

Lernblatt Molekularbiologie

Mutationen

Zur eindeutigen Bezeichnung von Mutationen in Genen und Proteinen nutzt man das von der Human Genome Variation Society (HGVS) vorgeschriebene System. Die Benennung der Mutation erfolgt dabei nach den im Folgenden beschriebenen Regeln. Die vollständige Nomenklatur ist nachzulesen unter varnomen.hgvs.org.

Zunächst gibt ein Präfix an, auf welcher Ebene die beschriebene Mutation liegt, dabei nutzt man:

- c. für eine Mutation auf dem codierenden Abschnitt eines Gens, also der cDNA (Diese beinhaltet keine Introns oder 5′-/3′-untranslatierte Regionen!)
- g. für eine Mutation auf Ebene einer linearen, meist genomischen DNA (inklusive Introns, regulatorischer und nichtcodierender Abschnitte)
- p. für eine Mutation auf Ebene eines Proteins
- r. für eine Mutation auf RNA-Ebene

Für Mutationen auf DNA- und RNA-Ebene wird zunächst die Position der Mutation angegeben und anschließend die Form der Mutation. Dabei verwendet man

- $N_1 > N_2$ für eine einzelne Substitution des Nukleotids N_1 in N_2 , mehrere voneinander getrennte Substitutionen werden durch ein ; getrennt
→ Bsp.: c.34G>T;62C>A für die Substitution des G an Position 34 durch ein T und des C an Position 62 durch ein A
- del für eine Deletion an der angegebenen Position (ohne genauere Angabe der deletierten Nukleotide)
→ Bsp.: c.125_128del für die Deletion der Nukleotide an Position 125 bis 128
- ins $N_1N_2N_3N_4$ für eine Insertion der Nukleotide N_1 bis N_4
→ Bsp.: c.47_48insATG für die Insertion der drei Nukleotide A, T und G zwischen Position 47 und 48
- delins $N_1N_2N_3N_4$ für eine Deletion an der angegebenen Position und deren Ersatz durch die Nukleotide $N_1N_2N_3N_4$
→ Bsp.: c.101_107delinsATG für die Deletion der Nukleotide an Position 101 bis 107 und deren Ersatz durch die drei Nukleotide A, T und G
→ Hinweis: Die Substitution mehrerer aufeinander folgender Nukleotide gilt ebenfalls als Deletion und Insertion, z.B. c.34_36delinsATG für die Substitution der Nukleotide an Position 34 bis 36 durch die drei Nukleotide A, T und G

Für Mutationen auf Proteinebene gibt man zunächst die ursprünglichen Aminosäuren mit ihren Positionen an und anschließend die Form der Mutation. Dabei verwendet man:

- AS_1XAS_2 für eine Substitution der Aminosäure AS_1 an Position X durch AS_2 , mehrere voneinander getrennte Substitutionen werden durch ein ; getrennt
→ Bsp.: p.C32S für die Substitution des Cysteins an Position 32 durch ein Serin
→ Hinweis: Eine Nonsense-Mutation, die ein Stoppcodon einfügt, wird als Substitution beschrieben, z.B. p.C32X für die Mutation des Cysteins an Position 32 in ein Stoppcodon
- $AS_1X_1_AS_2X_2del$ für eine Deletion der Aminosäuren AS_1 bis AS_2
→ Bsp.: p.D74_T78del für die Deletion der Aminosäuren 74 (D = Asparaginsäure) bis 78 (T = Threonin)

- $AS_1X_1_AS_2X_2insAS_3AS_4$ für eine Insertion der Aminosäuren AS_3 und AS_4 zwischen AS_1 und AS_2
 → Bsp.: p.A61_R62insAGQ für die Insertion der Aminosäuren AGQ zwischen Position 61 (A) und 62 (R) (aus ...AR... wird ...AAGQR...)
- $AS_1X_1_AS_2X_2delinsAS_3AS_4$ für eine Deletion der Aminosäuren AS_1 bis AS_2 und deren Ersatz durch die Aminosäuren AS_3 bis AS_4
 → Bsp.: p.W21_P22indelTYG für die Deletion der Aminosäuren an Position 21 bis 22 und deren Ersatz durch die drei Aminosäuren T, Y und G (aus ...WP... wird ...TYG...)

