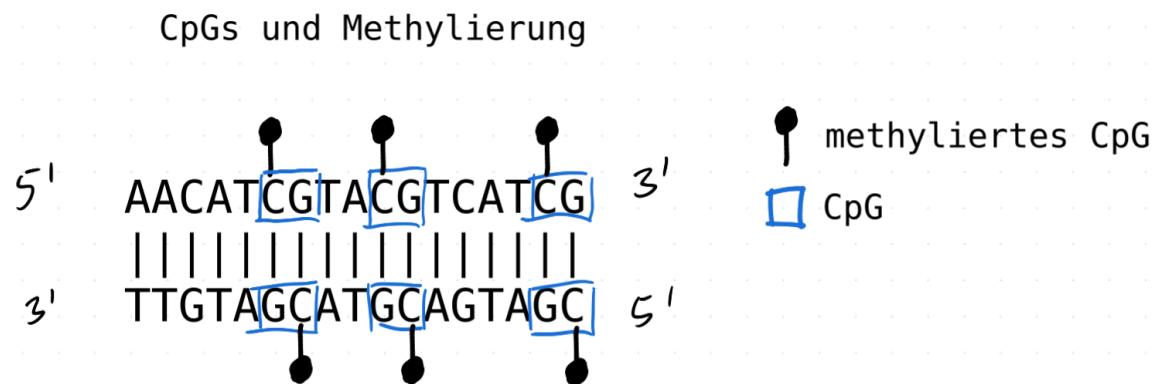


Methylome Masterclass

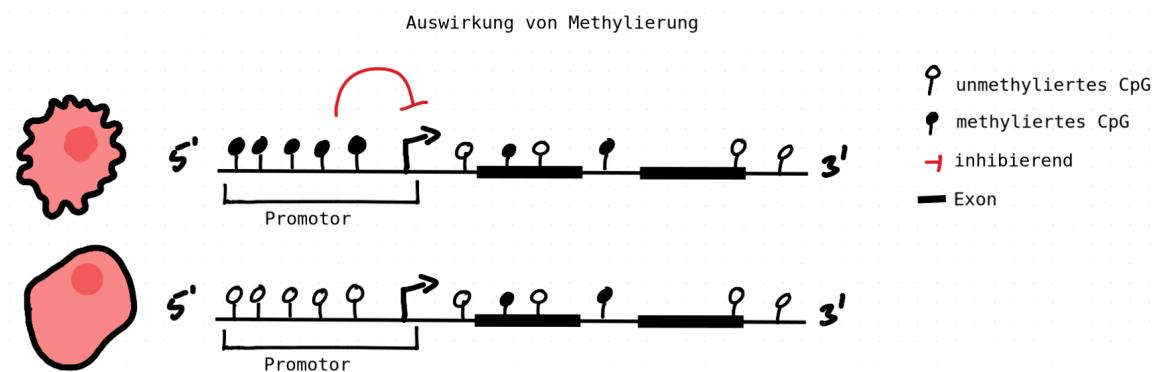
Methylierungsanalyse

Methylierung der DNA

In der DNA werden Modifikationen, die nicht die Sequenz der DNA verändern, genutzt, um die Expression von Genen zu regulieren. Diese Modifikationen fasst man unter dem Begriff Epigenetik zusammen. Eine dieser Modifikationen ist die Methylierung von Cytosin. Dabei wird enzymatisch eine Methylgruppe an das Cytosin eines 5'-CpG-3' Dinukleotids angehängt.



Methylierung ist meistens mit einer Herunterregulierung der Expression benachbarter Gene assoziiert. Dies ist im Allgemeinen darauf zurückzuführen, dass CpGs in Promotorregionen methyliert werden und die Methylgruppen das Binden von Transkriptionsfaktoren verhindern.



Das "p" in CpG steht für die Phosphatgruppe, die die beiden Basen in der DNA verbindet. DNA Regionen, die eine besonders hohe Dichte an CpGs haben (egal ob methyliert oder nicht) bezeichnet man als CpG Inseln. An jede CpG Insel schließen sich beidseitig eine Shore-Region gefolgt von einer Shelf-Region an.

Methylierungsanalysen haben typischerweise das Ziel, Methylierungslevel aller, einiger oder einzelner CpGs oder CpG Inseln in der DNA zwischen zwei oder mehreren Zuständen zu vergleichen. Beispielsweise kann von Interesse sein, welche CpGs in Promotorregionen von Genen sich zwischen gesundem Gewebe und Tumorgewebe unterscheiden. Hier könnte die vorhergehende Vermutung beispielsweise sein, dass im Tumor die Promotorregion eines Tumorsuppressorgens methyliert ist, wodurch die Expression des Gens reduziert und das regulierte Zellwachstum gestört wird.

