

Genetische Vielfalt beim Menschen: Focus auf SNPs

„The capacity to blunder slightly is the real marvel of DNA. Without this special attribute, we would still be anaerobic bacteria and there would be no music.“

- Lewis Thomas

Die Mechanismen durch die genetische Vielfalt entsteht, sind nicht für alle Spezies gleich

Die nachfolgenden Diskussionen beziehen sich auf Menschen und viele der hier erläuterten Fakten, Ideen und Mechanismen sind spezifisch für Menschen. Je weiter andere Organismen evolutionär vom Menschen entfernt sind, desto weniger ist die Situation übertragbar. Bei freilebenden Säugetieren z.B. ist vieles hier beschriebene zutreffend. Bei Säugetieren die stark gezüchtet werden (z.B. Hunden) ist schon vieles recht anders. Bei Einzellern und insbesondere bei Bakterien sind hingegen die Struktur der genetischen Vielfalt und die Mechanismen durch die diese Vielfalt entsteht völlig anders.

Die Grösse des menschlichen Genoms ermöglicht eine grosse Vielfalt trotz hoher inter-individueller Ähnlichkeit

Von einer Warte aus betrachtet sind sich Menschen genetisch sehr ähnlich. Untersucht man die Genome zweier nicht miteinander verwandter Menschen so findet man, dass die Sequenz der beiden Genome zu ~99.9% identisch ist.

Andersherum betrachtet, hat ein haploides menschliches Genom aber auch ca. 3 Milliarden Basenpaare (also 6 Milliarden bp im diploidem Genom). Die 0.1% Unterschiede entsprechen also immer noch ca. 6 Millionen genetischen Variationen zwischen den zwei untersuchten Personen. Trotz der insgesamt grossen Ähnlichkeit gibt es also eine Vielzahl von interindividuellen Variationen.

Genetische Vielfalt ist nicht zufällig verteilt, sondern folgt klaren Mustern

Durch die systematische Untersuchung der Genome von immer mehr Menschen hat man inzwischen 40 Millionen

Stellen im Genom gefunden, an denen sich Menschen besonders häufig unterscheiden.

Dabei hat sich herausgestellt, dass diese Variationen nicht zufällig auftreten, sondern ganz klaren Mustern bzw. Strukturen folgen.

Einige der im Kurs besprochenen experimentellen Methoden (z.B. genom-weite Assoziationsstudien GWAS) funktionieren nur, weil die genetische Variation im menschlichen Genom diesen Mustern folgt. Wir vertiefen hier also noch einmal die grundlegenden Konzepte und Definitionen die Sie bereits im ersten Semester gelernt haben. Dabei konzentrieren wir uns hauptsächlich auf die am häufigsten auftretenden Genetischen Variation, die sogenannten SNPs.

Single Nucleotide Polymorphismen (SNPs) sind die häufigste genetische Variation

Wie der Name sagt, ist ein SNP eine genetische Variation bei der ein einzelnes Basenpaar gegen ein anderes ausgetauscht ist (Abbildung 1).

Dabei ist es wichtig zu betonen, dass ein SNP keine „Beschädigung“ der DNA im Sinne eines *basepair mismatch* darstellt. Wenn also wie in Abbildung 1 gezeigt, bei einem

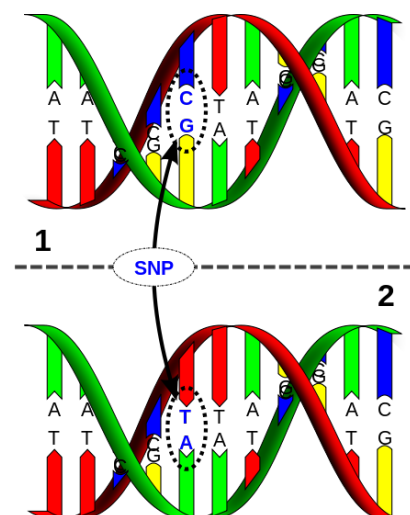


Abbildung 1 Gezeigt sind die einander entsprechenden Abschnitte der DNA Doppelstränge aus den Genomen zweier Personen. In diesem Abschnitt ist die Sequenz von Person 1 (oben) AAGCCTA wohingegen Person 2 (unten) die Sequenz AAGCTTA hat. Die beiden Sequenzen unterscheiden sich also nur in einem einzelnen Basenpaar. Eine solche genetische Variation bezeichnet man als Single Nucleotide Polymorphismus (SNP).