Die Aminosäuren können dabei sowohl im Drei-, als auch im Einbuchstabencode beschrieben werden. Die folgende Tabelle gibt dabei eine Übersicht.

Einbuchstabencode	Dreibuchstabencode	Aminosäure
A	Ala	Alanin
C	Cys	Cystein
D	Asp	Asparaginsäure/Aspartat
E	Glu	Glutaminsäure/Glutamat
F	Phe	Phenylalanin
G	Gly	Glycin
H	His	Histidin
I	Ile	Isoleucin
K	Lys	Lysin
L	Leu	Leucin
M	Met	Methionin
N	Asn	Asparagin
P	Pro	Prolin
Q	Gln	Glutamin
R	Arg	Arginin
S	Ser	Serin
T	Thr	Threonin
U	Sec	Selenocystein
V	Val	Valin
W	Trp	Tryptophan
X	-	Stoppcodon
Y	Tyr	Tyrosin

CRISPR/Cas9 – eine programmierbare Genschere

Die CRISPR/Cas9-Methode gehört heute zu den Standardmethoden zur gezielten Veränderung der genomischen DNA von Organismen, dem sogenannten *Genome Editing*. Für deren Entdeckung wurde im Jahre 2020 der Nobelpreis für Chemie an die beiden Wissenschaftlerinnen Emmanuelle Charpentier und Jennifer Doudna verliehen. Grundlage dafür ist das CRISPR-System, welches zuerst in *Escherichia coli* entdeckt wurde. Dieses verwendet Endonukleasen wie das Protein Cas9, um DNA an einer spezifischen Sequenz ähnlich einer programmierbaren Genschere zu schneiden.

Im Labor macht man sich das CRISPR/Cas9-System folgendermaßen zunutze: Die Endonuklease Cas9 kann positionsspezifisch einen Doppelstrangbruch in der genomischen DNA verursachen. Dazu benötigt sie eine sogenannte single guide RNA (sgRNA), die dem Code-Strang des zu verändernden DNA-Abschnitts entspricht. Für deren Zusammenstellung gelten folgende Prinzipien:

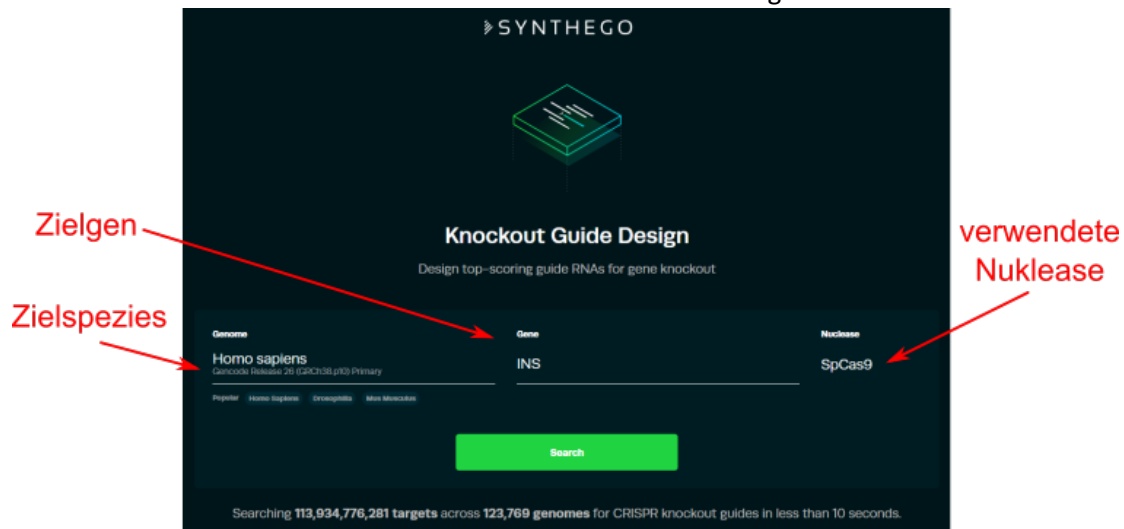
- Die sgRNA besteht immer aus einer 20 Nukleotide langen Sequenz.