Um diese Analyse durchführen zu können, sind mehrere Schritte nötig:

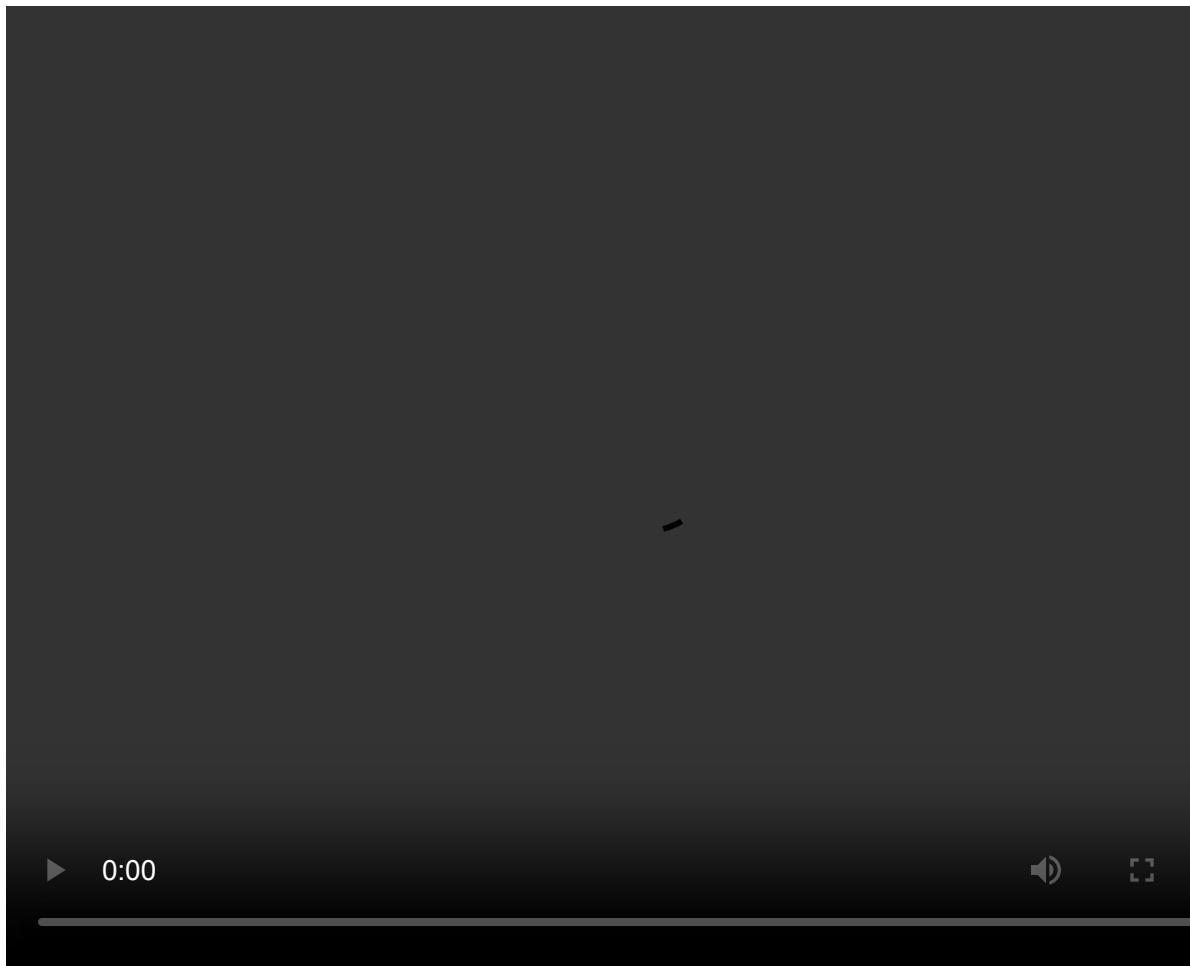
1. Vorbereitung der Zellen, deren Methylierung bestimmt und verglichen werden soll
2. Extraktion der zellulären DNA
3. Behandlung der DNA, um methylierte von unmethylierten Cytosinen unterscheiden zu können (Bisulfit-Conversion)
4. Amplifikation und Fragmentierung der DNA
5. Hybridisierung der inzwischen einzelsträngigen DNA mit einzelsträngiger DNA auf Microarrays
6. Auslesen von Fluoreszenzsignalen auf den Microarrays
7. Prozessierung der Fluoreszenzsignale mit bioinformatischen Tools in Methylierungslevel pro CpG pro Zellprobe
8. Normalisierung der Methylierungslevel
9. Überprüfung auf und Entfernen von Batch-Effekten
10. Vergleich der Methylierung zwischen Zellproben

Diese Schritte werden innerhalb dieses Moduls erklärt, veranschaulicht und teilweise selbst durchgeführt.

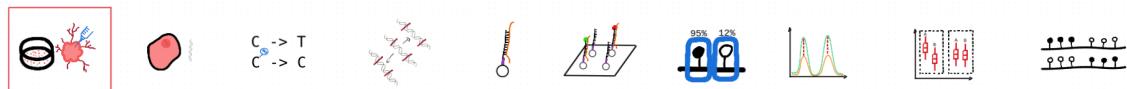
Von Zellen zur DNA

Die Gewinnung von DNA aus Proben findet experimentell statt. Um den Prozess von Anfang bis Ende abzubilden sind im Folgenden in drei Videos alle Schritte gezeigt, die im Labor ablaufen, um die DNA zu extrahieren.

