structural haemoglobinopathies are caused by CNVs affecting *HBD* and *HBB*, leading to a fusion of the genes, which gives rise to abnormal hybrid haemoglobin variants. Three of the five benign variants are duplications of the entire  $\beta$ -globin locus (African I duplication, IthaID: 2550; African II duplication, IthaID: 3112) or the *HBG2* ( $G\gamma$  duplication, IthaID: 3596), while the fourth is a deletion (Italian HS1, IthaID: 3936) involving only the HS1 region of the  $\beta$ LCR (Supplementary Figure S2A) leaving the functionality of  $\beta$ LCR and  $\beta$ -globin genes intact [16]. The last benign variant is a deletion involving part of *HBG1* and *HBG2*, leading to a single functional fusion-gene.

### 3. Discussion

The present study is the most comprehensive analysis of haemoglobinopathy-associated CNVs to date. All available CNV information related to haemoglobinopathies was investigated and analysed, and an open tool that facilitates the identification and diagnosis of these CNVs was developed.

The importance of fully characterising CNVs associated with haemoglobinopathies is still as relevant today as it was decades ago when identification and characterisation of large deletion/duplications first began. The combination of a rare CNV removing both *HBA1* and *HBA2* with more prevalent  $\alpha$ -globin gene pathogenic variants can result in clinically significant conditions, such as Hb H disease and the usually lethal Hb Bart's hydrops fetalis syndrome.

Furthermore, HPFH and  $\delta\beta$ -thalassaemia are molecularly similar conditions located at the end of the  $\beta$ -globin locus involving the *HBB* and *HBD*. As both conditions are characterised by elevated foetal haemoglobin (HbF) levels, the distinction between them is based on phenotypic differences and haematological indices. Heterozygotes of HPFH are characterised by normal red cell indices, normal HbA2 level and increased levels of HbF (15–30%) with homogeneous pancellular distribution, while  $\delta\beta$ -thalassaemia heterozygotes are characterised by hypochromic and microcytic red cells, a normal level of HbA2 and a less severe increase in HbF levels (5–15%) with heterogeneous intercellular distribution [17,18]. In addition, the compound heterozygosity of an HPFH deletion with a  $\beta^0$ -thalassaemia mutation can lead to an asymptomatic condition involving mild anaemia [19]; this is in contrast to the effect of compound heterozygosity of  $\delta\beta$ -thalassaemia along with another form of  $\beta$ -thalassaemia that leads to a severe clinical picture and a transfusion-dependent condition [18,20]. Most cases of HPFH and  $\delta\beta$ -thalassaemia are sporadic and rare; therefore, further evaluation and additional data are required for adequate categorisation. A complex network of factors and regulatory regions is known to affect the HbF levels which, along with the genetic characteristics that patients carry, favour a higher level of HbF expression [21]. Clarifying the breakpoints of deletions leading to either  $\delta\beta$ -thalassaemia or HPFH will shed light on the molecular mechanisms that enhance  $\gamma$ -globin production and the elevation of HbF levels [18]. For example, two CNVs that delete the entire *HBB* (South East Asian (SEA) deletion; IthaID: 1505 and Caucasian HPFH, IthaID: 2124) present a typical  $\beta$ -thalassaemia trait phenotype in the heterozygous state with microcytic, hypochromic erythrocytes and increased HbA2 level, but also with an elevated level of HbF. These two large deletions (>27 kb) were first characterised as HPFH because of significantly raised HbF levels of over 20%. Identification of the exact breakpoints in recent studies indicated that the deletions are more akin to  $\beta$ -thalassaemia syndrome with elevated HbF, as the 3' breakpoint downstream of the  $\beta$ -globin gene was found to contain the HS-1  $\gamma$ -silencing element, while *HBD* and erythroid-specific elements at the 5' breakpoint remained intact [22–24]. Considering the above deletions of HPFH, misleading genetic counselling for a clinically mild condition could occur in cases of compound heterozygosity with

$\beta^0$ -thalassaemia mutation. Therefore, complete characterisation and the easier detection and identification of CNVs is important not only for the correct diagnosis and treatment of the patients, but also for genotype–phenotype correlation and for comprehensive prenatal genetic counselling of the carriers.

