

2.

Quelle: <https://www.ncbi.nlm.nih.gov/nucore/MG388092.1>

```
ORIGIN
  1 tccccacccc acgaccccc ggattctgat ccacaaatcc cccctcccta tgttgagcct
 61 acagcccccc aagtccttcc agtcatgcac ccacatggtg cccctcccag ccacgccc
121 tggcaaatga aagacctaca ggccattaag caagaagtct cccaagcagc ccctggaagc
181 cccagtttta tgcagaccat cgggcttgca gtgcagcagt ttgacccccc tgccaaagac
241 ctccaagacc tctgcagta cctttgctcc tccctcgtgg cttccctcca tcaccagcag
301 ctatatagct ttatatcaga ggcagaaacc cgaggtatta caggttataa ccccttagcc
361 ggtccccctc gtgtccaagc caacaatcca caacaacaag gattaaggcg agaataccag
421 caactctggc tcgccgcctt cgcgccttg ccaggagtg ccaaagaccc ttcctgggcc
481 tctatcctcc aaggcctgga ggagccttac cagccttcg tagaacgcct caacatagcc
541 cttgacaatg ggctgccaga aggcacgccc aaagacccca ttttacgttc cttagcctac
601 tctaattgcaa acaagaatg ccaaaaatta ctac
```

Nur die ersten 100 Stellen:

```
tccccacccc acgaccccc ggattctgat ccacaaatcc cccctcccta tgttgagcct
acagcccccc aagtccttcc agtcatgcac ccacatggtg
```

3.

DNA_sequence

```
Tccccacccacgacccccggattctgatccacaaatccccctccctatgttgagcctacagcccccaagtcctccagtcatgca
cccacatggtg
```

Aminosäuresequenz:

5'3' Frame 1

SPTHDPPDSDPQIPPPYVEPTAPQVLPV**MHPHG**

5'3' Frame 2

PPPTTPRILIHKSPLP**MLSLQPPKSFQSCTH****MV**

5'3' Frame 3

PHPRPPGF-STNPPSLC-**AY**SPSPSSHAPT**W**

Open reading frames sind gelb hinterlegt; grün=erste AS einer Sequenz; rot=Start des open reading frames

a)

Ich persönlich hätte immer die Genomsequenz der Aminosäuresequenz vorgezogen, da diese genauer ist. Während der Transkription könnten Fehler entstehen und ebenfalls wird jede Aminosäure als Basentriplett codiert wobei auch die Wobble-Hypothese dazu kommt. Somit denke ich wäre die Genomsequenz genauer, jedoch ist die Aminosäuresequenz natürlich bei weitem kürzer und enthält daher weniger Daten. Sie besteht aus maximal 1/3 der DNA. Es wird aber meist nur der open reading frame beachtet, wodurch die Datensätze noch geringer sind.

b)

Es ist wichtig alle zu durchsuchen, da eine AS durch 3 Basen codiert ist und es je nach Lesemuster zu einer Rasterverschiebung kommen kann. Somit kann es in verschiedenen Frames überhaupt nicht zu open reading frames kommen. Zusätzlich gibt es einen Leit- und einen Folgestrang und auf jedem kann die Sequenz vorkommen.

4.

Durchgeführt mit vollständiger DNA bzw AS-Sequenz: SPTHDPDSD PQIPPPYVEP
TAPQVLPV **MH PHGAPPSHRP WQMKDLQAIK QEVSQAAPGS PQFMQTIRLA VQQFDPTAKD**
LQDLLQYLCS SLVASLHHQQ LDSFISEAET RGITGYNPLA GPLRVQANNP QQQGLRREYQ
QLWLAAFAAL PGSADPSWA SILQGLEEY HAFVERLNIA LDNGLPEGTP KDPIRLSLAY
SNANKECQKL L



Die Suchsequenz und das Logo haben kaum Übereinstimmungen. Selbst an charakteristischen Stellen wie C172 oder C192 gleichen sie sich nicht.

5.

Quelle: <https://www.ncbi.nlm.nih.gov/nucore/BC002404.2>

Homo sapiens claudin 5, mRNA (cDNA clone MGC:8543 IMAGE:2822745), complete cds

ORIGIN

```

1 gaggtgcgac agaccgcgag ggcaaacgga ctggggccaa gagccgggag cgcgggcgca
61 aaggcaccag ggcccgccca gggcgccgag cagcacggcc ttgggggttc tgcgggcctt
121 cgggtgcgag tctgcctct agccatgggg tccgcagcgt ttgagatcct gggcctgggtg
181 ctgtgccttg tgggttgggg ggggtctgat ctggcgtgag ggctgcccac gtggcagggtg
241 accgccttcc tggaccacaa catcgtgacg gcgcagacca cctggaaggg gctgtggatg
301 tctgtcgttg tgcagagcac cgggcacatg cagtgcacaa tgtacgactc ggtgtctggct
361 ctgagcaccg aggtgcaggc ggcgcgggag ctcaccgtga gcgcgtgct gctggcgttc
421 gttgcgctct tctgaccctt ggcgggagcg cagtgcacca cctgcgtggc ccggggcccg
481 gccaaggcgc gtgtggccct caccggaggc gtgctctacc tgttttgagg gctgtgtggc
541 ctctgtccac tctgtgtggt cgccaacatt gtcttccgag agttttacga cccgtctgtg
601 cccgtgtgag agaagtagca gctggggcga gcgctgtaca tgggtgtggc ggccaccgcg
661 ctgctcatgg tagggcgctg cctcttctgc tgcggcgctt ggggtctgac cggcgtctcc
721 gacctcagct tccccgtgaa gtactcagcg ccgcggcgag ccacggccac cggcgactac
781 gacaagaaga actacgtctg agggcgcttg gcacggccgg gcccctcctg ccagccacgc
841 ctgagcggcg ttggataagc ctggggatcc ccgcagggac cgcggcttcc gccgggtagc
901 gcggcgcgca ggctcctcgg aacgtccggc tctgcgcccc gacgcggctc ctggatccgc
961 tccgtgctgc gccgcagct gacctctcc tgcactagc ccggccctgc ccttaacaga
1021 cggaatgaag tttccttttc tgtgcgagcg gctgtttcca taggcagagc ggggtgtcaga
1081 ctgaggattt cgcttccctt ccaagacgct ggggtgtctt gctgtgtgct tacttccag
1141 aggtcctgag tgaattcgga gggcgagatg cagagccagc ggccccacc ggaagatgtg
1201 tacacctggt ctttactcca tggcgaggcg ccgagccagc ggaccagtga cttggcctgg
1261 acctcccggt ctaactccag catctcccca ggcaaggctt gtgggcaccg gagcttgaga
1321 gagggcgagg gtgggaaggc taagaatctg cttagtaaat ggtttgaact ctcaaaaaaa
1381 aaaaaaaaaa aaaaaaaaaa a

