**South-Italy β°-Thalassaemia: a novel deletion not removing the γ-globin silencing element and with 3' breakpoint in a hsRTVL-H element, associated with β° thalassemia and high levels of Hb F**

Maria De Angioletti[1-3], Vincenzo Sabato[4], Gennaro Musollino[1], Prezioso Romeo[1], Clementina Carestia[1†] and Giuseppina Lacerra[1]

[1]*Istituto di Genetica e Biofisica "Adriano Buzzati-Traverso", Consiglio Nazionale delle Ricerche, Napoli, Italy*

[2]*ICCOM Consiglio Nazionale delle Ricerche, Sesto fiorentino, Firenze, Italy*

[3]*Core Research Laboratory – Istituto Toscano Tumori, Firenze, Italy*

[4]*Hospital "Giovanni XXIII", Bari, Italy*

† This article is dedicated to the memory of Clementina Carestia.

---

## Online Supplementary Materials and Methods

***Restriction mapping of the β-globin gene cluster.*** DNA (10 μg) was digested with restriction enzymes (3 U/μg DNA) under conditions recommended by manufacturers. Digested DNA was separated by electrophoresis on 0.8% agarose gel (IBI, New Haven, USA). The DNA was transferred from the gel to a positive charged nylon membrane (hybond-N+, GE Healthcare). The nylon filters were hybridized with the following DNA probes: β-globin (Pst I 4,442 bp), δ-globin (Pst I 2,316 bp), ψβ gene (Bgl II 4,136 bp), γ-globin (Bam HI 4,936 bp), ε-globin (Bam HI 0,652 bp), and the probe 1,054 bp Sst I/Bam HI located at 73,699 bp 3' from the β-globin gene cap site (NCBI contig NT_009237). Labelled DNA bands were detected by autoradiography (Fuji RX films, Fuji, Japan) and Cronex-Quanta III intensifying screens (Dupont, Bad Homburg, Germany).

***Inverse-PCR (I-PCR) and sequencing of the deletion breakpoint.*** DNA digested with the Pvu II restriction enzyme was separated by electrophoresis on 1.2% NuSieve GTG agarose gel (FMC, Rockland, USA). The DNA fragments with length from 4.0 to 5.0 kb were excised, purified by chloroform-phenol extraction and circularized by DNA ligase (400.000 U/ml, New England Biolabs, Beverly, USA) under conditions recommended by the manufacturer. The inverse PCR (I-PCR) amplification was carried out with the primers located in the δ-globin gene reported in Table S1. The I-PCR fragment of 1,596 bp was cloned in pGEM-T vector. The plasmid DNA was sequenced with the primers used for the I-PCR and the primers pUC/M13 forward and reverse localized in pGEM-T vector sequence (Promega, Madison, USA).

