Задание 1.

```
annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/1
$ ./sum.sh

annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/1
$ ls
fin  fout  sum.sh*

annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/1
$ cat fout
20
```

сам скрипт:

```
#!/bin/bash

awk -F'\t' '{sum = $1 + $2 + $3; print sum > "fout"}' fin
~
```

Задание 2.

```
annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/1
$ vim hello_world.sh

annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/1
$ ./hello_world.sh

annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/1
$ ls
fin  fout  hello_world.sh*  sum.sh*

annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/1
$ cat fout
Hello, world!
```

```
#!/bin/bash

echo "Hello, world!" > fout
```

Задание 3.

```
annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/3
$ ./count_ten.sh

annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/3
$ ls
count_ten.sh*  fin  fout

annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/3
$ cat fout
f
```

```
#!/bin/bash

awk '{print substr($0, 10, 1)}' fin > fout
~
```

Задание 4.

```
annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ vim ticket.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ ./ticket.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ ls
fin  fout  ticket.sh*

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ cat fout
Sam
957779
Varun
077712
```

```
#!/bin/bash

while IFS= read -r name && IFS= read -r ticket; do
    if [[ "$ticket" == *"777"* ]]; then
        echo "$name" >> fout
        echo "$ticket" >> fout
    fi
done < fin
~
```

а можно через grep (видно насколько короче выходит)

```
annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ vim t2.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ ./t2.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ ls
fin  fout  fout2  t2.sh*  ticket.sh*

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/ДЗ8/4
$ cat fout2
Sam
957779
Varun
077712
```

```
#!/bin/bash
grep -B1 "777" fin | sed '/^--$/d' > fout2
~
~
```

Задание 5.

```
annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/5
$ vim sort_exom.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/5
$ ./sort_exom.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/5
$ ls
fin  fout  sort_exom.sh*

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/5
$ cat fout
ENSP102T003     1       1258    1679
ENSP103T002     1       2257    3199
ENSP102T004     2       1680    2134
ENSP102T002     3       935     1257
ENSP102T001     5       578     934
ENSP102T005     7       2135    2731
ENSP102T007     14      3402    4090
ENSP102T010     17      7157    7980
ENSP102T009     21      6432    7156
ENSP102T008     22      5968    6431
ENSP102T011     X       9672    10014
ENSP103T001     Y       1234    2256
