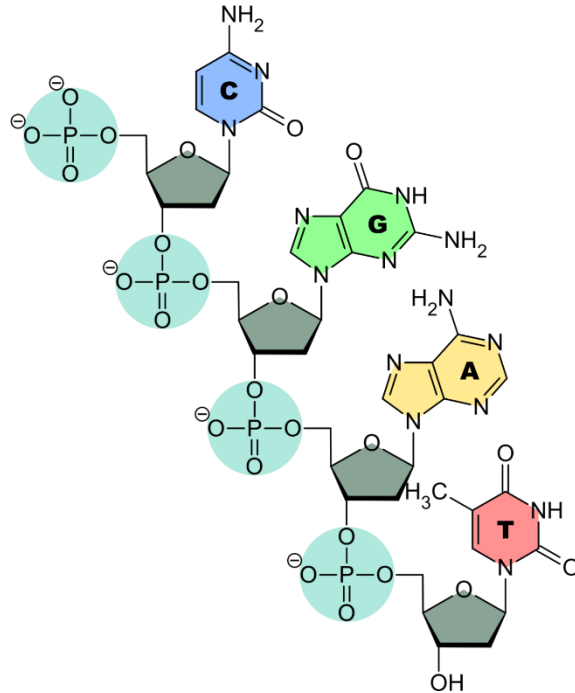


Nukleinsäuremetrologie

Eine kleine Reise durch die Zeit

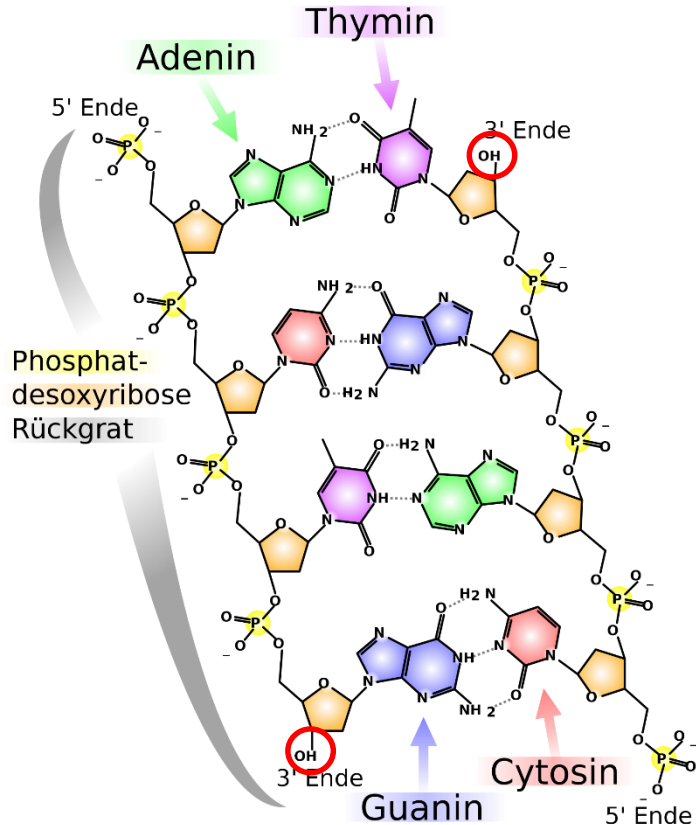
Dr. Matthias Rösslein

Nukleinsäuren



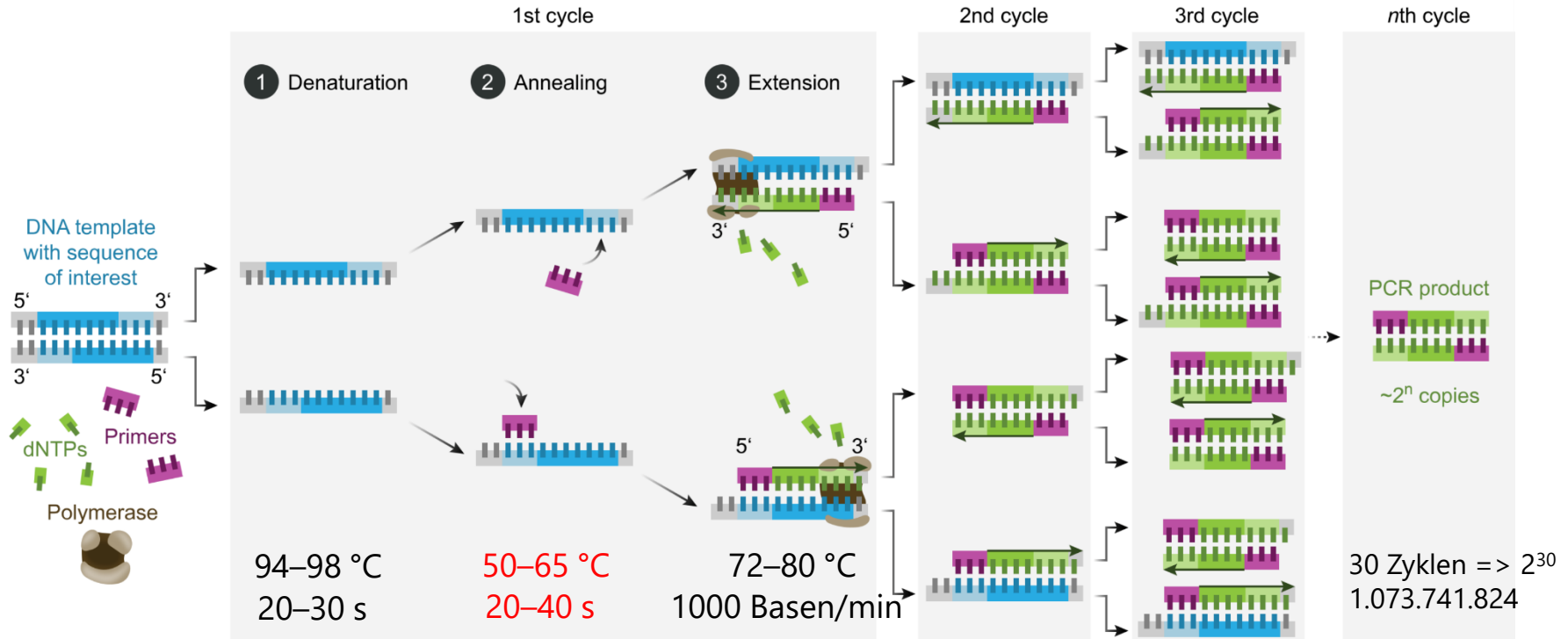
- Nukleinsäuren bestehen aus den vier Bausteinen **C**ytosin, **G**uanin, **A**denin und **T**hymin
- Diese Makromoleküle enthalten bei allen Organismen die genetische Information
- Sie sind die Informationsspeicher des Lebens (DNA) und
- dienen als Signalüberträger und katalysieren damit biochemische Reaktionen (RNA)

DNA - Informationsspeicher des Lebens

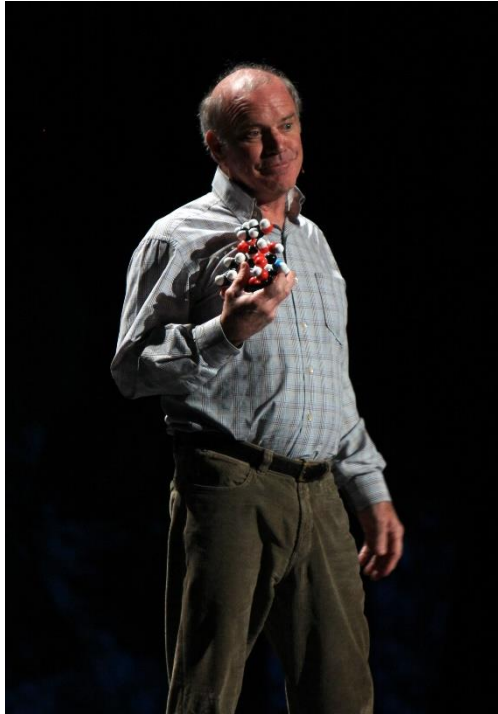


- Die DNA liegt als Doppelstrang vor, der um sich selbst gewunden eine Doppelhelix bildet
- Rechtsgängige Helix mit einer Ganghöhe von 3.54 nm und 10 Basenpaaren und einem Durchmesser von 2.37 nm
- Ein DNA-Stück aus 4 möglichen Grundbausteinen mit einer Gesamtlänge von 10 Basenpaaren
- ergibt $4^{10} = 1.048.576$ mögliche Kombinationen

DNA-Analytik - Polymerase Chain Reaction (PCR)



Kary Mullis – der Erfinder der PCR



- Kary Mullis arbeitet für Cetus Corporation in Emeryville, Kalifornien,
- wo er für die Synthese von kurzen Ketten von DNA Molekülen verantwortlich war
- Während er in einer Nacht, unter dem Einfluss von LSD stehend, auf dem Pacific Coast Highway in seinem Auto fuhr, ersann er die Idee für PCR
- Er spielte in seinen Gedanken einen neuen Weg zur Detektion von Veränderungen in der DNA durch
- Dabei wurde ihm klar, dass er stattdessen eine neue Methode zur Vervielfachung irgendwelchen Regionen auf der DNA durch wiederholte Zyklen von Verdoppelungen mit Hilfe der DNA Polymerase gefunden hatte
- 'Beginning with a single molecule of the genetic material DNA, the PCR can generate 100 billion similar molecules in an afternoon (1983)'

