

HW2

answer and explanation

TA 劉昕恩

r12630156@ntu.edu.tw

問題1

- How many exons does CYP2D6 have? (you are suggested to blastn NG_008376.4_cds with NG_008376.4_complete to find the answer)
- Ans: 9

Explanation:

- Tool: blastn
- Query sequence : NG_008376.4_cds
- Subject database / sequence : NG_008376.4_complete

問題1

| | | | | | | | | |
|-----------------------------------|-------------|---------|-----|---|---|------|------|------|
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 98.985 | 197 | 1 | 1 | 978 | 1174 | 8170 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 100.000 | 182 | 0 | 0 | 1313 | 1494 | 9056 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 100.000 | 180 | 0 | 0 | 1 | 180 | 5020 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 100.000 | 180 | 0 | 0 | 665 | 844 | 7459 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 100.000 | 176 | 0 | 0 | 178 | 353 | 5899 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 100.000 | 166 | 0 | 0 | 505 | 670 | 6866 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 99.375 | 160 | 1 | 0 | 347 | 506 | 6620 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 100.000 | 145 | 0 | 0 | 842 | 986 | 7826 |
| lcl NG_008376.4_cds_NP_000097.3_1 | NG_008376.4 | 100.000 | 145 | 0 | 0 | 1172 | 1316 | 8817 |

9 Matches

History



search datasets



imported: Shared History (HW2 data)

14 shown, 1 hidden

5.55 GB



14: megablast NG_0083
76.4_cds.fasta vs 'NG_0
08376.4_complete.fasta'



9 lines

format: **tabular**, database: ?

問題2

- Is CYP2D6 on the positive or negative strand of human chromosome 22? (you might find the answer directly on the web)

Ans: negative

Reference:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5591463/>

Yang Y, Botton MR, Scott ER, Scott SA. Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. Pharmacogenomics. 2017 May;18(7):673-685. doi: 10.2217/pgs-2017-0033. Epub 2017 May 4. PMID: 28470112; PMCID: PMC5591463.

問題3, 7

- Which sequence in HG002.1 is the one most likely to contain CYP2D6?
- Ans: HG002#1#JAHKSE010000065.1
- Which sequence in HG002.2 is the one most likely to contain CYP2D6?
- Ans: HG002#2#JAHKSD010000035.1

問題3, 7

Explanation:

- Tool: blastn
- Query sequence : NG_008376.4_complete
- Subject database / sequence : HG002.1.fa.gz / HG002.2.fa.gz
- Find the Match Contig in Tabular Result

問題3, 7

Match Contig



Subject Start / End



| | | | | | | | | | | |
|-------------|---------------------------|--------|-------|-----|----|------|-------|----------|----------|----------|
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 99.691 | 11317 | 29 | 3 | 1 | 11312 | 8715775 | 8727090 | 0.0 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 94.665 | 5717 | 209 | 44 | 4156 | 9849 | 8706274 | 8711917 | 0.0 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 89.275 | 3739 | 299 | 61 | 5548 | 9238 | 8698015 | 8701699 | 0.0 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 98.780 | 1475 | 6 | 3 | 9839 | 11312 | 8713471 | 8714934 | 0.0 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 89.359 | 780 | 54 | 13 | 4783 | 5557 | 8696340 | 8697095 | 0.0 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 98.929 | 467 | 5 | 0 | 1 | 467 | 8727930 | 8728396 | 0.0 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 87.978 | 549 | 33 | 9 | 3618 | 4161 | 8705729 | 8706249 | 3.29e-17 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 91.753 | 291 | 22 | 2 | 207 | 496 | 29318685 | 29318974 | 1.00e-10 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 88.959 | 317 | 30 | 5 | 192 | 506 | 23383190 | 23382877 | 1.01e-10 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 89.404 | 302 | 30 | 2 | 206 | 507 | 10102911 | 10102612 | 1.69e-10 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 92.720 | 261 | 19 | 0 | 208 | 468 | 38317 | 38577 | 6.07e-10 |
| NG_008376.4 | HG002#1#JAHKSE010000065.1 | 89.597 | 298 | 29 | 1 | 206 | 503 | 156349 | 156644 | 6.07e-10 |

