

Aufbau und Funktion von DNA, Genen und Chromosomen

Körper- und Keimzellen, Mitose und Meiose

Mutation, Rekombination und Crossing-Over

Replikation, Transkription, Translation und genetischer Code

Allele, InDels, SNVs, SNPs und Vielfalt

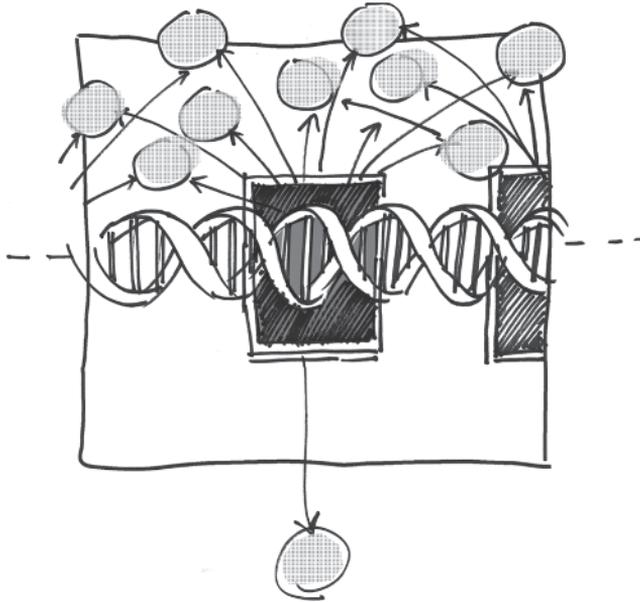
Krönung der Vielfalt: das Immunsystem

# Kapitel 1

## Die Welt der Gene

Sicherlich können Sie mit einer Überschrift wie *Die DNA der Firma* etwas anfangen, oder? – Was verbinden Sie damit? – Das Börsenblatt titulierte so am 6. Dezember 2016 einen Beitrag über Unternehmenskultur. Ich habe zahlreiche weitere Beispiele aus den unterschiedlichsten Bereichen gesammelt: Mal steht DNA für die Unternehmenskultur, mal für die Eigenschaften eines Autos, mal für die Moral einer Gesellschaft. Das Spannende daran ist, dass es zeigt, dass der Begriff in unseren Wortschatz eingegangen ist. DNA steht sinnbildlich für Eigenschaften, die gegeben sind. Klar. Der Begriff, der eigentlich ein Akronym ist – doch dazu gleich mehr – stammt ja auch aus dem Umfeld der Genetik, die früher einmal Vererbungslehre hieß. Und bei der Vererbung geht es ja um die Weitergabe von, naja, etwas. Genetik wiederum leitet sich aus den altgriechischen Wörtern *geneá* für »Abstammung« und *génesis* für »Ursprung« ab. In diesen Begriffen steckt somit auch Vererbung drin.

Was aber ist eigentlich DNA? Und wie wird sie vererbt? Welche Funktion hat das Erbe?



## Was denn nun: DNS, DNA oder Gene?

Das Wort **DNA** ist ein Akronym, also ein Kurzwort, das aus den Anfangsbuchstaben mehrerer Wörter zusammengesetzt ist. DNA steht für *deoxyribonucleic acid*. Dies ist die englische Bezeichnung eines Moleküls, das wir im Deutschen als Desoxyribonukleinsäure, kurz **DNS**, bezeichnen. Das deutsche Wort DNS wird kaum noch verwendet – und von mir überhaupt nicht. Für den Fall, dass Sie mal in die Versuchung kommen, französische Texte zu lesen: Dort begegnet Ihnen die DNA als ADN (*acide désoxyribonucléique*).



DNA, die Desoxyribonukleinsäure, ist das Baumaterial, aus dem die Gene aufgebaut sind. Wie Plastik und Kartoffelstärke ist die DNA ein Polymer, also eine lange Kette aus (vier) gleichartigen Molekülen.

Das Besondere an Nukleinsäuren ist, dass sie aus einzelnen Bausteinen, den Nukleotiden, aufgebaute Polymere sind. Ein DNA-Molekül besteht aus zwei Polymersträngen (DNA-Strängen), die sich umeinander winden. Daher sprechen wir von der **DNA-Doppelhelix** (Abbildung 1.1; siehe auch Abbildung 1.10).



Die Zeichnung der DNA in Abbildung 1.1 aus der Veröffentlichung von JAMES WATSON (1928–) und FRANCIS CRICK (1916–2004) hat CRICKS Frau ODILE CRICK (1920–2007) gezeichnet. Sie verwendete dabei mutmaßlich die Proportionen des Goldenen Schnitts. Die der Struktur zugrundeliegenden Daten wurden von ROSALIND FRANKLIN (1920–1958) per Röntgenstrukturanalyse gewonnen. Sie starb, bevor der Nobelpreis 1962 für die Strukturaufklärung vergeben wurde.

Die Bausteine der DNA-Moleküle lassen sich in noch kleinere Moleküle zerlegen, nämlich in Phosphat, den Zucker Ribose und eine der vier Nukleobasen **Adenin**, **Thymin**, **Guanin**



**Abbildung 1.1:** Modell der DNA-Doppelhelix.

und **Cytosin**. – Wer in Chemie aufgepasst hat, wird sich wundern: Eine Nucleinsäure besteht aus Basen? Hier nur soviel: Die sauren Eigenschaften verursacht das Phosphat. – Die Basen geben jedem Baustein eine »Identität«, weshalb sie auch einfach mit A, T, G oder C abgekürzt werden. Außerdem halten die Nucleobasen die beiden DNA-Stränge zusammen, indem sich immer ein A des einen Stranges mit einem T des Gegenstranges paart (Abbildung 1.2). Ebenso paart sich C mit einem G. Was heißt »paaren«? Durch chemische und physikalische Kräfte (vor allem ionische und van-der-Waals-Kräfte) ziehen sie sich gegenseitig an.



**Abbildung 1.2:** Schematische Darstellung der DNA.

Die Basenpaarungen führten zur Formulierung der bekannten **Chargaff'sche Regeln**, die der deutschstämmige, in die USA emigrierte, Biochemiker ERWIN CHARGAFF (1905–2002) in den 1950er Jahren feststellte:

- ✓ **Erste Paritätsregel** | In doppelsträngiger DNA entspricht der Anteil Adenin genau dem Anteil Thymin und der Anteil Cytosin genau dem Anteil Guanin: %A=%T und %G=%C.

- ✓ **Zweite Paritätsregel** | Außerdem entspricht in einzelsträngiger Nukleinsäure (DNA oder RNA) der Anteil Adenin etwa dem Anteil Thymin und der Anteil Cytosin etwa dem Anteil Guanin:  $%A \sim %T$  und  $%G \sim %C$ .

Damit entdeckte ERWIN CHARGAFF die grundlegende Eigenschaft der DNA als Erbsubstanz.



Auch das US-amerikanische Architekturbüro NBBJ, das einen Entwurf für die neue Konzernzentrale von *Amazon* in Arlington vorgelegt hat, bedient sich der DNA-Doppelhelix. Der Entwurf erinnert an eine riesige Kräuterspirale, deren Kern die »natürliche Schönheit einer Doppelhelix« zugrundeliegt – natürlich spiegelverkehrt (siehe auch *Wow* in Abschnitt *Einfluß von DNA-Strukturvarianten* in diesem Kapitel).

Da die DNA-Stränge durch A–T- und C–G–Paare zusammengehalten werden, reicht es, die Abfolge der Nukleotide des einen Strangs zu kennen – die Sequenz des Gegenstrangs ergibt sich dann daraus (siehe Abbildung 1.2). Die Stränge sind komplementär zueinander. Diese Komplementarität der DNA-Doppelhelix hat weitreichende Folgen, wie FRANCIS CRICK und JAMES WATSON, die Entdecker der DNA-Struktur, bereits im Schlusssatz ihrer Publikation von 1953 vermuten. Denn sie lässt einen einfachen Kopiermechanismus vermuten.



Der US-amerikanische Molekularbiologe JAMES WATSON ist übrigens eine sehr umstrittene Persönlichkeit. So hat er sich wohl heimlich die Röntgenaufnahmen von DNA-Kristallen, die ROSALIND FRANKLIN am selben Institut gemacht hat, besorgt, um die Struktur der DNA aufzuklären. Auch macht er immer wieder mit rassistischen Bemerkungen von sich Reden. Im Jahr 2007 hatte ihn das den Posten des Kanzlers des US-amerikanischen *Cold Springs Instituts* gekostet. 2019 wurden ihm deswegen alle akademischen Ehrentitel aberkannt. Wegen finanzieller Probleme hatte WATSON 2014 seine Nobelpreismedaille für 4,7 Millionen US-Dollar vom Auktionshaus *Christie's* in New York versteigern lassen.



DNA besteht aus Bausteinen, die mit A, T, G oder C abgekürzt werden. Die Abfolge wird als DNA-Sequenz bezeichnet und sieht zum Beispiel so aus: ACCGCTAGC. Die Sequenz des Gegenstrangs ergibt sich entsprechend.

Für die Genetik bedeutet das, dass es ausreicht, nur die Sequenz eines Strangs zu notieren. Wenn Sie eine DNA-Sequenz in einem Arztbericht oder sonst wo sehen, werden Sie genau eine Sequenz sehen – den Gegenstrang kann man sich der Übersichtlichkeit halber sparen.

## Und das Gen?

Als **Gen** bezeichnen wir einen Sequenzabschnitt auf der DNA, der ein Merkmal codiert. So wird es in der Schule gelehrt und das ist auch nicht verkehrt. Allerdings ist dies die Minimalversion einer Definition für ein Gen. Wir Genetiker stocken meist erst einmal, wenn wir nach einer Definition für das Gen gefragt werden. Soll zum Beispiel der Bereich auf der DNA, der für die Regulation der Aktivität eines Gens verantwortlich ist, dazugehören? Auch täuscht die Minimaldefinition vor, das es für jedes Merkmal ein Gen gibt. Die meisten

Merkmale werden indes durch mehrere Gene beeinflusst – vielleicht sogar alle (siehe *Erinnerung* in Kapitel 8, Abschnitt *Ökotypen der Buche*).



Ein Gen ist ein Abschnitt auf der DNA, der zur Ausprägung eines, möglicherweise sichtbaren, Merkmals beiträgt.

## Chromosomen als Horte der Gene

Die DNA befindet sich in fast jeder Zelle – sie ist ihr Erbgut und enthält ihren Bauplan. Allein die roten Blutzellen (Erythrozyten) enthalten keinen Zellkern und folglich keine DNA. Jede einzelne Zelle unseres Körpers hat ihr eigenes Erbgut. Ein siebzig Kilogramm schwerer Mensch besteht aus rund dreißig Billionen Zellen – jetzt dürfen Sie auf die Waage gehen und sich wiegen. Das ist eine Drei mit dreizehn Nullen. Aber: Wenn es um Zellen mit DNA geht, müssen wir die roten Blutzellen abziehen. Das sind 83 Prozent unserer Zellen, die keinen Zellkern und damit keine Chromosomen mehr enthalten. Demzufolge haben nur rund fünf Billionen unserer Körperzellen ein Erbgut.



Prokaryonten wie Eu- und Archaeobakterien (meist spricht man einfach von Bakterien) haben keinen Zellkern und keine Mitochondrien, Chloroplasten oder andere Organellen. Eukaryonten haben einen Zellkern, in dem sich die DNA befindet. Zu ihnen gehören unter anderem Algen, Pilze, Pflanzen und Tiere.

Bei Bakterien befindet sich das Erbgut einfach nackt in der Zelle.



Unser Körper beherbergt auf der Haut und vor allem im Darm mehr Zellen von Mikroorganismen als eigene Körperzellen. Das sind rund 38 Billionen Bakterien, die insgesamt allerdings nur etwa zweihundert Gramm wiegen und Teil unseres **Mikrobioms** sind.

Unser Erbgut (Genom) ist auf 22 **Autosomen** (die Chromosomen 1–22) und zwei **Gonosomen** (die Geschlechtschromosomen X und Y) verteilt (Tabelle 1.1). Hinzu kommt noch das Chromosom der Mitochondrien: Das sind quasi Organe von Zellen, sogenannte Organellen, die für die Energiegewinnung wichtig sind. Obwohl das winzige Mitochondrienchromosom bei einigen Erkrankungen und der Vererbung (Kapitel 8, Abschnitt *Eva, die Mutter aller*) eine Rolle spielt, vernachlässigen wir es erst einmal. Die Gonosomen bestimmen das Geschlecht, jedoch nicht so einfach wie Sie vielleicht meinen: Männliche Samenzellen enthalten 22 Autosomen und entweder ein X- oder ein Y-Chromosom. Weibliche Eizellen enthalten ebenfalls 22 Autosomen und ein X-Chromosom. Bei der Befruchtung der Eizelle kommen folglich 23 Chromosomen des Mannes und 23 Chromosomen der Frau zusammen und bilden die befruchtete Eizelle mit 46 Chromosomen, die sich zum Embryo und weiter zum Baby entwickelt.

Während der Befruchtung vereinigen sich die Chromosomensätze der Geschlechtszellen, wobei die XX-Konstellation zu einem weiblichen und die XY-Konstellation zu einem männlichen Individuum führt. Für die Regulation der Genaktivität (Abschnitt *Ein Gen wird aktiv* in diesem Kapitel) bringt das erhebliche Probleme mit sich, die ich im Kapitel 12, Abschnitt *Geschlechtsgerechtigkeit* beleuchte.