Der Schlüssel zur PCR



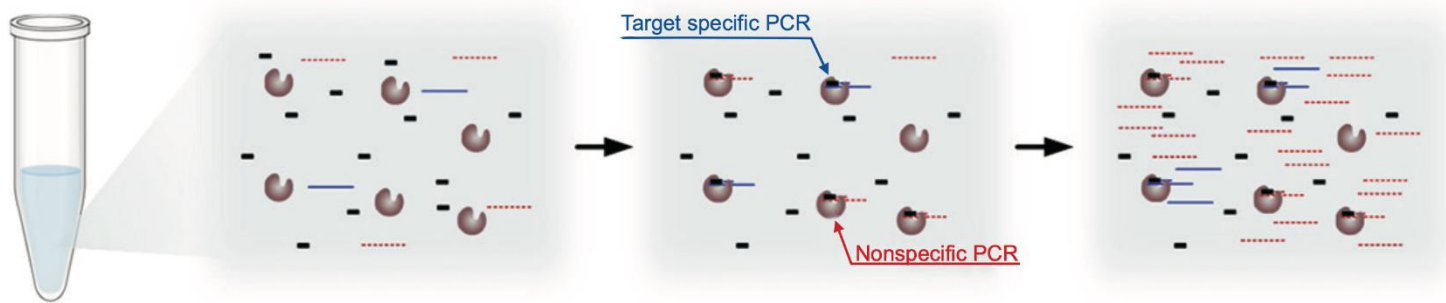
- Die Entdeckung der Taq Polymerase im Jahre 1976 hat das Verfahren der PCR dramatisch vereinfacht und verbessert
- Diese DNA Polymerase wurde aus einem thermophilen Bakterium, dem *Thermus Aquaticus*, welches in in der Natur in heißen Quellen bei 50 bis 80 °C lebt, gewonnen
- Die DNA Polymerase, die vom *T. aquaticus* isoliert wurde, ist als ein Protein sogar bei hohen Temperaturen in seiner Struktur stabil
- Somit bleibt sie auch nach dem Schritt der Denaturierung der DNA bei 98°C aktiv
- Dies ermöglicht, dass nicht mehr nach jedem Zyklus der Denaturierung neue DNA Polymerase zu gegeben werden musste
- Dies erlaubt einen automatisierten thermo-cyclischen Prozess für die Vervielfachung der DNA