ENSP102T006     Y       2732    3401
```

```
#!/bin/bash

sort -t$'\t' -k2,2V -k3,3n -k4,4n fin > fout
~
```

Задание 6.

```
annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/6/6.1
$ vim count.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/6/6.1
$ ./count.sh
85
```

```
annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/6/6.2
$ vim count.sh

annus@▓▓▓▓▓▓▓▓_▓▓▓▓ MINGW64 ~/Bioinformatics_2025/hw/hw_8/Д38/6/6.2
$ ./count.sh
57
```

```
#!/bin/bash

awk '
    BEGIN { total = 0 }
    /[0-9]/ && ! /[XxYy]/ {
        gsub(/\r|\n/, "");
        total += length($0)
    }
    END { print total }
' fin
```

Задание 7.

```
annus@[redacted] MINGW64 ~/Bioinformatics_2025/hw/hw_8
$ cat fout
Longest insertion in chr1:
Length: 7369
Sequence: TGCACTCCAGCCTGGGTGACAGAGCGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAGATGTTTGCTTGCTTGCTACTATGTGATACCACTGT
TAGAAAATAGAGGAAGCCTGCCTGGGCTTTCTGTGCCCCAAATGAGAAAGGGAAAATTGCAAGAAGATGTTATTATCATCAGTCTTCCTAAGTATAAAAGAGA
CAGGCAGATAATTTAAGGAAGAGAAGCATCCAGTGCTGCTTGATCATGGACAGGCCTGATTCGTGTGCTGCTTCCATGACAGCATCTGGGCAGGGAGTCACAG
ATCTCTGTGCCCATACCTTTGAAGCAGTTCTAGTTAAGGGTTCAAAAACCAGAGTAAGTGACATCTAATTATTTCTCTTTTTTTTTTTTCTTTTTGAGACGGA
GTCTCGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGCATTCTCGGCTCACTGCGACCTCTGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCAAGT
AGCTGGGACTACAGGTGACTGCCACCACGCCCCGCTAATGTTTTTGTATTTTTAGTAGAGATGGGGTTCTACCTTGTTGGCCAGGCTGGTCTTGAACTTCTGA
CCTCAGGTGATCCACCTGTCTCAGCCTCCCAACACCTAATTATTTCTTTTCTTTTCTTTTTTTTTTTGAGATAGAGGTTCACTCTTGTTGCCCAAGCTGGAGT
GCAATGGCATGATCTTGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCCGTTCCCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCTGCCACCA
CGCCCGGCTAATTTTTGTATTTTTAGTAGAGCTGGGGTTTCACCCTGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCTGCCCATCTCGGGCCTC
CCAAAGTGCGGGGATTACAGAGGTGAGCCACCGCACCTGGCTGACACCTAATTATCTCTTAATGTTATTTTGATTTTTTTAAACTTTATTTTCTTTTTTTTGT
TGGAGATGGGGCCTCGCTATGTTGCCTATGCTGACTGATCTCAAACTCCTGGCCTCAAGCAATCCTCCCACCTAGGCCTCCCAAAGTGCTGGGATTACAAATG
TGAGCTACCATGCCTGGCCTGTTGTTATCTTGAGTTTAAAATGAAAGGGCTTTCTGAAAAGCCTTTCTGTTTAGTATCTGTTCTTAGAGTTCTGAACTTTCCA
GTGACTGACTGTATCATTTAGGGTTCAGACTAGGAAATAGAAACCCACTCTGGATATTTTAAGTGGAAAGAGATTTAAGTACTTAGGCAAAAAGTACTTAAAA
GAGCTGAGGCCGGGCGTGGTAGCTCATGACTATAATCCCAGCACTTTGGGAGGCTGAGGCCGGCAGATCACCTGAGTCCAGTAGTTCAAGACAAGCCTGGCCA
ACATGGTGAAACCCTGTCTCTACTAAAAATACAAAATTAGCTGGGCCCATACCTGTAATCCCAGCTACTAGGGAGGCTGAGGCAGGAGAATTGCCTGAGCCTG
GGAGGCAGAGGTTACGGTGAACTGAAATTGTGCCACTGCACTCCAGCCTGGGCAGCAGAATGAGACTGTCTCAAAAAACAACAACAAAAAAAAGTAAAAGGGC
TGGAGAAGTGGAAGTCAGGGTCACTCCTCCGATGTTGCCGTCAGAGTTCAACAAACCAATTCCGCTTCTGCTGCTATTGAACTGTCTCAGTCTCCTGAAGTTG
GTGACGGACAGTGGACTCCAGCGTCTAGCCACCACACCCACACTAGAAAGGATTACTGGCTATAAATAGCGCAATAATAGGATCAGGCCGGGCATGGTGGCTC
ACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGATCGAGACCACCCTGGCCAACATGGCAAACCCCGTCTCTACTAA
