

Clinical implementation of Clinician-oriented locus specific mutation detection and deposition system

Soo-young Jung, M.T, Jin-Young, Choi, M.T, Ki-young Yoo, M.D

Korea Hemophilia Foundation, Seoul, Korea

Objectives

F8 gene contains 26 exons, 186kb long. Not only sequencing but also interpretation is time-consuming and labor intensive.

So we developed the automated mutation detection software of DNA sequence, called Kohemgene (COMUS: Clinician-oriented locus specific mutation detection and deposition system). This software uses UCSC genome browser as reference and is loaded in The Korea Hemophilia mutation database (www.kohemgene.org).

This study is designed to compare KOHEMGENE and Sequencher® in terms of time-saving and accuracy of interpretation.

Material and Methods

We analyzed with desktop computer (Pentium® IV, 512MB) 59 unrelated hemophilia A patients not having intron 22 inversion mutation. The time for the loading of data into software to completion of interpretation was recorded in seconds. Kohemgene and Sequencher® were used for the sample patient.

Meanwhile process of Sequencher® is composed 5 steps;

- 1) Input Ref. Sequence,
- 2) Input Data file,
- 3) Trimming,
- 4) data analysis
- 5) Report

Process of Kohemgene program,

- 1) Input Data file (AB1 or FASTA),
- 2) data analysis,
- 3) Report.

Process of Sequencher®, Two researchers participated in the analysis, who experienced more than five years as their professional.

Report from KOHEMGENE

1. All Query Sequence

> f8.minus.ab1.00.ab1 378 0 378 ABI trimmed : aligned : quality

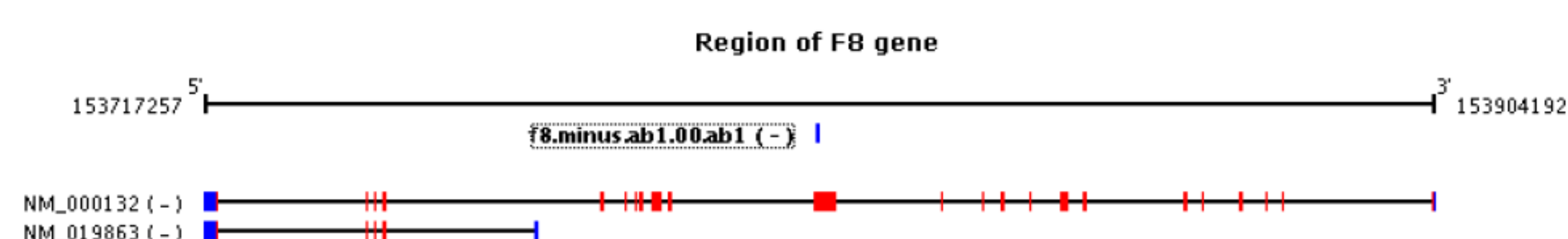
```

1  GCAATGGGTC TCCTGGCCAT CTGATCTDS TGAAGGGAG CCTTCTCAG GBAACABAGG GAGGATTAA GTGGAATBAA GCAACACAGC CTGAAAAAGT TCCCTTTCTG AGAGTAGCAA
121 CAGAAAGCTC TGCAGAAACT CCCTCAAGC TATTGATCC TCTTGTCTG GATAACCACT ATGTTACTCA GATACAAA GAAGAGTGA AATCCAGA GAAGTACCA GAAAAACAG
241 CTTTAAAGAA AAAGATACC ATTTTGTCC TGAAGCCTG TGAAGCAAT CATGCAATAG CAGCAATAAA TGAAGBACAA AATAAGCCG AAATAGAAGT CATCTAGCA AAGCAAGTA
361 GGACTGAAA GCTGTGCT
  
```

2. All Mutation Candidates

	Position on Reference	Source Query Sequence	Nucleotide on Reference	Nucleotide on Query	HGMD	dbSNP	Conservation Score	Protein Change
1	153810325	f8.minus.ab1.00.ab1	c	t	HGMD	-	0.6400	NM_000132.p.W1645*
2	153810328	f8.minus.ab1.00.ab1	g	a	-	-	0.0020	NM_000132.p.T1644I

3. Gene Region Diagram



In the gene region diagram, the top line represents a gene locus of a target gene, and the blue bar below the top line represents a mapped query sequence. The next two lines represent mRNAs in the target gene locus. The blue box is UTR regions, and the red box is coding regions.