Die Anfänge der PCR

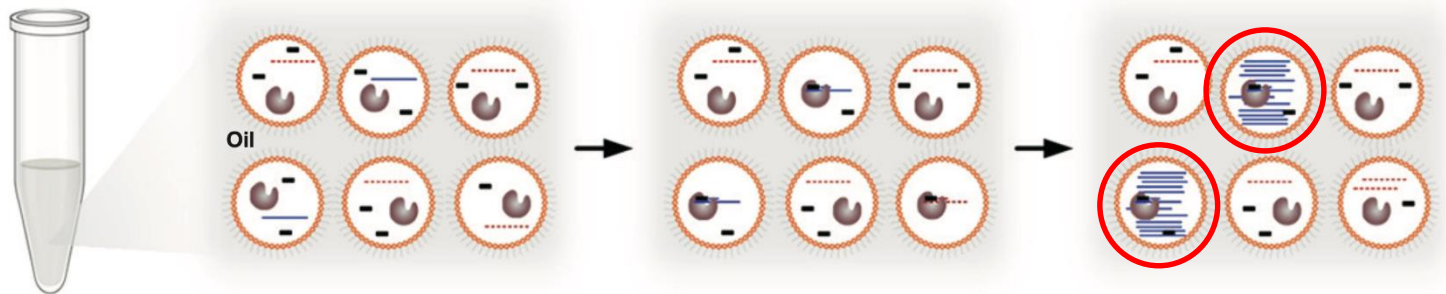


Der nächste wichtige Schritt

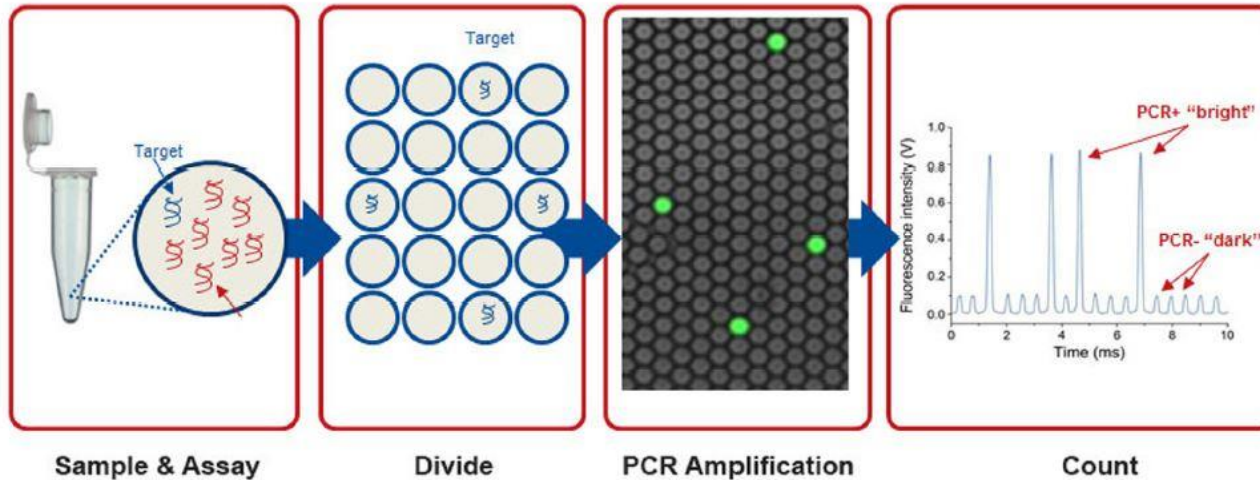
Conventional PCR



Emulsion PCR

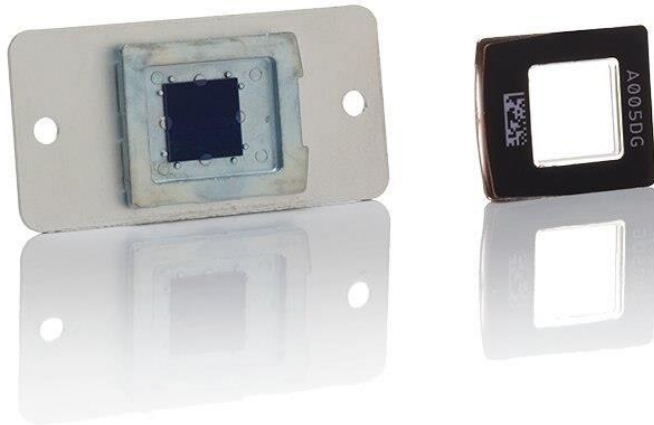


Der nächste wichtige Schritt – digital PCR



- Die digitale PCR verwendet eine Vereinzelung der DNA-Moleküle durch Grenzverdünnung in einer grossen Anzahl getrennter Reaktionsgefässe. Diese sind im Femtoliter Bereich
- Die Amplifikation mit der DNA-Polymerase erfolgt mit den vereinzelt DNA-Molekülen
- Daher kommt es in jedem Reaktionsgefäss zu einem digitalen Ergebnis (Amplifikation: ja oder nein)
- Durch die Auszählen einer grossen Anzahl von Reaktionsgefässen wird eine statistische Signifikanz erreicht

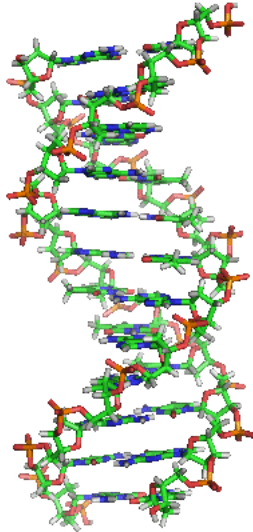
Der nächste wichtige Schritt – digital PCR



Zwei sich konkurrenzierende Prinzipien:

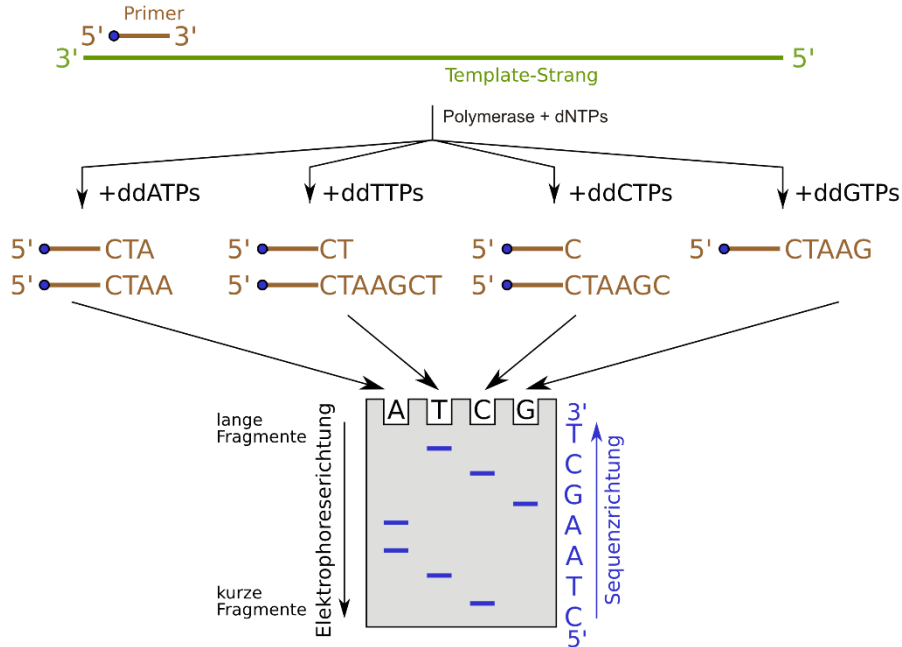
- Vereinzelung der DNA Fragmente auf dem Silikon Chip
- Vereinzelung der DNA Fragmente in einer öligen Emulsion