AAATACAAAAATTAGCCAGGCGTGGTGGCACATGCCTATAGTCCCAGCTACTTGGGAGGCTGAGGTGGGAGAATCCCTTGAACCCGGGAGGTGGAGGTTGCAG
TGAGCCGAGATTGCGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCGTCTCAAAAAAAAAGATTACTTGTCAGGCTCCACAGTCCCAGGATGATGGC
CTCTGTCCTCCTTCTGCCTTCCAAATCTCTCAGGAGAGAATCTCATTAGCAGAAATCGCAGTCAAGTCCACAGCCTGCTGCAAGGACCTCTGGGAAGTGTGTT
GGGCAGGATTGGGCACTGAGTCAGTAGGAAGTATCTCACACAGTGATCTCTCAAGAGGACTTCCTTTTCAGACTTGGCAGGGAGAAGAGTCTCCACTGCTTGA
TGGCAAAAAATTCTCCCATGAGGGAGCAAGTTCCTGGGCCACAGCCCTACCCGACCCTCTGGCTTTATAGTAACCTTATTGCCCATTAAATAGTCAACTCAGC
TGCTCAGAAGGGTTGGGATGTAGGGGAATAAAGGCCCACAGTTGGATCCCACAAGTGTAGTTGATGTTATTTGGGTAACATTGCACAGTTTGAAAACAGTGGG
CACCCTGCTGCCAGCCCAGAATAGAGCTATACACAAATCCTGCTTAGGGCCCCTTCCTTTAAGGCCTCAGGCCCTTGCCGCTCTAGGACCTTCTCCTTGGCGT
TCTGGTTCTTTGTCACAAGATTAAGCCACACTCTTCCATGGTCAAGAGTGTGGGCTCTGGACCAGACTGCTGGGGTTCAAGTCCTGGCTCTGCTATTAATACT
TTTAAGTTAAATGACCTTAGGCAAATTAGTCAGTGTCCTCTGTCAATGGGGTAATAAACAATTCTGATATGAGAGATGAGCAGCACATTAAAACCAGAAGTAT
TTTCAGTGAAACTAGCTGCAAGGTGGCAGTAATATAGTCGCTTTGTAATAACCAAAATGTTCAGGCTGATCATGGTGGCTCACATCTATAGTCCCAGCACTTT
GGGAGTCTGAGGCAGGAGGATTGCTTGAGGCCAGGAGTTCAAGACCAGTGTGGGCAACATAGTGAGACCTCATCTCTCCAAAAAACTAAAGTTAGCTGAACAT
GGTGGCACGTGCCTGTAGTCGCAGCTATCTGGGAGGCTGAGGCGGTAGGATCACTTGAGCCCAGGAGTTTTGGTTGCAGCGAGCTATGATTGTGCCACTGTGC
TCCATCCTGGGCAATAGAGTGAGCCTGTGTCTCTGAAAAACAAACCAAAAAAACCCCAAAACATTCACTCAATCAAATTGTTATTGAGGCTGGGCGCAGTGGC
TCACGCCTGTAATCCCAGCACTTTGGGAGGCCTAGGCGGGCAGATCACAAGGTCAGGAGATCGAGACCATCCTGGCTAACATGGTGAAACCCTGTCTCTACTA
AAAATACAGAAAAAATTAGCCGGGCGTGGTGGTGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAAGCAGGAGAATGGCGTGAACCTAGGAGACGGAGGTT
GCAGTGAGCTGAGATCGTGCCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAATTGTTATTGAATGCCTCCTCATAGA
CTCTGCTAGGAGCTGGAGTAGCAACTGTAAGCCTCCTTTCTCCTTCCCTTACGGAGATAATAATATTTATGATATAGGATGTTGAGAAGATTTTTTTTTTCTT
TTGAGGCGGAGTTTTGCTCTTGTTGCCCAGGCTGGAGTGCAATAGCGCAATCATGGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTCCTGCCTCA
GCCTCCCGAGTAGCTGGGATTATAGGCATGTGCCACCACACCCGGCTAATTTTCTATTTTTAGCAGAGACAGGGTTTCTCCATGTTGGTCAGGCTGTGTCTCG
AACTCCTGACCTCAGGTGATCCACCCGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCACGCCCAGTCGAGAAGATTTTTAATTTTTTTTAG
TCTTCCACAACAGTCATCTTATGTTGAGAAGACTTAATCAGATAATCCATGTAAAGGATTCAGTGTAATGTGTGGCATATGATAAATAATGTTGGCCAGGCAC
CTATAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCGCTTGAGGTCAGAAGTTCAAGACCAGCCTGGCCAACATGATGAAACCCCGTCACTACTAAAAA
TACAAAAAAAAATTGGCTGAGCGTGGTGGTGCGCCCCTGTAATCCCAGCTGCTCGGGAGGCTGAGGCAAGAGAATTGTTTGAACCCAGGAGCCAGAGGTTACA
GTGAGCTGAGATCACACACTGCACTCAGCCTGGGCGTCAGAGCAAGACACCATCTCAAAATAATAAATAAATAAATAATGTTAATGATATTCTTTTTTTTTTC
TTTTTTTGAGATGGAGTCTCACTCTGTCGCCCAGGTTGGAGTGCAGTGGTGCGGTCTTGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGC
CTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCCCGCCACCACACCCGGCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGG
TCTCGATCTTCTGACCTCGTGATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTGCCCAGCTATAAATAATGTTAATGATATA
TTCTAATGGAGCATGTGAGAAAGAGAAATATTCTATTGTTTTGTTGTTTTATAGACACCATCAAGAGAAATTGGAGGCTAGTGACTGTGATCACCAGCAGAAC
TCACCTACTCTAGAAAGGCCTGGAAGGAAGAGGAAGTGGACTGAAACACAAGATTCTAGTCAAAAGAAATCCCTAGAGCCAAAAACAAAAGGTTTGTATGTTT
TAGCAGGACCCACGGACTAGAGTTGATAAGATTCTGTAGCTGAGCAAACCCCTGTGTATTTTTTTTTTTTTAGATGGAGTCTCGCTCTATCACCCAGGCTGGA
GTGCAGTGGCATGATCTCAGCTCACTGCAAGCTCTGCCTCCTGGGTTCATGCCATTCTCCTGCCTCAGCCTCCTGAGCAGCTGGGACTACAGGTGCCTGCCAC
CACGCCTGTCTAATTTGTTTGTATTTTTAGTAGAGACGGGGTTGCACCATGTTAGCCAGGATGGTCTTGATCTCCTGACCTCATGATCCGCCCACCTCGGCCT
CCCAAAGTGCTGGGATTACAGGCGTGAGCCACAGCGCCTGGGCCTATTTTTTTTGTTTTATAGAGATGGGGTCTCACTATTTTGGCCAGGTTGGTCTTGAAC
TCCTGGCCCCAAGCAGTCCTCCTGCCTTGGCCTCCCAAAGTGTTAGGATTACAGGTGTCAGCCACTATGCCTGGCCTTTATTTTCATTTTATATTGCATACTT
GTGGGTGTCTGGAAAATGCCCCTTGTTAGTTGCATGAAGTGAAGGAGAATAAAAGGCACTTGGAGCTTATTTCTCCTGCAGTGGTCTCTAATCTTTGAGGTGT
CCTGGGTCTCTTTTGGGAATCTGATTCTCATAAAATAAAAGCTCATAAAAACTGTGAACTGTCATTCTTTTTTTTTTTTTGAGACAGAGTTTTGCTCTTGTTG
CCCAGGCTGGAGTGCGATGGCACAATCTCGGCTTACTGCAACCTCTGCCTCCCGAGTTCAAGTGATTCTCCTGCCTCAGCCTCTCAAGTATCTGGGGTTACAG
GCGTGCGCCACCACACCCGGCTAATTTTGTATTTTTAGTAGAAACGGTATTTCACCATGTTGGTCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTACCC
GCCTCGGCCTCCCAGAGTGCTGGGATTACAGGCTAGAGCCACCGCACTCGGCCTAAGCCCTCATCCTTGAAAAGCACATACGTATCTTCATTTTGATACAGTG
GGTTATTAGAAAAGTAAACCCCTTTCCACTTGGAGGAGGTAGAGGGGAAGAAAAAGCATACCCCAAAGCTCTGTTTTATTCTTCCCAGCTGTGCCAAAGGTCA
AGCTGCTGTGTGGGGCAGATTTATTGGAGTCCTTTGCTGTTCCCAATTTGTGGAAGAGTGAAGACATCACCCAAATCGTGGCCAACTATGGGCTCATATGTGT
TACTCGGGCTGGAAATGATGCTCAGAAGTTTATCTATGAATCGGATGTGCTGTGGAAACACCGGAGCAACATTCACGTGGTGAATGAATGGATCGCTAATGAC
ATCTCATCCACAAAAATCCGGAGAGCCCTCAGAAGGGGCCAGAGCATTCGCTACTTGGTACCAGATCTTGTCCAAGAATACATTGAAAAGCATAATTTGTACA
GCTCTGAGAGTGAAGACAGGAATGCTGGGGTCATCCTGGCCCCTTTGCAGAGAAACACTGCAGAAGCTAAGACATAGGAATTCTACAGCATGATATTTCAGAC
TTCCCATTTGGGGATCTGAAACAATCTGGGAGTTAATAACTGGGGAAAGAAGTTGTGATCTGTTGCCTAAACTAAAGCTTAAAAGTTTAGTAAAAATCGTCTG
GGCACAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGTGGATCACGGGGTCAAGAGATCGAGACCATCCTGGCCAATATGGTGAAACCC
CATCTCTACTAAAAATACAAAAATTAGCTGTGTGTGGTGGCACGTGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGACCCCAGGTGG
TGGAGGTTGCAGTGAGCCAAGATTGCACCATTGCACTCCAGCCTGGCGACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAATTTAGTAAAAATCAATGGTA
AGCTAAAATAAGTTTTTGTTTGTTTATTTGTTTTTGAGATGGAGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGCCGCATAACTATAATCCCAGC
TACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCACAGGTTCCAGTGAGCCAAGGTTGTGCCACTGCACTCCAGCCTGGGCAAAAAGCAAAA
CTCCATCTCAAAGAGAAAAAAAAAAAAAGACCGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCTAAGTGGGTGGATCACGTGAGGTCAAGA
GTTCAAGACCAGCCTGGCCAATATGGTGAAACCCCATCTCTACTAAGAATACAAAAAATTAGCTGAGCATGGTGGTGGGCTCCTGTAGTCCCAGCTACTTGGG
AGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTAAGCCAAGATCGTGCCAT
```

Longest deletion in chr2:
Length: 9582
Sequence: CAGGTAAAAACTACTGTCAAAAGTTGATATTTTTTGAAGTGAAGCAGTTATATCTTCTTTCTCTACCTATAGTACCTAACTCAACTTTTATGT
AAGAATCTTTTTACTTTCATTTAAACCTTTAGTTCATTGCTATCTAGTGTTTAGAAACATTATTAGCAGCAGGATGCAAATTAGGAATTTTTTTTTTAGAGT
ACTCCAGGTTTTCATCTGAGTGAAATTTAATTAATTGCCAGTAATAGCTTATCTAACTGTATTTTGTTTGGTCTCTGATTGCTTTAGCAGTAAACATTTTCCT
TATGCACCATCTGAGTGTTTGAGACAAGGTACACTTACGATGTTTACTGAGTTGAACCCAGAAGGAATTAGAAGCAATGAATTATTTGTAAACACTTTTTCCT
TTTCTTTTTTTTTCATGCTCCCTTTCTCCCTCCTTCCTTCTTTCCTTCTTTCTTCCTTCCTTCCTCTTCTCTCACCTTATCTCCTTCCCTTCCTCCTCTTTTC
TCTTCTTCCTTCCTTTCAATGAAAGAATACTCATTGGCCAGGTGCGGTGGCTCATGCCTATAATCCCAGCACTTTGTGAGGCCGAGGCTTGCGGATCACCTGA
GGTCAGGAGTTCGAGACCAACCTGACCAATATGGTGAAACCCCGTTTCTACTAATAATATGCACCTGTAATCCCAGCTGCTCAGGAGGCTGAGATAGGAGAAT
TGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCAAGATCTTGCCGTTGCACCCCAGCCTGGGCAACAAGAGCAAAACTCTGTCTCAAAAAAATAAAAAAAG
GCCTGGCGCGGTGGCTTACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGGCGGATCATGAGGTCAGAAGTTCGAGACCAGCCTGGCCAATATGGTGAA
ACCCCATGTCTACTAATAATACAAAAATTAGCCAGGCGTGGTGGCGAGCACCTGTAGTCCCAGCTACTCGGGAGGCTGGGGCAGAGGAATTGCTTGAACCCAG
TAGGCGGAGGTTGCAGTGAGCTGAGATCACGCCATGGCTCTCCAGCCTGGGTAACAGAGCAAGACTCCGTCTCAAAAAAAGGGAAAGAATACTCATTTAAGTG
AATGTTTGTTTGAGGCTTACTTGAAAATAGCCAAATGAATTTATAGTTTTCCCTTCTTAGGGATTACATCGTTGTAGTATCATCTAAAAACTTTAACTATAAAT
CTGGCATTGCATTCTTCTAGGTGCTATACAGATGATTTGATCTGTGTGGCCACTGTCTTTTAGGAGTAACACCATTAAAAGCATAGAAAAAGGAACTAGCATT
GGACAAATTTATGAAATAGAGCTATTTAATTTAAAAAGATTAATAATGGAGTTGAATTGTTTGGCTTTAAAGACATTAGGCATTACACATTAGTTTATATATA
TCAAGTGTGTGTATATATATATACACACACACACATATATATGTATCGTGTGTGTGTATATATATATAAATAAAATTTTTATTTATTTATTTATTTTA
TTTATTTATTTTTTATTGATAATTCTTGGGTGTTTCTCACAGAGGGGGATTTGGCTGGGTCATAGGACAATAGTGGAGGGAAGGTCAGCAGATAAACAAGTGA
ACAAAGGTCTCTGGTTTTCCTAGGCAGAGGACCCTGCGGCCTTCCGCAGTGTTTGTGTCCCTGGGTACTTGAGATTAGGGAGTGGTGATGACTCTTAACGAGC
ATGCTGCCTTCAAGCATCTGTTTAACAAAGCACATCTTGCACCGCCCTTAATCCATTTAACCCTGAGTGGGACACAGCACATGTTTCAGAGAGCACAGGGTTG
GGGGCAAGGTCACAGATCAACAGGATCCCAAGGCAGAAGAAGTTTTCTTAGTACAGAACAAAATGGAAAGTCTCCCATGTCTACTTCTTTCTACACAGACACG
GCAACCATCTGATTTCTCAATCTTTTCCCCACCTTTCCCGCCTTTCTATTCCACAAAGCCGCCATTGTCATCCTGGCCCGTTCTCAATGAGCTGTTGGGCACA
CCTCCCAGACGGGGTGGTGGCCGGGCAGAGAGGCTCCTCACCTCCCGGACGGGGCGGCTGGCCAGGCGGGGGGCTGACCCCCCCACCTCCCTCCCGGACGGGG
CGGCTGGCCGGGCGGGGGGCTGACCCCCCCACCTCCCTCCCGGACGGGGCGGCTGGCCGGGCGGGGGGGCTGACCCCCCCCACCTCCTGGCCGGGCGGGTGGCT
GACCCCCTGCCACCTCCCTCCCGGACGGGTCGGCTGGCCGGGCGGGTGGCTGACCCCCCCACCTCCCTCCCGGACAGGTCGGCTGGCCTGGCTGGGGGCTGAC
CCCCCCACCTCCCTCCCGGACGGGGCGGCTGGCCGGGCGGGGGGGCTGACCCCCCCACCTCCCTCCCGGATGGGGCGGCTGGCCTGGCGGGGGGCTGACCCCCC
CCACCTCCCTCCCGGACGGGGTGGCTGCCGGGCGGAGACACTCCTCACTTCCCAGACGGAGCGGCTGCTGGGCGGAGAGGCTCCTCACTTCTCAGACGGGGCG