Chromosom	Länge (mm)	Basenpaare	Variationen	Proteine
1	85	248.956.422	12.151.146	2.058
2	83	242.193.529	12.945.965	1.309
3	67	198.295.559	10.638.715	1.078
4	65	190.214.555	10.165.685	752
5	62	181.538.259	9.519.995	876
6	58	170.805.979	9.130.476	1.048
7	54	159.345.973	8.613.298	989
8	50	145.138.636	8.221.520	677
9	48	138.394.717	6.590.811	786
10	46	133.797.422	7.223.944	733
11	46	135.086.622	7.535.370	1.298
12	45	133.275.309	7.228.129	1.034
13	39	114.364.328	5.082.574	327
14	36	107.043.718	4.865.950	830
15	35	101.991.189	4.515.076	613
16	31	90.338.345	5.101.702	873
17	28	83.257.441	4.614.972	1.197
18	27	80.373.285	4.035.966	270
19	20	58.617.616	3.858.269	1.472
20	21	64.444.167	3.439.621	544
21	16	46.709.983	2.049.697	234
22	17	50.818.468	2.135.311	488
X	53	156.040.895	5.753.881	842
Y	20	57.227.415	211.643	71
mtDNA	0,0054	16.569	929	13
	<b>1.052</b>	<b>3.088.286.401</b>	<b>155.630.645</b>	<b>20.412</b>

**Tabelle 1.1:** Ausgewählte Daten zu den Chromosomen eines Mannes. Das weibliche Erbgut ist mit 3.187.099.881 Basenpaaren rund hunderttausend Basenpaare größer.



Es ist wichtig, sich klar zu machen, dass ein Mensch nicht ein Genom hat, sondern zwei Genome: eines von der Mutter und eines vom Vater geerbt. Die Anzahl der vollständigen Chromosomensätze (Genome) in jeder Zelle wird als **Ploidiegrad** bezeichnet. Der Mensch und die meisten anderen Tiere sind diploid (2N), haben demnach zwei Genome beziehungsweise Chromosomensätze. Viele Pflanzen haben einen höheren Ploidiegrad: Zum Beispiel ist der Saatweizen hexaploid (6N; siehe auch Abbildung 3.6) und die meisten Handelserdbeeren sind dekaploid (10N).

Die Anzahl der Chromosomenpaare erhöht nicht nur die Gesamtmenge der DNA in einem Genom, sondern auch die Komplexität des Genoms. Obwohl der Großteil der DNA-Sequenzen jedes Chromosomenpaares identisch sind, sind es die Unterschiede, die die Breite der biologischen Variation innerhalb einer Art ausmachen.



Die größte Zelle, die Sie vermutlich jemals gesehen haben ist – na? – das Straußenei. Vogeleier sind Eizellen mit Schale. Ein Ei ist demzufolge eine riesige Zelle.

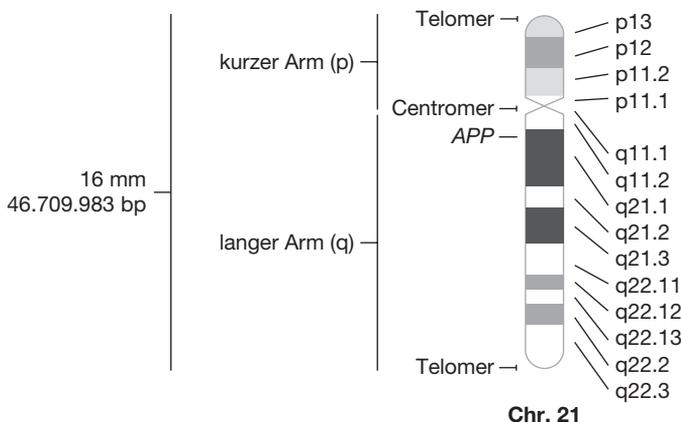
Unsere Körperzellen enthalten somit einen doppelten Chromosomensatz, was wir als **diploid** bezeichnen. Ein Satz stammt vom Vater, einer von der Mutter. Ei- und Samenzellen,

die sogenannten Geschlechts- oder **Keimzellen**, haben jeweils einen einfachen Chromosomensatz, was wir als **haploid** bezeichnen. Bei der Bildung der haploiden Keimzellen aus diploiden Zellen im Eierstock beziehungsweise im Hoden kommt es damit zu einer Reduktion der Chromosomensätze. Dafür ist der Prozess der Meiose verantwortlich (siehe Abschnitt *Meiose und Crossing-Over* in diesem Kapitel).



Unsere 46 Chromosomen haben gemeinsam eine Gesamtlänge von etwa 2,1 Meter (Tabelle 1.1). Nochmal: Zwei-komma-eins Meter DNA pro Zelle! Nähmen wir nun die fünf Billionen Körperzellen mit Erbgut (schließen also die roten Blutzellen aus) und könnten alle Chromosomen aneinanderbinden, dann ergäbe dies einen Faden von 10,7 Milliarden Kilometer. Der Faden würde 72-mal zur Sonne und zurück reichen oder 11,3-mal der Bahn der Erde um die Sonne entsprechen. Diese riesigen Fäden sind in den Zellkernen Ihrer Zellen zusammengeknäult. Nur die gerade aktiven Bereiche werden sozusagen entknäult.

Chromosomen werden häufig als schematisches **Idiogramm** dargestellt. Wie das aussieht, ist in der (Abbildung 1.3) dargestellt. Dieses resultiert aus der gebänderten Färbung, die ein Chromosom nach der Behandlung mit dem **Giemsa-Farbstoff** annimmt. Das Verfahren wurde im Jahr 1904 von dem deutschen Chemiker und Bakteriologen GUSTAV GIEMSA (1867–1948) vorgestellt und ist seitdem ein Standardverfahren in der Histologie. Diese Bänderung mit helleren G-Bändern und dunkleren R-Bändern ist unter dem Mikroskop sichtbar und entsteht infolge regional unterschiedlicher physikalischer und chemischer Eigenschaften des Chromosoms. Dazu muss man wissen, dass am Aufbau des Chromosoms auch **Proteine** (siehe Abschnitt *Das Protein als Genprodukt* in diesem Kapitel) beteiligt sind – Chromosomen bestehen demzufolge nicht aus nackter DNA. In der Abbildung 1.3 befindet sich das Gen *APP* an Position 21q21.3 (Basenpaare 25.880.550 bis 26.171.128). Es codiert für das *Amyloid Precursor Protein*, das an der Alzheimererkrankung beteiligt ist.



**Abbildung 1.3:** Idiogramm des Chromosoms 21, auf dem auch das Gen *APP* liegt.

Zwei Bereiche stechen bei der Giemsa-Färbung besonders hervor: die Enden, die als **Telomere** bezeichnet werden und ein zentraler Bereich, das **Centromer**. Das Centromer trennt den kurzen und langen Arm des Chromosoms, diese werden auch als **p-Arm** ( $p = \textit{petit}$ ) und **q-Arm** ( $q = \textit{hmm}$ , naja,  $q$  folgt auf  $p$ ) bezeichnet (siehe Abbildung 1.3). Die

Giemsa-Bänderung wird noch heute für die Lokalisierung von Genen auf den Chromosomen verwendet. So befindet sich das Gen *APP*, welches an der Alzheimererkrankung beteiligt ist, an **Position 21q21.3**. Seitdem die Sequenz des Chromosoms 21 bekannt ist, können wir auch schreiben, dass das Gen *APP* den Abschnitt von Basenpaar 25.880.550 bis 26.171.128 auf dem Chromosom einnimmt.



Die Bezeichnung 21q21.3 für ein Gen besagt, dass es auf dem langen Arm (q) des Chromosoms 21 in der Giemsa-Bande 21, Unterbande 3 liegt. Diese Art der Beschreibung eines Genortes ist heute noch weit verbreitet.

## Chromosomen, Sex und Mutationen

Eine wirkliche Sensation war die Evolution der Prozesse, die zum Sex geführt haben. Damit meine ich nicht den Geschlechtsakt. Der Vorgang, den ich meine, findet bereits früher statt, nämlich bei der Entstehung der Geschlechtszellen. Bei der ständig stattfindenden Zellteilung, wird der Chromosomensatz während eines Prozesses namens **Mitose** zunächst verdoppelt und dann auf zwei Tochterzellen aufgeteilt. Die Mitose ist die Grundlage des Wachstums durch Zellvermehrung. Das beherrschen schon die einfachsten Bakterien – ja, es war die Grundlage dafür, dass sich Leben nach seiner Entstehung überhaupt verbreiten konnte. Genetische Vielfalt entsteht dabei kaum. Die Verdopplung des Chromosomensatzes, ein Prozess namens **Replikation** und katalysiert von der **DNA-Polymerase** (siehe Kapitel 5: *DNA-Polymerase*), soll ja gerade möglichst fehlerfrei erfolgen. Gott (oder dem Zufall) sei Dank, arbeitet die DNA-Polymerase jedoch nicht ganz perfekt. So entsteht bei ihrer Arbeit etwa alle hunderttausend Basenpaare ein Fehler. Davon werden über neunundneunzig Prozent nach der Replikation von der Zelle »erkannt« und korrigiert. Die verbleibenden Fehler bei der »Abschrift« der Erbinformation durch die DNA-Polymerase bezeichnen wir als **Mutation** (siehe auch Kapitel 6: *Bändigung des Zufalls*); die Häufigkeit des Auftretens von Mutationen als Mutationsrate.



Der niederländische Botaniker HUGO DE VRIES (1848–1935) gilt als Begründer der Mutationstheorie: »*Als Mutationstheorie bezeichne ich den Satz, dass die Eigenschaften der Organismen aus scharf voneinander unterschiedenen Einheiten aufgebaut sind. Diese Einheiten können zu Gruppen verbunden sein, und in verwandten Arten kehren dieselben Einheiten und Gruppen wieder.*« Jene scharf voneinander unterschiedenen Einheiten nennen wir heute Gene.

Schlussendlich haben Bakterien **Mutationsraten** von nur einem Fehler pro hundert Millionen bis einer Milliarde Nukleotide. Beim Menschen liegt sie etwa um den Faktor zehn höher.



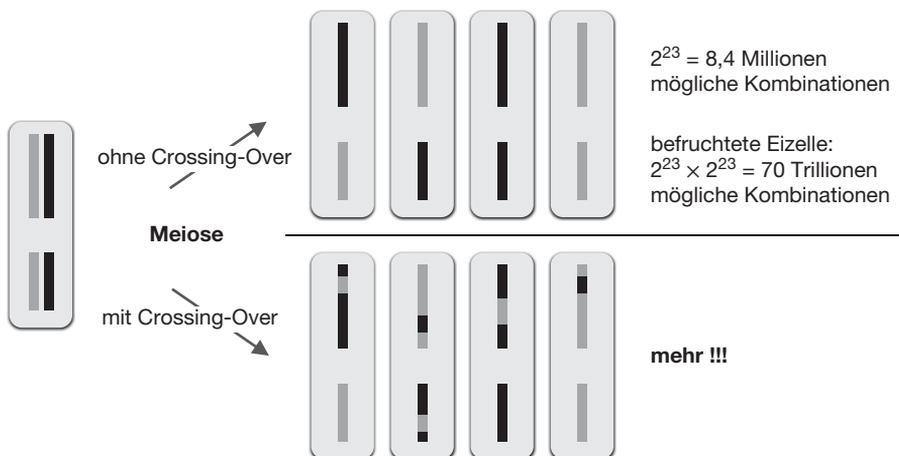
Kleine Mutationsraten bei Bakterien können wegen ihrer schnellen Vermehrung große Wirkung entfalten. Das Genom unseres Darmbakteriums *Escherichia coli* hat etwa 4,2 Millionen Basenpaare beziehungsweise 8,4 Millionen Nukleotide. Geht man von einem Fehler pro einer Milliarden Nukleotide aus, dann weist jede Tochterzelle bei jeder Teilung von *E. coli* im Durchschnitt 0,0084 neue Mutationen auf. Anders gesagt: Ungefähr ein Prozent der Tochterzellen enthält eine neue Mutation. Das klingt wenig. Da sich *E. coli* alle dreißig Minuten teilen kann,

entstehen in zehn Stunden allerdings rund eine Million ( $2^{20}$ ) Zellen. Das bedeutet, zehntausend dieser Bakterien tragen mindestens eine Mutation – die beispielsweise zu einer Antibiotikaresistenz führen kann.

Berechnungen zeigen, dass bei jeder Zellteilung rund dreißig bis neunzig Mutationen eingeführt werden. Bei Stammzellen kann diese Zahl auf bis zu drei Mutationen pro Zellteilung reduziert sein, da sie besonders sorgfältige Reparaturmechanismen besitzen. Es erfordert etwa fünfzigtausend Mutationen, bevor eine Zelle zu einer **Krebszelle** entartet. Im Alter von sechzig Jahren ist statistisch gesehen jedes Nukleotid in wenigstens einer Körperzelle mutiert.

## Meiose und Crossing-Over

Mutationen führen also zu einer genetischen Mannigfaltigkeit, die für die **Evolution** essenziell ist. Nur durch genetische Variationen zwischen Individuen können neue Lebensformen und -räume erschlossen werden. Je komplexer Lebewesen werden, desto aufwendiger wird deren Vermehrung. Einige Tiere und Pflanzen können sich ungeschlechtlich (vegetativ) beispielsweise durch Sprossung vermehren. Das geht schnell. Aber nur die Erweiterung der Mitose um die **Meiose** eröffnet Lebewesen eine weitere Möglichkeit, genetische Vielfalt zu erzeugen: durch Mischen (Abbildung 1.4). Wie indes findet diese Durchmischung statt? Dazu wird der doppelte (diploide) Chromosomensatz in zwei einfache (haploide) Chromosomensätze aufgeteilt. Der diploide Chromosomensatz wird jedoch nicht einfach geteilt, sondern erst verdoppelt (**Meiose 1**) und dann geteilt (**Meiose 2**; siehe auch Abbildung 1.5).