SNP auf einem der DNA-Stränge ein C gegen ein T ausgetauscht ist, ist auf dem Gegenstrang die komplementäre Base auch immer ausgetauscht. Mit anderen Worten, das G auf dem komplementären Strang ist dann durch ein A ersetzt. Beide Varianten haben ein perfektes Basenpaar.

SNPs vs. private Punktvarianten

Wenn man nicht nur zwei, sondern mehrere Genomsequenzen miteinander vergleicht (Abbildung 2) stellt man fest, dass Basensubstitutionen an bestimmten Positionen in der Genomsequenz besonders häufig auftreten, während sie an den meisten anderen Positionen sehr selten sind.

Im strengeren Sinne bezeichnet man nur solche Einzelbasensubstitution als SNPs, die an einer dieser häufig substituierten Positionen auftreten. Dabei hat sich eine Häufigkeit von 1% als *cutoff* eingebürgert. Dieser Wert von 1% ist sicherlich mit einer gewissen Willkür gesetzt. Wichtig ist aber, dass man SNPs konzeptionell von solchen genetischen Variationen trennt, die nur in einer Einzelperson oder einer einzelnen Familie auftreten. Die letzteren Variationen bezeichnet man gemeinhin als private (Punkt-)Varianten.

Der entscheidende Unterschied zwischen SNPs und privaten Punktvarianten ist dabei das Alter dieser Variationen. SNPs im strengeren Sinne sind in der Regel sehr alt (>100'000 Jahre). Diese Varianten sind also schon durch viele Generationen vererbt worden und sind mehrfach

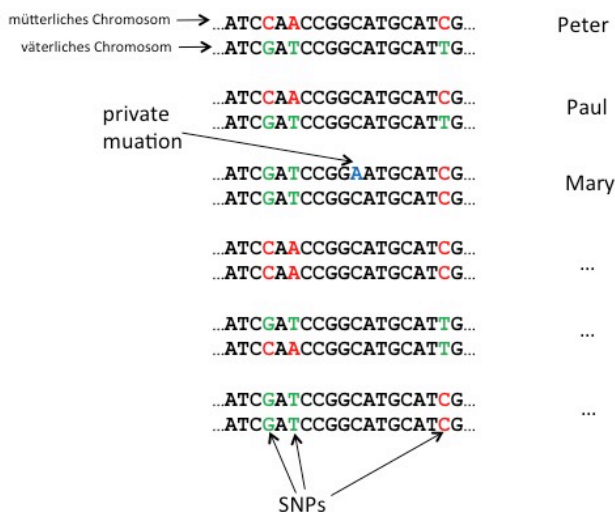


Abbildung 2 Im Gegensatz zu sogenannten privaten Mutationen findet man bei SNPs dieselbe Basensubstitution bei vielen Personen. Strenggenommen bezeichnet man die Substitution eines Basenpaares nur dann als SNP, wenn sie in mindestens 1% der Gesamtbevölkerung zu finden ist. Bitte beachten: In dieser Abbildung wird von jedem Chromosom nur ein Strang des DNA Doppelstrangs gezeigt.