History



search datasets



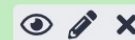
imported: Shared History
(HW2 data)

6 shown

2.16 GB



6: megablast NG_00837
6.4_complete.fasta vs 'H
G002.1.fa.gz'



問題4, 8

- Is CYP2D6 on the positive or negative strand of the sequence in your answer of Q3?
- Ans: Positive
- Is CYP2D6 on the positive or negative strand of the sequence in your answer of Q7?
- Ans: Positive

問題4, 8

Explanation:

- We can use the order qstart – qend and sstart – send in tabular blastn results to find out it is positive or negative

• E.g.

| | qstart | | qend | sstart | | send |
|--------|--------|---|-------|---------|---|---------|
| Case 1 | 1 | ➡ | 11312 | 8715775 | ➡ | 8727090 |
| Case 2 | 1 | ➡ | 11312 | 8727090 | ⬅ | 8715775 |

- In Case1 we get +/+ result, which means it is in positive strand
- In Case2 we get +/- result, which means it is in negative strand
- Pairwise Result explanation:

<https://support.nlm.nih.gov/knowledgebase/article/KA-05228/en-us>

問題5, 9

- Which variant does HG002.1 NOT have?
- Ans: 2577 C>T

- Which variant does HG002.2 NOT have?
- Ans: 1658 C>T

問題6, 10

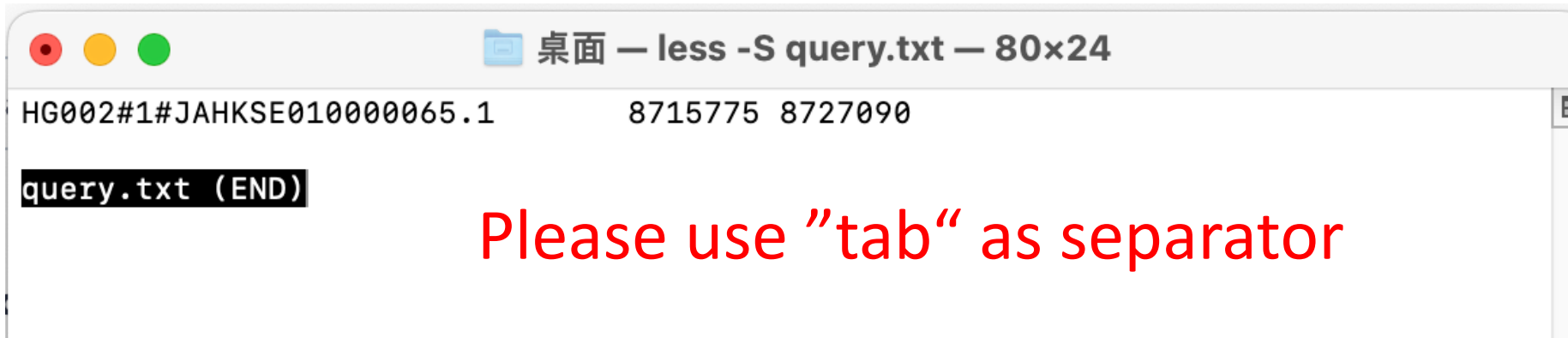
- What is the allele type of CYP2D6 in HG002.1?
- Ans: CYP2D6*4.015

- What is the allele type of CYP2D6 in HG002.2?
- Ans: CYP2D6*2.001

問題5, 6, 9, 10

Explanation

- Using bedtools to get the required fasta region
- Tool: bedtools
- Query: query.txt
- Fasta file: HG002.1.fa.gz



A terminal window titled "桌面 — less -S query.txt — 80x24" displays the output of a bedtools query command. The output shows a single line: "HG002#1#JAHKSE010000065.1 8715775 8727090". Below this line, the text "query.txt (END)" is displayed. The terminal window has a standard macOS-style title bar with red, yellow, and green window control buttons.

```
HG002#1#JAHKSE010000065.1      8715775 8727090
query.txt (END)
```