6. Nucleotide Alignment

- Matching region on NM_000132 region of reference nucleotide sequence

```

      N Q S C L R E T R B Q K A W T V E I E P K N Q B E N I A A I A H N S E C A N L S L I T D K
... GGGTTTGGAGGACAGCAGCCTTTCAGTCTACCTTGTCTTGGCAGTACTTCTATTTCGGGCTTATTTGTCCCTCATTATTCGCTGCTATTGCAATGATTCCTTTCACAAAGCTTCAGBACAAATGGTATCCTT
      ABCACAGCCTTTCAGTCTACCTTGTCTTGGCAGTACTTCTATTTCGGGCTTATTTGTCCCTCATTATTCGCTGCTATTGCAATGATTCCTTTCACAAAGCTTCAGBACAAATGGTATCCTT
      S C L R E T R B Q K A W T V E I E P K N Q B E N I A A I A H N S E C A N L S L I T D K
  
```

```

R L F P V K G P R N A E N W K I A G E T G Q L L S G E V L D L H G P S G N S T E T
TCAGAAAGBAACTTTCCAGGCTGTGTTGCTTCCACTTAATGCTCCCTCTGTCCTGAAAGAGGCTCCCTCCACAGATCCAGATGGCCAGBAGACCCATTGCTAGTTCCBT...
TCAGAAAGBAACTTTCCAGGCTGTGTTGCTTCCACTTAATGCTCCCTCTGTCCTGAAAGAGGCTCCCTCCACAGATCCAGATGGCCAGBAGACCCATTGCTAGTTCCBT...
R L F P V K G P R N A E N W K I A G E T G Q L L S G E V L D L H G P S G N
  
```

7. Protein Sequence

> NM_000132
> mutation of NM 000132

```

RKKDSVDES SHFLQAKKN NLSLAILTLE MTGDQREVBS LGTSATNSYT YKVVNTVLP KPDLPKTSBK VELLPKYHY QKDLFPTETS NGSPHLDLV EGSLLQTEB AIKWEANRP
RKKDSVDES SHFLQAKKN NLSLAILTLE MTGDQREVBS LGTSATNSYT YKVVNTVLP KPDLPKTSBK VELLPKYHY QKDLFPTETS NGSPHLDLV EGSLLQTEB AIKWEANRP
GKVPFLRVAT ESSAKTPSKL LDPLAWNHY GTQIPKEEWK SOEKSPEKTA FKIKDTL LSL NACESNHAI A INEONKPE I EVYAKQGR TEHLCSNPP VIKKHOREIT RTLQSDQEE
GKVPFLRVAT ESSAKTPSKL LDPLAWNHY GTQIPKEEWK SOEKSPEKTA FKIKDTL LSL NACESNHAI A INEONKPE I EVYAKQGR TEHLCSNPP VIKKHOREIT RTLQSDQEE
IDYDQTSVE MKKEDFDIY EDENSPRSP QKTRHYFIA AVERLWQVM SSSPHLRNR AOSGVPQFK KYVDFEFTDS SFTQPLRYGE LNEHLGLLP YVRAEVDNI MYVFRNDSR
IDYDQTSVE MKKEDFDIY EDENSPRSP QKTRHYFIA AVERLWQVM SSSPHLRNR AOSGVPQFK KYVDFEFTDS SFTQPLRYGE LNEHLGLLP YVRAEVDNI MYVFRNDSR
  
```

In the nucleotide alignment part, red boxes represent mutations. In the protein sequence part, red box represents mutation and yellow region represents untranslated protein sequences due to nonsense mutation.

RESULT

The result with 59 participants was matched between two programs and analysis time is reduced to 1/30 using the KOHEMGENE program.

There is no difference of interpretation between the two programs. Analysis time by using the KOHEMGENE program was 56 seconds on average. And the analysis time with Sequencher® program was 30 minutes on average.

Conclusions

The KOHEMGENE program (COMUS) provides a DNA sequence mutation analysis that helps to identify the hemophilia mutation. Also, the KOHEMGENE program (COMUS) functions automatically to search known mutation or unknown mutation in HGMD® & HAMSTeRS (Haemophilia A Mutation, Search, Test and Resource Site).

Then, researchers do not need the procedure by its help. Conclusively, the KOHEMGENE program (COMUS) facilitates research in the field of Hemophilia mutation analysis.