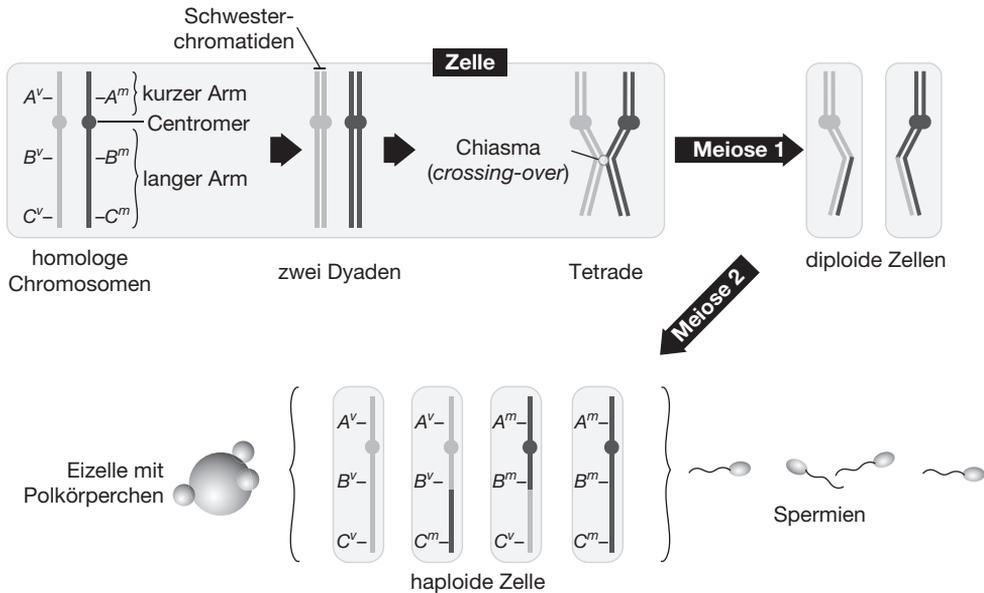
**Abbildung 1.4:** Die Meiose führt zur Entstehung von Samen und Eizellen.

In der Abbildung 1.4 sind nur zwei Chromosomen gezeigt. Bei 23 Chromosomen gibt es 8,4 Millionen Kombinationsmöglichkeiten. Bei der anschließenden Befruchtung mit Keimzellen eines anderen Organismus quadriert sich die Zahl der möglichen Kombinationen. Findet zusätzlich noch Rekombination in Form von Crossing-Over statt, werden einzelne Abschnitte der Schwesterchromosomen ausgetauscht. Die Anzahl der Kombinationen – ganz ohne Mutationen – ist unermesslich. Es lebe der Sex.



Durch Mutationen entstehen neue Allele. Die Rekombination während der Meiose durchmischt vorhandene Allele.

Während der doppelte Chromosomensatz bei der Meiose auf einen einfachen reduziert und auf die vier Geschlechtszellen aufgeteilt wird, kommt es zur **Rekombination**: Zwischen den homologen väterlichen und mütterlichen Chromosomen kann es beim **Crossing-Over** zu einem Austausch von Erbinformation kommen (Abbildung 1.5).



**Abbildung 1.5:** Tanz der Chromosomen während der Meiose am Beispiel eines einzelnen Chromosomenpaares.

Stellen wir uns ein mütterliches Chromosom  $m$  vor: Es trägt die Gene  $A^m$ ,  $B^m$  und  $C^m$ . Die väterliche Variante, Chromosom  $v$ , trägt die Gene  $A^v$ ,  $B^v$  und  $C^v$ . Während der **Meiose 1** wird das Chromosomenpaar verdoppelt. Diese Kopie wird **Chromatid** genannt. Es entsteht je Chromosom eine Dyade (griechisch für Zweiheit) mit zwei **Schwesterchromatiden**. Jedes Chromosom bekommt demzufolge ein Schwesterchen. Diese Dyaden bilden daraufhin eine Tetrade (griechische Doppelkopfrunde – Scherz – eine Tetrade ist eine Art Vierpack), bei der sich Nichtschwesterchromatide (quasi Cousinenchromatide) aneinander anlagern und dabei ein sogenanntes Chiasma (griechisch für Kreuzung; benannt nach dem griechischen Buchstaben  $\chi$ ) bilden können. Hier kommt es dann zum Crossing-Over, also dem Austausch zwischen den beiden DNA-Strängen. Während der **Meiose 2** werden alle Chromatiden voneinander getrennt und auf einzelne Zellen, die Keimzellen, verteilt. Als Ergebnis der Meiose mit Crossing-Over entstehen vier Geschlechtszellen. In dem in Abbildung 1.5 gezeigten Beispiel wurde dabei ein Teil des langen Chromosomenarms ausgetauscht. Daher kam es zu einem Austausch des  $C$ -Allels, also der mütterlichen und väterlichen Variante des Gens  $C$ .



Sie können die Rekombination auch mit einem Kartenspiel simulieren: Nehmen Sie nur die roten Farben Herz und Karo. Das soll dem weiblichen (Herz) und männlichen (Karo) Chromosom mit jeweils den Genen, Ass, König, Dame, Bube, Zehn ... entsprechen – der diploide Chromosomensatz. Jetzt mischen Sie und teilen die Karten auf zwei Stapel so auf, dass jeder Stapel von jedem Bild nur eine Karte enthält – die Meiose. Die Farben werden dabei gemischt – Rekombination.

Den Austausch von Genabschnitten durch Crossing-Over habe ich stark vereinfacht dargestellt. In Wirklichkeit kann die Rekombination an mehreren Stellen parallel stattfinden. Es können sogar nur Teilabschnitte eines Gens zwischen dem mütterlichen und väterlichen Chromosom ausgetauscht werden. Dann entstehen sogenannte **chimäre Allele**.

Wir müssen uns auch vor Augen führen, dass beim Crossing-Over beide beteiligten Chromosomen (also die Cousinenchromatide, siehe Abbildung 1.5) jeweils DNA-Doppelstrangbrüche erleiden, die dann wieder fehlerfrei repariert werden. Es darf nicht ein Nukleotid entfernt oder zusätzlich eingebaut werden, da eine solche Mutation sonst beispielsweise zu einem defekten Protein führen könnte.

Liegen Gene auf den Chromosomen eng beieinander, dann werden sie beim Crossing-Over mit hoher Wahrscheinlichkeit gemeinsam ausgetauscht. Solche Gene scheinen folglich miteinander verbunden zu sein: Wir sprechen von einer **Genkopplung**. Dadurch können sie nicht frei mit anderen Genen ausgetauscht werden, wie GREGOR MENDEL (1822–1884) es uns lehrt (siehe auch Kapitel 2, Abschnitt *Die Mendel'schen Regeln*). Die Auswirkungen der Genkopplung sind Thema im Kapitel 2, Abschnitt *Nicht-Mendel'sche Vererbung*.



Die US-amerikanische Genetikerin BARBARA McCLINTOCK (1902–1992) hat im Jahr 1983 »für ihre Entdeckung mobiler genetischer Elemente« den Nobelpreis erhalten. Zu diesem Zeitpunkt lag ihre Entdeckung schon über dreißig Jahre zurück und weder die Struktur der DNA noch der genetische Code waren bekannt. Ihre Entdeckung der »springenden Gene«, so nennen wir die mobilen genetischen Elemente auch, ist eng mit der Aufklärung des Crossing-Over verbunden. Seit den 1920er Jahren forschte BARBARA McCLINTOCK daran, wie sich Chromosomen während der Reproduktion von Maispflanzen verändern. Dazu entwickelte sie Techniken zur mikroskopischen Visualisierung von Maischromosomen und publizierte 1931 die erste genetische Karte von Mais: die Position von drei Genen auf Chromosom 9. Aus ihren Beobachtungen entwickelte sie, gemeinsam mit ihrer Doktorandin HARRIET CREIGHTON (1909–2004), die Idee der genetischen Rekombination durch Crossing-Over während der Meiose. Da sie eine Frau war, wurde ihre wissenschaftliche Leistungen lange ignoriert und erst nach Jahrzehnten honoriert.

Die Entwicklung dieser **geschlechtlichen Fortpflanzung** hatte demzufolge den riesigen Vorteil, Vielfalt in das Erbgut zu bringen, indem Vorhandenes immer wieder neu gemischt wird. Die Grundlage hierfür bildet der Wechsel zwischen diploiden und haploiden Zellen mit der damit verbundenen Meiose. Bakterien können das nicht. Sie sind alleine auf **Mutationen** angewiesen.

Für die Regulation der Genaktivität (siehe Abschnitt *Ein Gen wird aktiv* in diesem Kapitel) brachte die Evolution von Geschlechtschromosomen allerdings auch erhebliche Probleme mit sich, die ich im Kapitel 12, Abschnitt *Geschlechtsgerechtigkeit* beleuchte.



Wissenschaftler können bei spezialisierten Firmen Mäusemutanten bestellen, bei denen ein komplettes Gen »ausgeschaltet«, wir sagen ausgeknockt, ist. Fachsprachlich spricht man daher von einer **Knockout-Mutante**. Die meisten Knockouts verhalten sich neutral: Sie haben keinen erkennbaren Effekt auf den Träger. Der britische Entwicklungsbiologe LEWIS WOLPERT (1929–2021) sagte dazu: »Bei Mäusen gibt es viele Gene, die man ausknockt ohne eine Veränderung zu sehen, und man schließt daraus, dass sie redundant (verzichtbar) sind. Ich frage: Haben Sie Ihre Mäuse in die Oper gebracht? Können sie noch Wagner von Mozart unterscheiden?« Was WOLPERT meint: Es geht letztlich darum, den Beitrag des infrage stehenden Gens auf die **Fitness** in einer bestimmten Umwelt zu bestimmen. Die Fitness wiederum beschreibt hier nicht den Trainingszustand, sondern den Überlebens- und Fortpflanzungserfolg.

## Gene sind mehr als Gene

Als der Augustinermönch und spätere Abt GREGOR MENDEL 1866 die nach ihm benannten Vererbungsregeln publizierte (siehe Kapitel 2, Abschnitt *Die Mendel'schen Regeln*), da sprach er von Elementen: »Die unterscheidenden Merkmale zweier Pflanzen können zuletzt doch nur auf Differenzen in der Beschaffenheit und Gruppierung der Elemente beruhen, welche in den Grundzellen derselben in lebendiger Wechselwirkung stehen.« MENDELS Veröffentlichung wurde zunächst kaum wahrgenommen und erst im Jahr 1900 unabhängig voneinander durch den niederländischen Botaniker und Begründer der Mutationstheorie HUGO DE VRIES, den deutschen Botaniker CARL CORRENS (1864–1933), den österreichischen Pflanzenzüchter ERICH VON TSCHERMAK (1871–1962) und den britischen Biologen WILLIAM BATESON (1861–1926) wiederentdeckt. Letzterer bezeichnete MENDELS »Elemente« als **Allelomorphe**, also Merkmalseinheiten, die bei der Vererbung durch andere ersetzt werden können. Daraus entwickelte sich der Begriff **Allel**, der heute für Genvarianten steht (siehe Abschnitt *Ein Gen kommt selten allein: Allele* in diesem Kapitel).



Wieso beschäftigte sich ein Mönch wie GREGOR MENDEL eigentlich mit Genetik? Im neunzehnten Jahrhundert stand die Frage nach der Entstehung der Arten im Mittelpunkt des Interesses vieler Naturwissenschaftler, Philosophen und Theologen. Der Botaniker CARL VON LINNÉ (1707–1778) hatte zunächst angenommen, dass es nur soviel Tier- und Pflanzenarten gebe, wie von Gott geschaffen wurden. Auf dem Titelblatt seiner *Systema Naturae* von 1758, in dem er die noch heute gültige Systematik von Tieren und Pflanzen begründet, zitierte er den Psalm 104:24: »Herr, wie sind Deine Werke so groß und viel! Du hast sie alle weislich geordnet, und die Erde ist voll Deiner Güter.« Im Gegensatz zu heute waren vor der Zeit der Aufklärung *religio et ratio*, Glaube und Wissenschaft, keine Gegensätze. GALILEO GALILEI (1564–1642), welcher der Erde eine zentrale Funktion im Weltgeschehen abstritt, plante in den Benediktinerorden eintreten. Mit seiner Forschung wollte er, wie auch LINNÉ und MENDEL, die Werke Gottes durch Experimente und Logik erklären und somit preisen.

Zur Zeit der Wiederentdeckung von MENDELS Arbeiten gab es das Wort **Gen** noch nicht. Dieses wurde 1909 von WILHELM JOHANNSEN (1857–1927) geprägt. Der Begriff ist unter Genetikern indes nicht so eindeutig, wie Sie vielleicht meinen. In der klassischen Genetik bezeichnet das Gen eine **Vererbungseinheit**. Das ist ein alter und abstrakter Begriff, der zunächst kein physisches Pendant hatte. Das moderne Bild entwickelte sich erst mit der Entdeckung, dass die aus DNA bestehenden Chromosomen die Erbträger sind und somit die Vererbungseinheiten tragen müssen. Mit der Aufklärung der molekularen DNA-Struktur als einer Abfolge von Nukleotiden wurde das Bild schärfer: Gene sind eine Abfolge von Nukleotiden in der DNA. In dieser Abfolge der vier Buchstaben A, T, G und C steckt Information: die **Blaupause** des Lebens, der Bauplan der Organismen und sogar die Flugroute einiger Zugvögel – andere lernen die Route von ihren Eltern. Die Dechiffrierung des Codes (siehe Abschnitt *Der genetische Code* in diesem Kapitel) führt zu dem häufig bemühten Bild, dass ein Gen für ein Protein, respektive ein Merkmal, codiert: die **Ein-Gen-Ein-Enzym-Hypothese**. Enzyme sind Proteine, die biochemische Reaktionen in den Zellen katalysieren. Sie sind somit für den Stoffwechsel und damit für die Lebensprozesse verantwortlich. Diese Annahme ist allerdings aus mehreren Gründen viel zu stark vereinfacht.



