GENETICS

PHENOTYPE OF THE CRISPR CAS BABIES

13.12.2018

CCR5 annotation, including the known delta32 Mutation and the three mutations introduced by He Jiankui. PAM=Protospacer adjacent motif, DSB=putative double strand break.

From the CCR5 sequence we get the following amino acid sequences

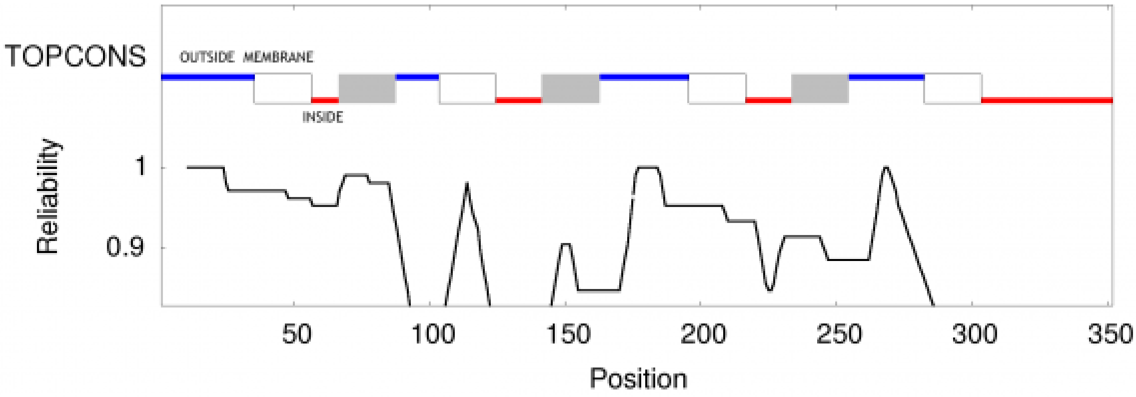
```
>sp|WT.P51681|CCR5_HUMANC-Cchemokinereceptortype5OS=HomosapiensOX=9606
GN=CCR5PE=1SV=1
MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVILILINCKR
LKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFII
LLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFASLPGIIFTRSQKEGLHYTCSS
HFPYSQYQFWKNFQTLKIVILGLVLPLLVMVICYSGILKTLLRCRNEKKRHRAVRLIFTI
MIVYFLFWAPYNIVLLLNTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHCCINPIIYAFV
GEKFRNYLLVFFQKHIAKRFCKCCSIFQQEAPERASSVYTRSTGEQEISVGL

>sp|-32.P51681|CCR5_HUMANC-Cchemokinereceptortype5OS=HomosapiensOX=960
6GN=CCR5PE=1SV=1
MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVILILINCKR
LKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFII
LLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFASLPGIIFTRSQKEGLHYTCSS
HFPY
IKDSHLGAGPAAACHGHLLLGNPKNSASVSK
```

```
>sp|-15.P51681|CCR5_HUMANC-Cchemokinereceptortype5OS=HomosapiensOX=960
6GN=CCR5PE=1SV=1
MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVILILINCKR
LKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFII
LLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFASLPGIIFTRSQKEGLHYTCS
SQYQFWKNFQTLKIVILGLVLPLLVMVICYSGILKTLLRCRNEKKRHRAVRLIFTI
MIVYFLFWAPYNIVLLLNTFQEFFGLNNCSSSNRLDQAMQVTETLGMTHCCINPIIYAFV
GEKFRNYLLVFFQKHIAKRFCKCCSIFQQEAPERASSVYTRSTGEQEISVGL

>sp|-4.P51681|CCR5_HUMANC-Cchemokinereceptortype5OS=HomosapiensOX=9606
GN=CCR5PE=1SV=1
MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVILILINCKR
LKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFII
LLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFASLPGIIFTRSQKEGLHYTCSS
HFP
YSINSGRISRH

>sp|+1.P51681|CCR5_HUMANC-Cchemokinereceptortype5OS=HomosapiensOX=9606
GN=CCR5PE=1SV=1
MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFGFVGNMLVILILINCKR
LKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQLLTGLYFIGFFSGIFFII
LLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFASLPGIIFTRSQKEGLHYTCSS
HFPY
KSVSILEEFPDIKDSHLGAGPAAACHGHLLLGNPKNSASVSK
```

