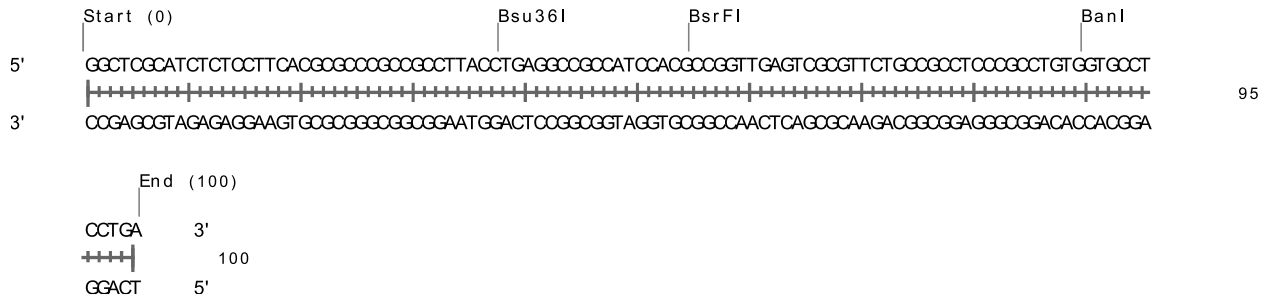


Aufgabe 2: Genom-Sequenz von Human T-cell leukemia virus type I (NC_001436)

Ersten 100 Basen:



Aufgabe 3: ersten 30 Aminosäuren des 1. 5'3' Frames

5'-MKFYTYSSKPKPRSRPAPRRRRRHPPPTTP-3'

- Warum ist die Suche in Aminosäuresequenzen der Suche in Genomsequenzen vorzuziehen?
Für die 20 Aminosäuren gibt es 64 verschiedene Codons. Falls nur die Ähnlichkeit des Proteins von Interesse ist, beziehungsweise es um die konservierte Natur von funktionell relevanten Motifen geht, ist deshalb die Aminosäuresequenz vorzuziehen. So geht unnötige Komplexität verloren (Kürzere Sequenzen und Verlust der potenziell verschiedenen Codons) und man vergleicht nur Elemente, die besonders konserviert sind (aufgrund des Selektionsdrucks). Es ist sinnvoll, die DNA-Sequenzen zu vergleichen, wenn man z.B. evolutionäre Distanzen (Abstammungen) bestimmen möchte, da hier auch Mutation in Sequenzen interessant sind, die keinen Selektionsdruck erfahren.
- Warum ist es sinnvoll, alle 6 möglichen Übersetzungs-Frames durchzusuchen?
Solange nicht bekannt ist, welcher Reading Frame im Organismus der tatsächlich exprimierte ist (oder hauptsächlich exprimierte), sollten alle Reading Frames geprüft werden, so dass man sicher sein kann, den korrekten gefunden zu haben.