Another characteristic example is  $\epsilon\gamma\delta\beta$ -thalassaemia, a rare thalassaemia syndrome caused by large deletions affecting the  $\beta$ -globin gene cluster. To date, over 30 such deletions have been described in heterozygous form only [25]. In general, homozygous deletions affecting the entire  $\beta$ -globin gene cluster are very rare presumably because they are not compatible with foetal survival [7]. Severe neonatal haemolytic anaemia is a characteristic clinical manifestation of  $\epsilon\gamma\delta\beta$ -thalassaemia, with affected neonates distinguished from  $\beta$ -thalassaemia patients due to the requirement of intrauterine and often perinatal red blood cell transfusions until the age of six months old. Beyond the first few months of life, anaemia is spontaneously normalised due to the elevated expression and production of  $\beta$ -globin in the first year of life [25,26]. Adult heterozygote haematological parameters resemble the  $\beta$ -thalassaemia trait phenotype, but due to normal HbA2 levels,  $\epsilon\gamma\delta\beta$ -thalassaemia is commonly undiagnosed or misdiagnosed [27]. Furthermore, hypochromic and microcytic indices with normal HbA2 levels could be due to compound heterozygous variants in *HBD* and *HBB*, as well as in  $\alpha^0$ -thalassaemia carriers, leading to a differential diagnosis of  $\epsilon\gamma\delta\beta$ -thalassaemia [25]. Without proper testing, the diagnosis may be overlooked, and the patient may be improperly treated.

Furthermore, determining the exact breakpoints of CNVs can lead to delineation of novel regulatory regions and other genes related with the expression of the  $\alpha$  and  $\beta$ -globin clusters. Intergenic enhancer and repressor, miRNAs, non-coding RNAs, transcriptional and epigenetic regulators, and epigenetic factors, may regulate the suppression and activation of globin genes as found in other genes in human cells [18,19]. Moreover, many CNVs trigger human genetic diseases associated with developmental disabilities [28]. In thalassaemia syndromes, CNVs are associated with the occurrence and severity of the disease [13]. Additionally, a comparison of different CNV breakpoints can provide insights into the biological processes that create them.

The MLPA technique has been proven to identify the structure and estimate the size of the arrangement region of known and unknown CNVs in unsolved cases after performing conventional techniques. MLPA is routinely used as a diagnostic screening tool in many laboratories, for many hereditary and acquired disorders triggered by gene dosage changes, and requires only a thermocycler and capillary electrophoresis equipment [22,29]. Nevertheless, the MLPA method can omit small deletions as the probes are not continuous or overlapping. In addition, the distance between many MLPA probes ranges between 3 and 10 kb; hence, if the CNV breakpoints extend beyond the range of MLPA probes, the use of array CGH analysis is required for size and localisation determination and also the use of GAP-PCR/CBAS and DNA sequencing for complete identification of the breakpoints [13,30]. The whole procedure is time-consuming and labour-intensive, making the characterisation of unknown CNVs highly complicated. Moreover, the presence of large repetitive sequences, which characterise both  $\alpha$ - and  $\beta$ -globin loci, makes the identification of CNV breakpoints even more difficult as the repetitive sequences restrict the efficiency of the methods [31].

The majority of the CNVs are extremely rare and unique to the families in which they have been described. As more of such CNVs are identified with universal screening, there is a need for a comprehensive single methodology akin to next-generation sequencing that can fully characterise breakpoints and genomic recombination events in a single analytical process [32]. Many studies suggest that WGS can be used to accurately locate the breakpoints of large segment deletions/duplications [33]. Furthermore, a recent study presented the development of a CNV application based on quantitative real-time PCR (qPCR) for the suspected cases of  $\alpha$ -thalassaemia following normal gap-PCR results [34].

Due to the high percentage of genomic deletions involving the  $\alpha$ -globin gene cluster and the high heterogeneity of the deletions within the  $\beta$ -globin cluster, the IthaCNVs tool

can facilitate laboratory staff and clinicians to confirm the suspected diagnosis of CNVs based on molecular DNA screening and analysis. Although in some cases the pathogenicity of a CNV can be inferred from the molecular analysis (e.g., CNVs removing both *HBA1* and *HBA2* causing  $\alpha$ -thalassaemia), the interpretation of CNVs should combine evidence about the disease mechanism, phenotype, and mode of inheritance, as suggested by the ACMG/AMP guidelines. Fully integrated with other ITHANET databases, IthaCNVs provides a manually curated list of all known CNVs associated with haemoglobinopathies and, through its diverse filtering options, it can facilitate the diagnosis of rare and diagnostically challenging cases and can potentially lead to standardised classification of such cases [35,36].