Der Weg zur DNA Sequenzierung



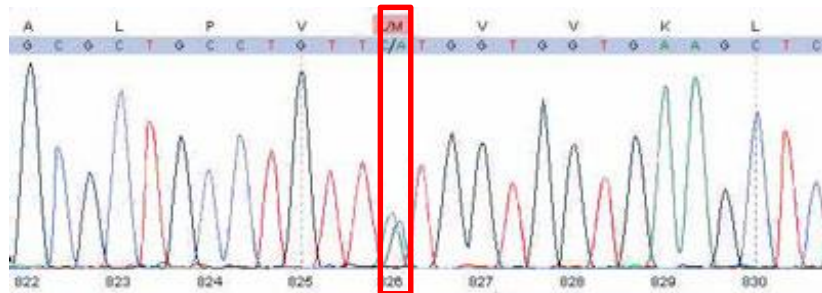
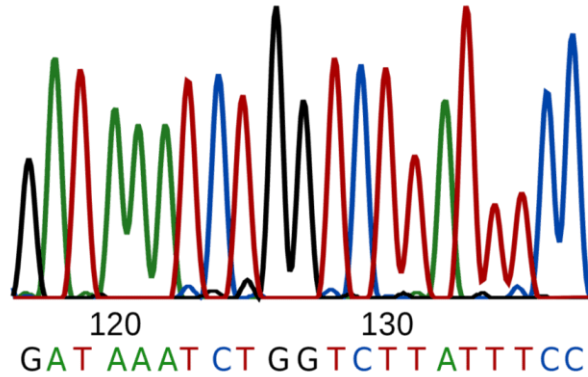
- Die DNA-Sequenzierung als Ablesen der Nukleotid-Abfolge in der DNA war über Jahrzehnte hinweg bis in die Mitte der 1970er Jahre ein ungelöstes Problem
- Die Herausforderungen einer Genomsequenzierung beschränken sich jedoch nicht nur auf das direkte Ablesen der Nukleotidsequenz
- In den meisten Verfahren werden in jeder einzelnen Sequenzierreaktion auf Grund technischer Beschränkungen nur kurze DNA-Abschnitte bis maximal 1000 Basenpaare abgelesen
- Nach Erhalt der Sequenz wird dann der nächste Primer (mit einer Sequenz aus dem Ende der vorigen Sequenzierung) hergestellt, was als **'Primer Walking'** bezeichnet wird

Der Weg zur DNA Sequenzierung – Sanger



- Die Kettenabbruch Synthese nach Sanger
- In vier sonst gleichen Ansätzen wird je eine der vier Basen zum Teil als Didesoxynukleosidtriphosphat (ddNTP) zugesetzt
- Diese Kettenabbruch-ddNTPs besitzen keine 3'-Hydroxygruppe
- zudem sind sie sind radioaktiv markiert
- Werden sie in den neusynthetisierten Strang eingebaut, ist eine Verlängerung der DNA durch die DNA-Polymerase nicht mehr möglich,
- da die OH-Gruppe am 3'-C-Atom für die Verknüpfung mit der Phosphatgruppe des nächsten Nukleotids fehlt
- In der Folge entstehen DNA-Fragmente unterschiedlicher Länge, die in jedem einzelnen Ansatz stets mit dem gleichen ddNTP enden
- Diese enden je nach Ansatz mit A, C, G oder T

Der Weg zur DNA Sequenzierung – Sanger



- Seit Anfang der 1990er Jahre werden vor allem mit Fluoreszenz-Farbstoffen markierte Didesoxynukleosid-triphosphate eingesetzt
- Jedes der vier ddNTPs wird mit einem unterschiedlichen Farbstoff gekoppelt
- Diese Modifikation erlaubt es, alle vier ddNTPs in einem Reaktionsgefäß zuzugeben
- Eine Aufspaltung in getrennte Ansätze und der Umgang mit Radioisotopen entfällt
- Die entstehenden Kettenabbruchprodukte werden mittels Kapillarelektrophorese aufgetrennt und mit Hilfe eines Lasers zur Fluoreszenz angeregt
- Die ddNTPs am Ende jedes DNA-Fragmentes zeigen dadurch Fluoreszenz unterschiedlicher Farbe und können so von einem Detektor erkannt werden
- Das Elektropherogramm (die Abfolge der Farbsignale, die am Detektor erscheinen) gibt direkt die Sequenz der Basen des sequenzierten DNA-Stranges wieder