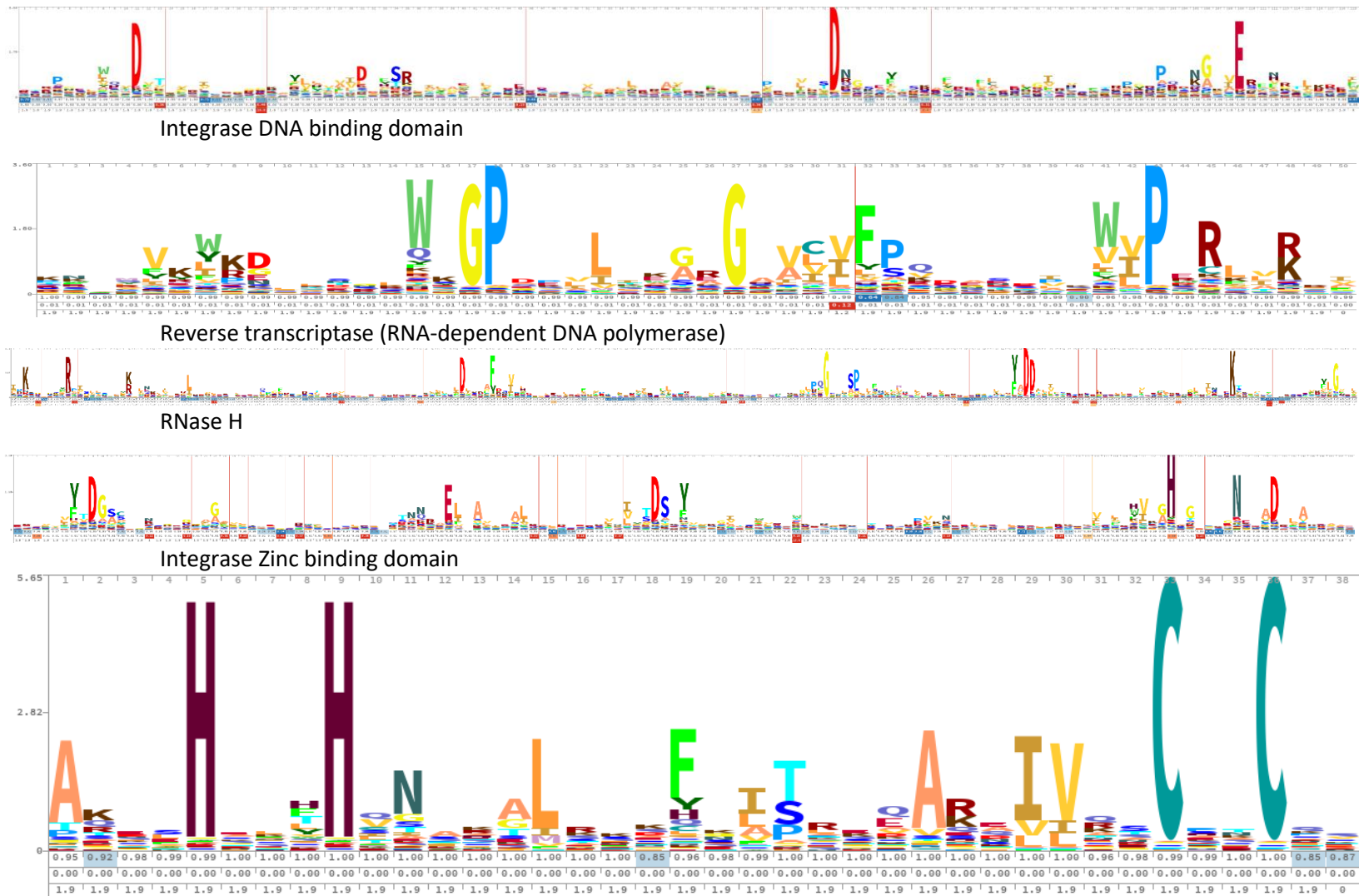
Aufgabe 4: Motive in Sequenz

Es konnten Integrase_Zn (Integrase Zinc binding domain), RNase_H, RVT_1 (Reverse transcriptase (RNA-dependent DNA polymerase)), IN_DBD_C (Integrase DNA binding domain), rve (Integrase core domain) domänen gefunden werden.

Family id	Family Accession	Clan	Bit Score	Ind. E-value	Cond. E-value	Description
rve	PF00665.25	CL0219	94.55	4.30E-27	1.30E-30	Integrase core domain
IN_DBD_C	PF00552.20	-	70.29	5.60E-20	1.70E-23	Integrase DNA binding domain
RVT_1	PF00078.26	CL0027	44.24	1.40E-11	4.20E-15	Reverse transcriptase (RNA-dependent DNA polymerase)
RNase_H	PF00075.23	CL0219	43.18	4.00E-11	1.20E-14	RNase H
Integrase_Zn	PF02022.18	-	38.68	6.50E-10	2.00E-13	Integrase Zinc binding domain

Logos:

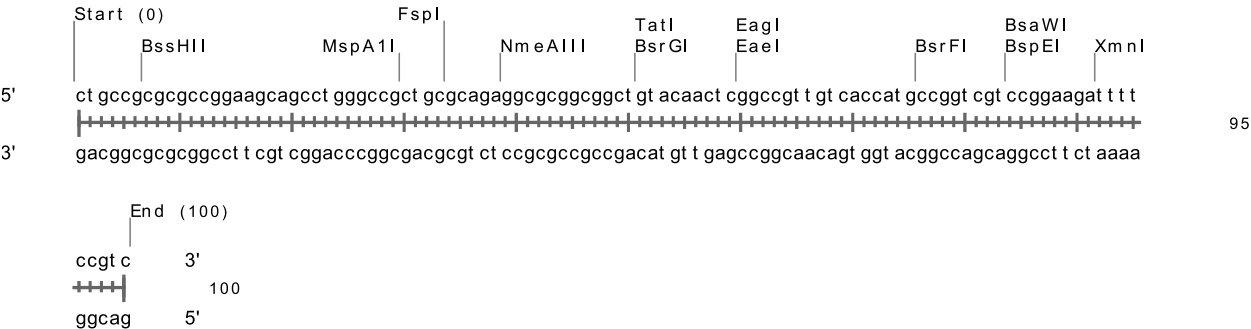
Integrase core domain:



Die E-Werte liegen alle zwischen 10^{-10} und 10^{-50} welches darauf hindeutet, dass die Sequenz sehr ähnlich zu den Logos ist. Dabei nimmt die Ähnlichkeit zwischen Sequenz und Logo von rve zu IN_DB_D_C zu RVT_1 zu RNase_H zu Integrase_Zn ab.

Aufgabe 5: Erneute durchführung mit C9orf78 (Homo Sapiens)

Ersten 100 Basen:



ersten 30 Aminosäuren des 1. 5'3' Frames:

5' – MPVVRKIFRRRRGDSESEEEDEQDSEEVRLK–3'

Es konnte Hep_59 (Hepatocellular carcinoma-associated antigen 59) als Motif gefunden werden.

Family id	Family Accession	Clan	Bit Score	Ind. E-value	Cond. E-value	Description
Hep_59	PF07052.10	-	103.68	7.70E-30	4.60E-34	Hepatocellular carcinoma-associated antigen 59

Logo:

Wie ähnlich ist das Logo zur Suchsequenz?

SESAETNRRDEDADMMKYIETELKKRKG-----VEHEEQVKPKNAEDCLYELPENIRVSSAKKTEEMLSNQMLSGIPEVDLGIDAKIKNIISTEDAKARL
rftketgkrdrvdkhMekYIeeELakrkkgkskqeqdeeeekeseassksedaedelyelpelhkvsskeeeeeealsnqmlggIPEVDLgieaklkNietekakrkL

Erstellt mit <https://www.genome.jp/tools/motif/>

übereinstimmende Aminosäuren: Hoch konservierte AAs

Sequence ID: Query_35643 Length: 106 Number of Matches: 1

Range 1: 2 to 106 Graphics			▼ Next Match ▲ Previous Match		
Score	Expect	Method	Identities	Positives	Gaps
93.2 bits(230)	3e-31	Compositional matrix adjust.	58/105(55%)	71/105(67%)	9/105(8%)
Query 2	FSAETNRRDEDADMMKYIETELKKRKG-----IVEHEEQVKPKNAEDCLYELPEN	52			
Sbjct 2	FTKETGKRVDVKHMEKYIEELAKRKKGSKQEQDEEEKESEASSKSEDAEDELYPEH	61			
Query 53	IRVSSAKKTEEMLSNQMLSGIPEVDLGIDAKIKNIISTEDAKARL	97			
Sbjct 62	LKVSSKEEEELSALNQMLGGIPEVDLGIEAKLKNIETEKAKRKL	106			

<https://blast.ncbi.nlm.nih.gov/Blast.cgi> 1 Blast von Logosequenz (jeweils AA mit höchster relative Entropie) und übereinstimmender Sequenz von C9orf78

Der Sequenzabschnitt von C9orf78 in dem Hep_59 gefunden wurde stimmt zu einem hohen Grad mit der Logosequenz von HEP_59 überein. So ist der Sequenzabschnitt in allen hoch konservierten Aminosäuren identisch und stimmt insgesamt zu 55% mit der Logosequenz überein. Der größte Unterschied kommt von einer Lücke in der Sequenz von C9orf78 mit einer Länge von 9 Aminosäuren. Der Domain-E-Wert von $4.6 * 10^{-34}$ gibt weiter an das die Sequenz in C9orf78 und die des HEP_59 Motifs eng miteinander verwandt sind.