#### 4. Materials and Methods

The list of all CNVs (Supplementary File S1) associated with haemoglobinopathies was retrieved from the IthaGenes database (Access Date: 7 November 2022). Specifically, the dataset of this study includes 291 CNVs (>50 bp), including pathogenic, disease-modifying, and benign variants. CNVs with known breakpoints were reviewed to either confirm or correct the exact location of the upstream (5') and downstream (3') breakpoints using available evidence and the latest literature, while a rigorous literature review was performed for CNVs with missing breakpoint information. In many cases, MLPA was the only method that was used for CNV clarification, resulting in breakpoint locations between 5' and 3' instead of the exact genomic locations. This range is the gap between the two probes in which the breakpoint occurred. At the 5' breakpoint, the first genomic location of the range is not affected and represents the last normal amplified probe, while the second location is affected and represents the first affected probe. At the 3' breakpoint, the first genomic location of the range is affected and represents the last affected probe, while the second location is not affected and represents the first normal amplified probe (Figure 1). In addition, the breakpoints of some CNVs reported in the past were not available because they were either characterised with restriction fragment length polymorphism (RFLP) analysis or identified during screening for other disorders and only the relevant region or deleted genes were reported. The updated CNV dataset was subsequently independently reviewed by three diagnostic laboratories from Cyprus, the Netherlands, and Portugal, with members co-authoring this study, to help identify discrepancies and provide the missing information. Subsequently, the available MLPA probes involved in each CNV were identified using information about their positions and names from the MRC-Holland commercial kits (<https://www.mrcholland.com/>) for both globin loci, accessed on 7 November 2022. The SALSA MLPA Probemix P140-C1 HBA and P102-D1 HBB kit were used for the  $\alpha$ - and  $\beta$ -globin locus, respectively.

We developed IthaCNVs (<https://ithanet.eu/db/ithacnv>) using the existing infrastructure of the ITHANET portal based on PHP/MySQL, as described by Kountouris et al., 2014 [4]. All CNVs are linked to their corresponding variant page on the IthaGenes database, where additional details about associated phenotypes, context sequence, external links and relevant publications are provided.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232415920/s1>.

**Author Contributions:** A.M., P.K. and M.K. conceived and designed the study; S.T. designed and developed the online tool; M.X. analysed the data and produced the figures; C.S., C.B. and C.L.H. curated and validated data; A.M. and P.K. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Project: EXCELLENCE/1216/256).

**Data Availability Statement:** IthaCNVs and the associated data are freely available on the ITHANET portal at <https://ithanet.eu/db/ithacnv>.