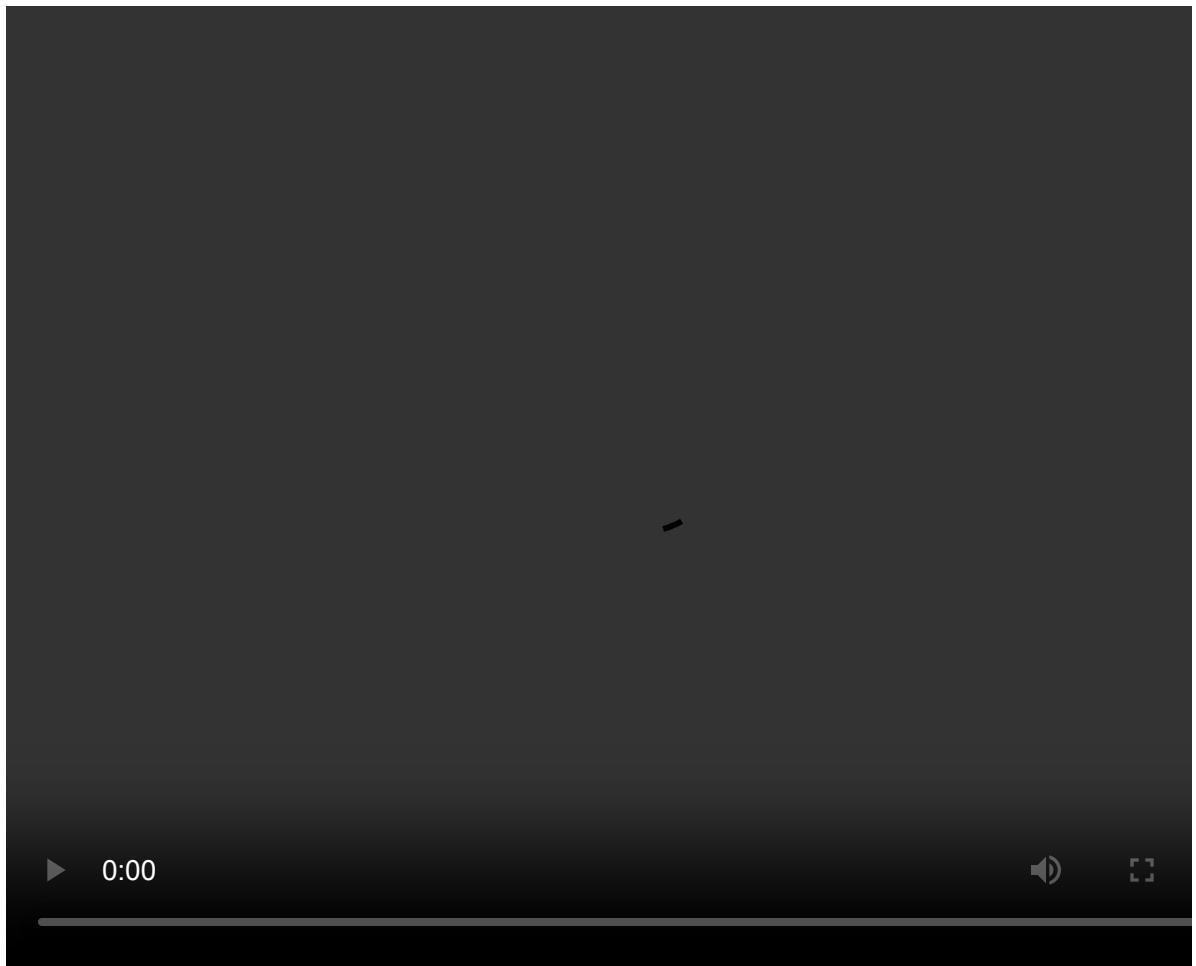
1 Intro



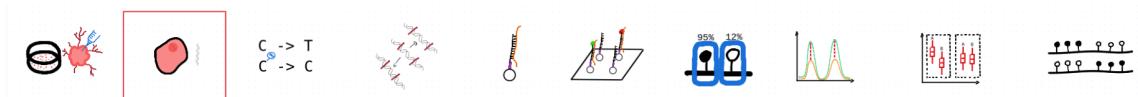
2 Zellkultur



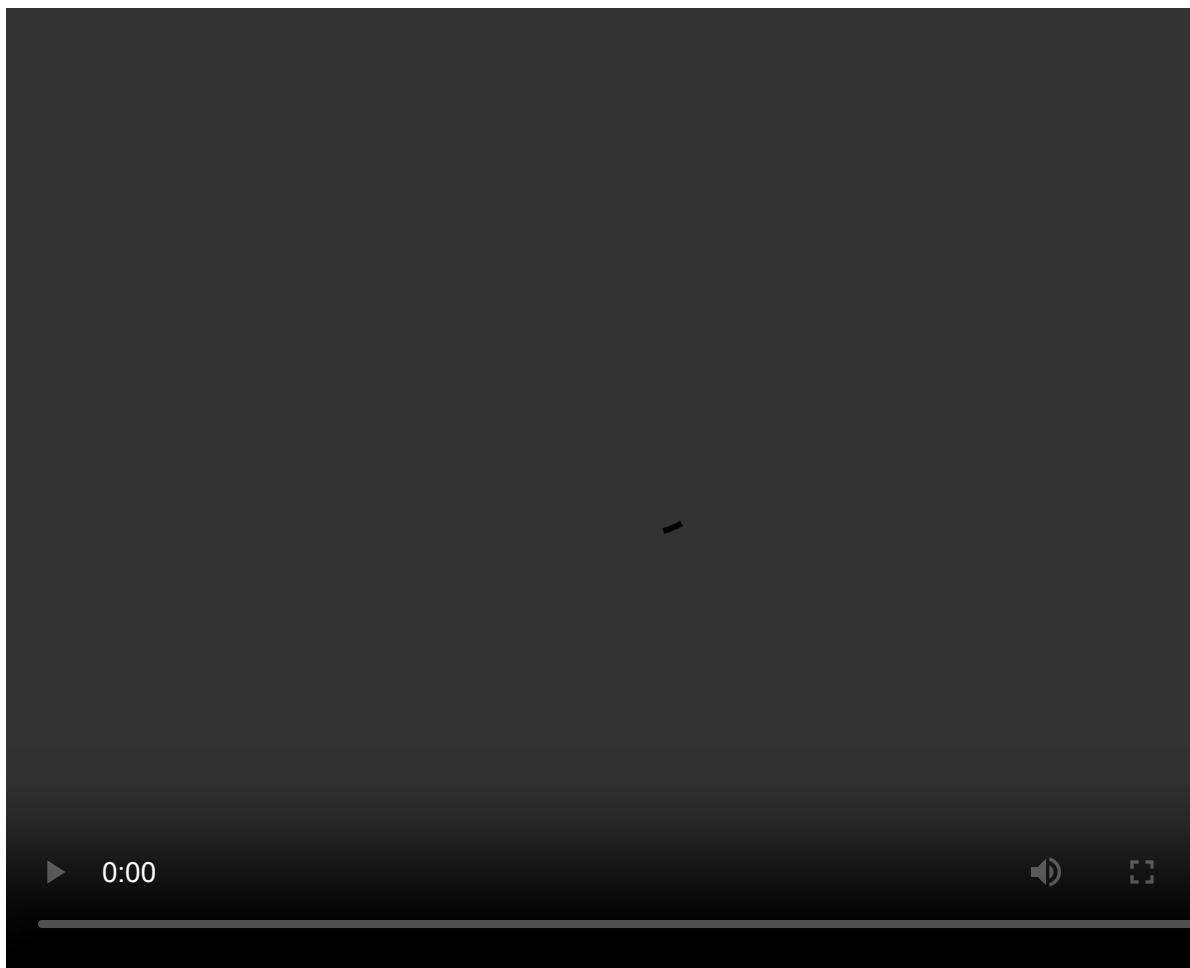
Die Zellen, aus denen die DNA extrahiert wird, werden im Labor bei optimalen Bedingungen aufbewahrt. Dadurch wird sichergestellt, dass vor der Extraktion der DNA ausreichend Zellen vorhanden sind und die Zellen möglichst unbeeinflusst von Stressoren sind, die Einfluss auf die Methylierung haben könnten.



3 DNA Extraktion



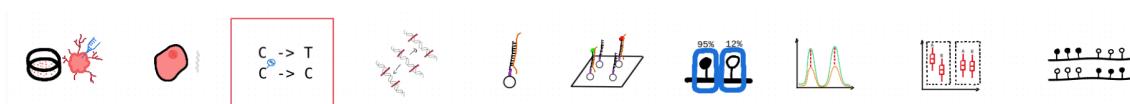
Nachdem die Zellen von der Schale abgelöst und zu einem Pellet abzentrifugiert wurden startet die Extraktion der DNA. Dabei ist wichtig, dass die DNA möglichst gut von Proteinen und RNAs getrennt wird ohne zu viel DNA zu verlieren.



Nachdem die DNA aus den Zellen extrahiert wurde, folgen weitere experimentelle Schritte, die dazu dienen, die Methylierung der DNA innerhalb einer Probe festzustellen. Da die Methylgruppen der methylierten Cytosine nicht ohne weiteres detektiert werden können, nutzen wir einen experimentellen Ansatz, der die Methylierung als epigenetische Modifikation der DNA in der Sequenz der DNA sichtbar macht.