## Online Supplementary Results

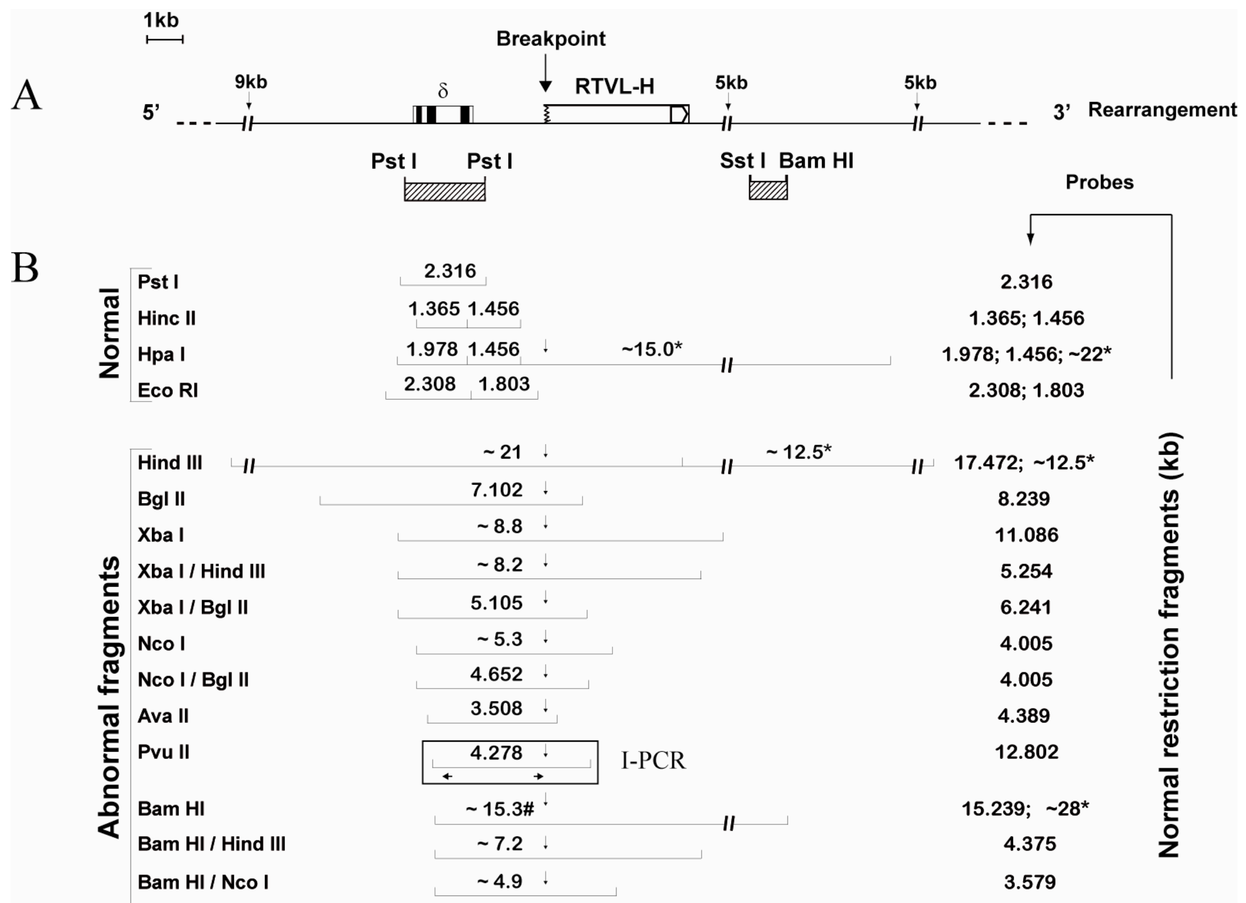
***Localization of the deletion breakpoints by restriction enzyme mapping.*** The restriction mapping analysis of the β-globin gene cluster confirmed the absence of the β-globin gene; in fact, hybridization with the β-globin probe gave no abnormal fragments, but only normal fragments with 50% reduction of intensity. On the contrary, the δ-globin gene was present, and the hybridization with the Pst I δ-probe revealed abnormal fragments, namely the fragments Hind III ≈ 21 kb, Xba I ≈ 8.8, Bgl II ≈ 7.1 kb, Nco I ≈ 5.3 kb, Pvu II ≈ 4.2 kb, Ava II ≈ 3.5 kb, whereas the Bam HI fragment was apparently of normal size (Supplementary Figure S1). No abnormal fragments were obtained with the enzymes Pst I, Hinc II, Hpa I and Eco RI. The restriction map of the rearranged region (Supplementary Figure S1) indicated that the 5' breakpoint was in a 723 bp sequence between the Eco RI (last conserved) and the Nco I (first deleted) restriction site, at 1,624 and 2,347 bp downstream the poly A of the δ-globin gene respectively. To characterize the 3' breakpoint region we used a set of probes localized downstream the β-globin gene. By using the Sst I/Bam HI 1.08 kb probe (at 6,923 bp downstream the 3' end of the RTVL-H element) we found two abnormal fragments: the 15.3 kb Bam HI fragment (detected also with the δ-probe) and the 15.0 kb Hpa I fragment (Supplementary Figure S1).