Gene codieren für ein Protein oder eine RNA. Zu der proteincodierenden Sequenz gehören die voranstehenden Promotoren und die nachgestellten Terminatoren sowie die tausende Basenpaare entfernten Enhancer und Silencer. Diese nehmen Einfluss auf die Aktivität eines Gens und gehören funktional zum Gen, wie der nächste Abschnitt genauer zeigt.

## Moderne Sicht auf die Gene

Heute wissen wir, dass an dem Aufbau vieler Enzyme und nicht-enzymatischer Proteine mehrere Gene beteiligt sind. Zudem gibt es Gene, die gar nicht für Proteine codieren, sondern für **RNA** (siehe Abschnitt *Ein Gen wird aktiv* in diesem Kapitel). Einige Wissenschaftler zählen auch jene Abschnitte auf der DNA zu den Genen, die deren Aktivität regulieren (siehe Tabelle 1.2); dazu gehören:

- ✓ **Promotoren** | Sie liegen unmittelbar vor den proteincodierenden DNA-Abschnitten. Hier startet die Transkription.
- ✓ **Terminatoren** | Sie markieren das Ende des zu transkribierenden Bereichs.
- ✓ **Enhancer** | Enhancer können mehrere tausend Basenpaare vom Promotor entfernt liegen und fördern die Transkription.
- ✓ **Silencer** | Silencer können ebenfalls mehrere tausend Basenpaare vom Promotor entfernt liegen und hemmen die Transkription.

Die DNA-Sequenzen in den Promotorabschnitten werden von sogenannten **Transkriptionsfaktoren** erkannt, die daran binden und die Transkription beeinflussen. Sie fragen sich, wie ein »weit entfernter« DNA-Abschnitt auf die Aktivität eines Gens wirken kann? Dazu müssen Sie sich vor Augen halten, dass die DNA wie ein Wollknäuel in der Zelle liegt und somit auch distante DNA-Regionen unmittelbar nebeneinander liegen können.

Komponente	Anteil [%]
Inaktive Transposons	45
Aktive Transposons	0,1
Introns in proteincodierenden Genen	22
<b>Proteincodierende Gene</b>	<b>1</b>
Inaktive virale Sequenzen	9
Virale Sequenzen	0,1
Introns in RNA-codierenden Genen	6
RNA-codierende Gene	1
Pseudogene	5
Regulatorische Sequenzen	0,2
Unbekannte Funktion	9
Sonstige	1,6
	<b>100</b>

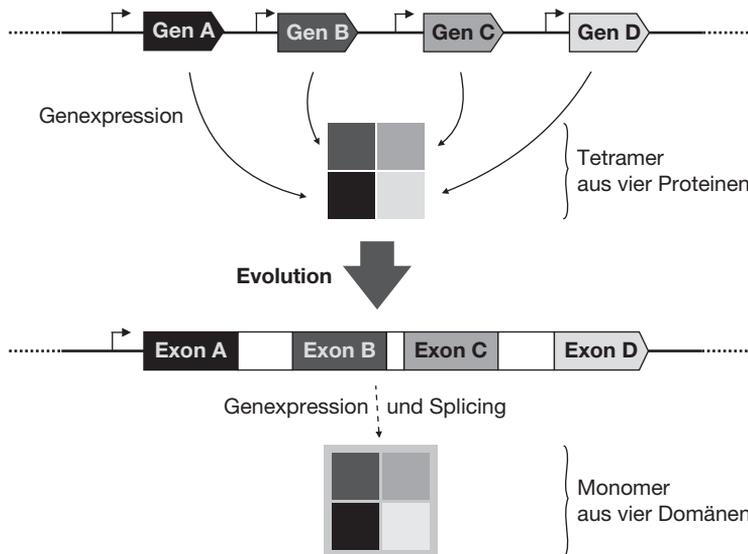
**Tabelle 1.2:** Anteile funktionaler Komponenten, die in unserem Genom codiert sind.

Sie sehen: Es gibt nicht nur Gene, die unmittelbar für ein Protein und damit für einen Phänotyp codieren, sondern auch welche, deren Produkte gar keine Proteine sind. Das hat Konsequenzen: Je mehr regulative DNA-Abschnitte es gibt, desto mehr kann mit einer relativ geringen Anzahl an proteincodierenden Genen verwirklicht werden. Das passt zu der Beobachtung, dass mit einer zunehmenden Komplexität der Lebewesen, die Größe des Erbguts nicht proportional zunimmt (siehe Kapitel 3, Abschnitt *Das Genom und Omics*). Es wird demzufolge mit relativ wenig Erbinformation mehr Komplexität erzeugt. Dieses Phänomen ist als **C-Wert-Paradox** bekannt (siehe auch Abbildung 3).



Nur etwa ein bis zwei Prozent unseres Erbguts codieren für Proteine. Dieser Anteil wird als sogenannte Exom bezeichnet. Der größte Teil besteht aus inaktiven springenden Genen (Transposons) und Introns.

In den proteincodierenden Sequenzen von höheren Lebewesen, den sogenannten Eukaryonten (siehe *Erinnerung* in Abschnitt *Chromosomen als Horte der Gene* in diesem Kapitel), wechseln sich in der Regel codierende und nicht-codierende Abschnitte ab. Diesen werden als **Exons** und **Introns** bezeichnet. Ein wichtiger Schritt bei der Übersetzung einer Gensequenz in ein Protein ist die Entfernung der Introns. Diesen Prozess bezeichnen wir als **Splicing**. Warum Gene aus Exons und Introns bestehen, die dann relativ kompliziert wieder rearrangiert werden müssen, ist ein Rätsel der Evolution. Eine gängige Theorie, die **Exontheorie der Gene**, besagt, dass in der Evolution zunächst kurze Gene entstanden, die für kleine Proteine mit einfachen Funktionen codierten. Proteine, die »Hand-in-Hand« zusammenarbeiteten, haben sich dann zu Proteinkomplexen aus mehreren Proteinen zusammengelagert (Abbildung 1.6). Im Laufe der Evolution haben sich deren Gene zu einem Gen zusammengefunden und konnten so einfacher als eine Einheit reguliert werden. So war sichergestellt, dass immer gleich viele Untereinheiten für den Proteinkomplex zur Verfügung standen. Statt vier Regulationseinheiten (Promotoren, siehe Abbildung 1.6) war nur noch eine Regulationseinheit und ein Splicingapparat notwendig. Der Splicingapparat kann aber auch andere Gene splicen: die Investition »lohnt« sich. Nach dem Splicen entsteht ein Protein mit mehreren funktionellen **Domänen**.



**Abbildung 1.6:** Die Exontheorie der Gene.



Die Exontheorie der Gene lässt sich gut mit dem weihnachtlichen Keksebacken vergleichen. Der Teig wird erst ausgerollt und dann mit verschiedenen Formen ausgestochen (Funktion 1, 2, 3 ...). Es gibt auch »Förmchenroller« mit vielen integrierten Förmchen zu kaufen. Dies beschleunigt die Arbeit, da zeitgleich viele Förmchen ausgestochen werden können. Ähnlich effizient ist es, mehrere Proteine, die an demselben Prozess arbeiten (Multienzymkomplex), zu einem Protein mit mehreren funktionalen Domänen (Multidomänenprotein) zusammenzuführen.

Für ein spätes Auftreten der Introns in der Evolution, die sogenannte **Intron-Late-Hypothese** spricht auch die Beobachtung, dass evolutionär ältere Bakterien in der Regel keine Introns haben, die höher entwickelten Eukaryonten dagegen fast immer. Naja, wenn ich »in der Regel« und »fast« schreibe, dann ist schon alles gesagt: Es gibt wieder einmal Ausnahmen. Entsprechend gehen Verfechter der **Intron-Early-Hypothese** davon aus, dass Introns von Anfang an vorlagen und nur bei einigen evolutionären Linien verloren gegangen sind. Diese Debatte ist über vierzig Jahre alt und immer noch nicht geklärt – vielleicht finden Sie eine Lösung?

## Krebs, p53 und das berühmteste Gen

Die berühmtesten Gene der Bakterien, zumindest aus meiner Sicht und in Hinblick auf gentechnologische Anwendungen, stelle ich Ihnen im Kapitel 5: *Drei Gene verändern die Welt* vor. Wie sieht es hingegen beim Menschen aus?

Ein Kandidat wäre das Gen *INS* auf dem Chromosom 11, aus dem das erste gentechnische **Medikament** produziert und 1982 in den USA zugelassen wurde: Humaninsulin. Das Medikament namens Humulin wurde in Partnerschaft von den US-amerikanischen

Unternehmen *Genentech* und *Eli Lilly* entwickelt. Dazu wurde das Gen aus dem menschlichen Genom kopiert und in das Bakterium *Escherichia coli* eingefügt. Aus letzterem wurde schließlich das **Insulin** gewonnen. Zuvor wurde Schweineinsulin, das man in einem aufwendigen Verfahren aus Bauchspeicheldrüsen geschlachteter Schweine gewann, zur Pille gedreht.

Ein weiterer Kandidat wäre das Gen *ADA* auf Chromosom 20. Es war das Ziel der ersten genehmigten klinischen **Gentherapie** (siehe Kapitel 13: *Gentherapien*).

Eine Studie aus dem Jahr 2017 kommt dagegen zu einem anderen Ergebnis: Das berühmteste Gen ist *TP53* auf Chromosom 17 (Tabelle 1.3). Der Bioinformatiker PETER KERPEDEJIEV (1984–) durchsuchte dafür vierzigtausend wissenschaftliche Publikationen nach Nennung von Gennamen. Dabei fand er den Topkandidat, der für das Tumorsuppressorprotein p53 codiert. Dieses Protein ist bei der Hälfte aller Tumore mutiert. Wegen dieser herausragenden Bedeutung wurde das vom Gen *TP53* codierte Protein p53 vom wissenschaftlichen Top-Journal *Science* 1993 zum »Molekül des Jahres« gekürt.



Der Einfluss wissenschaftlicher Fachzeitschriften wird durch den *Journal Impact Factor* beschrieben. Dieser sagt aus, wie oft die Artikel einer Fachzeitschrift in wissenschaftlichen Veröffentlichungen anderer Fachzeitschriften pro Jahr zitiert werden. Zu den Top-Journalen gehören *PLOS ONE*, die *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *Nature*, das *Journal of the American Chemical Society* und das US-amerikanische Pendant zu *Nature*, die ebenfalls interdisziplinäre, naturwissenschaftliche Fachzeitschrift *Science*.

Gen	Paper	Chromosom	Krankheit
<i>TP53</i>	10.312	17	Krebs
<i>TNF</i>	5.994	6	Entzündungen
<i>EGRF</i>	5.967	7	Krebs
<i>VEGFA</i>	4.799	6	Krebs
<i>APOE</i>	4.570	19	Fettstoffwechsel

**Tabelle 1.3:** Gene des Menschen mit den meisten Erwähnungen in der wissenschaftlichen Primärliteratur (im Wissenschaftsjargon *Papern*) und deren wichtigste Beteiligung an Krankheiten. Stand: 02. Juni 2021.

Das Protein **p53** wird auch als **Wächter des Genoms** bezeichnet. Seinen Namen hat es nach seiner Entdeckung 1979 erhalten. Die »53« steht für das molekulare Gewicht des Proteins: 53 Kilodalton. Die Einheit **Dalton** ist nach dem englischen Chemiker JOHN DALTON (1766–1844) benannt. Ein Dalton (Da) entspricht etwa dem Gewicht eines Wasserstoffatoms. Damit wiegt p53 etwa soviel wie dreiundfünfzigtausend Wasserstoffatome. Das Protein ist wirklich ein Tausendsassa und es gibt populärwissenschaftliche Bücher alleine über p53, seine Entdeckung und seine Funktion, zum Beispiel *p53: The Gene that Cracked the Cancer Code* von SUE ARMSTRONG.

Als Wächter des Genoms hat p53 weitreichende Funktionen im Erbgut. Zuallererst ist es ein **Tumorsuppressorgen**. Ein was? Ein: Tumor – Suppressor – Gen! Es unterdrückt (supprimiert) also Krebs (Tumore). Das ist schon mal eine wichtige Aufgabe. In den 1960er Jahren dachte man, dass Krebs durch Mutationen ausgelöst wird, die den mutierten Genen

eine neue Funktion geben (*gain-of-function*). Das bedeutet, dass die Mutation eines Allels (siehe Abschnitt *Ein Gen kommt selten allein: Allele* in diesem Kapitel) ausreicht, um eine Zelle zu einer Krebszelle zu machen. Solch ein Allel nennen wir **Proto-Onkogen**, wenn es noch nicht durch eine Mutation aktiviert wurde und **Onkogen**, nachdem es aktiviert wurde. Proto-Onkogene sind häufig Gene, deren Regulation (siehe auch Abschnitt *Ein Gen wird aktiv* in diesem Kapitel) verändert ist, sodass sie im falschen Gewebe oder zum falschen Zeitpunkt aktiv sind. Der US-amerikanische Genetiker ALFRED KNUDSON (1922–2016) war der Erste, der ganz richtig vermutete, dass Krebs auch durch Funktionsverlust (*loss-of-function*) infolge einer Mutation ausgelöst werden kann. Im intakten Zustand verhindern diese Gene folglich Krebs. Das war natürlich schwer nachweisbar, da, wenn ein Allel mutiert und somit defekt war, noch das zweite Allel ein intaktes Protein codierte. Dieses konnte die Funktion des defekten Allels übernehmen. ALFRED KNUDSON schlug daher die **Two-Hit-Hypothese** vor, wonach beide Allele durch Mutationen defekt sein mussten.