Please use "tab" as separator

問題5, 6, 9, 10

Bedtools Result

```
>HG002#1#JAHKSE010000065.1:8715775-8727090
```

```
TGTCTGCCAAGGGTGCCAGCACAGCCGCCCCACTCCAGGGGAAGAGGAGTGCCAGCCCTTACCCACCTGAGTGGGCACAGTGTAGCATTATTTCATTAGCCCCACACTGGC
CTGACCATCTCCCCGTGGGCTGCATGACAAGGAGAGAGAACAGGCTGAGGTGAGAGCTACTGTCAACACCTAAACCTAAAAATCTATAATTGGGCTGGGCAGGGTGGCTC
ACGCTGTAATCCCAGCACTTTGGGAGGCCGAGATGGGTGGATCACCTGAGGTGAGATGTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAA
AAAATTAGCTGGGCGTGGTGGTGGGTGCCGTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCACAC
CATTGCACTCCAGTCTGGGTGATAAGTATGAAACGCCATCTCCAAAACAAAAGAAAAGCCTAATCCCCAAGAACTGTCAGTCTTTCACCTGTCTGCTAGCTCCCAGGGAGA
CCCCACTTGGCAGGGCTGTCTACATTTGTCCTGAGATCTCTTCTGGTGGGAACAGCACTTTCCTCAGGAAAGTTGTTGAAAGTCATCAGATCCATGATTGAAAATCGAAGC
TGCCTGTGGTGATGGATAACAGCTGGGGTTAAAAAGCAGCAGCTGGGGCATGAGCGGTCCACAGTGAGTTTTTGTTGTTTGTGTTTTTGGGTGGGGGATGGGGTCTT
GCTAGGTCTCAAACCTCTGGCCTCAAGTCATCTCTCCATTACAGCCTTCTGAGTCACCTGACACTACAGGTGTGAGCCACCATGTCCAGCTTGTAGTGGTTTTGAACAGCTCT
TGCCCCTTCTTGGGAATCTAGGTGCCCTGCACGTGGGTAAGGCTGTCTGCAGCTGTGCCATATTCAGGAAGGCCGCAAGGCCCTGAGCCCTCACCCGTGACTGACCTGAG
GTGCTGTGCAGACAGCAGGTGACGGCTAAGGGAAGTTGAGCACTGCCTAGCCGAGCACTGAAGCCACGCCCGGCACACAGAGAGAGACCCACTCGGCAAAGACTTCGCTTC
CAGGCACCTAAGGAACTCTCTGACCAGTCATTAGCTGACCCTGCCGTAACCTGAAGAGCGGCTTCAGTGGCCACAGCTCGCAGGGAATGGAGACATTAATGCTTAGTCAGAA
TTAGTTTCAGAAAGTCAACCAGCAAAGAAACAGCTCCAACAGGCAACAACAACACATCTTGGCAGGAGAAGAAATCTGACTTCCGGAGTTGCCACATTATCGCCCGTG
AAATGTCCAGGTTTTAAACAAATTATGAGACATGGAAGGAAACCGAAAGGACGACCCAGCAACGGAAGGACCAATGGGACACAGCCGATGCTGCAATTGCTAGACAAA
GACGTTCACTGAGCTCATTTAAATATGTTCAAAGACCTAAACATGCTGCATCTGAGGCTGCACCGGCTGGAACCTGCTGATCTCGGAAGCTAAGCATGGTCAGGCCTGGCT
AGTACTTCAAAGGGAGAAACCAGCTGTAGGCCTGGTGCAGTGGCTCACACCTATAATCTTAGCACTCTGGGAAGCTGAGGCCGTGGATTGCTTGAGCCAGGAGTTTGAGA
GCAGCTTGGGAAATGTGGTGAGACCCCCATCTCTACAAAAAATTTAAAAAATTAGCTGGCTGCCATATGGTCCCAGCCTCTCAGGATGCTGAGGTAGGAGGATCACTTCAGCC