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

## References

- Modell, B.; Darlison, M. Global Epidemiology of Haemoglobin Disorders and Derived Service Indicators. *Bull. World Health Organ.* **2008**, *86*, 480–487. [\[CrossRef\]](#) [\[PubMed\]](#)
- Weatherall, D.J. The Thalassemias. *BMJ* **1997**, *314*, 1675–1678. [\[CrossRef\]](#) [\[PubMed\]](#)
- Old, J. Haemoglobinopathies. *Prenat. Diagn.* **1996**, *16*, 1181–1186. [\[CrossRef\]](#)
- Kountouris, P.; Lederer, C.W.; Fanis, P.; Feleki, X.; Old, J.; Kleanthous, M. IthaGenes: An Interactive Database for Haemoglobin Variations and Epidemiology. *PLoS ONE* **2014**, *9*, e103020. [\[CrossRef\]](#)
- Kountouris, P.; Michailidou, K.; Christou, S.; Hadjigavriel, M.; Sitarou, M.; Kolnagou, A.; Kleanthous, M.; Telfer, P. Effect of HBB Genotype on Survival in a Cohort of Transfusion-Dependent Thalassemia Patients in Cyprus. *Haematologica* **2021**, *106*, 2458–2468. [\[CrossRef\]](#)
- Harteveld, C.L.; Voskamp, A.; Phylipsen, M.; Akkermans, N.; den Dunnen, J.T.; White, S.J.; Giordano, P.C. Nine Unknown Rearrangements in 16p13.3 and 11p15.4 Causing Alpha- and Beta-Thalassaemia Characterised by High Resolution Multiplex Ligation-Dependent Probe Amplification. *J. Med. Genet.* **2005**, *42*, 922–931. [\[CrossRef\]](#)
- Thein, S.L. The Molecular Basis of  $\beta$ -Thalassemia. *Cold Spring Harb. Perspect. Med.* **2013**, *3*, a011700. [\[CrossRef\]](#)
- Colosimo, A.; Gatta, V.; Guida, V.; Leodori, E.; Foglietta, E.; Rinaldi, S.; Cappabianca, M.P.; Amato, A.; Stuppia, L.; Dallapiccola, B. Application of MLPA Assay to Characterize Unsolved  $\alpha$ -Globin Gene Rearrangements. *Blood Cells Mol. Dis.* **2011**, *46*, 139–144. [\[CrossRef\]](#)
- Rooks, H.; Clark, B.; Best, S.; Rushton, P.; Oakley, M.; Thein, O.S.; Cuthbert, A.C.; Britland, A.; Ruf, A.; Thein, S.L. A Novel 506kb Deletion Causing  $\text{E}\gamma\delta\beta$  Thalassemia. *Blood Cells Mol. Dis.* **2012**, *49*, 121–127. [\[CrossRef\]](#)
- Zarrei, M.; MacDonald, J.R.; Merico, D.; Scherer, S.W. A Copy Number Variation Map of the Human Genome. *Nat. Rev. Genet.* **2015**, *16*, 172–183. [\[CrossRef\]](#)
- Efremov, G.D.; Simjanovska, L.; Plaseska-Karanfilska, D.; Stanojevic, E.; Petkov, G.H. Hb Jambol: A New Hyperunstable Hemoglobin Causing Severe Hemolytic Anemia. *Acta Haematol.* **2007**, *117*, 1–7. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kim, S.Y.; Lee, S.H.; Cho, S.I.; Song, S.H.; Hattori, Y.; Park, S.-K.; Song, J.; Choi, Y.; Yang, M.; Park, H.; et al. Molecular Identification of the Novel  $\text{G}\gamma$ - $\beta$  Hybrid Hemoglobin: Hb  $\text{G}\gamma$ - $\beta$  Ulsan ( $\text{G}\gamma$  through 13;  $\beta$  from 19). *Blood Cells Mol. Dis.* **2010**, *45*, 276–279. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hu, L.; Shang, X.; Yi, S.; Cai, R.; Li, Z.; Liu, C.; Liang, Y.; Cai, D.; Zhang, F.; Xu, X. Two Novel Copy Number Variations Involving the  $\alpha$ -Globin Gene Cluster on Chromosome 16 Cause Thalassemia in Two Chinese Families. *Mol. Genet. Genom. MGG* **2016**, *291*, 1443–1450. [\[CrossRef\]](#) [\[PubMed\]](#)