- In der Zielsequenz muss nach diesen 20 Nukleotiden direkt die Basensequenz NGG folgen (N steht dabei für eine beliebige Base).
- Die Effizienz der sgRNA kann erhöht werden durch ein G an Position 1, sowie ein A oder T an Position 17.

Die sgRNA kann dabei an einen beliebigen der beiden zueinander komplementären DNA-Stränge binden, es ist daher möglich, aber nicht notwendig, bei für Proteine codierenden Genabschnitten eine zur codierenden Sequenz komplementäre sgRNA zu finden.

Heutzutage nutzt man für die Erstellung von sgRNAs computerbasierte Algorithmen, die außerdem viele weitere Kriterien berücksichtigen, um die am besten passende Sequenz zu ermitteln. Ein Beispiel hierfür ist die frei zugängliche Software des sgRNA-Anbieters Synthego. Deren Benutzung wird im Folgenden genauer erläutert:

- Öffnen Sie die Seite <https://design.synthego.com> in einem Internetbrowser.
- Geben Sie zunächst das Genom der Spezies an, für die Sie die sgRNA planen, sowie anschließend die Abkürzung des gewünschten Zielgens. Als verwendete Nuklease lassen Sie die Standardeinstellung (SpCas9) ausgewählt. Im folgenden Beispiel soll das für Insulin codierende Gen *INS* in einem menschlichen Genom ausgeschaltet werden.



- Durch einen Klick auf „Search“ starten Sie die Suche nach passenden sgRNAs.
- Als Ergebnisse werden automatisch vier sgRNAs empfohlen. Diese können durch ein Scrollen nach unten angezeigt werden. Außerdem ist in einer Grafik die Position dieser vier sgRNAs gezeigt.



- Weitere mögliche sgRNA-Sequenzen können durch einen Klick auf „All Guides“ angezeigt werden. Dort kann außerdem ausgewählt werden, welche Referenzsequenz als Grundlage für den Algorithmus verwendet wird, sowie in welchem Exon des Gens mögliche sgRNAs angezeigt werden sollen.

The screenshot shows the SYNTHGO interface for the *INS* gene. At the top, there are tabs for 'Recommended Guides' and 'All Guides'. Below this, the gene information is displayed: Homo sapiens, Gene: *INS*: ENST00000250971 (primary), Mediator: SpCas9. A genomic map shows the gene structure on chromosome 11, with Exon 3 (142 bp) highlighted in purple. Below the map, a table lists 'RECOMMENDED GUIDES FOR KNOCKOUT'.

RANK	SEQUENCE	CUTSITE	EXON	ON TARGET SCORE	OFF TARGETS
1	CGUCCUGGCAGUGGGGCAGG	2359,992	Exon 3	0.470	0,0,2,42,473
2	GCAGGGGGCAGGUGGAGCU	2359,985	Exon 3	0.425	0,0,12,89,902

Red arrows in the image point to the 'All Guides' button, the 'Exon 3 - 142bp' label on the genomic map, and the 'SEQUENCE' column in the table. A red arrow also points to the 'Sequenzen' label at the bottom right.

- Nach der Anzeige der sgRNA-Sequenzen können Sie das Fenster schließen.

Sequenzanalyse mittels BLAST

Um die mittels CRISPR/Cas9 veränderten DNA-Abschnitte mit den ursprünglichen Sequenzen zu vergleichen, kann man den internetbasierten BLAST-Algorithmus (*Basic Local Alignment Search Tool*) der National Institutes of Health in den USA verwenden. Dessen Benutzung wird im Folgenden genauer erläutert:

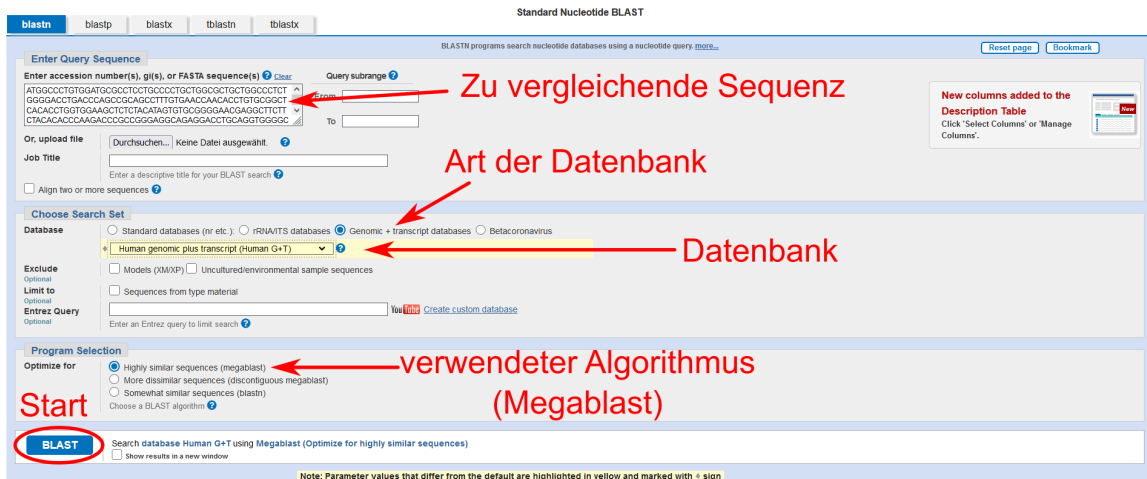
- Rufen Sie die Website der BLAST-Suchmaschine unter <https://blast.ncbi.nlm.nih.gov/> auf.
- Zum Vergleichen von DNA-Sequenzen wählen Sie links das Nucleotide BLAST aus.

The screenshot shows the NCBI BLAST website. At the top, there is a 'Basic Local Alignment Search Tool' header and a news banner. Below this, the 'Web BLAST' section is visible, featuring four buttons: 'Nucleotide BLAST' (nucleotide to nucleotide), 'blastx' (translated nucleotide to protein), 'tblastn' (protein to translated nucleotide), and 'Protein BLAST' (protein to protein). A red arrow points to the 'Nucleotide BLAST' button with the text 'Suche nach DNA-Sequenzen'. Below the buttons is a 'BLAST Genomes' search bar with a dropdown menu showing 'Human', 'Mouse', 'Rat', and 'Microbes'.

- Kopieren Sie die zu vergleichende Sequenz in das Eingabefeld links oben. Im vorliegenden Beispiel wurde eine mutierte Sequenz der humanen *INS*-cDNA gewählt.
- Das Optimieren der Suchparameter ermöglicht eine schnellere Suche und präzisere Ergebnisse. Für menschliche Gene und die der Maus existieren spezielle Datenbanken, die deren gesamtes

Genom und Transkriptom umfassen. Dazu wählen Sie als Art der Datenbank „Genomic + transcript databases“, sowie als Datenbank in diesem Fall „Human genomic plus transcript“. Für eine umfassendere Suche kann unter den „Standard databases“ die „Nucleotide collection“ als umfassendste Datenbank gewählt werden, die neben verschiedenen Organismen auch synthetische DNA-Sequenzen beinhaltet.

- Durch die Auswahl des zu verwendenden Algorithmus wird die Suche für den Grad der Übereinstimmung zwischen den zu vergleichenden Sequenzen optimiert. Ein Megablast sucht nur nach Sequenzen, die eine sehr hohe Übereinstimmung zur eingegebenen Sequenz aufweisen, was für die Analyse von Punktmutationen am besten geeignet ist.
- Durch einen Klick auf „BLAST“ starten Sie den Algorithmus, der nun für einige Sekunden bis Minuten läuft.



- Im Ergebnissenfenster werden anschließend verschiedene Suchparameter angezeigt. Die verglichenen Sequenzen lassen sich unter dem Reiter „Alignments“ anzeigen. Die Darstellung „Pairwise with dots for identities“ erlaubt eine übersichtlichere Anzeige der Unterschiede zwischen der Eingabesequenz (obere Zeile: „Query“) und der zugrundeliegenden Vergleichssequenz der Datenbank (untere Zeile: „Sbjct“). Unterschiede in der Wildtyp-Sequenz werden dabei rot hervorgehoben, während identische Nukleotide als schwarze Punkte abgekürzt werden. Es wird standardmäßig die Sequenz des codierenden Stranges in 5′-3′-Richtung angegeben.