CAGGAAGTTGAGGCTGCAGTGAGCCATGACTGCATCACTGCACCTCCAGCTTGGGCGACAGAGAGACCCCTCTCCAAGAAAAAGAAAAGAACCATGTCAAAGAACTAACGAA
AGTGTGGGAACAATGTCTCACCAATTAGAGAATATCAATAATGGGATGAACCTTATAAAAAGGGGCTGGGCATGGTGGCTCATGCCATAATCCCAGCACTTTGGGAGGCTG
AGGCGGGCATATCATGAGGTCAAGAGATTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTTAAAAATACAAAAATTAGCCGGGCGTGGTGGCAGCTGCCTGTAA
TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGAGAGGCAGAGATTGCAGTGAGCCGAGATTGCACCACTGCACCTACAGCCTGGGTGACAGAGCGATACT
CCAAAAACAAAACAAAACAAAACAAAAGTTTAAAAAGGAACCAATAAAAAATCTTGGAGTTGTAGGGTAAAAATAATGAAATTCATCCCAGGGGCCAAGAGC
AGATTGGAACAATTGGAAGAAAGAGCCTGTGACTATGGAGAGAGGCCACCTGAGGTAGTCCCCCTCTGAGGAACAGGAACAGCATGAAGAGCAATGCACAGAGATCCAGAGA
CCTGGAGACGCCGTCAAGCTTTCCGACATACACAAATGGGAGTCCCAGGAAGAAGACAGGGAGAAAGGAGTAAAGGAATAGTTGAAGAATTAATGGCTGAAAAACCTCCC
AAATCTGATGAAAAATATTAATCCGTACATCCAAAAAGCTCATCAAACTCAAGTAGGGTAAACTCAAAGAGATCTTCAGGCATACGCATCATCAATCACTGTCAAAGA
CAGATTTTTCTTTTTTTTAGAATTTTAAATGTACCTTTTAATTTGCTCCTGGGGCAAAGAGCCAGGACTGGTACTAGAGCAGTGTCTGGGATGAGAAGAATTTAATAAAATGG
GATTAGGTCCAATGGTTGGGTTAGGGGAGGCAACCTGCTCGGAAGGATCAGCCTCAACCTATCCATGCAGCAGGGCCTCCACCTGTCCCTCTCCGTAGTCCCACACCTGGAA
CCCAGAGCCATCTGCCTCTTCCAGATCATGGCCGACAGCACTCCACCGACTGCTGCTGGAGCAGGCACAGGATTCACTTATTGAGGGCTGTGGCCTGGCACAGATCATAG
CCTATACCCAGGGACAGTTGTGTCACTTCTGCCACCACCACATCCGCCTTCTGCAGCCACATCAAGTACCACCTCATGGATGAGCCCGTCACCCCCAGCGGACTTATCAACCC
CGCGTCCAGCTCCACAGCCGCCACGTGCTCGGTGAGCACTGGCTCCAAGCATGGCAGCTGCCATACAATCCACCTGTAGAGGGCCCCGGTCCCTCTGTCTCAGTGGATGATC
CCGTAGAAGTCCAGAGCTCGGACGCTGCCCTCCCACAAAAGACAGGATTTTGAAGCAGCAAGAGAGAAGAGACGTATCAGGTAGTACAGTGGCTCAGGCCGTGTAATCCCA
GCACCTTTGGGAGGCCAGGTGGGAGGATCGCTTACCCCAGGAATTCAGACCAGCCTGGACAACCTTGAAGAACCCTGGTCTCTACAAAAAATACAAAATTAGCTGGGATTG
GGTGGGCTGGCTCATGCCTATAATCCAGCACTTTGGGAGCCTGAGGTGGGTGGATCACCTGAAGTCAGGAGTTCAAGACTAGCCTGGCCAACATGGTGAAACCCCTATCTCT
ACTGAAAAATATAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGCTGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCG
```