Von der DNA zur Methylierung

4 Bisulfit-Konversion

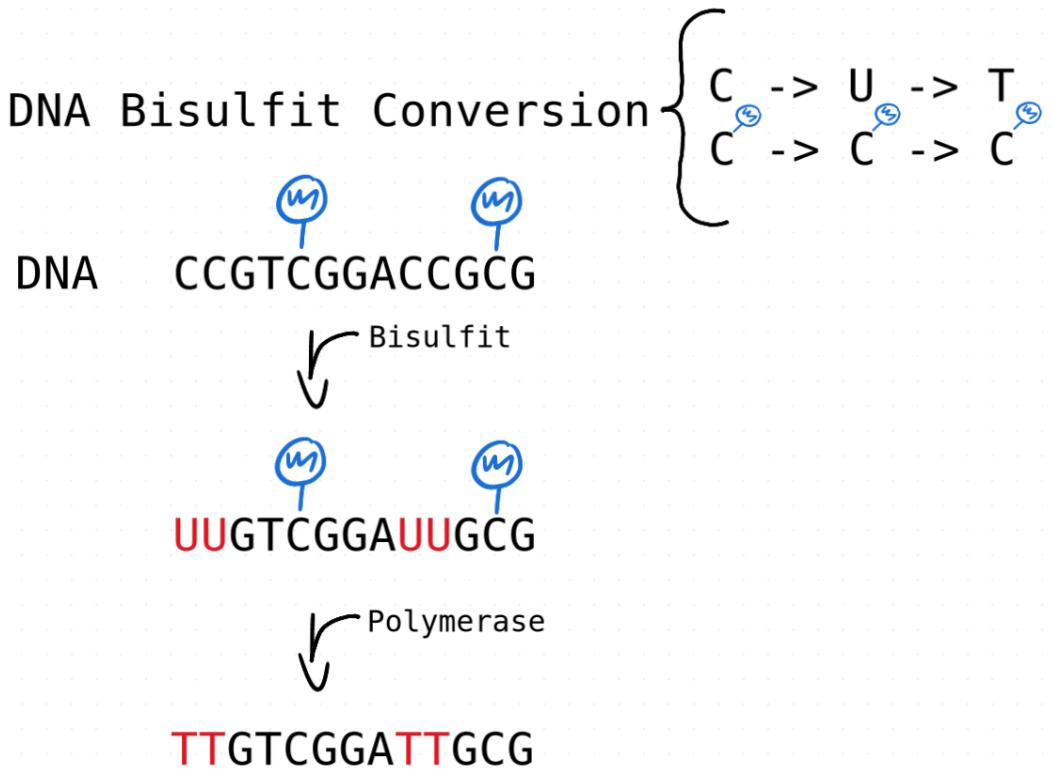


Um die Methylierung aller CpGs überprüfen zu können, nutzen wir die sogenannte Bisulfit-Konversion. Die DNA wird dabei mit Bisulfit behandelt, wodurch unmethylierte Cytosine chemisch zu Uracil verändert und anschließend durch Replikation mit Thymin ersetzt werden.

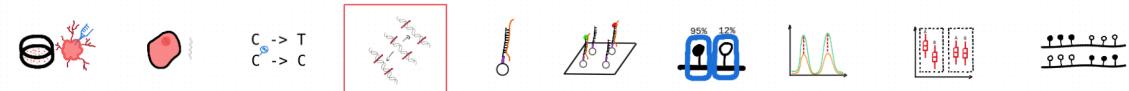
Dadurch können für jedes potenziell methylierte Cytosin zwei Sequenztypen entstehen:

1. das Cytosin ist methyliert und bleibt dadurch ein Cytosin
2. das Cytosin ist unmethyliert und wird durch ein Thymin ersetzt

Wie dieser Prozess genau funktioniert, muss uns an dieser Stelle nicht interessieren. Wir legen den Fokus darauf, zu verstehen, wozu die Bisulfit Conversion führt



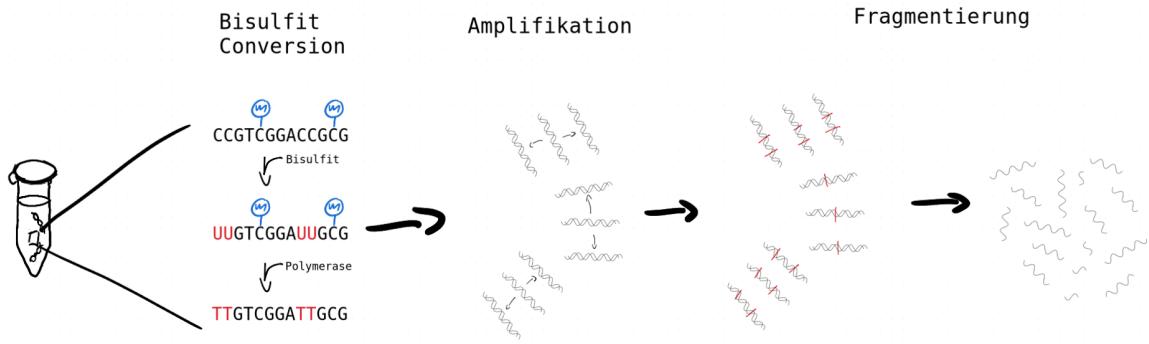
5 Library Vorbereitung



Nach der Bisulfit Conversion muss die DNA weiter vorbereitet werden, um mit dem Illumina Methylation BeadChip (eigentlich ein Microarray) kompatibel zu sein. Das Ziel hierbei besteht darin:

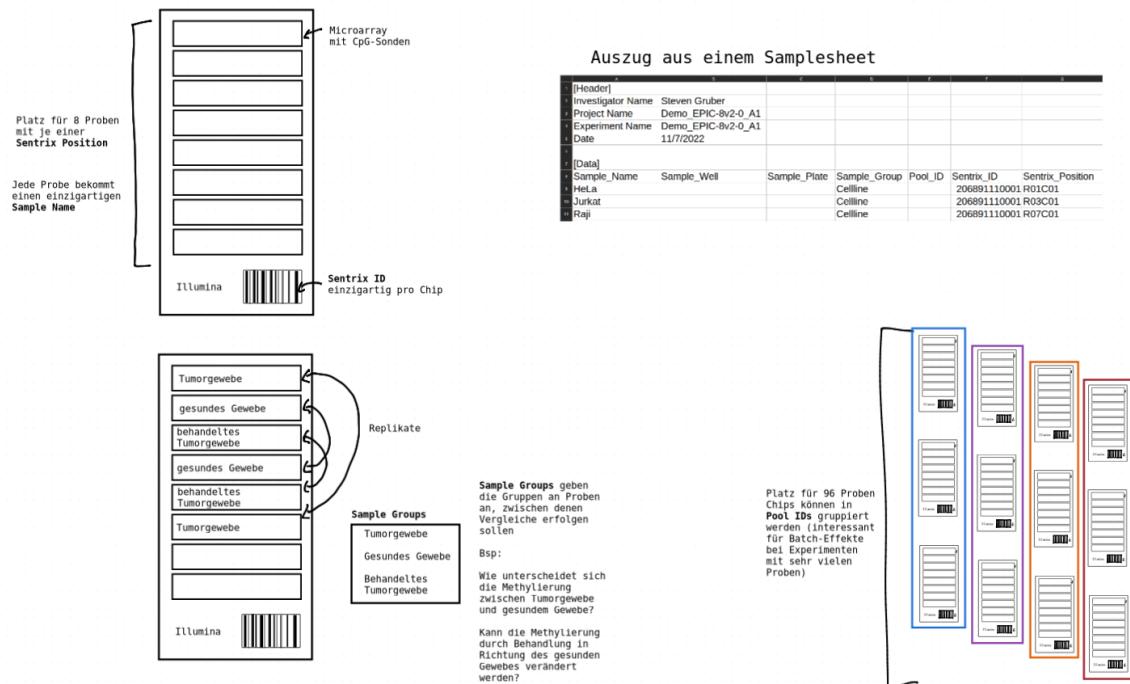
1. Die DNA zu amplifizieren, um viele Kopien jeder genomischen Region zu haben
2. Die DNA zu fragmentieren, um kleine DNA Stücke zu erhalten

Anschließend haben wir eine große Menge relativ kleiner einzelsträngiger DNA Fragmente mit idealerweise einzigartiger Sequenz, in denen methylierte Cytosine weiterhin Cytosin sind, während unmethylierte Cytosine durch Thymin ersetzt wurden. Die Gesamtheit der gewonnenen DNA Fragmente bezeichnet man als Library.

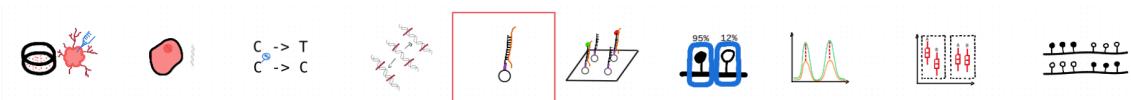


6 Experimentelles Design und Samplesheets

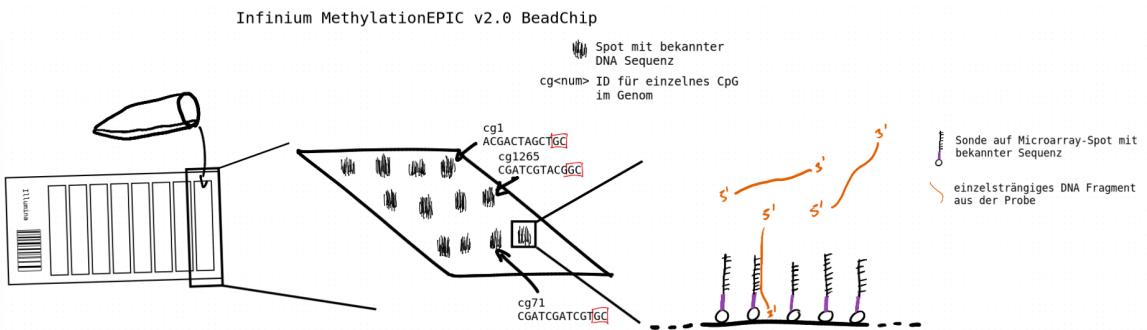
Jedem Experiment liegt eine konkrete Fragestellung zugrunde. Nach der Fragestellung richtet sich, wie viele und welche Zustände miteinander verglichen werden und wie viele Proben benötigt werden. Das Experimentelle Design ist in einem Samplesheet festgehalten. Es beginnt mit einem Header, der Metadaten enthält, gefolgt von einer Auflistung aller enthaltenen Proben mit einzigartigen **Sample_Names**. Jeder Probe wird über die **Sentrix_ID** ein Beadchip und über die **Sentrix_Position** eine Position auf dem Beadchip zugeordnet. Die Unterteilung in zu vergleichende Zustände wird als **Sample_Group** angegeben. Um repräsentative Zustände zu gewährleisten sollte es zu jeder Sample_Group mehrere Proben (Replikate) geben. Bei Experimenten mit sehr vielen Proben kann es sinnvoll sein, eine **Pool_ID** zuzuordnen, um festzuhalten, welche Proben zusammen prozessiert wurden (beispielsweise am gleichen Tag, von der gleichen Person). Dadurch können unerwartete Einflüsse auf die Daten gefunden werden.