Das Projekt zur Aufklärung des menschlichen

Gen



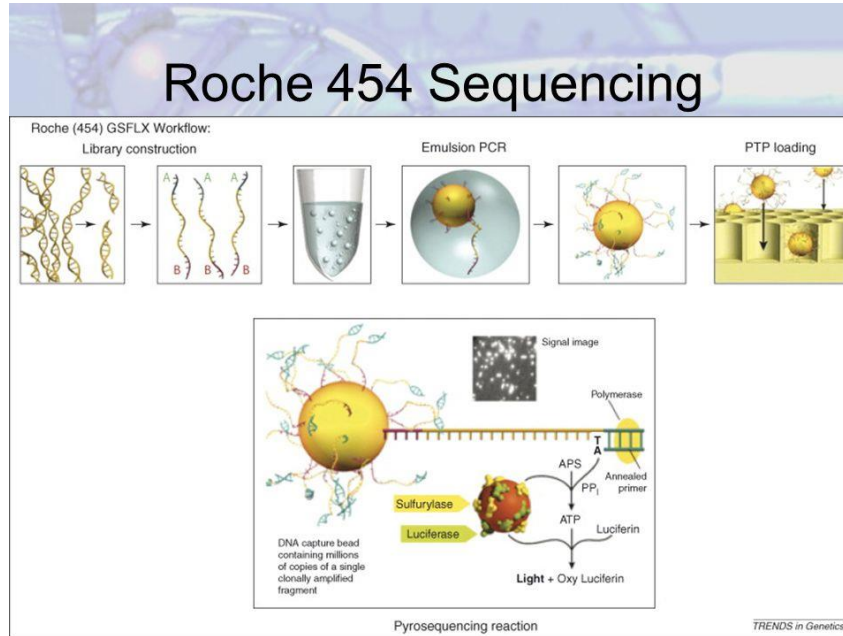
- Das Projekt zur Aufklärung des menschlichen Genoms (englisch Human Genome Project, HGP) war ein grosses internationales Forschungsprojekt
- Es wurde im Herbst 1990 mit dem Ziel gegründet, das Genom des Menschen vollständig zu entschlüsseln
- Das heisst die Abfolge der Basenpaare der menschlichen DNA auf ihren einzelnen Chromosomen durch Sequenzieren zu identifizieren
- Die vollständige Sequenzierung des Genoms dient der Erforschung vieler biologischer Prozesse
- Angestrebt wurde ein besseres Verständnis von Erbkrankheiten und der molekularen Mechanismen der Krebsentstehung
- Durch Vergleich des menschlichen Erbguts mit dem anderer Lebewesen erhoffen sich Wissenschaftler zudem Erkenntnisse zur Evolution

Das Humangenomprojekt (HPG)



- Das Projekt wurde in den USA im Oktober 1990 im Rahmen eines öffentlich finanzierten internationalen Forschungsverbunds gegründet
- Am Projekt nahmen zu Beginn über 1.000 Wissenschaftler in 40 Ländern teil
- Ziel war die Sequenzierung des menschlichen Genoms bis 2005
- Die Human Genome Organisation bekam 1998 durch die neu gegründete US-Firma Celera private Konkurrenz
- Das HGP fand in der Öffentlichkeit einen vorläufigen Höhepunkt, als im Februar 2001 beide Forschungsunternehmen die Sequenzierung des menschlichen Genoms verkündeten
- Erfasst wurden bis dato nur 83 % des Genoms
- 2003 wurde die Fertigstellung im Rahmen der angelegten Maßstäbe verkündet
- Erst seit Mai 2021 gilt das menschliche Genom als vollständig entschlüsselt
- Es umfasst 19.969 Gene
- Zu Beginn wurden mindestens 100.000 Gene erwartet, um alle Merkmale des menschlichen Körpers kodieren zu können