GCTGCCGGGCGGAGGGGCTCCTCACTTCTCAGACGGGGCGGTTGCCAGGCAGAGGGTCTCCTCACTTCTCAGACGGGGCGGCCGGGCAGAGACGCTCCTCACC
TCCCAGACGGGGTCTCGGCTGGGCAGAGGCGCTCCTCACATCCCAGACGGGGCGGCGGGGCAGAGGCGCTCCCCACATCTCAGACGATGGGCGGCCGGGCAGA
GACGCTCCTCACTTCCTAGATGTGATGGCGGCCAGGAAGAGGTGCTCCTCACTTCCTAGATGGGATGGCGGCCGGGCGGAGACGCTCCCCACTTTCCAGACTG
GGCAGCCGGGCAGAGGGGCTCCTCACATCCCAGATGATGGGCGGCCAGGCAGAGACGCTCCTCACTTCCCAGACGGGGTGGCGGCCGGGCAGAGGCTGCAATC
TCAGCACTTTGGGAGGCCAAGGCAGGCGGCTGGGAGGTGGAGGTTGTAGCGAGCCGAGATCACGCCACTGCACTCCAGCCTGGGCACCATTGAGCACTGAGTG
AACGAGACTCCGTCTGCAATCCCGGCACCTCGGGAGGCCGAGGCTGGCGGATCACTCGCGGTTAGGGGCTGGAGACCGGCCTGGCCAACACAGCGAAACCCCG
TCTCCACCAAAACCAGTCAGGCGTGGCGGCGCGTGCCTGCAATCGCAGGCACTCGGCAGGCTGAGGCAGGAGAATCAGGCAGGGAGGTTGCAGTGAGCTGAGA
TGGCAGCAGTACAGTCCAGCTTCGGCTCGGCATGAGAGGGAGACCGTGGAAAGAGAGGGAGAGGGAGACCGTGGGGAGAGGGAGAGGGAGAGGGAGCTAAATA
AAATTTTTTTAAGAGACAGAGCCTCACTCTGTCACCTAGGCTGGAGTGCAGTGGTGTGATCACAGCTCACTGCAGCCTTGACCTCCCCAGGCTCAGGTGATTC
TCCCACCTCAGCCTCTGAGTAGCTGGGACTACAGGCATGCACCACCAAGCCTGGTTCATTTTTGTATTTTTTTGTAGAGACAAGGTTTCACCATGTTGCTCAG
GCTGGTCTCAAACTCCTGGGCTCAAGTGATCCACCCACTTCAGCCTCCAAAAGTGCCTGGATTACAGGAGTGAGCCACCGCACCAGGCCAAACATTAGTTTGT
AATTCAAAAATATAATGAAATAAATCAGTGATTAGTAGATTAAGTACAGGTTAATCAGTTATCTAGAAGTATATTTTCTAGTTAATTAAAAGTACAGAGATTC
AATTGAGGCATTTTTCTGAAAATAGCAGTAATTGCAGTAGTATCAACAAGTACTTTTTTTGGCCGGGTGCGGTGGCTCACGCCTGTAATCCCAGCACCTTAGG
AGGCCGAGGCAAGTGTATCACCTGAGGTCGGGAGTTCAAGACCAGCCTGACCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAATTAGCTGGATGTGG
TGGCACATGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAAGGCTGAGATTGCACCATTGCACTCC
AGCCTGGGCAACAAGAGTGAAACTCGGTCTCAAAAAAAAAAAAAAAAAAAAAAGAAGTACTTTTTTTGTTTTGTTTTTAAGAGACAGGGTCTCACTCCTGCCCAG
GCCGGAGTGCAGTAGTGCAATCATAGCTCACTGCAGCACTGAACTCTTGGGCTCAAGCAGTCCTCCTGCTCAGCCTCCTGAGTAGCTAGGACTACAGGTGCAT
GCCACCACGCCCGGCTAGTATTTTATTTTTTGTCGAGACAGTGTTTTGCTGTATTGCCCAGGCTGGCCTCGAACTCCTGGGCTCAAGTGATCCTCCTGCCTCA
GCCTCCCAAAATGCTGTGATTGCAGGCATGAGCCACCACACCCAAACCCACAAGTATTGTATTATCAGAACTTTGACCACAGTTTGGTAATTTCTTTGTTAAT
AGTCTCATGTAGGATTTTTGTTTTAGAAGCATGTCTGACTGAAATGATTTTTTCATCAGATTAGTACACTGCTACAAACAATATATTTTGAAAAACTACTGTG
TTCAAAGAATGACTATCGCTTTTGATCTGATAAATAATAAAGAATGATATTAAATGTTTGAATAGGGAAAGGCAATTAAATCTTGTTTTCAGAAGATTATTTT
AGCAAATATGTAGGATTAATTAAAATAGAAATAGAACTAGAAGGGGCTTAAAAAGGAAAAAGAAATGGAGAGTTACTGTTTAATGGGTACAGAGTTTCAGTTT
GGGAAGATAGAAACAGTTCTAGAAATGGTTAGTAGTAATGGTTGTACAATAGTGTGAATGTACTTAATACCACAGAATGGTGCACTTAAAAATGGTTAAAATG
ATCAGTTTTATGTTATGTATATTCTACCACGATAAAAAAAAAAATAAAGGAATAAGAAGAGAGTTGCTTTAAAATACCCCGTAATCTTGGATATGGAAATGGGG
AGAGGCAAATCTGGGAGAAAATGCAGAGAAGAAATCTGTACTTGTGGTTATCTGTTAAAAGAATGAGATGTCAGGCCAGGTGTGGCTCACGCCTGCAAACCCA
GCACTTTAGGAGGCTGAGAGGGGTGGGGGTGGATCTCCAGGTCAGGAGTTCGAGACCAGCCTGGCCAGCATGGTGAAACCCCGCTCTACTACAAATACAAAAA
TTAGACAGGTGCAGTGGCACACGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACCTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGAT