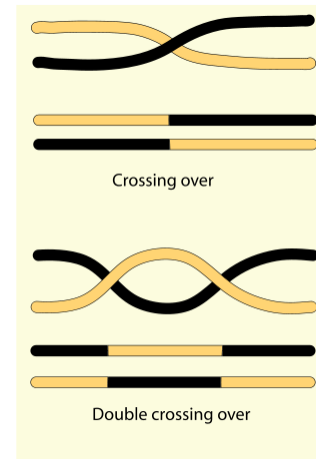


Abbildung 3 Ein Crossing Over Event ist ein Prozess bei dem zwei Chromosome untereinander einen Teil ihrer Doppelstränge gegeneinander austauschen. Dies ist kein „Unfall“, sondern ein genau choreographierter und kontrollierter Prozess, der während der Prophase der Meiose stattfindet. Durch solche Crossing Overs entsteht Diversität durch die Rekombination von existierenden genetischen Variationen.

durch *crossing over events* (Abbildung 3) zwischen homologen Chromosomen ausgetauscht worden. Die Verteilung von SNPs in einer Population und die Wahrscheinlichkeit, dass bestimmte SNPs gemeinsam auftreten, folgen also ganz bestimmten statistischen Mustern die durch diese, viele Generationen dauernden Prozesse geformt wurden. Private Punktvarianten sind noch sehr jung und folgen nicht diesen statistischen Mustern.

SNPs starten als private Punktvarianten

Sowohl SNPs als auch private Punktvarianten haben ihren Ursprung in einer Punktmutation in einer Keimbahnzelle. In jeder Generation treten ca. 30 solcher neuen Mutationen auf und werden dann – vielleicht – an die nächste Generation weitergegeben. Wie die nachfolgende Überlegung zeigt, werden die meisten dieser Varianten aber wieder aus der Population verschwinden– selbst wenn diese genetische Variante keine phänotypischen Nachteile verursacht. Geht man z.B. davon aus, dass ein Elternpaar 2 Kinder hat und dass einer der beiden Eltern für eine autosomale Variante heterozygot ist, so besteht bereits in der ersten Generation eine 25% Chance, dass keines der Kinder diese Variante erbt. In der nächsten Generation besteht dasselbe Risiko wieder. Und wenn eine Variante erst einmal aus dem Genpool verschwunden ist, bleibt die Chance, dass sie weitervererbt wird Null. Durch Zufall oder Selektion kommt es aber immer wieder vor, dass einige Punktvarianten sich in einer Population so weit ausbreiten, dass die Wahrscheinlichkeit, dass sie

verloren gehen, sehr stark abfällt. Während dieser Ausbreitung, die einige hundert oder tausend Generationen dauern kann, wird diese Variante in einer Vielzahl von crossing-over events gemeinsam mit benachbarten SNPs transferiert. Dabei entwickeln sich die entsprechenden Co-Vererbungsmuster und aus der privaten Punktvariante ist ein SNP geworden.

In der Folge driftet die Frequenz mit der dieses SNP in der Population auftritt, über einen sehr langen Zeitraum bis die Frequenz per Zufall irgendwann auf null abfällt. Dann verschwindet das SNP wieder. Oder die Frequenz mit der die „neue“ Variante in der Population auftritt, steigt auf 100% und die ursprüngliche Variante verschwindet. In diesem Fall spricht man davon, dass das SNP fixiert ist.

Wie lange dauert es von der Entstehung einer Punktmutation bis zur Fixierung?

Basierend auf der Anzahl von SNPs die zurzeit in menschlichen Populationen anzufinden sind und der Häufigkeit mit der dieselben SNPs in nahe verwandten Spezies (Schimpansen und Orang-Utans) anzutreffen sind, kann man abschätzen, dass der Zufallsprozess von der Entstehung bis zu Fixierung eines menschlichen SNPs im Mittel ca. 250'000 Jahre dauert.