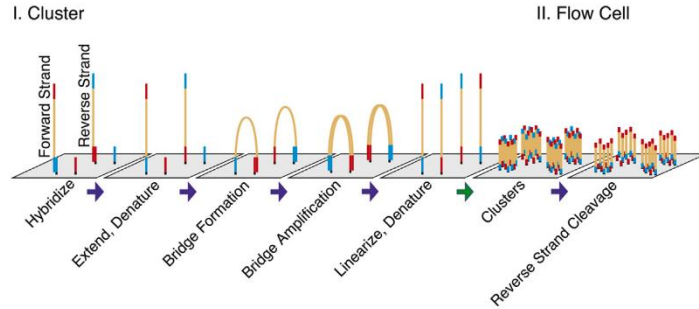
Next Generation Sequencing - 454



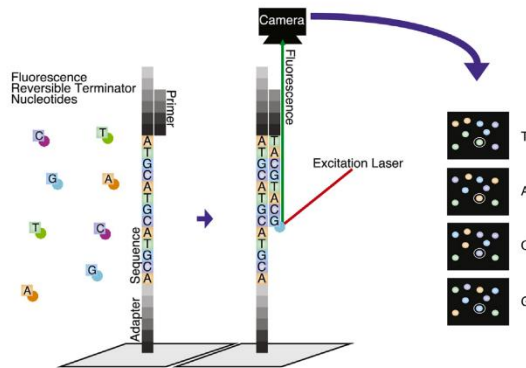
- 454 veröffentlicht im Jahre 2005 die GS20 Sequenziermaschine
- Es ist der erste Next Generation DNA Sequenzer (NGS) auf dem Markt
- Im 2008 ergibt sich mit der Einführung, des Genome Sequenzer FLX die Möglichkeit 400-600 Millionen Basenpaare pro Durchlauf mit einer Länge 400-500 Basen zu sequenzieren
- Einzelne kurze einzelsträngige Basenpaare (300-600) werden an einer Emulsion verankert
- Diese Emulsionstropfen werden einzeln in jeweils ein Locher einer Picotiterplatte gespült
- Die vier DNA Nukleotide werden sequenziell in einer fixierten Abfolge jeweils über die Picotiterplatte bei einer Sequenzierung gespült
- Wenn eines der vier Nukleotide komplementär zum Einzelstrang ist, wird es zu diesem Strang dazu addiert
- Die Addition von einem oder mehreren Nukleotide erzeugt ein Lichtsignal, dass mit der CCD Kamera aufgezeichnet wird
- Diese Technik basiert auf der Sequenzierung durch Synthese
- Im Oktober 2013 verkündet Roche, dass sie die Plattform 454 auslaufen lässt und den Support Mitte 2016 beendet

Next Generation Sequencing - Illumina

A. Clustering



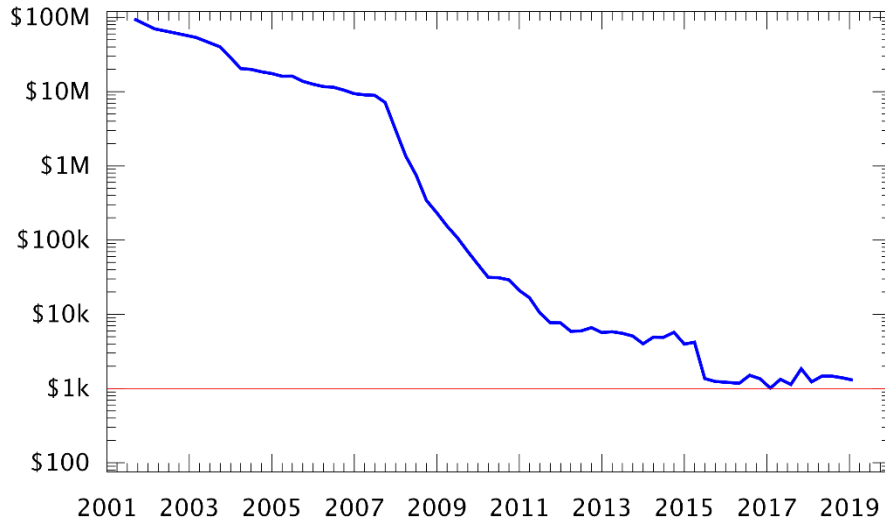
B. High-throughput sequencing



- Illumina Sequenzierung beginnt mit der Fragmentierung der DNA
- An diese DNA Fragmente werden mit einem Adapter versehen, der als Referenzpunkte bei der Sequenzierung dienen
- Die modifizierten DNA Fragmente werden anschliessend in einer Flusszelle an ein Oligonukleotid, das sich in einem Nanometer grossen Vertiefung verankert ist, gebunden
- Jetzt werden aus jedem Fragment ein Cluster gebildet, das aus rund tausend Kopien des Fragments besteht
- Diese Clusterbildung wird mit Hilfe der Brückenverstärkung dank der PCR durchgeführt
- Als nächstes werden die Primer und modifizierten Nucleinsäuren dazu gefügt
- Diese Nucleinsäuren haben einen reversible gekoppelten und fluoreszierenden Blocker, der dazu führt, dass nur jeweils eine Nucleinsäure zum DNA Fragment synthetisiert wird
- Nach jeder Syntheserunde nimmt eine Kamera ein Bild der Flusszelle auf
- Ein Computer ermittelt mittels der Wellenlänge des fluoreszierenden Lichtes welche Nucleinsäure an jedem Punkt auf dem Chip dazugekommen ist
- Anschliessend werden in einem chemischen Prozess die fluoreszierenden Blocker entfernt
- Dieser Prozess wird fortgesetzt, bis das ganze DNA Fragment sequenziert wurde
- Mittels dieser Technologie können hunderttausende Teile eines Genoms parallel sequenziert werden