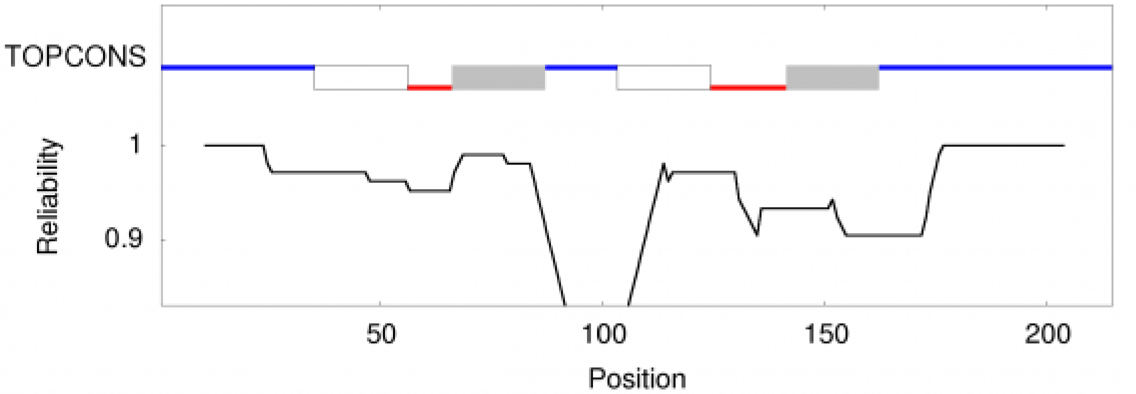
The -15/WT genotype will have both, a normal and a slightly shortened CCR5 giving no HIV protection at all. The -4/+1 genotype may suffer the fate of non-sense mediated decay. Here are the amino acid predictions for each genotype (I reverted to Topcocns as Protter had some problems in generating correct plots).

Predicted signal peptide and TM-helix positions (position starting from 1):

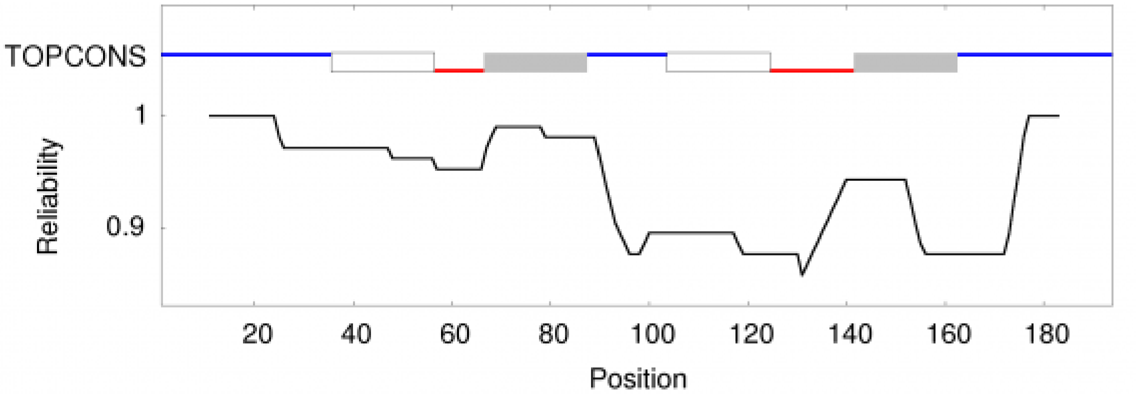
TOPCONS	TM1: 36-56,	TM2: 67-87,	TM3: 104-124,	TM4: 142-162,	TM5: 196-216,	TM6: 234-254,	TM7: 283-303
OCTOPUS	TM1: 35-55,	TM2: 67-87,	TM3: 96-126,	TM4: 129-159,	TM5: 196-216,	TM6: 231-251,	TM7: 283-303
Philius	TM1: 37-58,	TM2: 67-88,	TM3: 104-124,	TM4: 142-165,	TM5: 198-218,	TM6: 234-255	
PolyPhobius	TM1: 36-56,	TM2: 68-88,	TM3: 106-129,	TM4: 142-165,	TM5: 198-220,	TM6: 235-257,	TM7: 280-300
SCAMPI	TM1: 36-56,	TM2: 67-87,	TM3: 104-124,	TM4: 146-166,	TM5: 195-215,	TM6: 236-256,	TM7: 284-304
SPOCTOPUS	TM1: 35-55,	TM2: 67-87,	TM3: 96-126,	TM4: 129-159,	TM5: 196-216,	TM6: 231-251,	TM7: 283-303
4mbsA	TM1: 32-56,	TM2: 67-88,	TM3: 103-124,	TM4: 145-166,	TM5: 194-215,	TM6: 237-257,	TM7: 279-300



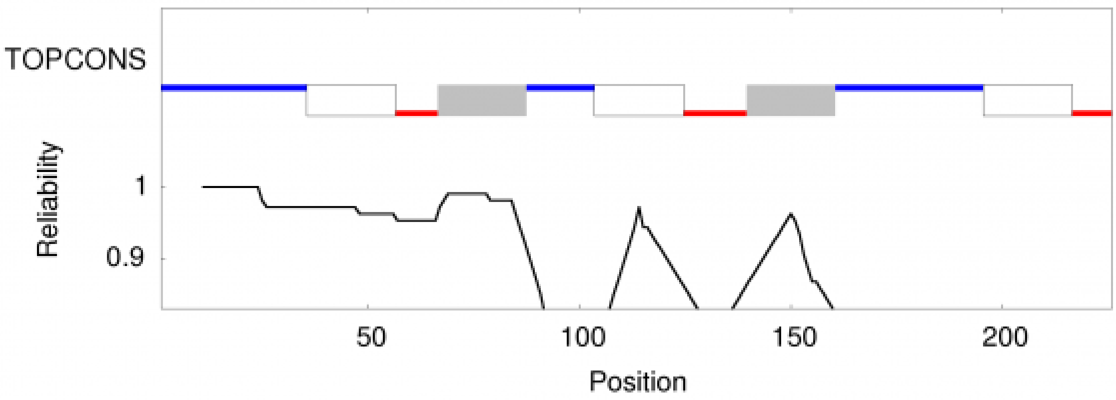
Wildtype: There are 7 transmembrane alpha helices I to VII, connected by three extracellular loops (ECL1–3) and three intracellular loops (ICL1–3). ECL2, forms a β-hairpin structure ([Tan 2013]).