- Amid, A.; Cheong, M.; Eng, B.; Hanna, M.; Hohenadel, B.-A.; Nakamura, L.M.; Walker, L.; Odame, I.; Kirby-Allen, M.; Waye, J.S. Hb S/B+Thalassemia Due to Hb Sickle and a Novel Deletion of DNase I Hypersensitive Sites HS3 and HS4 of the  $\beta$  Locus Control Region. *Haematologica* **2015**, *100*, e166–e168. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhu, F.; Wei, X.; Cai, D.; Pang, D.; Zhong, J.; Liang, M.; Zuo, Y.; Xu, X.; Shang, X. A Novel 223 Kb Deletion in the Beta-Globin Gene Cluster Was Identified in a Chinese Thalassemia Major Patient. *Int. J. Lab. Hematol.* **2019**, *41*, 456–460. [\[CrossRef\]](#)
- Kulozik, A.E.; Bail, S.; Bellan-Koch, A.; Bartram, C.R.; Kohne, E.; Kleihauer, E. The Proximal Element of the Beta Globin Locus Control Region Is Not Functionally Required in Vivo. *J. Clin. Investig.* **1991**, *87*, 2142–2146. [\[CrossRef\]](#)
- Calzolari, R.; McMorro, T.; Yannoutsos, N.; Langeveld, A.; Grosveld, F. Deletion of a Region That Is a Candidate for the Difference between the Deletion Forms of Hereditary Persistence of Fetal Hemoglobin and Deltabeta-Thalassemia Affects Beta- but Not Gamma-Globin Gene Expression. *EMBO J.* **1999**, *18*, 949–958. [\[CrossRef\]](#)
- Mayuranathan, T.; Rayabaram, J.; Das, R.; Arora, N.; Edison, E.S.; Chandy, M.; Srivastava, A.; Velayudhan, S.R. Identification of Rare and Novel Deletions That Cause  $(\Delta\beta)0$ -Thalassaemia and Hereditary Persistence of Foetal Haemoglobin in Indian Population. *Eur. J. Haematol.* **2014**, *92*, 514–520. [\[CrossRef\]](#)
- Sankaran, V.G.; Xu, J.; Byron, R.; Greisman, H.A.; Fisher, C.; Weatherall, D.J.; Sabath, D.E.; Groudine, M.; Orkin, S.H.; Premawardhena, A.; et al. A Functional Element Necessary for Fetal Hemoglobin Silencing. *N. Engl. J. Med.* **2011**, *365*, 807–814. [\[CrossRef\]](#)
- Bollekens, J.A.; Forget, B.G. Delta Beta Thalassemia and Hereditary Persistence of Fetal Hemoglobin. *Hematol. Oncol. Clin. N. Am.* **1991**, *5*, 399–422. [\[CrossRef\]](#)
- Ghedira, E.S.; Lecerc, L.; Faubert, E.; Costes, B.; Moradkhani, K.; Bachir, D.; Galactéros, F.; Pissard, S. Estimation of the Difference in HbF Expression Due to Loss of the 5'  $\delta$ -Globin BCL11A Binding Region. *Haematologica* **2013**, *98*, 305–308. [\[CrossRef\]](#) [\[PubMed\]](#)
- So, C.C.; So, A.C.Y.; Chan, A.Y.Y.; Tsang, S.T.Y.; Ma, E.S.K.; Chan, L.C. Detection and Characterisation of Beta-Globin Gene Cluster Deletions in Chinese Using Multiplex Ligation-Dependent Probe Amplification. *J. Clin. Pathol.* **2009**, *62*, 1107–1111. [\[CrossRef\]](#) [\[PubMed\]](#)
- Pissard, S.; Raclin, V.; Lacan, P.; Garcia, C.; Aguilar-Martinez, P.; Francina, A.; Joly, P. Characterization of Three New Deletions in the  $\beta$ -Globin Gene Cluster during a Screening Survey in Two French Urban Areas. *Clin. Chim. Acta Int. J. Clin. Chem.* **2013**, *415*, 35–40. [\[CrossRef\]](#) [\[PubMed\]](#)