Die mooresches Gesetzmässigkeit für die Sequenzierung

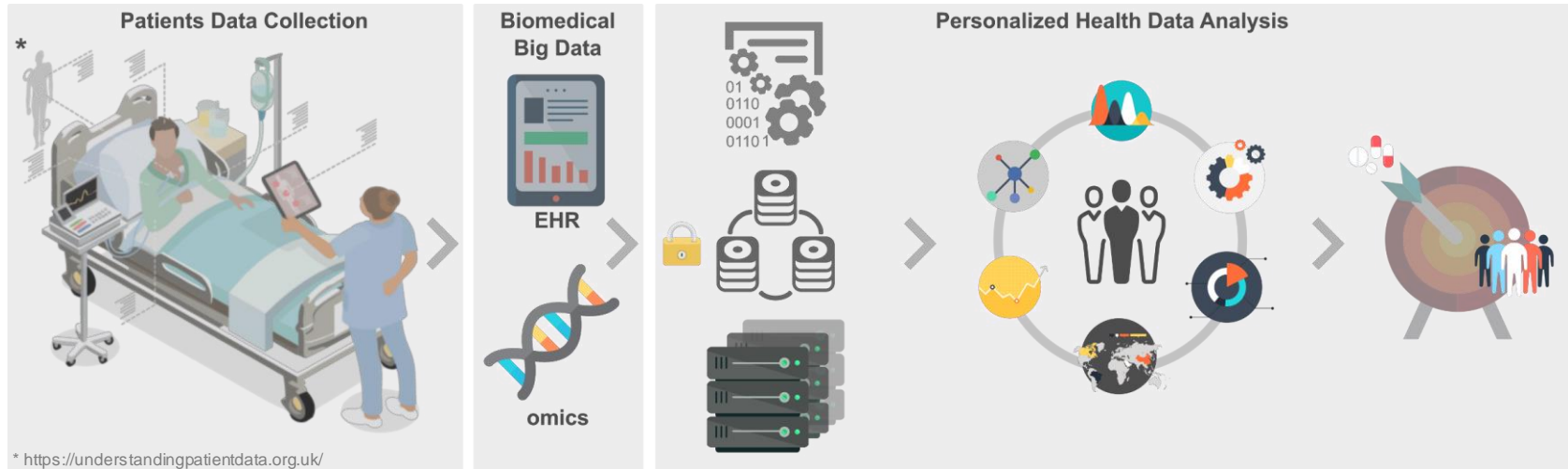
Cost to sequence human genome in USD



- Die mooresche Gesetzmässigkeit besagt, dass sich die Komplexität integrierter Schaltkreise mit minimalen Komponentenkosten regelmäßig alle 12 bis 18 Monate verdoppelt
- Es gab auch eine mooresche Gesetzmässigkeit bei der Sequenzierung
- Hier haben sich die Kosten pro menschliches Genome um einen Faktor 10 und nicht einen Faktor 2 pro Jahr für die gleiche Leistung gesenkt
- Dies hiess auch, dass jedes Jahr die Sequenzer 10 Mal mehr leisteten
- Seit ca. 2016 wurde ein gewisses Plateau vom Preis her erreicht (1000 US\$ für ein menschliches Genome)
- In der Zwischenzeit fallen die Kosten bei der Analyse des menschlichen Genome an, und nicht mehr bei der Sequenzierung. Das heisst bei der Bioinformatik
- Die Sequenzierung führt zu einer riesigen Datenflut an Informationen, die in den kommenden Jahre noch massiv zunehmen wird

Use case: Personalized Health Research

Provide the **right treatment**, at the **right moment** to the **right patients** (precision medicine) and ensure that as many people as possible **stay healthy** (prevention; personalized health).

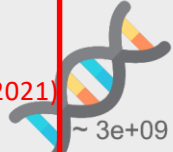



Research on human data: **sensitive personal data**
Hereafter, per convention: **confidential research data**

5 Vs: **Volume** **Velocity** **Variety** **Veracity** **Value**

“We are drowning in information but starved for knowledge” John Naisbitt (1982)

- omics (gen-, epigen-, transcript-, etc.) >> **Bioinformatics**: domain specific and/or integrative analyses
- eHealth (electronic health records, histopathological and radiological images, etc.) >> **AI**: machine learning
- citizen data (digital footprints, mobile apps) >> **AI**: machine learning

Knowledge	Data	1x Human Genome Sequencing
<p>> 6,000 Mendelian disorders studied at genetic level</p> <p>Most with unknown role in health and disease</p>	<p>> 9,797,389,622,108,464 total bases (SRA@NIH, 9e+15 in 2017)</p> <p>> 57,328,564,591,416,100 (5.7e+16 in 2021)</p> <p>2025 prediction: 1e+21</p>	 <p>~ 3e+09 bases</p>  <p>100 gigabytes</p>

High Performance Computing (HPC)



80 % Unstructured Data
20 % Structured Data