Proto-Onkogene werden zu Onkogenen, indem ihre Genprodukte zum falschen Zeitpunkt oder am falschen Ort aktiv werden. Tumorsuppressorgene (Anti-Onkogene) führen meist durch Verlust ihrer Aktivität zu Krebs.

## Funktionsweise eines Genomwächters

Die Molekularbiologie geht immer der wichtigen Frage nach, wie Form und Funktion zusammenhängen. Die Form ist im Falle des p53 sein Proteinstruktur, die wiederum von seiner Aminosäuresequenz bestimmt wird. Die 393 Aminosäuren des p53 codieren für vier funktionale **Proteindomänen**, die uns schon einiges über seine Arbeit verraten:

- ✓ **Signaldomäne** | Diese führt zu einem Transport des Proteins an seinen Wirkort, den Zellkern.
- ✓ **Oligomerisierungsdomäne** | Sie bewirkt, dass immer vier p53-Proteine gemeinsam als Viererpack (Tetramer) zusammenarbeiten.
- ✓ **DNA-Bindedomäne** | Hiermit erkennt und bindet p53 an spezifische DNA-Abschnitte. Mehr als neunzig Prozent der bekannten p53-Mutationen in menschlichen Tumoren betrifft diese Domäne.
- ✓ **Transaktivierungsdomäne** | Sie dient der Aktivierung anderer Gene.

Nach seiner Herstellung im Zytoplasma wird p53 in den Zellkern transportiert, in dem sich die Chromosomen befinden. Hier tut es sich mit drei weiteren p53ern zusammen und bildet ein funktionsfähiges Tetramer. Dieses erkennt bestimmte Abschnitte auf den Chromosomen und bindet daran. Von hieraus können auch weitere Gene aktiviert werden.



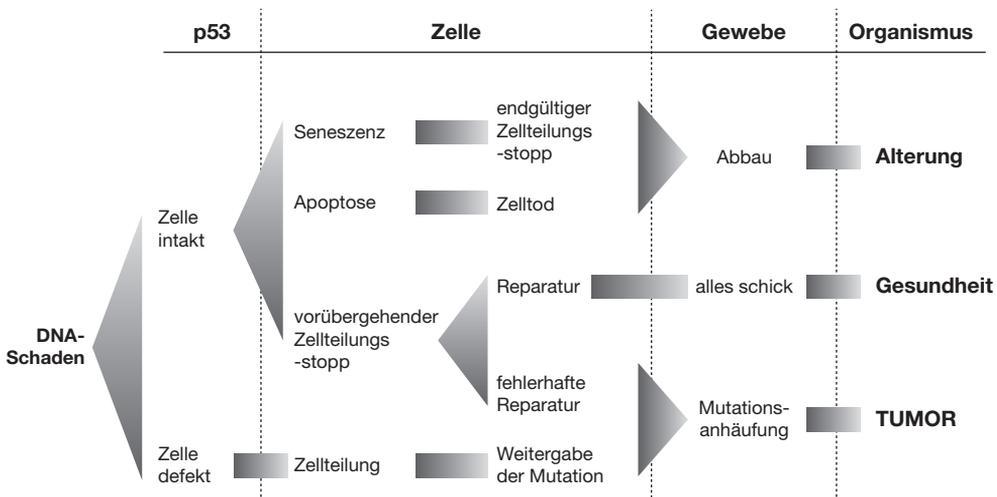
Man dachte zunächst, dass p53 ein Proto-Onkogen ist, da schon ein mutiertes Allel zu Krebs führt. Dies liegt daran, dass bereits ein defektes p53-Monomer in einem Tetramer zu dessen Funktionsverlust führt. Dies ist wie bei einem Auto, wo ein platter Reifen bereits das ganze Auto lahmlegt, auch wenn die anderen drei Reifen noch intakt sind. Erst 1989 wurde endgültig bewiesen, dass p53 ein dominantes Tumorsuppressorgen ist.

Sehen wir uns die **Funktion von p53** einmal genauer an. Wie kann ein Gen beziehungsweise sein Genprodukt über die Integrität des Erbguts wachen? Die Produktion von p53 steht mit seinem Abbau in der Zelle im Gleichgewicht. Seine Halbwertszeit beträgt etwa zwanzig Minuten. Sein Abbau kann beispielsweise durch die Wirkung von **Mutations-sensorproteinen** gestört werden. Diese erkennen DNA-Schäden, die durch die Wirkung ultravioletter Strahlung entstanden sein können (siehe auch Kapitel 5, Abschnitt *DNA-Doppelstrangbruchreparatur*). Dadurch kommt es zu einer Anreicherung von p53, infolgedessen es seine Wirkung voll entfalten kann. So wird beispielsweise der Eintritt der Zelle in die Zellteilung (Mitose) gehindert, indem p53 die Bildung entsprechender »Bremsproteine« fördert. Erst wenn der Schaden behoben wurde, wird das Auf- und Abbau-Gleichgewicht wieder hergestellt und die Zelle darf sich teilen und ihr »repariertes« Erbgut weitergeben. Letztlich bewirkt p53 sogar, dass wir nach regelmäßigen Sonnenbädern braun werden.



Es ist elementar, sich für das Verständnis der Prozesse in einer Zelle deren Dynamik zu vergegenwärtigen. Solange eine Zelle lebt, halten sich Auf- und Abbauprozesse, Ordnung und Chaos die Waage. Auch Moleküle wie Proteine sind nicht einfach »da«. Sie »schwirren« in den Zellen umher und »haften« an manchen Strukturen stärker oder schwächer und werden von anderen sogar abgestoßen.

Tatsächlich ist p53 ein **Master Regulator**, also ein Protein, das über eine Vielzahl zellbiologischer Prozesse wacht und diese reguliert (Abbildung 1.7). Daher ist es auch nicht verwunderlich, dass wir das Protein in vielen anderen Lebewesen finden. Die ersten p53-Proteinvarianten konnten schon in einzelligen Kragengeißeltierchen und mehrzelligen Seeanemonen nachgewiesen werden, mit denen wir vor über einer Milliarde Jahre gemeinsame Vorfahren hatten. Und dort hat es dieselbe Aufgabe wie bei uns. Auch gibt es fiese Viren und Bakterien, welche die Bildung von p53 stören und damit zum Ausbruch von Krebs beitragen können.



**Abbildung 1.7:** Wirkung des »Genomwächters« p53 auf zellbiologische Prozesse, die über Leben und Tod einer Zelle und ihres Gewebes entscheiden.



Nacktmulle haben nicht nur kein Fell und sind, naja, von umstrittener Schönheit – sie bekommen vor allem kein Krebs. Das ist erstaunlich, da diese kleinen Nagetiere über dreißig Jahre alt werden können. Man sieht ihnen ihr Alter nicht einmal an; den einzigen Hinweis liefern ihre Zähne. Sie leben in unterirdischen Bauten, sind recht unempfindlich gegenüber Sauerstoffmangel, können im Gegensatz zu anderen Nagetieren ihre Körpertemperatur nicht regulieren und leben in einer Monarchie mit einer dominierenden Königin. Spätestens, seitdem das Erbgut des Nacktmulls im Jahr 2011 entschlüsselt wurde, ist man auf der Suche nach dem heiligen Gral: der Krebsresistenz. Eine Ursache scheint wiederum p53 zu sein, das bei Nacktmullen eine viel größere Stabilität aufweist.

## Ein Gen kommt selten allein: Allele

Sie ahnten es vielleicht schon: Die beiden Chromosomensätze unserer Körperzellen tragen dieselben Gene, die jedoch unterschiedlich ausgeprägt sein können, wie JOHANN WOLFGANG VON GOETHE (1749–1832) auf seine Weise schon anmerkte:

*»Vom Vater hab ich die Statur, des Lebens ernstes Führen, Vom Mütterchen die Frohnatur und Lust zu fabulieren. Urahn herr war der Schönsten hold, das spukt so hin und wieder; Urahn frau liebte Schmuck und Gold, das zuckt wohl durch die Glieder. Sind nun die Elemente nicht aus dem Komplex zu trennen, Was ist denn an dem ganzen Wicht Original zu nennen?«*

Was GOETHE hier beschreibt ist genau jene Wirkung unterschiedlicher Genvarianten. Oft wird auch von unterschiedlich geprägten Genen gesprochen. Es gibt jedoch unterschiedliche Formen der Prägung (siehe auch Kapitel 12, Abschnitt *Prägung von Genen*). Hier will ich bei Genvarianten bleiben, die wir als **Allele** bezeichnen. Wir tragen von jedem Gen zwei Allele, ein väterliches und ein mütterliches. Diese beiden Allele können identisch sein: Die Abfolge der DNA-Bausteine wäre dann identisch. Aber bereits ein unterschiedlicher Baustein, also ein einzelner Basenpaarunterschied, führt zu unterschiedlichen Allelen (Abbildung 1.8).

**Allel 1:** ... G C T A **A** T C G C ...

**Allel 2:** ... G C T A **C** T C G C ...

**Abbildung 1.8:** Ein einzelner Unterschied in der Abfolge der Nukleotide, folglich der Gensequenz, reicht, um Allele voneinander zu unterscheiden.

Beachten Sie, dass ich in der Abbildung 1.8 nur einen DNA-Strang von jedem Allel aufgeschrieben habe. Das abweichende Nukleotid habe ich grau unterlegt.

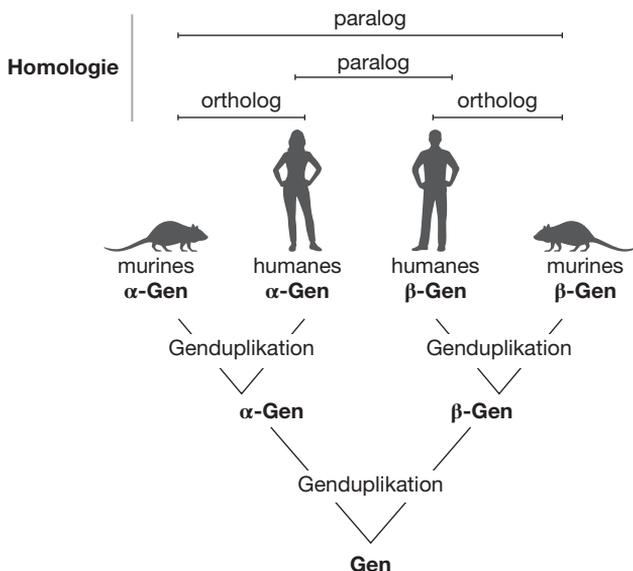


- ✓ Unterschiedliche Sequenzvarianten eines Gens werden als **Allele** bezeichnet.
- ✓ Jeder Mensch trägt zwei Allele, ein mütterliches und ein väterliches.
- ✓ An einen Nachkommen wird jeweils eines der beiden Allele weitergegeben.

- ✓ In einer Population können tausende Allele eines Gens auftreten.
- ✓ Allele entstehen durch Mutationen und Rekombinationen.

## Genverwandtschaften

Neben unterschiedlichen Genvarianten in Form der Allele gibt es noch eine weitere wichtige Einteilung, nämlich **homologe Gene**. Ein homologes Gen (oder einfach Homolog) ist ein Gen, das in zwei oder mehr Arten (Spezies) vorkommt und von einem »Ur-Gen« eines gemeinsamen Vorfahren abstammt (Abbildung 1.9). Befinden sich die Duplikate in dem haploiden Erbgut einer Art, so spricht man von paralogen Genen oder Proteinen. Homologe Gene und Proteine bei unterschiedlichen Arten werden dagegen als ortholog bezeichnet. Der englische Paläontologe RICHARD OWEN (1804–1892) definierte Homologie als »das gleiche Organ in unterschiedlicher Form und Funktion«. Er verstand unter homologen Strukturen solche, die sich zwar im Detail unterschieden, jedoch vom gleichen Körperplan, dem sogenannten »Archetyp«, abgeleitet waren. Im Gegensatz dazu sind analoge Strukturen solche, die zwar ähnliche Funktionen erfüllen, indes nicht vom selben Archetyp abgeleitet wurden, sondern unabhängig voneinander entstanden.



**Abbildung 1.9:** Konzept der Homologie von Genen und Proteinen.

Was bedeutet dies in Bezug auf DNA-Sequenzen?

- ✓ **Homologe Gene** | können in ihrer DNA-Sequenz einander ähnlich sein. Ähnliche Sequenzen sind jedoch nicht unbedingt homolog.
- ✓ **Orthologe Gene** | sind homologe Gene, die nach der Aufspaltung einer Art auf beide Arten aufgeteilt wurden, aber ihre Hauptfunktion erhalten haben. Das Konzept orthologer Gene bezieht sich auf unterschiedliche Arten.

- ✓ **Paraloge Gene** | entstehen durch Gendverdopplungen innerhalb desselben Genoms einer Art. Das Konzept paraloger Gene bezieht sich auf Genduplikate innerhalb des haploiden Genoms einer Art.
- ✓ **Xenologe Gene** | wurden durch horizontalen Gentransfer (Austausch) zwischen verschiedenen Arten erworben.
- ✓ **Allele** | beziehen sich immer auf den Genpool einer Art; sie haben nichts mit Homologie zu tun. Allele haben immer sehr ähnliche DNA-Sequenzen, homologe Gene dagegen nicht zwangsläufig.