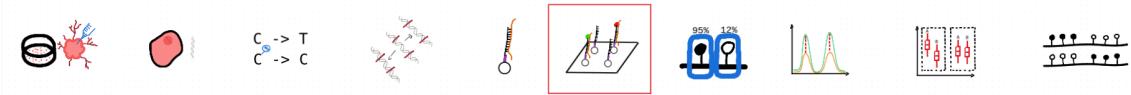
7 Infinium MethylationEPIC v2.0 BeadChip



Der Infinium MethylationEPIC v2.0 BeadChip ist ein Microarray. Das bedeutet, auf dem BeadChip befinden sich Sonden bzw. Probes für über 935 000 Positionen im menschlichen Genom. Als Sonde bezeichnet man ein Stück einzelsträngige DNA, das auf dem BeadChip befestigt ist und eine einzigartige Sequenz im menschlichen Genom abdeckt. Für jede der 935 000 Positionen gibt es eine Vielzahl identischer Sonden, die sich auf einem Spot befinden. Die Aufgabe der Sonden ist es, mit Fragmenten aus der Library zu hybridisieren und daraufhin ein erfassbares Signal zu erzeugen, über das die Menge hybridisierter DNA eines Spots mit anderen Spots verglichen werden kann.

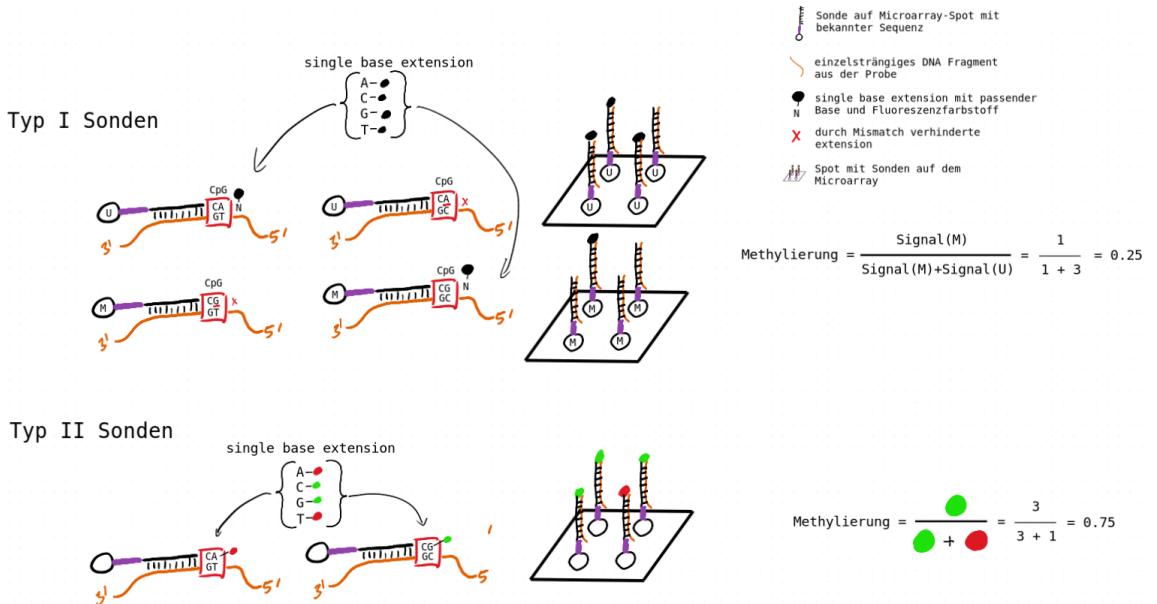


8 Sonden (Probes) und Sondentypen



Nicht alle Sonden auf dem BeadChip sind gleich. Die grundlegende Funktionsweise beruht bei allen Sonden darauf, spezifisch die methylierten bzw. unmethylierten DNA Fragmente einer genomischen Region zu binden. Der Unterschied liegt darin, wie aus der gebundenen DNA das Methylierungslevel der spezifischen Region bestimmt wird. In allen Fällen wird ein Fluoreszenzsignal der methylierten Sonden mit dem Fluoreszenzsignal der unmethylierten Sonden verglichen.

Es gibt Typ I und II Sonden.



Von allen Typ I Sonden gibt es zwei Versionen: Eine Version hybridisiert mit der unmethylierten, die andere mit der methylierten DNA nach Bisulfit-Konversion. Nach der Bindung wird die Sonden-DNA um eine einzelne Base erweitert, die mit einem Fluoreszenzfarbstoff markiert ist. Da die Positionen beider Sondenversionen auf dem Chip bekannt sind, kann nach Hybridisierung und Verlängerung das Fluoreszenzsignal der beiden Sondenversionen verglichen und daraus das Methylierungslevel bestimmt werden.