***Identification of the breakpoint sequence***

The identification of the breakpoint sequence was obtained by Inverse PCR strategy. We chose the fragments obtained by Pvu II digestion because the substantial difference in length between the normal and abnormal fragments (about 4.3 instead of 12.8 kb) and because it includes, as inferred by Eco RI digestion, about 900 bp of the downstream portion of the 3' breakpoint (Supplementary Figure S1), a length that is optimal for the subsequent sequencing and identification.

The Pvu II fragments were subjected to inverse-PCR amplification by using primers with inverted directions (Supplementary Table S1). The reverse primer was localized at 56-75 bp downstream the 5' extremity of the Pvu II fragment; the forward primer was localized 6 bp downstream the internal conserved Eco RI site. A fragment of about 1,600 bp was obtained, cloned into the pGEM vector and sequenced. The 5' breakpoint was at 2,138 bp dowstream the polyA of the δ-globin gene; the 3' breakpoint was inside a hsRTVL-H element (1,555/1,560 bp from its start) located about 70 kb dowstream the δ-globin gene (Figure 2). The same exanucleotide 5'-GTCTGA-3' is present in both sequences closely flanking the deletion: the presence of this microhomology may have facilitated the deletion and prevents the exact attribution of the exanucleotide 5'-GTCTGA-3' to one of the flanking regions. The breakpoints were also confirmed by the sequencing of a 1,293 bp specific fragment derived from the gap-PCR (Supplementary Table S1). The analysis of the

deletion specific amplicon also revealed the presence of 13 SNP base substitutions in the 3'-arm of the deletion (Supplementary Figure S2).




**Online Supplementary Figure S1. (A)** The structure of the β-globin gene cluster resulting from the ~66 kb deletion (South-Italy β°-thal)**.** An arrow indicates the position of breakpoint. The dashed boxes show the position of the two probes used for the restriction analysis: 5' probe: Pst I fragment containing the δ-globin gene; 3' probes the 1,080 bp Sst I/Bam HI fragment located at 73,699 bp from the 3' cap site of β-globin gene (NCBI, contig NT_009237). **(B)** Restriction mapping of the breakpoint region of the deletion. The restriction fragments without asterisk have been detected by the 5' probe; the restriction fragments indicated with an asterisk (*) have been detected by the 3' probe; the restriction fragment indicated with a # has been detected by both probes. Vertical arrows indicate the position of the breakpoint. The expected length of the normal restriction fragments is reported on the left. The 4.2 kb abnormal Pvu II restriction fragment used for the Inverse PCR is highlighted by a box. The horizontal arrows show the position of the primers used for the I-PCR.

```
      ********** ********** ********** ********** **********
5' N  TTCTTTTCTC ACTATGGCGT TGCTTTATAA ACTATGTTCA GTATGTCTGA
Del   TTCTTTTCTC ACTATGGCGT TGCTTTATAA ACTATGTTCA GTATGTCTGA
3' N                                                  GTCTGA
                                                      ******

Del   GGTGCCTGAC GTCCAGGCAT TCTTTTACAC ATTGGTCCCT CCCTAGTCTC
3' N  GGTGCCTGAC GTCCAGGCAT TCTTTTACAC ATTGGTCCCT CCCTAGTCTC
      ********** ********** ********** ********** **********

Del   TTCCCAATGC AACTTGTCCC AAATCTTCCT TCTTTCCCTC CCGCCTGTCC
3' N  TTCCCAATGC AACTTGTCCC AAATCTTCCT TCTTTCCCTC CCGCCTGTCC
      ********** ********** ********** ********** **********

Del   CCTCAGTCCC AACCCCAAGC GTCGCTGAGT CTTTCTACTC TTCCTTTTCT
3' N  CCTCAGTCCC AACTCCAAGC GTCGCTGAGT CTTTCTAATC TTCCTTTTCT
      ********** *** ****** ********** ******* ** **********

Del   ACAGACCCAT CTGACGTCTC CCCTCCTCCC CAGGCTGCTC CTTGCCAGGC
3' N  ACAGACCCAT CTGACGTCTC CCCTCCTCCC CAGGCTGCTC CTTGCCAGGC
      ********** ********** ********** ********** **********

Del   TGAGCTAGGT CCCAATTCTT CCTCAGCCTC AGCTCCTCCA CCCTGTAATC
3' N  CGAGCTAGGT CCCAAATCTT CCTCAGCCTC AGCTCCTCCA CCCTGTAATC
       ********* ***** **** ********** ********** **********

Del   TTTTTATCAC CTCCCCTCCT CACACCCGGT CTGGTTTACC ATTTCATTCC
3' N  TTTTTATCAC CTCCCCTCCT CACACCCGGT CCGGTTTACA GTTTCATTCC
      ********** ********** ********** * *******    *********

Del   GTGAGTAGCC CTCCCCAACC TGCCCAGCAA TTTCCTCTTA AGAAGGTGGC
3' N  GTGACTAGCC CTCCCCAACC TGCCCAGCAA TTTCCTATTA AGAAGGTGGC
      **** ***** ********** ********** ****** *** **********

Del   TGAAGCTAAA GGCGTAGACA AGGTTAATGC TCCTTTTTCC TTATCAGACC
3' N  TGAAGCTAAA GGCGTAGTCA AGGTTAATGC TCCTTTTTCT TTATCCGACC
      ********** ******* ** ********** *********  ***** ****

Del   TCTCCCAAAT CAGAGCGTTT AGGCTCTTTC ATCAAATATG AAAAACCCAG
3' N  TCTCCCGAAT CAGAGCGTTT AGGCTCTTTC ATCAAATATG AAAAACCCAG
      ****** *** ********** ********** ********** **********

Del   CCCAGTTCAT GGCTGGTTCG GCAGCAACCC TGAGACGCTT TACAGCCC
3' N  CCCAGTTCAT GGCTCGTTCG GCAGCAACCC TGAGACGCTT TACAGCCC
      ********** **** ***** ********** ********** ********
```