## Ein Gen wird aktiv

Nur rund ein bis zwei Prozent des Erbguts codieren für Proteine. Einen relativ großen Anteil am Erbgut haben regulative DNA-Abschnitte (Tabelle 1.2). Besonders wichtig für die Regulierung der Aktivität von Genen sind Promotoren, Enhancer und Silencer. Das ist wie immer im Leben: Um richtig aktiv zu werden, braucht es häufig Unterstützung.

Unmittelbar vor den codierenden Abschnitten auf der DNA liegen die **Promotoren**. Im engeren Sinne codieren sie auch für etwas, jedoch nicht für Proteine oder RNA, sondern für die Aktivität. Ein Promotor ist in der Regel rund vierzig Basenpaare lang. An bestimmte Abschnitte innerhalb dieser vierzig Basenpaare binden **Transkriptionsfaktoren**. Davon gibt es eine große Vielfalt. Diese Transkriptionsfaktoren, die selbst wiederum Proteine sind, wirken wie Angeln: Sie fischen sich eine **RNA-Polymerase**. Dieses Enzym beginnt daraufhin die DNA-Doppelhelix zu öffnen und liest einen DNA-Strang ab. Was heißt das? Das bedeutet, dass ein DNA-Strang, der codogene Strang, als Vorlage für die Biosynthese eines RNA-Moleküls dient. An Promotoren gebundene Transkriptionsfaktoren rekrutieren RNA-Polymerasen. Dieses Verfahren ermöglicht auf einfache Weise Vielfalt: Die Transkriptionsfaktoren sind spezifisch für bestimmte DNA-Sequenzen, während das entscheidende Enzym, das die Arbeit macht, immer dasselbe ist. Tatsächlich ist die RNA-Polymerase ein Multienzymkomplex, da es mehrere Aufgaben gleichzeitig zu erledigen hat.



An Mäusen konnte gezeigt werden, dass 1.063 Gene auch nach dem Tod gesunder erwachsener Tiere noch abgelesen (transkribiert) werden. Während die meisten Transkripte innerhalb von dreißig Minuten nach dem Tod zunahmen, stiegen einige erst nach ein oder zwei Tagen an. Die Daten deuten darauf hin, dass beim Tod des Organismus ein schrittweiser Shutdown stattfindet.

Die **Enhancer** liegen teilweise mehrere tausend Basenpaare vom Promotor entfernt. Durch die dreidimensionale Anordnung der DNA als »Knäuel« im Zellkern können aber auch in der Sequenz weit auseinanderliegende DNA-Abschnitte strukturell eng beieinanderliegen (siehe auch Abschnitt *Gene sind mehr als Gene* in diesem Kapitel). Enhancer fördern die Aktivierung eines Gens, indem sie den Transkriptionsfaktor bei der Rekrutierung der RNA-Polymerase unterstützen. Es gibt natürlich auch gegenteilige Elemente, die **Silencer**. Diese hindern an den Promotor gebundene Transkriptionsfaktoren bei ihrer Arbeit.

Ein Gen kann durchaus unter dem Einfluss von mehreren Promotoren, Enhancern und Silencern stehen. Damit wird alleine schon die Regulierung des Starts der Genaktivierung, der Transkription, beliebig komplex. Aber mit der Initiation der Transkription ist es ja noch nicht getan. Zu jeder Zeit kann die Transkription, also die Abschrift der DNA in ein RNA-Molekül, gestört und unterbrochen werden. Dann entstehen unvollständige Transkripte, die von der Zelle wieder abgebaut und die Nukleotide auf diese Weise recycled werden. Schafft es die RNA-Polymerase bis zum Terminator, dann liegt eine vollständige Abschrift des codogenen DNA-Strangs vor. Dieses **Transkript** kann bereits das Endprodukt sein: eine funktionale RNA, die in der Zelle, zum Beispiel, als **Transfer-RNA** (tRNA) Aminosäuren zu den Ribosomen transportiert oder als **CRISPR-RNA** (crRNA) in Bakterien an der Bekämpfung von Viren beteiligt ist.



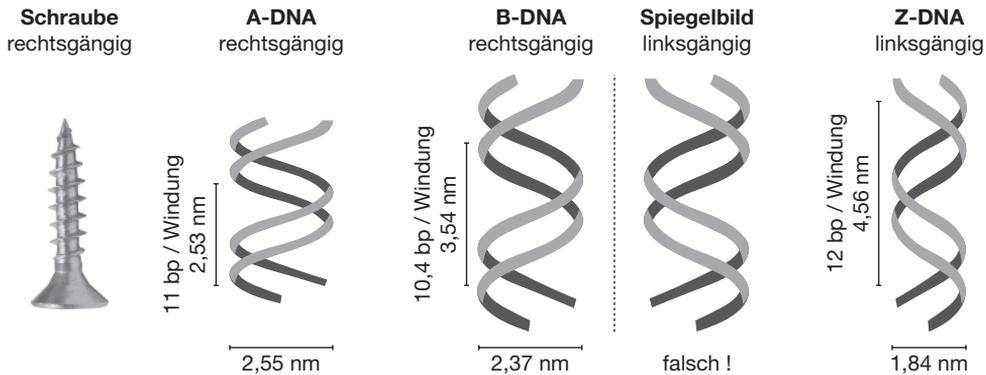
Das angestaubte zentrale Dogma der Molekularbiologie besagt, dass aus einem Gen auf der DNA eine mRNA wird, die in ein Protein übersetzt wird. Aber aus einer RNA-Vorlage kann auch wieder DNA entstehen (reverse Transkription) und RNA kann das funktionale Endprodukt sein.

Meistens denken wir bei Transkripten an die **Boten-RNA** (englisch: *Messenger-RNA*, mRNA) und somit ein Zwischenprodukt, das an den Ribosomen in ein Protein übersetzt (translatiert) wird. Die **Ribosomen** sind riesige Protein-RNA-Komplexe in den Zellen. Sie bestehen aus zwei Teilkomplexen, einer großen und einer kleinen Untereinheit. Ein bestimmter Bestandteil der kleinen Untereinheit erkennt einen kurzen Abschnitt, die **Shine-Dalgarno-Sequenz**, auf der mRNA und bindet daran. Daraufhin kommt die große Untereinheit dazu und die **Translation** kann starten.

## Einfluß von DNA-Strukturvarianten

Wann immer jemand über DNA spricht, haben wir meistens die DNA-Doppelhelix wie in der Abbildung 1.1 vor Augen. Es gibt allerdings noch weitere **Strukturvarianten** der Doppelhelix, die in der Abbildung 1.10 gezeigt sind. Die B-DNA ist jene von JAMES WATSON und FRANCIS CRICK beschriebene Form (siehe auch Abbildung 1.1). Wie eine Schraube mit rechtssteigendem Gewinde ist die Drehrichtung der A- und B-DNA im Uhrzeigersinn. Die Zickzackform der Z-DNA entspricht einem Linksgewinde. Diese Varianten spielen insbesondere bei der Genregulation eine wichtige Rolle. Denn nicht nur die dreidimensionale Verteilung im Zellkern beziehungsweise der Bakterienzelle ist von Bedeutung, sondern auch die Form der DNA-Doppelhelix. Diese bestimmt unter anderem, wie sich bestimmte Proteine an die DNA anlagern können, da sich die Zugänglichkeit der Basen im Inneren der Helix ändert.

Zudem hat die DNA-Doppelhelix, ähnlich wie ein Schneckengehäuse oder eine Schraube, eine Drehrichtung. Na, in welche Richtung dreht man einen Wasserhahn zu? – Oder eine Glühbirne in die Fassung? Im Uhrzeigersinn. Dieselbe Drehrichtung haben Schrauben und die **A-Form** sowie die **B-Form** der DNA. Bei diesem sogenannten rechtssteigendem Gewinde (Rechtsgewinde) verlaufen die Gewindeflanken von links unten nach rechts oben (siehe Abbildung 1.10). Die **Z-DNA**, benannt nach dem Zickzackverlauf des Rückgrats der Helix, ist dagegen linksgängig und strukturell sehr unterschiedlich von der A- und B-Form.



**Abbildung 1.10:** Strukturvarianten der DNA, die unter anderem eine Rolle bei der Regulation der Genaktivität spielen.



Die in Firmenlogos und Abbildungen dargestellte DNA-Doppelhelix sollte rechtsgängig sein. Leider wird sie vielfach, sogar in Lehrbüchern, spiegelbildlich dargestellt. Der Wissenschaftler TOM SCHNEIDER sammelt »falsch gewickelte« DNA in einer *Hall of Fame*. Sogar ein Cover des renommierten Wissenschaftsmagazins *Nature* vom 9. September 2010 ist dabei. Ergo: Designer aufgepasst!

## Das Protein als Genprodukt

Das Resultat eines aktiven Gens ist letztlich ein **Protein**. Die Proteine aller Lebewesen sind aus zwanzig unterschiedlichen Aminosäuren aufgebaut. Manche davon kennen Sie bestimmt. Glutamat ist beispielsweise der gerne verwendete Geschmacksverstärker E620 – aber auch ein sogenannter Neurotransmitter in unserem Nervensystem. Aber keine Sorge, das eine hat nichts mit anderem zu tun. Wichtig für uns ist nur, dass Glutamat ein der zwanzig Aminosäuren ist, aus deren Aneinanderreihung ein Protein entsteht (siehe auch Abbildung 10.2). Diese Abfolge bestimmt die Gensequenz (siehe Kapitel 1, Abschnitt *Der genetische Code*).

Aus der Chemie ist von vielen Molekülen bekannt, dass sie in spiegelbildlichen Varianten vorkommen können. Wir sprechen dann von **Stereoisomeren**. Das gilt auch für Aminosäuren, die ebenfalls in zwei räumlichen Anordnungen vorkommen. In Lebewesen finden sich ausschließlich die sogenannten L-Aminosäuren. Zu jeder L-Aminosäure gibt es eine spiegelverkehrte D-Variante (siehe Abbildung 9.3). Wie fatal es ist, wenn das falsche Stereoisomer zur Anwendung kommt, hat unter anderem der **Contergan-Skandal** Anfang der 1960er Jahre gezeigt. Das Beruhigungsmittel führte bei Schwangeren zu Schwangerschaftsabbrüchen oder schweren Schädigungen des Embryos. Später stellte sich unter anderem heraus, dass der Wirkstoff Thalidomid in zwei stereoisomeren Formen auftritt. Die eine Variante hat die gewünschte sedierende Wirkung, die andere Form ist dagegen fruchtschädigend.

Der US-amerikanische Wissenschaftler GEORGE CHURCH (1954–) hat Überlegungen angestellt, ob es nicht möglich wäre, ein **spiegelbildliches Lebewesen** zu erzeugen. Statt aus L-Aminosäuren müssten die Proteine aus R-Aminosäuren gebildet werden und auch

die DNA-Drehung müsste umgekehrt sein. Ein solches Lebewesen wäre dann nicht mehr kompatibel zu herkömmlichen Organismen; eine echte Parallelwelt entstünde (siehe auch Kapitel 9, Abschnitt *Parallelwelten*).



Das Rechtsgewinde hat sich bei Schrauben – nicht bei der DNA – aus ergonomischen Gründen durchgesetzt, da die meisten Menschen Rechtshänder sind. Rechtshänder können mit einer Drehbewegung im Uhrzeigersinn eine größere Kraft und somit ein größeres Drehmoment, aufbringen. Das erklärt vermutlich, weshalb wir die Schrauben dann nicht mehr heraus bekommen. Fragen Sie das nächste mal einen Linkshänder!

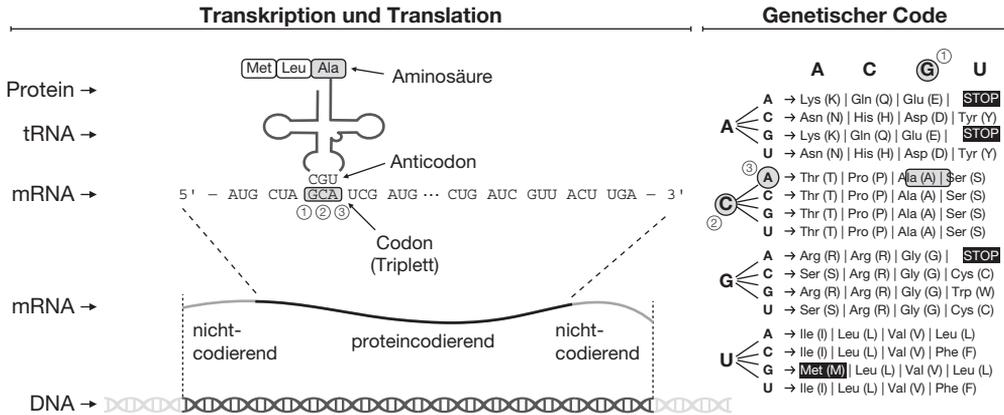
## Der genetische Code

Wir haben nun schon viel über Gene und deren Produkte, die Proteine (Eiweiße), gesprochen. Wie indes wird aus einem Gen ein Protein? Naja, eigentlich ist die Frage nicht präzise gestellt, denn ein Gen »verwandelt« sich nicht in ein Protein, sondern codiert für ein Protein – oder eine regulatorisch oder katalytisch aktive RNA (siehe Abschnitt *Ein Gen wird aktiv* in diesem Kapitel). Der Weg vom Gen zum aus Aminosäuren aufgebauten Protein führt über die Boten-RNA, die wir nun fachgerecht als **mRNA** (*Messenger-RNA*) bezeichnen wollen. Sie entsteht während des von der RNA-Polymerase katalysierten Prozesses der **Transkription** (siehe Abschnitt *Ein Gen wird aktiv* in diesem Kapitel). Das ist zunächst noch, molekular betrachtet, recht langweilig. Richtig spannend wird es bei der nun folgenden **Translation**, bei der auf Grundlage der Nukleotidabfolge der mRNA ein Protein gebildet (synthetisiert) wird (Abbildung 1.11). Dabei codieren immer drei Nukleotide auf der mRNA eine Aminosäure des Proteins. Diese drei Nukleotide nennen wir ein **Triplet** oder Codon. Die Decodierung, um in der Sprache der Kryptologie zu bleiben, übernimmt ein weiteres RNA-Molekül, die Transfer-RNA, kurz **tRNA**. Die tRNA wird zweidimensional als dreiblättriges Kleeblatt dargestellt, dreidimensional sieht sie eher wie ein Amboss aus. An dem einen Ende ist eine Aminosäure angehängt, an dem anderen Ende befindet sich das **Anticodon**, das die »spiegelbildliche« (revers-komplementäre) Sequenz zum Codon aufweist. Wie in der Abbildung 1.11 dargestellt, kommt es zu einer Basenpaarung zwischen der tRNA und einem Triplet auf der mRNA. Dies geschieht allerdings nicht so »nackig« wie dargestellt, sondern findet im Zentrum von riesigen Molekülaggagaten, den **Ribosomen**, statt.

