Ve složce Data jsou k dispozici nasimulovaná data z analýzy target sequencing. Pro každého pacienta jsou k dispozici výsledky zvlášť v csv souboru (Cislo\_protokolu\_Jmeno.out.csv). Dále máte k dispozici excel s informacemi o pacientovi (KLIENTI.xlsx).

Z výsledků analýzy cíleného sekvenování Vás budou zajímat výsledky analýzy trombofilních mutací (TROMBO\_full a TROMBO\_basic ve sloupečku analysis).

Napište skript, který bude automaticky tvořit reporty pro toto vyšetření do samostatné složky Reporty. Report se bude ukládat jako Word a bude pojmenovaný tak, jak je pojmenovaný ukázkový report. Ukázkový report nazvaný 002\_19\_Novotna\_TROMBO\_CZE.pdf je k dispozici ve složce Ukazka, společně s obrázky hlavicka.png a repromeda\_line.png. Programovací jazyk si zvolte podle vlastních preferencí.

Podmínky k reportování jednotlivých genotypů jsou uvedené v tabulce:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Název varianty | Pozice | Záměna | Variant frequency | Genotyp |
| F5; G1691A (Leiden) | 169519049 | T>C | >0.8 | G/G(WT) |
|  |  |  | 0.2-0.8 | G/A(HET) |
|  |  |  | <0.2 | A/A(HOM) |
| F2; G20210A (Protrombin) | 46761055 | G>A | >0.8 | A/A(HOM) |
|  |  |  | 0.2-0.8 | G/A(HET) |
|  |  |  | <0.2 | G/G(WT) |
| MTHFR; C677T | 11856378 | G>A | >0.8 | T/T(HOM) |
|  |  |  | 0.2-0.8 | C/T(HET) |
|  |  |  | <0.2 | C/C(WT) |
| MTHFR; A1298C | 11854476 | T>G | >0.8 | C/C(HOM) |
|  |  |  | 0.2-0.8 | A/C(HET) |
|  |  |  | <0.2 | A/A(WT) |

Pozici, záměnu a variant frequency máte k dispozici v csv souborech daných pacientů.

Následující úkoly zpracujte v R:

1) V csv souborech jednotlivých pacientů máte k dispozici i výsledky GJB2 analýzy. Např. u 002\_18\_Novakova.out.csv najdete nalezenou variantu v tomto genu. Vytvořte v R následující větu, která se vypíše pouze do konzole:

*Molekulárně genetickým vyšetřením genu GJB2 byla u pacienta zjištěna přítomnost patogenní varianty c.101T>C (p.M34T) v heterozygotním stavu.*

Výstup maximálně zautomatizujte, hranice variant frequency pro určení heterozygotního/homopzygotního stavu jsou následující: heterozygot 0.2-0.8, homozygot >0.8. zápis mutace včetně proteinového zápisu čerpejte taky automaticky z csv sloupeček AAChange\_refGene. Některé varianty jsou pokryté více amplikony a tudíž mohou být zacallované dvakrát, pro zjednodušení pracujte s prvním záznamem ve výsledné tabulce.

Pokud u pacienta není žádná varianta a nebo je zacallovaná varianta s četností pod 0.2, pak se vypíše věta:

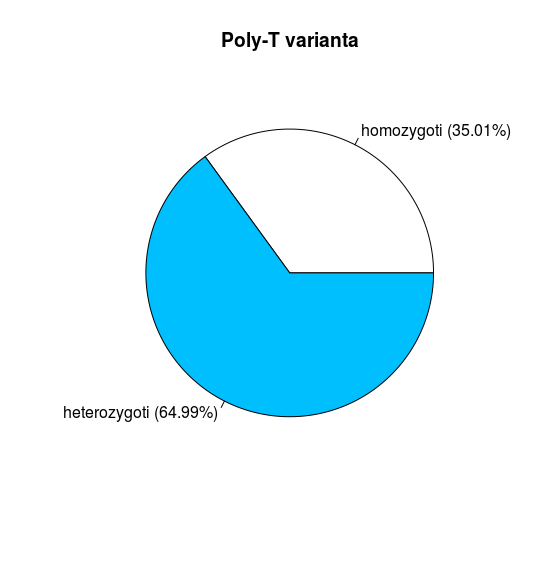
*Nebyla detekovaná varianta v GJB2 genu.*

2) K dispozici máte datový soubor cftr.csv, který obsahuje výsledky z CFTR analýzy. V jednom sloupečku jsou uvedené výsledky poly-T alelické varianty a ve druhém sloupečku jsou uvedené záměny, inzerce či delece ve standardizovaném zápisu. V případě, že je u pacienta uvedeno c.[=];[=], pak u tohoto pacienta nebyla nalezena žádná záměna, inzerce či delece. V případě heterozygotů u poly-T varianty jsou jednotlivé nálezy oddělené středníkem, v případě homozygotů je za středníkem uvedena pomlčka. Pokud pacient nese variantu 5T a počet TG repetic je alespoň 12 (tento zápis by byl jako 5T12TG), pak je přenašečem pro onemocnění. Vytvořte následující statistiky a tabulky (stačí výstupy v konzoli, není třeba výsledky ukládat do wordu/ppt apd).

a) Zjistěte počet pacientů pouze s poly-T variantou nebo pouze se záměnou/inzercí/delecí a nebo s poly-T variantou i záměnou/inzercí/delecí.

|  |  |
| --- | --- |
| 5T(>=12TG) | 20 |
| MUT | 96 |
| MUT + 5T | 1 |

b) Vykreslete koláčový graf ukazující zastoupení homozygotů a heterozygotů pro poly-T variantu.



c) Zjistěte počet jednotlivých variant ve sloupečku CFTR\_mutace.

d) Zjistěte alelickou četnost jednotlivých poly-T variant.