CGCGCCACTGCACTCCAGCCTGGGTGACAGAACGAGACTCCATGTCAAAAACTAAAAAGAATGGGATTCAAAGGAATTGAGAAATATTCAGAGGTTGGCCGGG
CACGGTGGCTCAGGCCTGTAATCCTAGCACTTTGGGAGGCTGAGGTGGGTGGATCATTTGAGGTCAGGAGTTCGAAACCAGCCGGCCAACATGGTGAAAGCC
TATCTCTACTAAAAATATTTAAAAATCAGCTGGTCGTGGTGGCAGTCACCTGTAATTCTAGCTACTTGGGAGTCTGAGGCAGGAGAACTGCTTGAACTAGGGA
GGCGGAGGTTGCAGTGAGCCGAGATTGCCTCCATTGCCCTCCATCCTGGGCAACAGAGCGAGACTCCATCTCTTAAAGAAAAAAAAGAAGGAACTATTCAGAG
GTCTTAAAACACAGTGAGGGTGGGCCAGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGCTAGTGGATCACCTGAGGTCAGGAGTTTG
AGACCAGCCTGACCAACGTGGTGAAACCCCAACTGTACTAAAAGTACAAAAAAATTAGCCGGGCTGGTTGTGGGCACCTGTAATCCCAGCTACTCGGGAGGC
TGAGGCAAGAGAATTGCTTGAACCCGGGACGCAGAGGTTGCAGTGAGCAGAGATCGCGCCACCGCACTCCAGCCTGGGTGACAAAGCGAGACTCTGTCTCAAA
AAAAAAAAAAAAAAAAAAAAGTGAGGGTCAGAGTGGCAACCAGAAAGATGAGTTAATCCTGGAGACTGTTCGATGGGGTATTGACAGACAACTAAATAAATTTT
TAAAGACAATTAAAAATAATAAATAAGATGGAAATTTAAGATTAAAATGGTTATTTGGGAGTTGTCTACATAGAGATGGTAACTGAGAAGTTATGAGTCGATG
CACTCCTTTAACTGGAACATAAAAAGAAAGAAAAATAGGTGAATAATGACTGAATATTGGGGTACTACCCCCATGAAGTAGGTTAATTAAGAAGACAAATGAT
CAGAAAAGTAGTGAGAAAACTGAAAAATCCTAATGCCTTGTAAGTCAGTATTAGAAAACTTTAAGAACTAAGATTTCAGCCTGAGGAAGCTGATTAAAAAATA
AAAATAAAAAACCACACCAAGATTGATGGCAATAGATGTAATTGAAAAGAGGAAACAGAACTGAGGAATGATTTTTGAATTTGAGTAGGAGGAGGTCATTGGT
ATTTTTAGAAGAGTTTCTGTGGAATACAGAACACAGAAATCATGTTACAGAGTGTATATAGGAGGGGAAAGTATTTAAAAAGCACAGTCAAGGGATACAGGGC
ACTCTTGAAAATTTTAACAATTAAATGAGGTGGAAGTAGAAAAGAAAAAAGGATCATATAAATATATTAATAAACCTAGAGAATAGTGTATACTAACTATAAA
ATTCTTATAAAACATTTTTTAAACTTTATGATATACCTCAAAATGCAGTATGTGCTGTATTATTAAATGGAAACACTATGTAATGAAGGTTACTTCTCTATA
AAATGTTAGTAAAATAAATGATTTATCTGTATTACAGAGGAAGAGTGAAATGCTACAAAGTTTAAATAGAAGGGGGAGAAGGAAAAAAAAGAAGAGATAGA
AAAAGGAAGAAAGAATAAGTGTAATTTGAGAAGGGAGCTAAAAAAAGGTCTGTAACTGGTAGAGACATAGTTAACAAGTGGCACATGGGGTGAGAAGCAGGTA
ATGGTATCTGATAGTGCTAGATATTACTATAAGATATTTTGGGCAGTTCTAAAGAAAATGAGATTATAAAACTTTCCTTGAGCAATCCAGCGATAGTAAGGAA
ATTCAGTAAGGTTTTAAGGAAATCCAATAAGGTTATTATCTTCCATGATGATGATTTTTGTCATTAACATTCACCTTTGTATCTCTAATACCTAACACTAAGT
TAGCACTAAACAAGTTTCTATTGAACAATTGAATTTTGAACTGTGAAGATTTAGTGGAAATTTATTTGGTAATTCTTAACCATTGCATTCATAAAAGATTGTG
CTTCAATTTCCTTTAGCAACTTTTAAAATTATAATTTGGACCTAATTTTTTAAAAGAGTGAAATTGTGAATTTTTAGTTATTCTACTAAACCAGCATTTGAAA
AAAAAATAATAAATAACACTTGGAAGGTATGTGTATGTGTTGTGTGTATGTCTCAGAGACAGATCAGATACATTAGTAAATACCTTGCAATTCACAGATGGTA
TAGGATAAGGTTTTATCACATGGGGTTTCAAGGAAGAAAGTGATATACTAACTGCCATGATTATTAAATCCTTTAATTTTATCTCTCAAGTTCAATGTTGATG
TTATTGATCAACACTAAGTATCTGCTTAAATGTTTACAGTATGTGAAACTCACTTGTATTTATACATAAGGCAGTATTACAACAAACTTGGGCGATGCTGCAC
AGATAATAGAATTAATCTTTGTAAATCTTTTTTAATATTGGTGGAGGGCAAGTATTTATTAGTTGTTTTGAGTATTCAAATAGAAACAAAACTTAAATGAAAC
ATAAACTTATTTGTAATGTGGTTTCCATCTGGGTCATTATAACTTTGTATTATCTCCATGTAAGTGATATAAGGAAGGCTGCAGACATTGCCATTTTGCTTAC
AGGAAGTCTGAAACACAGAATAGTTAAATGACTTTTCCAAGGTCAAGATGAAGTGAAGATTAGAACCTAAATGCCTTATCCTCTGTCTTGGCGTTCTTCCTTC