Der genetische Code gilt, von kleinen Abweichungen abgesehen, für alle Lebewesen. Auf der rechten Seite der Abbildung 1.11 sind die Aminosäuren in ihrer Dreibuchstaben- und, in Klammern, Einbuchstabenabkürzung aufgeschrieben. Das Startcodon AUG und die Stopcodons UGA, UAA und UAG sind schwarz unterlegt.



Erstaunlich, dass zwischen sechs und neun Wasserstoffbrückenbindungen zwischen den Nukleotiden eines Triplets der tRNA und mRNA über die korrekte Übersetzung in eine Aminosäure entscheiden. Die dahinterstehende Bindungsenergie ist winzig klein. Ohne die strukturegebende Hilfe der Ribosomen wäre der Prozess undenkbar.



**Abbildung 1.11:** Der genetische Code.

## Translation am Ribosom

Ribosomen sind Multienzymkomplexe, das heißt, dass sie zum einen aus vielen Komponenten aufgebaut sind und, zum anderen, viele chemische Reaktionen gleichzeitig katalysieren. Die Komponenten sind Proteine und ribosomale RNA, die sogenannte **rRNA**. Je nach Dichte, folglich Schwere und Größe, werden die rRNA-Bausteine voneinander unterschieden. Da die Dichte in der Molekularbiologie mit der von dem schwedischen Chemiker und Nobelpreisträger THEODOR SVEDBERG (1884–1971) benannten Einheit gemessen wird, sprechen wir beispielsweise von der 16S rRNA. Deren Nukleotidsequenz wird für die Berechnung der Verwandtschaft von Bakterien (genauer: Prokaryonten) verwendet. Das Gegenstück dazu ist die 18S rRNA bei Tieren, Pflanzen und Pilzen, den Eukaryonten.



Da Ribosomen bei allen bekannten Lebewesen die Übersetzung (Translation) des genetischen Codes vornehmen, sind sie hoch konserviert. Das heißt, dass sie und ihre Komponenten sich von Lebewesen zu Lebewesen kaum unterscheiden. Dies hat sich der US-amerikanische Mikrobiologe CARL WOESE zunutze gemacht, um als Erster durch den Vergleich von 16S rRNA-Sequenzen, einen Stammbaum der Bakterien zu erstellen. Dabei entdeckte er *en passant* einen neuen Bakterienstamm, die Archaeobakterien.

Der **Ablauf der Translation** an den Ribosomen ist in drei Schritte einteilbar:

- ✓ **Initiation** | Ein unbeladenes Ribosom erkennt eine freie mRNA und bringt sich am ersten Codon, dem Startcodon AUG (entsprechend ATG auf der DNA), in Position.
- ✓ **Elongation** | Nacheinander präsentiert das Ribosom Codon für Codon an einer für die Transfer-RNA zugänglichen Stelle. Die jeweils zum präsentierten Codon passende Transfer-RNA bindet an den mRNA-Ribosom-Komplex und die von der Transfer-RNA herangezogene Aminosäure wird an die bereits bestehende Aminosäurekette »angeheftet«. Die entladene Transfer-RNA wird wieder entlassen und das Ribosom präsentiert das nächste Codon.

- ✓ **Termination** | Erkennt das Ribosom am Ende der Boten-RNA das Stopcodon, so wird die Boten-RNA freigegeben und das Ribosom kann die nächste Boten-RNA übersetzen.

Das **Startcodon** bezeichnet den Beginn des Bereichs auf der mRNA, der in ein Protein übersetzt werden soll. Die Sequenz des Startcodons lautet auf der mRNA AUG. Da AUG für die Aminosäure Methionin codiert (siehe Abbildung 1.11), beginnt jedes Protein augenscheinlich mit der Aminosäure Methionin (Met, M). Das Ribosom »wandert« dann in Dreierschritten Codon für Codon auf der mRNA entlang, während gleichzeitig die Aminosäurekette, also das Protein, Aminosäure für Aminosäure verlängert wird. Proteine sind im Durchschnitt etwa dreihundert Aminosäuren lang – das entspricht logischerweise neunhundert Nukleotiden. Trifft das Ribosom auf eines der drei Stopcodons UGA, UAA oder UAG, dann bindet daran keine tRNA, sondern ein Protein, das die Beendigung (Termination) der Translation einleitet.



Eine Analyse aller Proteine, des Proteoms, hat schnell gezeigt, dass entgegen der Erwartung nicht alle Proteine mit der Aminosäure Methionin beginnen. Es ist eine der vielen möglichen posttranslationalen Modifikationen, also Veränderungen am Protein nach dessen Geburt bei der Translation, dass das Methionin entfernt wird.

Diese vereinfachte Darstellung darf nicht davon ablenken, dass der Prozess der Translation an den Ribosomen sagenhaft komplex ist und an vielen Stellen reguliert wird. Zwei Dinge sind aber wesentlich:

- ✓ **Codon-Anticodon-Wechselwirkung** | Ganz offensichtlich ist die exakte Codon-Anticodon-Wechselwirkung (Hybridisierung) des tRNA-Anticodons an das mRNA-Codon im Herzen des Ribosoms der Schlüssel für die **Dechiffrierung**.
- ✓ **Beladung der tRNA** | Weniger offensichtlich ist, dass die korrekte Beladung einer entladenen tRNA mit der zum Anticodon passenden Aminosäure entscheidend für die korrekte Übersetzung ist. Dafür sind Enzyme mit dem kompliziert klingenden Namen **Aminoacyl-tRNA-Synthetase** verantwortlich.

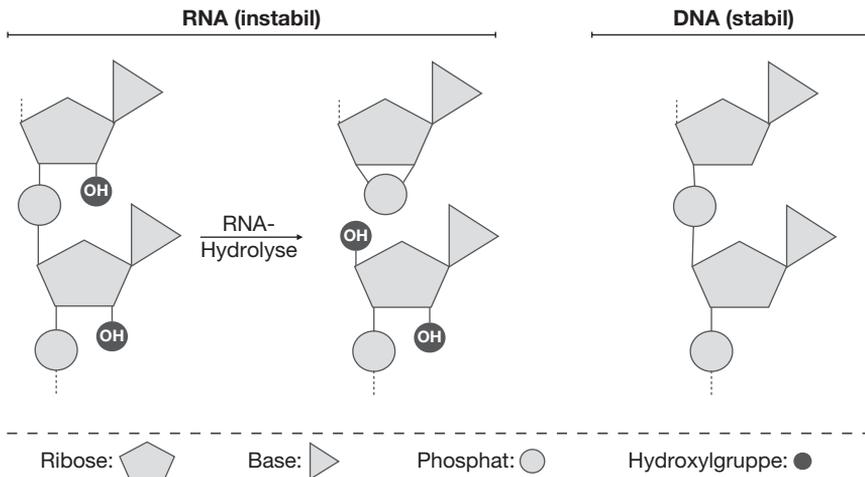


Das Ribosom und die Aminoacyl-tRNA-Synthetase sorgen seit Milliarden von Jahren für die korrekte Übersetzung des genetischen Codes – bei allen Lebewesen. So wichtig wie die korrekte Translation des genetischen Codes am Ribosom ist, so wichtig ist die korrekte **Beladung der tRNA** mit der zum Triplet gehörigen Aminosäure. Der Aminoacyl-tRNA-Synthetase kommt demzufolge eine wichtige und meist unterschätzte *Gatekeeper*-Funktion bei der Erhaltung des genetischen Codes zu.

## Ursprung der mRNA

Warum gibt es überhaupt eine mRNA neben der DNA? Warum so kompliziert? Dafür gibt es zwei Erklärungsansätze. Zum einen wird davon ausgegangen, dass im Laufe der Evolution zunächst RNA als alleiniger Informations- und Funktionsträger vorlag. **Ribozyme** sind RNA-Moleküle, die biochemische Reaktionen katalysieren können. Gleichzeitig kann

dieses Molekül kopiert werden. Zwei Aufgaben sind hier folglich vereint. Die RNA ist aber chemisch labil und zerfällt schnell (Abbildung 1.12). Bei dieser **RNA-Hydrolyse** wird die freie Hydroxylgruppe (-OH) an dem Zucker Ribose der RNA attackiert und die RNA aufgespalten (hydrolysiert). Daher war die Entwicklung von DNA als langfristiger Datenspeicher sehr hilfreich. Das ist die evolutionäre Erklärung. Die Evolution hat daher aus der Not eine Tugend gemacht und die Instabilität der RNA für die **Regulation der Genexpression** genutzt. Die RNA ist ein gutes Molekül, um Daten vorübergehend (transient) zu »präsentieren« (exprimieren). Der Preis für diese Form der Regulation ist die notwendige ständige Produktion der mRNA für Proteine, die benötigt werden.



**Abbildung 1.12:** RNA ist instabil und wird schnell abgebaut.



DNA ist stabil und kann tausende Jahre überstehen. RNA ist instabil und eignet sich daher für Regulationsprozesse.

## Mehrdeutigkeiten

Wer sich den Code in der Abbildung 1.11 anschaut, stellt fest, dass es für manche Aminosäuren mehrere Triplets gibt. So wird die Aminosäure Alanin (Ala) durch die vier Codons GCA, GCC, GCG und GCU codiert. Das heißt, der Code ist **degeneriert**, also mehrdeutig. Für andere Aminosäuren gibt es nur ein einziges Triplett. Ganz grob gilt, dass für Aminosäuren, die häufig in Proteinen verbaut sind, mehr Triplets vorhanden sind. Dies entlastet die zugehörigen tRNAs-Pools, die ja jeweils neu beladen werden müssen. Jeder Organismus hat eigene **Codonpräferenzen** (englisch: *codon usage*). Wenn ein Gen des Menschen in Hefezellen gebildet (exprimiert) werden soll, muss eine **Codonoptimierung** vorgenommen werden, um eine gute Ausbeute zu erzielen.



Die Codonpräferenz gleicht einem Fingerabdruck für eine bestimmte Art. Sie dient auch als Indiz für den horizontalen Austausch von Genen zwischen Arten (siehe Kapitel 3, Abschnitt *Gen-Shopping – Horizontaler Gentransfer*). Brisanz erlangte das Thema jüngst (Juni 2021) in der Diskussion darüber, ob das

Coronavirus SARS-CoV-2 aus dem Labor stammt oder natürlich entstanden ist. Menschliche Zellen codieren die Aminosäure Arginin gerne mit den Codons CGT, CGC oder CGG. Doch CGG ist bei Coronaviren das unbeliebteste Codon für Arginin. SARS-CoV-2 besitzt sogar ein Doppelcodon, also: CGGCGG. Nur fünf Prozent der Arginincodons von SARS-CoV-2 sind CGG und ein Doppelcodon wurde bisher in keinem anderen Beta-Coronavirus gefunden. Zufall? Das gilt es noch zu klären.



Auf der Webseite zum Buch <http://bit.do/gen-dummie> können Sie den genetischen Code »in Aktion« sehen.

## Mutationen: Quelle genetischer Unterschiede

Ohne Mutationen würde die Evolution stillstehen. Es könnte nichts Neues mehr entstehen, sondern »nur« Vorhandenes neu kombiniert werden (siehe auch Abschnitt *Chromosomen, Sex und Mutationen* in diesem Kapitel). Mutationen führen dagegen dazu, dass nicht die Kombination der Gene im Erbgut verändert wird, sondern deren DNA-Sequenz (Abbildung 1.13). Der häufigste Prozess bei der **Entstehung neuer Gene** ist die vorhergehende Verdopplung eines vorhandenen Gens. Dabei wird in der Regel nicht gezielt ein Gen, sondern ein ganzes Chromosom oder ein großer Abschnitt auf einem Chromosom, zum Beispiel, durch die Wirkung von Transposons (siehe Kapitel 3, Abschnitt *Gene-Hopping – Das dynamische Genom*) verdoppelt. Nehmen wir beispielhaft an, dass das Gen A die Funktion A erfüllt und nun doppelt vorliegt. Dann kann die eine Kopie des Gens gefahrlos mutieren, ohne dem Organismus zu schaden, da die andere Kopie die Funktion noch zu erfüllen vermag.