```

Nur die ersten 100 Stellen:

gaggtgcgac agaccgcgag ggcaaacgga ctggggccaa gagccgggag cgcgggcgca
aaggcaccag ggcccgccca gggcgccgag cagcacggcc ttgggggttc

Aminosäuresequenz:

5'3' Frame 1

EVRQTRGANGLGPRAGSAGAKAPGPAQGAAQHGLGGSAGLRVRVSPLA **MGSAALEILGLVLCVGVGGGLIACG**
LPMWQVTAFLDHNIVTAQTTWKGLWMSCVVQSTGHMQCKVYDSVLALSTEVQAARALTSAVLLAFVAFVTLA
GAQCTTCVAPGPAKARVALTGGVLYLFCGLLALVPLCW FANIVVREFYDPSVPVSQKYELGAALYIGWAATALLMV
GGCLCCGAWVCTGRPDLSFPVKYSAPRRPTATGDYDKKNYV-GRWARPGPSCQPRLRGVG-AWGSPHGPRLPP

GSAARRLLGTSGSAPRRGWSIRSLRPQLTFSCH-PGALNRRNEVSFSVRGAVSIGRAGVRLRISLPLQDAGGLGCC
LTSQRLLLTSEGRMQSPGPPPEDVYTWSLLHRQGSPGTSDLAWTSRSHSSISPGKACGHRSLREGGSGKAKNLLSK
WFELSKKKKKKKKK

5'3' Frame 2

RCDRPAQTWDWGQEPGARAQRHQGPAPRSTALGVLRAFGCASRL-PWGPQRWRSWAWCCAWWAGGV-S
WRAGCPCGR-PPSWTTTS-RRRPPGRGCGCRAWCRAPGTCSAKCTTRCWL-APRCRRRGRSP-APCCWRSRSLSS-P
WRARSAPPAWPRARPRRVWPSREACSTCFAGCWRSCHSAGSPTLSSASFTTRLCPCRRSTSWAQRCTSAGRPPRC
SW-AAASCAAAPGSAPAVPTSASP-STQRRGGPRPPATTTTRTTTSEGAGHGRAPPASHACEALDKPGDPRMDRGF
RRVARRAGSSERPALRPDAAPGSAPACARS-PSPATSPALPLTDG MKFPFLCAALFP-AERVSD-GFRFP SKTLGVLA
ALLPRGSC-LRRGGCRAQGP HRK MCTPGLYSIGRARAQGPVTWPGPPGLTPASQARLVGTGA-ERAGVGRLRICL
VNGLSQKKKKKKKK

5'3' Frame 3

GATDPRGKRTGAKSRERGRKGTRARPGRRAARPWGFCGSPGARLASSHGVRVSGDPGPGAVPGGLGGSDPGVR
AAHVAGDRLPGPQHRDGADHLEGAVDVVRGAEHRAHAVQSVRLGAGSEHRGAGGAGAHRRERRAAGVRCALRD
PGGRAVHHLRGP GPGQACGPHGRRALPVLRAAGARATLLVRQHCRPRVLRPVCAEVRAGRS AVHRLGGH
RAAHGRRLLPLVLRRLGLHRPSRPQLPREVLSAAAAHGHRRLRQEELRLRALGTAGPLPATPARRWISLGIPAWTAA
SAG-RGAQAPRNVRLCAPTRLLDPLLPAPAADLLLPLARPCP-QTE-SFLFCARRCFHRQSGCQTEDFASPPRRWGS
WLLPYFPEAPADFGGADAEPAPTGRCVHLVFTPSAGPEPRDQ-LGLDLPVSLQHLPRQGLWAPELERGREWEG-E
SA--MV-TLKKKKKKKK

Durchgeführt mit AS-Sequenz:

EVRQTRGANGLGPRAGSAGAKAPGPAQGAAQHGLGGSAGLRVRVSPLAMGSAALEILGLVLCLVGVGGLILACG
LPMWQVTAFLDHNIVTAQTTWKGLWMSCVVQSTGHMQCKVYDSVLALSTEVQAARALT VSAVLLAFVALFVTLA
GAQCTTCVAPGPAKARVALTGGVLYLFCGLLALVPLCW FANIVVREFYDPSVPVSQKYELGAALYIGWAATALLMV
GGCLCCGAWVCTGRPDLSFPVKYSAPRRPTATGDYDKKNYV



Ebenfalls in diesem Beispiel gibt es kaum Übereinstimmungen. Selbst spezifische Stellen wie W28, W43 und C46 sind nicht übereinstimmend.