History



imported: Shared History (HW2 data)

8 shown

4.95 GB



8: bedtools getfasta on d ata 2 and data 7



1 sequences

format: **fasta**, database: ?

index file input.fasta.fai not found,
generating...



display with IGV local





```
>HG002#1#JAHKSE010000065.1:8715775-8727090
TGTCTGCCAAGGGTGCCAGCACAGCCGCCCCACTCCAGGGG/
GATGGTGGATCACCTGAGGTGAGTGTTCAGAGACGACCTC/
```



問題5, 6, 9, 10




- Using blast again to check the allele
- Tool: blastn
- Query sequence : fasta file after bedtools
- Subject database / sequence : CYP2D6.haplotypes.fasta




| | | | | | | | | | | |
|---|--------------|--------|-------|----|---|------|-------|------|-------|----|
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 99.850 | 11316 | 11 | 3 | 1 | 11315 | 2 | 11312 | |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 88.346 | 266 | 26 | 4 | 1967 | 2228 | 207 | 471 | 9. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 88.346 | 266 | 26 | 4 | 206 | 470 | 1966 | 2227 | 9. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 84.647 | 241 | 31 | 4 | 3478 | 3713 | 218 | 457 | 7. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 84.647 | 241 | 31 | 4 | 217 | 456 | 3477 | 3712 | 7. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 84.100 | 239 | 33 | 2 | 3478 | 3713 | 1977 | 2213 | 4. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 84.100 | 239 | 33 | 2 | 1978 | 2214 | 3477 | 3712 | 4. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 79.348 | 184 | 28 | 7 | 3339 | 3519 | 219 | 395 | 2. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.015 | 79.348 | 184 | 28 | 7 | 18 | 394 | 2339 | 3518 | 2. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 99.841 | 11316 | 12 | 5 | 1 | 11315 | | 11312 | |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 88.346 | 266 | 26 | 4 | 1967 | 2228 | 207 | 471 | 9. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 88.346 | 266 | 26 | 4 | 206 | 470 | 1966 | 2227 | 9. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 84.647 | 241 | 31 | 4 | 3478 | 3713 | 218 | 457 | 7. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 84.647 | 241 | 31 | 4 | 217 | 456 | 3477 | 3712 | 7. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 84.100 | 239 | 33 | 2 | 3478 | 3713 | 1977 | 2213 | 4. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 84.100 | 239 | 33 | 2 | 1978 | 2214 | 3477 | 3712 | 4. |
| HG002#1#JAHKSE010000065.1:8715775-8727090 | CYP2D6*4.014 | 79.348 | 184 | 28 | 7 | 3339 | 3519 | 219 | 395 | 2. |

Possible Alleles









History    

imported: Shared History
(HW2 data)
9 shown, 1 hidden
4.96 GB   

9: megablast bedtools g   
etfasta on data 2 and dat
a 7 vs 'CYP2D6.haplotypes.fasta'
4,500 lines
format: **tabular**, database: ?

'num_threads' is currently ignored
when 'subject' is specified.

問題5, 6, 9, 10

- We need to test which possible allele is exactly match in our genome, HG002.1.fa.gz / HG002.2.fa.gz. We use the pairwise output of blastn to check if the variants can be found at the exact location.
- Reference: <https://www.pharmvar.org/gene/CYP2D6>

問題問題5, 6, 9, 10 If we choose this variant



| | | |
|--------------------------------|---------|---|
| ↓ CYP2D6*4.015 | PV00387 | -1426C>T , -1235A>G , -1000G>A , 100C>T (P34S), 310G>T , 745C>G , 842T>G , 973C>A (L91M), 983A>G (H94R), 996C>G , 1662G>C , 1847G>A (splice defect), 2098A>G , 3385A>C , 3436C>A , 3583A>G , 4181G>C (S486T), 4402C>T |
|--------------------------------|---------|---|



If we want to test this allele

We may find the exact position at given Reference Sequence (in this case NG_008376.4)

| Variant Positions | | | | × |
|--------------------|----------|----------------------------|---------|--------------------|
| Gene | | NG_008376.4:g.3594C>T | | |
| Transcript | | NM_000106.6:c.-1426C>T | | |
| GRCh37 | | NC_000022.10:g.42528224G>A | | |
| GRCh38 | | NC_000022.11:g.42132217G>A | | |
| Reference Sequence | Position | Reference | Variant | |
| NG_008376.4 | | | | |
| Sequence Start | 3594 | C | > T | ← Expected variant |
| ATG Start | -1426 | C | > T | |



問題5, 6, 9, 10

```
> CYP2D6*4.015 PV00387 NG_008376.4 PharmVar Version:5.2.19.1
Length=11312
```

```
Score = 20798 bits (11262), Expect = 0.0
Identities = 11299/11316 (99%), Gaps = 6/11316 (0%)
Strand=Plus/Plus
```