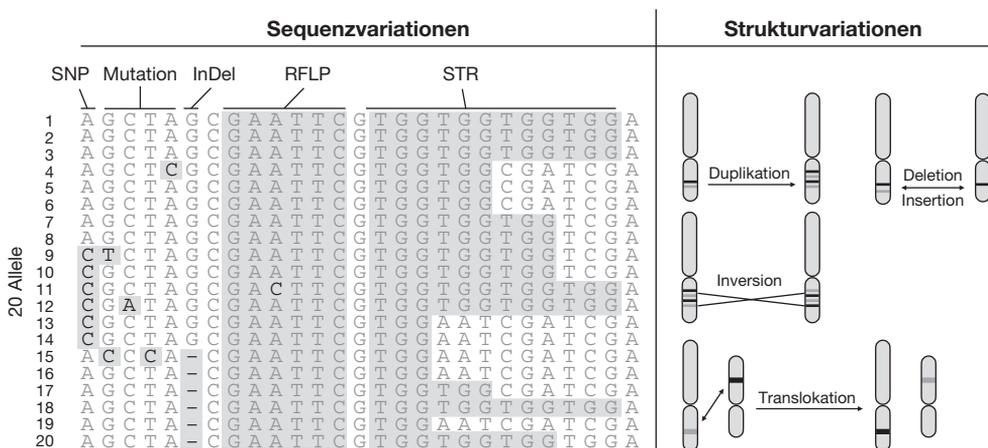


Abbildung 1.13: DNA-Sequenz- und Strukturpolymorphismen.

Mutationen sind somit der Motor der Evolution. In der Abbildung 1.13 sind typische Mutationen, sowohl die DNA-Sequenz als auch die Chromosomenstruktur betreffend, dargestellt. Von zwanzig Allelen ist jeweils ein DNA-Strang gezeigt. Einzelnukleotid-**Variationen** (SNPs, englisch: *single nucleotide polymorphisms*) unterscheiden sich von Mutationen dadurch, dass erstere in mindestens einem Prozent der Allele auftreten – es ist demnach eine Sache der Definition. Bei Insertionen und Deletionen (InDels, englisch: *insertions/deletions*) fehlen in einigen Allelen Basenpaare. Eine Mutation (A > C im Allel 11) in der Erkennungssequenz für das Restriktionsenzym EcoRI (GAATTC, siehe auch Kapitel 5, Abschnitt *Restriktionsenzyme*) bewirkt einen Restriktionspolymorphismus (RFLP, englisch: *restriction fragment length polymorphism*, siehe Kapitel 5, Abschnitt *DNA-Schnipsel*). Die Mikrosatellitensequenz TGG führt durch unterschiedlich häufige Wiederholungen (STR, englisch: *short tandem repeat*, siehe auch Kapitel 8, Abschnitt *Forensische DNA-Tests*) zu einem Längenpolymorphismus. Strukturvariationen betreffen immer lange DNA-Abschnitte.

Die wichtigsten die DNA-Sequenz betreffenden **Mutationen** sind:

- ✓ **Basenaustausche** | Einfache Basenaustausche, auch Punktmutationen genannt, machen die häufigste Zahl der Mutationen aus. Betreffen sie einen proteincodierenden Bereich, dann können sie den genetischen Code verändern und zu einer anderen Aminosäure im Protein führen. Dies kann einen Defekt und eine veränderte Funktion des Proteins zur Folge haben. Werden Basenaustausche an die Nachkommen weitergegeben und verteilen sich in einer Population, dann werden sie ab einer Häufigkeit von einem Prozent aller Allele als **SNPs** (*single nucleotide polymorphisms*) bezeichnet. Diese spielen für die Diagnostik eine riesige Rolle (siehe Abschnitt *Viele kleine Unterschiede: SNPs* in diesem Kapitel). Andernfalls sind es **SNVs** (*single nucleotide variations*).
- ✓ **Deletionen** | Verluste von Nukleotiden in der DNA können groß oder klein sein. Selbst die Deletion eines einzelnen Nukleotids kann einen großen Einfluss haben, etwa wenn sie einen proteincodierenden Bereich betrifft. Dadurch kommt es zu einer Verschiebung im genetischen Code und zu einem veränderten Protein (siehe auch Kapitel 13, Abschnitt *Nana und Lulu läuten eine neue Ära ein*).
- ✓ **Insertionen** | Für Einfügungen gilt das Gleiche wie für Deletionen.
- ✓ **Wiederholungen** | Wiederholungen (englisch: *repeats*) entstehen in der Regel durch Fehler der DNA-Polymerase beim Kopieren der DNA. Durch »Stottern« der DNA-Polymerase können vielfache Wiederholungen entstehen, die als *Short Tandem Repeats* (**STRs**) die Grundlage für den Vaterschaftstest und den genetischen Fingerabdruck legen (siehe auch Kapitel 8, Abschnitt *Forensische DNA-Tests*).

## Viele kleine Unterschiede: SNPs

Zwei zufällig ausgewählte menschliche Genome sind in 99,9 Prozent der DNA-Sequenz identisch. Die restlichen 0,1 Prozent der DNA enthalten Sequenzvariationen. Mit rund 335 Millionen beobachteten Unterschieden beim Vergleich menschlicher Genome verschiedener Populationen, sind **Einzelnukleotidaustausche** (englisch: *single nucleotide polymorphism*) oder **SNPs** (gesprochen *snippets*) der häufigste Typ solcher Variationen; gefolgt

von kurzen Insertionen oder Deletionen. Etwa alle ein- bis zweitausend Nukleotide findet sich beim Vergleich zweier menschlicher Genome ein SNP. Diese Variationen bilden eine Grundlage der Diversität in der Bevölkerung, der Individualität, der Anfälligkeit für Krankheiten und der individuellen Reaktion auf Medikamente (siehe auch Tabelle 8.3). Die Bestimmung, welcher SNP mit welcher Krankheit in Verbindung steht, gilt derzeit als heiliger Gral der genetischen Diagnostik und wird uns unter anderem in Kapitel 10: *Künstliche Intelligenz und echte Erkenntnisse* beschäftigen.



SNPs sind Polymorphismen, also Vielfältigkeiten. Wenn wir einen einzelnen haploiden Chromosomensatz betrachten, dann können wir keine Vielfältigkeit entdecken. Erst der Vergleich mit weiteren Chromosomensätzen beziehungsweise Allelen lässt genetische Variationen wie SNPs hervortreten.

Wie können wir einen SNP von einer Mutation unterscheiden? Oder, ist eine Mutation bereits ein SNP? Nein. Eine **Mutation** entsteht durch einen Fehler bei der Verdopplung des Erbguts vor der Zellteilung. Das kommt selten vor, aber es passiert. Erfolgt diese Mutation in einer unserer Körperzellen, dann werden alle nachfolgenden Körperzellen ebenfalls diese Mutation tragen. Ultimativ nehmen wir sie mit ins Grab. Nur wenn die Mutation in einer Keimzelle, also Ei- oder Samenzelle vorliegt, kann sie an nachfolgende Generationen weitergegeben werden. Wenn sich die Mutation so im Laufe der Generationen in einer Population ausbreitet und schließlich in mehr als einem Prozent aller Individuen einer Population vorhanden ist, dann bezeichnen wir die Mutation als SNP.



Mutationen von einzelnen Basenpaaren in der Keimbahn, die an Nachkommen weitergegeben werden und sich in einer Population verteilen, nennen wir SNP. Betreffen diese Mutationen Körperzellen, wird meist von SNVs (englisch: *single nucleotide variation*) gesprochen.

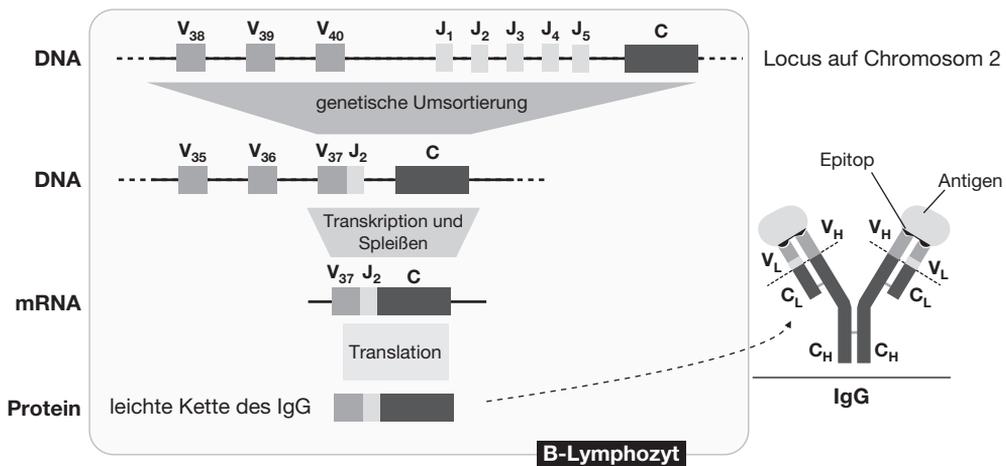
SNPs in proteincodierenden DNA-Regionen können zu einer Änderung des genetischen Codes und damit zu einer anderen Aminosäure in einem Protein führen. Die wiederum führt häufig zu einer Beeinträchtigung der Funktion des Proteins: eine Krankheit entsteht. Genetische Erkrankungen wie die Sichelzellenanämie,  $\beta$ -Thalassämie (siehe auch Kapitel 13, Abschnitt *Geneditierung beim Menschen*) und Mukoviszidose sind typische Beispiele. Auch der Schweregrad von Krankheiten und die Art und Weise, wie der Körper auf Behandlungen reagiert, sind Ausdruck von genetischen Variationen, die durch SNPs verursacht werden. Zum Beispiel ist eine Ein-Basen-Mutation im Apolipoprotein-E Gen *apoE* mit einem geringeren Risiko für die Alzheimer-Krankheit verbunden (siehe auch *Anekdote* in Kapitel 5, Abschnitt *Die zweite Generation*).

## Immunsystem: Umsortieren und ändern

Das Variabelste, das unser Körper hervorbringt, ist vermutlich unser Immunsystem (siehe auch Kapitel 15, Abschnitt *Unser Immunsystem*). Es ist in der Lage, über hundert Milliarden verschiedene **Antikörper** gegen körperfremde Moleküle (**Antigene**) zu produzieren. Und das ganz ohne Mutationen.

Es gibt eine Reihe verschiedener Antikörpertypen, die wir auch als **Immunglobine** bezeichnen. Das häufigste und wichtigste Immunglobin ist der Antikörper IgG. Er ist auch der einzige Antikörper, der während einer Schwangerschaft in den **Embryo** gelangt und diesem eine passive Immunität verleiht. Gebildet werden die **Immunglobine G** in den weißen Blutkörperchen (**Lymphozyten**), genauer in den sogenannten B-Lymphozyten.

Das IgG besteht aus jeweils zwei identischen schweren (H, englisch: *heavy*) und leichten (L, englisch: *light*) Proteinketten, die miteinander verbunden sind. Jede Proteinkette hat zwei Bereiche: einen konstanten Bereich (C, englisch: *constant*) und einen variablen Bereich (V, englisch: *variable*) (siehe Abbildung 1.14). Dieser variable Bereich ist für die Erkennung von Fremdmolekülen verantwortlich.



**Abbildung 1.14:** Herstellung und Aufbau eines IgG-Antikörpers.

Die Besonderheit bei der Synthese der Antikörper ist der zugrundeliegende Umbau auf der Ebene des Erbguts. Während der Reifung der B-Lymphozyten werden auf Chromosom 2 DNA-Segmente ausgetauscht und teilweise verworfen. So werden im Falle der leichten Proteinkette von vierzig V-Segmenten (*variable*) und fünf J-Segmenten (englisch: *join*, verbinden) je eines miteinander verknüpft (rekombiniert) und mit einem C-Segment (*constant*) zusammengeführt. Das ergibt zweihundert Kombinationsmöglichkeiten. Durch zusätzliche, in der Abbildung 1.14 nicht dargestellte D-Segmente (*diversity*) und kleine Fehler bei der Verknüpfung der Segmente, steigt die Variabilität noch einmal erheblich an. Ähnliches geschieht bei der Bildung der schweren Proteinkette. So entstehen Milliarden von Antikörpern, die sich an den Armen des »Y« unterscheiden.

Wenn nun ein Antikörper erfolgreich ein Antigen bindet, dann wird der B-Lymphozyt, der den Sechser im Lotto hatte, gezielt vermehrt; ein Prozess, den wir als **klonale Selektion** bezeichnen. Außerdem kann der B-Lymphozyt »gespeichert« werden, was dann das Gedächtnis des Immunsystems ausmacht.



## Wissen(ge)schaf(f)t → Dumme war gestern ...

- ✓ Die Struktur der DNA ist variabel und nimmt Einfluss auf die Aktivität von Genen.
- ✓ Rund 3,2 Milliarden Basenpaare verteilen sich auf je 23 väterliche und mütterliche Chromosomen. Nur knapp drei Prozent sind codierend, dem Rest werden regulative Aufgaben zugeschrieben.
- ✓ Jedes Gen liegt in zwei Allelen vor. Sie unterscheiden sich überwiegend durch SNPs.
- ✓ Ganz ohne Mutationen sorgen Rekombination und Crossing-Over bei der Meiose für genetische Vielfalt.
- ✓ Gene codieren funktionale RNA oder Proteine, können Introns besitzen und werden in ihrer Aktivität unter anderem durch Promotoren, Enhancer, Silencer und Terminatoren beeinflusst.
- ✓ Das Immunsystem der Säugetiere produziert unzählig viele unterschiedliche Antikörper durch aufwendige genetische Rekombinationen.