24. He, S.; Wei, Y.; Lin, L.; Chen, Q.; Yi, S.; Zuo, Y.; Wei, H.; Zheng, C.; Chen, B.; Qiu, X. The Prevalence and Molecular Characterization of ( $\Delta\beta$ )0-Thalassemia and Hereditary Persistence of Fetal Hemoglobin in the Chinese Zhuang Population. *J. Clin. Lab. Anal.* **2018**, *32*, e22304. [[CrossRef](#)] [[PubMed](#)]
25. Cardiero, G.; Prezioso, R.; Dembech, S.; Del Vecchio Blanco, F.; Scarano, C.; Lacerra, G. Identification and Molecular Characterization of a Novel 163 Kb Deletion: The Italian ( $E\gamma\delta\beta$ )(0)-Thalassemia. *Hematol. Amst. Neth.* **2016**, *21*, 317–324. [[CrossRef](#)]
26. Shalev, H.; Landau, D.; Pissard, S.; Krasnov, T.; Kapelushnik, J.; Gilad, O.; Broides, A.; Dgany, O.; Tamary, H. A Novel Epsilon Gamma Delta Beta Thalassemia Presenting with Pregnancy Complications and Severe Neonatal Anemia. *Eur. J. Haematol.* **2013**, *90*, 127–133. [[CrossRef](#)]
27. Repnikova, E.; Roberts, J.; Mc Dermott, S.; Farooqi, M.S.; Iqbal, N.T.; Silvey, M.; Nolen, J.; Taboada, E.; Li, W. Clinical and Molecular Characterization of Novel Deletions Causing Epsilon Gamma Delta Beta Thalassemia: Report of Two Cases. *Pathol. Res. Pract.* **2019**, *215*, 152578. [[CrossRef](#)]
28. Kaminsky, E.B.; Kaul, V.; Paschall, J.; Church, D.M.; Bunke, B.; Kunig, D.; Moreno-De-Luca, D.; Moreno-De-Luca, A.; Mulle, J.G.; Warren, S.T.; et al. An Evidence-Based Approach to Establish the Functional and Clinical Significance of Copy Number Variants in Intellectual and Developmental Disabilities. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2011**, *13*, 777–784. [[CrossRef](#)]
29. Munkongdee, T.; Chen, P.; Winichagoon, P.; Fucharoen, S.; Paiboonsukwong, K. Update in Laboratory Diagnosis of Thalassemia. *Front. Mol. Biosci.* **2020**, *7*, 74. [[CrossRef](#)]
30. Wang, Y.; Xiong, Y.; Liu, C.; Lu, J.; Wang, J.; Qin, D.; Liu, L.; Wu, J.; Zhao, X.; Fang, L.; et al. Characterisation of Two Unusual Cases of Haemoglobin Bart's Hydrops Foetalis Caused by -SEA and Large Novel  $\alpha$ -Globin Gene Cluster Deletions. *J. Int. Med. Res.* **2021**, *49*, 300060521993642. [[CrossRef](#)]
31. Clark, B.; Shooter, C.; Smith, F.; Brawand, D.; Steedman, L.; Oakley, M.; Rushton, P.; Rooks, H.; Wang, X.; Drousiotou, A.; et al. Beta Thalassaemia Intermedia Due to Co-Inheritance of Three Unique Alpha Globin Cluster Duplications Characterised by next Generation Sequencing Analysis. *Br. J. Haematol.* **2018**, *180*, 160–164. [[CrossRef](#)] [[PubMed](#)]
32. Fan, D.-M.; Yang, X.; Huang, L.-M.; Ouyang, G.-J.; Yang, X.-X.; Li, M. Simultaneous Detection of Target CNVs and SNVs of Thalassemia by Multiplex PCR and Next-generation Sequencing. *Mol. Med. Rep.* **2019**, *19*, 2837–2848. [[CrossRef](#)] [[PubMed](#)]
33. Jiang, F.; Xu, L.-L.; Tang, X.-W.; Li, D.-Z. A Novel Deletion of the Major Regulatory Element Flanking the  $\alpha$ -Globin Gene Cluster as a Cause of A0-Thalassemia. *Int. J. Lab. Hematol.* **2021**, *43*, O190–O192. [[CrossRef](#)] [[PubMed](#)]
34. Grimholt, R.M.; Fjeld, B.; Klingenberg, O. Hemoglobinopathy Gone Astray-Three Novel Forms of  $\alpha$ -Thalassemia in Norwegian Patients Characterized by Quantitative Real-Time PCR and DNA Sequencing. *Scand. J. Clin. Lab. Investig.* **2021**, *81*, 670–678. [[CrossRef](#)]
35. Riggs, E.R.; Andersen, E.F.; Cherry, A.M.; Kantarci, S.; Kearney, H.; Patel, A.; Raca, G.; Ritter, D.I.; South, S.T.; Thorland, E.C.; et al. Technical Standards for the Interpretation and Reporting of Constitutional Copy-Number Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2020**, *22*, 245–257. [[CrossRef](#)]
36. Kountouris, P.; Stephanou, C.; Lederer, C.W.; Traeger-Synodinos, J.; Bento, C.; Harteveld, C.L.; Fylaktou, E.; Koopmann, T.T.; Halim-Fikri, H.; Michailidou, K.; et al. Adapting the ACMG/AMP Variant Classification Framework: A Perspective from the ClinGen Hemoglobinopathy Variant Curation Expert Panel. *Hum. Mutat.* **2022**, *43*, 1089–1096. [[CrossRef](#)]