- Tool: blastn
- Query sequence : fasta file after bedtools
- Subject database / sequence : CYP2D6.haplotypes.fasta
- Output format: Pairwise

```
Query 1      TGTCTGCCAAGGGTGCCAGCACAGCCGCCCCACTCCAGGGGAAGAGGAGTGCCAGCCCTT  60
            |||
Sbjct 2      TGTCTGCCAAGGGTGCCAGCACAGCCGCCCCACTCCAGGGGAAGAGGAGTGCCAGCCCTT  61

Query 61     ACCCACCTGAGTGGGCACAGTGTAGCATTTATTCATTAGCCCCCAGCTGGCCTGACCAT  120
            |||
Sbjct 62     ACCCACCTGAGTGGGCACAGTGTAGCATTTATTCATTAGCCCCCAGCTGGCCTGACCAT  121
```

}

```
Query 3540   AGGAGTTCAAGACTAGCCTGGCCAACATGGTGAAACCCTATCTCTACTGAAAATATAAAA  3599
            |||
Sbjct 3539   AGGAGTTCAAGACTAGCCTGGCCAACATGGTGAAACCCTATCTCTACTGAAAATATAAAA  3598
```



3594

We may find out at position 3594, genome HG002.1 have T, just like CYP2D6*4.015

問題5, 6, 9, 10

> NG_008376.4 Homo sapiens cytochrome P450 family 2 subfamily D
member 6 (CYP2D6), RefSeqGene (LRG_303) on chromosome 22
Length=11312

Score = 20698 bits (11208), Expect = 0.0
Identities = 11281/11316 (99%), Gaps = 6/11316 (0%)
Strand=Plus/Plus

- Tool: blastn
- Query sequence : fasta file after bedtools
- Subject database / sequence : NG_008376.4_complete.fasta
- Output format: Pairwise

```
Query 1      TGTCTGCCAAGGGTGCCAGCACAGCCGCCCCACTCCAGGGGAAGAGGAGTGCCAGCCCTT 60
          |||
Sbjct 2      TGTCTGCCAAGGGTGCCAGCACAGCCGCCCCACTCCAGGGGAAGAGGAGTGCCAGCCCTT 61

Query 61     ACCCACCTGAGTGGGCACAGTGTAGCATTTATTCATTAGCCCCCAGCTGGCCTGACCAT 120
          |||
Sbjct 62     ACCCACCTGAGTGGGCACAGTGTAGCATTTATTCATTAGCCCCCAGCTGGCCTGACCAT 121

}
```

We may find out at position
3594, genome HG002.1 have
T, while the reference is C

```
Query 3540   AGGAGTTCAAGACTAGCCTGGCCAACATGGTGAAACCCTATCTCTACTGAAAATATAAAA 3599
          |||
Sbjct 3539   AGGAGTTCAAGACTAGCCTGGCCAACATGGTGAAACCCTATCTCTACTGAAAATACAAA 3598
```



3594

問題5, 6, 9, 10

| | | | |
|-------|------|---|------|
| Query | 7560 | CTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCGAGACCTGACTGAGGCC | 7619 |
| | | | |
| Sbjct | 7557 | CTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCGAGACCTGACTGAGGCC | 7616 |

Sequence Start 7596, ATG start 2577

Taking the fifth question as an example, use 2577 as the ATG start number, correlate it with the sequence start number 7596 (use PharmVar), and then verify it on the NTU Galaxy.

Variant Positions

Gene

NG_008376.4:g.7596C>T

Transcript

NM_000106.6:c.802C>T

GRCh37

NC_000022.10:g.42524217G>A

GRCh38

NC_000022.11:g.42128215G>A

Reference Sequence

Position

Reference

Variant

NG_008376.4

Sequence Start

7596

C

>

T

ATG Start

2577

C

>

T

問題5, 6, 9, 10

- The variants in the allele should all be tested, since we are not sure about the exact variant by only using tabular output. That is, the values (e.g. bit score, evalue...) in blastn output can only determine the similarity between sequence. However, these values in tabular output can narrow down the alleles to a few, that we are most confident about.