rs61880381 A>G; rs61880380 T>G; rs61880379 G>A; rs61880378 T>A; rs71459835 G>A; rs71472165 delinsTG; rs71459834 G>C; rs77608860 T>G; rs77207445 A>T; rs76886256 A>G; rs76957522 G>T; rs72482800 C>T; rs138562095 G>C

**Online Supplementary Figure S2.** Sequence of the 3' breakpoint of the fragment amplified by gap-PCR in the South-Italy β°-thalassemia. The region of microhomology (5'-GTCTGA-3') present in the sequences closely flanking the breakpoints is underlined. The sequence at 3' of the breakpoint, within the RTVL-H, includes various known SNP (indicated by box). The NCBI code for the each SNPs is reported.

**Online Supplementary Table S1.** Oligonucleotides used as primers in the Inverse-PCR, Gap-PCR and PCR reactions; sequence and position of the primers were from GenBank, sequence NT_009237.18 (chr 11); NT_022184.15 (chr 2); NT_025741.15 (chr 6).

| | Primers | bp | Chromosome or HBG genes position | Amplicon length (bp) | Oligo P mol | Ref |
|---|---|---|---|---|---|---|
| **A) Inverse-PCR** | | | **chr 11** | | | |
| 1 | R-AAGTGAAGCATCTCCTGGAC | 20 | 5195197-5195216 | 1596 | 7.5 | p.a. |
| 2 | F-AACATCACCTGGATGGGACA | 20 | 5192452-5192433 | | 7.5 | p.a. |
| **B) Gap-PCR and sequencing** | | | | | | |
| 3 | F-CAGAGAATTCATGAGAACATCACC | 24 | 5192444-5192467 | 1293 | 10 | p.a. |
| 4 | R-GTTTAGGCATTTGGAAGTTCTTGT | 24 | 5125024-5125047 | | 10 | p.a. |
| 5 | F-CTAGGAGTGTGGGAGCTAAATGAT | 24 | 5192270-5192293 | | | p.a. |
| 6 | R-TAAAGTGGCCTTTTGTCCTTTTAG | 24 | 5125260-5125283 | | | p.a. |
| **C) Primers for PCR synthesis of the $^{A}\gamma$-globin gene sequences** | | | | | | |
| 7 | F-TGAAACTGTGGCTTTATAGAAA | 22 | HBG1 -617/-596 | 696 | 10 | 5 |
| 8 | R-TTGTCCTCCTCTGTGAAATGAC | 22 | HBG1 79/58 | | 12 | 5 |
| 9 | R-TTGCCTTGTTCCGATTCAGTCA | 22 | HBG1 -240/-261 | | | 5 |
| **D) Primers for sequencing of the $^{A}\gamma$-globin gene sequences** | | | | | | |
| 10 | R-CAGCTCACACCCTGCTGTGCTC | 22 | HBG1 562/583 | 618 | 10 | p.a. |
| 11 | F-ATGCAGGTAGTTGTTCCCCTTCA | 23 | HBG1 1208/1186 | | 10 | 5 |