Bietet die neue Variante aber einen evolutionären Vorteil, bzw. Nachteil so kann sich die Fixierung bzw. Elimination sehr stark beschleunigen. Ein eindrucksvolles Beispiel hierfür ist der dramatische Anstieg einer genetischen Variation im regulatorischen Bereich des menschlichen Lactase Gens, der in Nordeuropäern stattgefunden hat. In nicht-europäischen Bevölkerungen wird das *LCT* Gen welches das Lactase Enzym encodiert nur im Säuglingsalter produziert. Dieses Enzym erlaubt es dem Säugling den Milchzucker Lactose zu verdauen und sich so von der Muttermilch zu ernähren. Im Alter von 3-4 Jahren wird dieses Gen abgeschaltet und die resultierende Unfähigkeit die Lactose in der Milch zu verdauen führt zu starken Verdauungsstörungen, die mit dazu beitragen, dass sich das Kind auf andere Nahrungsquellen umstellt.

Durch Zufall entstanden zwei Punktmutationen, die dazu führten, dass dieses Gen nicht mehr abgeschaltet wird und so selbst Erwachsene Lactase produzieren und daher Milch verdauen können. In den meisten Regionen der Welt haben Träger dieser Variante eher einen kleinen Fitness Nachteil. Vor einigen Tausend Jahren in Nordeuropa erlaubten diese Variationen aber ihren Trägern eine auf Milchproduktion basierende Landwirtschaftsform zu entwickeln, mit der sie diese Region mit einer viel höheren Bevölkerungsdichte besiedeln konnten, als dies mit anderen landwirtschaftlichen Methoden möglich war. Dieser

erhebliche Fitness Vorteil führte dazu, dass sich die entsprechende genetische Variante innerhalb von wenigen Tausend Jahren über fast die gesamte Nordeuropäische Population ausgebreitet hat.

Haplotypen sind Kombinationen von genetischen Varianten die häufig gemeinsam vererbt werden

Wie oben beschrieben, werden benachbarte SNPs häufig gemeinsam vererbt. Der zugrundeliegende Mechanismus sind sogenannten *crossing over events*, die während der Meiose stattfinden. Bei einem *crossing over event* werden einander entsprechende Abschnitte zweier einander entsprechenden Chromosomen physisch gegeneinander ausgetauscht. Das Resultat eines *crossing over events* ist, dass das von der Mutter geerbte Chromosom nun einen Abschnitt des väterlichen Chromosoms trägt und umgekehrt.

Dieser Austausch ist kein „Unfall“, sondern ein genau choreographierter Prozess, durch den immer neue Kombinationen der in einem Genpool vorhandenen Variationen generiert werden. Die so erzeugte Vielfalt erlaubt es Populationen dieser Organismen, sich schnell an neue Umweltbedingungen anzupassen.

Diese *crossing over events* führen dazu, dass benachbarte SNPs häufig gemeinsam zwischen den Chromosomen übertragen werden und so häufig in Kombination mit einander auftreten. In einer solchen Situation ist die Verknüpfung (engl. *linkage*) von SNPs also in einem (Zufalls)Ungleichgewicht (engl. *disequilibrium*). Man bezeichnet diese Tendenz von zwei oder mehreren SNPs gemeinsam aufzutreten als *linkage disequilibrium* oder kurz *LD*.

Es gibt dabei Orte im Genom an denen solche *Crossing over events* besonders häufig stattfinden. Man nennt diese Orte Rekombinations *hot spots*. Das Resultat ist, dass Gruppen von benachbarten SNPs häufig gemeinsam in klar definierten Blocks vererbt werden. Solche Blocks nennt man LD-Blocks. Die Länge solcher Blocks variiert stark. Die kleinsten LD-Blocks können einige wenige Tausend Basenpaaren kurz sein, grosse Blocks können aber eine Länge von einigen hunderttausend Basenpaaren erreichen.

Eine praktische Konsequenz dieser Struktur in der genetischen Variation ist, dass genetische Variationen als sogenannte Haplotypen (Abbildung 4) auftreten, also als bestimmte Kombinationen von Genotypen an den SNPs in diesem Block.

Um das Gros der genetischen Variationen im Genom eines Menschen zu bestimmen, reicht es daher aus, einige wenige strategisch ausgewählte SNPs innerhalb eines

