**Teoretická část:**

1. Data z NGS sekvenování (získaná ze vzorku lidské krve) ve formě fastq souborů mají být namapována na gen HBB.

a) Kde a jak byste získali fasta sekvenci celého HBB genu (pro verzi hg19)?

Nejprve přes search engine a projít databáze NCBI; RGD a ENSEMBL, a zjistit, kde mají záznamy o dané verzi genu.

Jelikož verze hg19 je sysonymum pro verzi genomu GRCh37, nejjednodušší je použít ENSEMBL databázi, kde je na výběr ve fasta jak celý gen, cDNA, tak i introny a exony.

b) Vložte níže sekvenci fasta, kterou byste použili jako referenci pro mapování.

Pokud bychom mapovali na celý gen.

>11 dna:chromosome chromosome:GRCh37:11:5246694:5250625:-1

TTGGAAAAGCAACCCCTGCCTTGAAGCCAGGATGATGGTATCTGCAGCAGTTGCCAACAC

AAGAGAAGGATCCATAGTTCATCATTTAAAAAAGAAAACAAAATAGAAAAAGGAAAACTA

TTTCTGAGCATAAGAAGTTGTAGGGTAAGTCTTTAAGAAGGTGACAATTTCTGCCAATCA

GGATTTCAAAGCTCTTGCTTTGACAATTTTGGTCTTTCAGAATACTATAAATATAACCTA

TATTATAATTTCATAAAGTCTGTGCATTTTCTTTGACCCAGGATATTTGCAAAAGACATA

TTCAAACTTCCGCAGAACACTTTATTTCACATATACATGCCTCTTATATCAGGGATGTGA

AACAGGGTCTTGAAAACTGTCTAAATCTAAAACAATGCTAATGCAGGTTTAAATTTAATA

AAATAAAATCCAAAATCTAACAGCCAAGTCAAATCTGCATGTTTTAACATTTAAAATATT

TTAAAGACGTCTTTTCCCAGGATTCAACATGTGAAATCTTTTCTCAGGGATACACGTGTG

CCTAGATCCTCATTGCTTTAGTTTTTTACAGAGGAATGAATATAAAAAGAAAATACTTAA

ATTTTATCCCTCTTACCTCTATAATCATACATAGGCATAATTTTTTAACCTAGGCTCCAG

ATAGCCATAGAAGAACCAAACACTTTCTGCGTGTGTGAGAATAATCAGAGTGAGATTTTT

TCACAAGTACCTGATGAGGGTTGAGACAGGTAGAAAAAGTGAGAGATCTCTATTTATTTA

GCAATAATAGAGAAAGCATTTAAGAGAATAAAGCAATGGAAATAAGAAATTTGTAAATTT

CCTTCTGATAACTAGAAATAGAGGATCCAGTTTCTTTTGGTTAACCTAAATTTTATTTCA

TTTTATTGTTTTATTTTATTTTATTTTATTTTATTTTGTGTAATCGTAGTTTCAGAGTGT

TAGAGCTGAAAGGAAGAAGTAGGAGAAACATGCAAAGTAAAAGTATAACACTTTCCTTAC

TAAACCGACATGGGTTTCCAGGTAGGGGCAGGATTCAGGATGACTGACAGGGCCCTTAGG

GAACACTGAGACCCTACGCTGACCTCATAAATGCTTGCTACCTTTGCTGTTTTAATTACA

TCTTTTAATAGCAGGAAGCAGAACTCTGCACTTCAAAAGTTTTTCCTCACCTGAGGAGTT

AATTTAGTACAAGGGGAAAAAGTACAGGGGGATGGGAGAAAGGCGATCACGTTGGGAAGC

TATAGAGAAAGAAGAGTAAATTTTAGTAAAGGAGGTTTAAACAAACAAAATATAAAGAGA

AATAGGAACTTGAATCAAGGAAATGATTTTAAAACGCAGTATTCTTAGTGGACTAGAGGA

AAAAAATAATCTGAGCCAAGTAGAAGACCTTTTCCCCTCCTACCCCTACTTTCTAAGTCA

CAGAGGCTTTTTGTTCCCCCAGACACTCTTGCAGATTAGTCCAGGCAGAAACAGTTAGAT

GTCCCCAGTTAACCTCCTATTTGACACCACTGATTACCCCATTGATAGTCACACTTTGGG

TTGTAAGTGACTTTTTATTTATTTGTATTTTTGACTGCATTAAGAGGTCTCTAGTTTTTT

ATCTCTTGTTTCCCAAAACCTAATAAGTAACTAATGCACAGAGCACATTGATTTGTATTT

ATTCTATTTTTAGACATAATTTATTAGCATGCATGAGCAAATTAAGAAAAACAACAACAA

ATGAATGCATATATATGTATATGTATGTGTGTATATATACACACATATATATATATATTT

TTTCTTTTCTTACCAGAAGGTTTTAATCCAAATAAGGAGAAGATATGCTTAGAACCGAGG

TAGAGTTTTCATCCATTCTGTCCTGTAAGTATTTTGCATATTCTGGAGACGCAGGAAGAG

ATCCATCTACATATCCCAAAGCTGAATTATGGTAGACAAAACTCTTCCACTTTTAGTGCA

TCAACTTCTTATTTGTGTAATAAGAAAATTGGGAAAACGATCTTCAATATGCTTACCAAG

CTGTGATTCCAAATATTACGTAAATACACTTGCAAAGGAGGATGTTTTTAGTAGCAATTT

GTACTGATGGTATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGTTTGAAGTCC

AACTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTC

ACCCTGTGGAGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGA

GCCAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACA

CAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAA

GTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT

GGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCATG

TGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCT

ATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCT

TTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCA

AGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCT

TTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGG

TGAGTCTATGGGACGCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTC

ATAGGAAGGGGATAAGTAACAGGGTACAGTTTAGAATGGGAAACAGACGAATGATTGCAT

CAGTGTGGAAGTCTCAGGATCGTTTTAGTTTCTTTTATTTGCTGTTCATAACAATTGTTT

TCTTTTGTTTAATTCTTGCTTTCTTTTTTTTTCTTCTCCGCAATTTTTACTATTATACTT

AATGCCTTAACATTGTGTATAACAAAAGGAAATATCTCTGAGATACATTAAGTAACTTAA

AAAAAAACTTTACACAGTCTGCCTAGTACATTACTATTTGGAATATATGTGTGCTTATTT

GCATATTCATAATCTCCCTACTTTATTTTCTTTTATTTTTAATTGATACATAATCATTAT

ACATATTTATGGGTTAAAGTGTAATGTTTTAATATGTGTACACATATTGACCAAATCAGG

GTAATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTA

TCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATC

ATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCA

ATATCTCTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAGGTTTCATAT

TGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGC

TGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTTCATACCTCTTATCTTC

CTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAA

TTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTG

GCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTG

TTCCCTAAGTCCAACTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTC

TGCCTAATAAAAAACATTTATTTTCATTGCAA

c) Co považujete za vhodný první krok ke zpracování fastq souborů?

Důkladný quality control: read coverage, kvalita readů, GC poměr v očekávané mezi, zjistit přítomnost overrepresented sekvencí.

d) Jaký tool byste použili k samotnému mapování?

Bowtie2 nebo STAR align

f) Jaké onemocnění se váže s genem HBB a jakou má dědičnost?

beta thalassemia – autosomální recesivní (obě alely musé být přítomny aby se nemoc plně projevila)

methemoglobinemia - autosomální recesivní

HbSS disease - autosomální recesivní

a další jako produkce hemoglobinu C a E

2. U jednoho z pacientů byly nalezeny varianty právě v HBB genu zobrazené v tabulce níže.

U každé varianty popište, zda byste ji považovali za patogenní a brali ji v potaz v další fázi reportování. Dle čeho jste se rozhodli? Označili byste daného pacienta jako zdravého, přenašeče či postiženého onemocněním vázaným s genem HBB?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Start | Ref | Alt | ExonicFunc\_refGene | AAChange\_refGene | CLINSIG | amplicon | variant\_depth | read\_depth | variant\_frequency |
| 5248233 | T | A | stopgain | exon1:c.20A>T:p.E7X | . | HBB\_exon1 | 1751 | 3535 | 0.5 |
| 5248244 | A | G | nonsynonymous SNV | exon1:c.8A>C:p.H3P | Pathogenic | HBB\_exon1 | 3557 | 3564 | 1 |
| 5247792 | C | G | nonsynonymous SNV | . | . | HBB\_exon2 | 8 | 602 | 0.01 |

M1. Jako patogenní ji nepovažuji, i když se jedná o změnu na stop kodón, protože VF je 0.5 značí, že druhá alela je v pořádku, ale je přenositelná.

M2. Tuto mutaci považuji za patogenní; VF = 1 obě alely jsou zasažené.

M3. Zde je VF blízká nule což znamená, že tato část je blízká referenci.

Jelikož M1 stopgain je před M2, znamenalo by to, že se M2 nemusí projevit u jedné z alel. To znamená, že jedna alela bude kódovat zkrácený gen a druhá s M2. To by znamenalo, že daný jedninec je bude i tak postižený onemocněním vázaným s genem HBB.