| **E) Primers for sequencing of the $^{G}\gamma$-globin gene sequences** | | | | | | |
| 12 | F-TGAAACTGTTGCTTTATAGGAT | 22 | HBG2 -619/-598 | 698 | 10 | p.a. |
| 13 | R-TTGTCCTCCTCTGTGAAATGAC | 22 | HBG2 79/58 | | 12 | 5 |
| 14 | F-GAGTAACTGGAAGATACTGATAA | 23 | HBG2 -1513/-1491 | 319 | 10 | 5 |
| 15 | R-AACTCGATCCATGACCTTGG | 20 | HBG2 -1195/-1214 | | 10 | 5 |
| 16 | F-ATGCAGGTAGTTGTTCTTCTTGC | 23 | HBG2 1228/1206 | 667 | 10 | 5 |
| 17 | R-CAGCTCACACCCTGCTGTGCTC | 22 | HBG2 562/583 | | 10 | p.a. |
| **F) HS-111 PCR and sequencing** | | | | | | |
| 18 | F-GCTTGGTGAAGTAGGAGATTC | 21 | 5341770-5341789 | 481 | 10 | 6 |
| 19 | R-GAGAACCCTGTGAGTAAGGA | 20 | 5342230-5342250 | | 10 | 6 |
| **G) 3'HS1 PCR and sequencing** | | | | | | |
| 20 | F-GCCTACTTCAGGTTTGTGG | 19 | 5166399-5166417 | 364 | 10 | 6 |
| 21 | R-ACATTCCTATTTGCCAAGG | 19 | 5166054-5166072 | | 10 | 6 |
| **H) BCL11A PCR and sequencing** | | | **chr 2** | | | |
| 22 | F-ACACGTCCACCAGTCTAGAAAGC | 23 | 39541956-39541978 | 284 | 10 | p.a. |
| 23 | R-GTTGAATGGATAAACGCCCTATG | 23 | 39542217-39542239 | | 10 | p.a. |
| **I) HBS1L-MYB PCR and sequencing** | | | **chr 6** | | | |
| 24 | F-AGCCCGTCCAGACACTCATTGTT | 23 | 39587846-39587868 | 431 | 10 | 8 |
| 25 | R-CTCAGTGATGGTATTTCTGGAGAC | 24 | 39588253-39588276 | | 10 | 8 |

p.a.= present article. Ref = references

**Online Supplementary Table S2**. Restriction fragment length polymorphism (RFLP), single nucleotide polymorphisms (SNP) and microsatellite in the β-globin genes cluster; SNP in the HBS1LMyb rs66650371 –TAC and in the BCL11A rs11886868 T>C present in all the family members. The less frequent allele of these polymorphisms has been associated with a high Hb F expression level. The family members are indicated as in the pedigree in Figure 1. The RFLP presence/absence was indicated with + or – respectively. The position of the SNP were from GeneBank, NT_009237.18 for Chr 11, NC_000002.11 for Chr 2, NC_000006.11 for Chr 6. The oligonucleotide sequences used for the PCR reactions are available upon request.