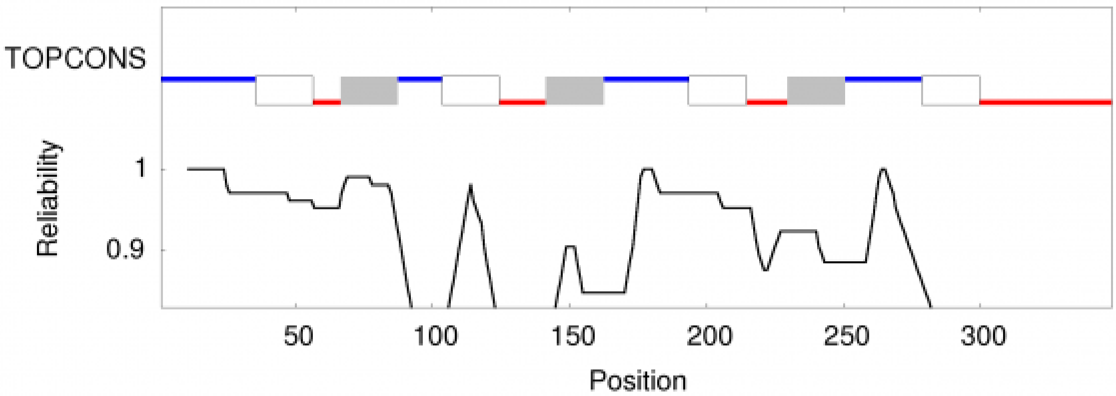
Known delta 32



Nana -4 genotype has a much shorter predicted amino acid sequence with missing G protein coupling. Resembles delta 32.



Nana +1 genotype is somewhat longer but missing helices VI and VII. No analogue known.



Lulu -15: resembles wild type, however, will be sensitive to allosteric changes by small-molecule CCR5 inhibitors .

From the [original AJHG paper](), however, also delta 32 carriers may be HIV infected. As there exists also non-CCR5 dependent virus replication, also Nana is at risk of HIV infection when being virus exposed.

Table 2

Clinical Description of Individuals with *CCR5* Alleles Altering Amino Acid Sequences

Variant (Patient)	Risk Group	Race	HIV-1 Status	AIDS Status[a]	No. of Years HIV−1 Positive and AIDS Negative[b]	*CCR5*-Δ*32* Genotype[c]
I12L	Hemophilia	Caucasian	Positive	Negative	13.6	+/*CCR5*-Δ*32*
C20S(1)	Homosexual	Caucasian	Positive	Negative	15.1	+/+
C20S(2)	Homosexual	Caucasian	Positive	Positive	12.2	+/+
A29S	IV drug user	African American	Positive	Negative	5.7	+/+
I42F	Homosexual	Caucasian	Negative	Negative	NA	+/*CCR5*-Δ*32*
L55Q(1)[d]	Hemophilia	African American	Positive	Negative	8	+/+
L55Q(2)	IV drug user	African American	Positive	Negative	7.4	+/+
L55Q(3)	IV drug user	African American	Positive	Negative	8.5	+/+
L55Q(4)	IV drug user	African American	Negative	Negative	NA	+/+
L55Q(5)	IV drug user	African American	Negative	Negative	NA	+/+
R60S	IV drug user	African American	Negative	Negative	NA	+/+
A73V(1)	Hemophilia	Caucasian	Positive	Positive	11.9	+/*CCR5*-Δ*32*
A73V(2)	IV drug user	Caucasian	Negative	Negative	NA	+/+
A73V(3)	Homosexual	Caucasian	Positive	Positive	7.8	+/*CCR5*-Δ*32*
C101X	IV drug user	African American	Positive	Negative	6.3	+/+
R223Q	Homosexual	Caucasian	Positive	Positive	13.1	+/+
228delK	IV drug user	Caucasian	Positive	Negative	11.3	+/+
G301V	Homosexual	Caucasian	Positive	Negative	15.2	+/+
A335V[e]	Hemophilia	Caucasian	Positive	Negative	14.9	+/+
Y339F(1)	Hemophilia	African American	Positive	Positive	10.5	+/+
Y339F(2)	IV drug user	African American	Positive	Negative	7.3	+/+
Y339F(3)	IV drug user	African American	Positive	Negative	2.8	+/+