TAATTTATCTTGTCATTTTAATAGGGTGGGAGGACGGGTAAAGAGACAGGAACAGTCTCTGTGTTCTTTGCCATGGTTTCTACTAGATCATTGAGTGTTGGTC
CCAGGGGTCTTCTATGTGGTTCAATATATTATATACATAAACTTAATTTTATTATCATATTTTGTAAAAATCAAAAGCCTAACATCAGAGTTGATTGTGCTTA
TTGGCAATGAATAGGGTTATTAAATAGTGATAAAATAGTTAATAATATTTAAGTCCATGATCTAAATACTACCTGTTAGTTAAAAGGAATTTTTAGAGCTGGG
CACAGTGGTTCACACCTGTAATCTCAGCACCTTGGGAGGCTGAGGCAGGAAAATCATTTGAGGGCAGAAGTATGAGACCAGCCTGGGCAACACGGCAAGACCT
TATCTCTACAAAAATATTTTTAAAAATTAGCTGGGCATGGTGCCACACTCTTGTAGTCTTAGCTATTAGGGAGGCCACTTGAGTGCAGGAGTTCAAGTTACAG
TGAGCTATGATAATGCCACCGTACTCCATCCTGGGCAACAAAGTGAGAGCGTTTCTTTGAAAAAAAATTTTTTTAAAGTATATTGATTTTTCTTTCAAATGAC
CATTATCATAAATATTTTTGTAACATTTGATACAGTCTGGCTGGACTATAATAGAACCAATGATGTAATGAGACATAGTGACTGTTCAACCCACCCACTTACA
TTTGGTTCTGTAGGCATATTTGTTGTACTAGGTCCTGGAGTTTTTCCTCATCTTTACCCGCCTCCATGGGATCAGTGTTCAGGCATGTTTTCAAGTATTAAGT
ATGCCTTATAAACATAATAGTCTTTTTTCCCTGAATTTAGTAATACCAGAAACCATTTAAAAATACCTATCCATCATTGAAACTACCATTTGACTTATCAACA
AATATTTGAGACTATTCTAAGCAAAACATTCATTAATAATGCATTAGACTGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTAGGAGGCCGAGGCTGGT
GGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTATCTCTACTAAAAATACAAAATTAGCTGGACGTGGTGGTGTGCGCCTGT
AATCCAAGCTACTCGGGAGGCTGAGGCGAGAGAATCACTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCAAC
AGAGTAAGACTCCGTCTAAAAAAAAAAAAAGAAAAAAAAATCATTAAACCTACAGCAAGGTCTCTTTAGGATTTCCAAATGTGCCTGAAGGGGATGAAAAAAT
AAAAAGATTGTGACACAATTTTTTTTTGCTTACTTGGTATCAGAAATACCCCTGTTATTCTATCATTTATGATAGTTAGAAATTTTATTCTGTCATTCTTTTCT
ACAAATCCACAGG

```
#!/bin/bash

INSERTION_CHROM="1"
DELETION_CHROM="2"

awk -v ins_chrom="$INSERTION_CHROM" -v del_chrom="$DELETION_CHROM" '
BEGIN {
    max_ins_len = 0
    max_del_len = 0
    longest_ins = ""
    longest_del = ""
}
/^#/ { next }

{
    chrom = $1
    ref = $4
    alt = $5


    if (chrom == ins_chrom && length(ref) < length(alt)) {
        ins_len = length(alt) - length(ref)
        if (ins_len > max_ins_len) {
            max_ins_len = ins_len
            longest_ins = substr(alt, length(ref) + 1)
        }
    }

    # Поиск делеций в другой хромосоме (длина REF > ALT)
    if (chrom == del_chrom && length(ref) > length(alt)) {
        del_len = length(ref) - length(alt)
        if (del_len > max_del_len) {
            max_del_len = del_len
            longest_del = substr(ref, length(alt) + 1)
        }
    }
}

END {
    print "Longest insertion in chr" ins_chrom ":"
    if (longest_ins == "") {
        print "NOT FOUND"
    } else {
        print "Length:", max_ins_len
        print "Sequence:", longest_ins
    }

    print "\nLongest deletion in chr" del_chrom ":"
    if (longest_del == "") {
        print "NOT FOUND"
    } else {
        print "Length:", max_del_len
        print "Sequence:", longest_del
    }
}
' clinvar_20250330.vcf > fout
~
```