| Current nomenclature | HUGO nomenclature | Chr nt position | SNP accession n. | I-1 | I-2 | II-1 | II-2 |
|---|---|---|---|---|---|---|---|
| **RFLP** | | **Chr 11** | | | | | |
| Hind III/ $^{G}\gamma$ | HBG2:c.369-82G>T | 5214717 | rs2070972 | -/- | -/- | -/- | -/- |
| Hind III/ $^{A}\gamma$ | HBG1:c.369-82T>G | 5209799 | rs28440105 | -/- | -/- | -/- | -/- |
| Hinc II/ ψβ | chr11:g.5220259G>A | 5203683 | rs10128556 | -/- | -/- | -/- | -/- |
| Hinc II/ 3'ψβ | chr11:g.5217034A>G | 5200458 | rs968857 | -/- | -/- | -/- | -/- |
| Ava II/ β | HBB:c.315+16G>C | 5187791 | rs10768683 | +/Δ | +/+ | +/Δ | +/+ |
| Bam HI/ 3' β | chr11:g.5195011G>A | 5178435 | rs9704069 | +/Δ | +/+ | +/Δ | +/+ |
| **Microsatellite** | | | | | | | |
| $^{A}\gamma$ IVS-II nt 571/619$(TG)_n(CG)_n$ | HBG1:c.368+568 $TG_{13-21}CG_{0-6}$ | 5209987/ 5210030 | not assigned | AC | AC | AC | nt |
| $^{G}\gamma$ IVS-II nt 572/624$(TG)_n(CG)_n$ | HBG2:c.368+572 $TG_{11-22}CG_{2-3}$CACG | 5214950/ 5214899 | not assigned | GB/GF | GA/GC | nt | nt |
| **SNP** | | | | | | | |
| $^{G}\gamma$ -1450 T>G | HBG2:c.-1503A>C | 5217461 | rs10128653 | T | T | T | nt |
| $^{G}\gamma$ -1280 A>G | HBG2:c.-1333G>A | 5217291 | rs2855121 | G | G | G | nt |
| $^{G}\gamma$ -1225 A>G | HBG2:c.-1278A>G | 5217236 | rs2855122 | A | A | A | nt |
| $^{G}\gamma$ -309 A>G | HBG2:c.-362A>G | 5216320 | rs112479156 | A | A/G | A/G | A |
| $^{G}\gamma$ -158 C>T[#] | HBG2:c.-211G>A | 5216169 | rs7482144 | C(-) | C(-) | C(-) | C(-) |
| $^{A}\gamma$ -588 A>G | HBG1:c.-641A>G | 5211671 | rs2855039 | G | G | G | G |
| $^{A}\gamma$ -499 T>A | HBG1:c.-552T>A | 5211582 | rs57966301 | T | T | T | T |
| $^{A}\gamma$ -369 C>G | HBG2:c.-422C>G | 5211452 | rs2855040 | G | G | G | G |
| $^{A}\gamma$ -222/5 -AGCA | HBG1:c.-275_278delAGCA | 5211309^ 5211310 | rs35720845 | +AGCA | +AGCA | +AGCA | +AGCA |
| **CH haplotype** | | | | I/I | I/I | I/I | |
| HS-111 (-21 G>A) | chr11:g.5358783G>A | 5342207 | rs2723375 | G/G | G/G | G/G | nt |
| HS-111 (+126 G>A) | chr11:g.5358637G>A | 5342061 | rs75685457 | G/G | G/G | G/G | nt |
| 3'HS1 (+179 C>T) | chr11:g.5182894G>A | 5166318 | rs12805212 | C/Δ | C/C | C/Δ | nt |
| **SNP *in trans* to the β-globin gene cluster** | | **Chr 2** | | | | | |
| BCL11A C>T | chr2:g.60720246C>T [7] | 60720246 | rs11886868 | T/C | T/C | T/C | T/C |
| | | **Chr 6** | | | | | |
| HBS1L-Myb | chr6:g.135418633_135418635delTAC [8] | 135418633/ 135418635 | rs66650371 | +TAC | +TAC | +TAC | +TAC |
| | chr6:g.135418635C>T [8] | 135418635 | rs77755698 | C | C | C | C |

Δ= deleted allele; nt= not tested. The numbers in parenthesis are relative to the reference articles.