Von Typ II Sonden gibt es lediglich eine Version. Die Sonde enthält die einzelsträngige DNA der genomischen Region und schließt mit der Base vor dem Cytosin von Interesse ab. Hybridisiert die Sonde, wird die Sequenz ebenfalls um eine Base verlängert. Dabei können zwei Fälle eintreten:

1. Die hybridisierte Sequenz aus der Probe hat an der entsprechenden Position ein C oder G -> das Cytosin von Interesse ist methyliert
2. Die hybridisierte Sequenz aus der Probe hat an der entsprechenden Position ein A oder T -> das Cytosin von Interesse ist unmethyliert

Je nach Methylierungszustand des Cytosin in der Zelle wird mit A/T bzw. C/G verlängert. A/T und C/G haben dabei unterschiedliche Fluoreszenzfarbstoffe und das Methylierungslevel berechnet sich aus dem Vergleich der Intensitäten beider Fluoreszenzfarbstoffe an der gleichen Position auf dem Chip.

Outputs

Wenn das Hybridisieren der Library DNA mit den Sonden der Microarrays gefolgt von der Fluoreszenzmarkierung abgeschlossen ist, werden die Fluoreszenzsignale ausgelesen.

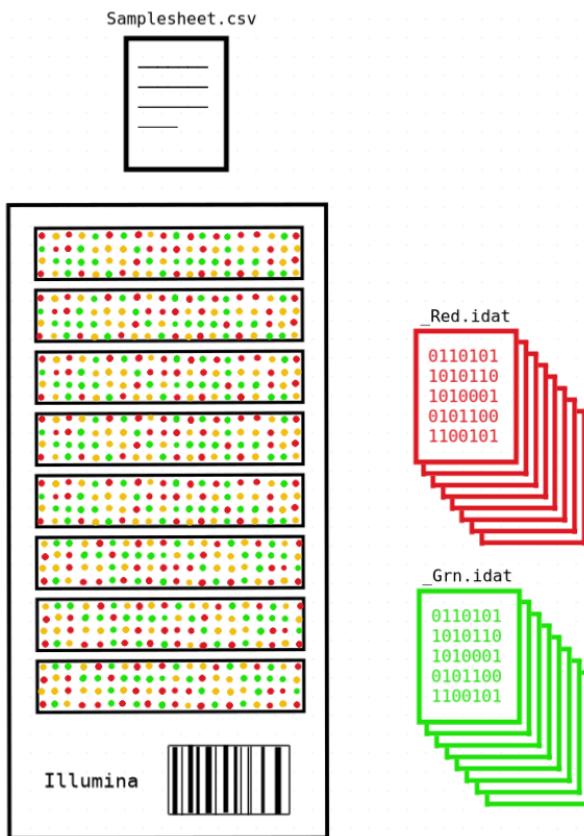
Dafür wird ein spezieller Scanner genutzt, der den Beadchip erst mit Lichtsignalen anregt, was dazu führt, dass die Fluoreszenzfarbstoffe der Sonden ein Lichtsignal abgeben. So

kann von jedem Spot auf jedem Microarray die Fluoreszenzintensität ausgelesen werden. Zusätzlich gibt es auf jedem Beadchip Kontrollsonden, die beispielsweise das Hintergrundfluoreszenzsignal auslesbar machen, was bei der späteren Prozessierung der Fluoreszenzsignale berücksichtigt wird.

Nach erfolgreicher Durchführung des Experiments erhalten Bioinformatiker zwei Arten von Daten:

1. Das Samplesheet, das Informationen über das durchgeführte Experiment gibt
2. Eine Menge von .idat Dateien

Die .idat Dateien enthalten jeweils die ausgelesenen Fluoreszenzsignale einer bestimmten Wellenlänge (rot und grün) einer Probe. Es gibt also insgesamt .idat Dateien in Höhe der doppelten Anzahl von Proben. Die Benennung der Dateien folgt dem Muster
<Sentrix_ID>_<Sentrix_Position>_Grn.idat bzw. *<Sentrix_ID>_<Sentrix_Position>_Red.idat*



Glossar

- Amplifikation - Vermehrung von DNA, häufig durch PCR
- BeadChip - Chip für Methylierungsanalysen, der mehrere Microarrays enthält
- Beta-Wert - Wert, der die relative Methylierung eines CpGs angibt
- CpG - CG Dinukleotid, dessen Methylierung von Interesse sein kann
- Differentially methylated probe (DMP) - Sonde, deren Methylierung sich zwischen zwei Zuständen signifikant unterscheidet

- DMR - Menge nahe zusammenliegender Sonden, deren Methylierung sich zwischen zwei Zuständen signifikant unterscheidet
- Enzym - Protein oder RNA mit katalytischer Funktion zur Umsetzung eines Substrats (bsp. Restriktionsenzym -> Zerschneiden von DNA)
- Eppendorf-Gefäß (Eppi) - Reaktionsgefäß für Versuche im Labor
- Fluoreszenzsignal - Lichtsignal, das von Fluoreszenzfarbstoffen abgegeben wird, wenn sie angeregt werden
- Fragmentierung - Zerschneiden/Zerbrechen eines DNA Moleküls in mehrere kleine Fragmente
- Genomische Region - Spezifische Region der DNA im Genom (der Gesamtheit der DNA im Zellkern einer Zelle)
- Hybridisieren - Verfahren, bei dem zwei komplementäre einzelsträngige DNA oder RNA Moleküle sich zu einem Doppelstrang verbinden
- Isotonisch
- Komplementär - In Bezug auf DNA eine einzelsträngige DNA, die sich aufgrund ihrer Basenfolge mit einer weiteren einzelsträngigen DNA zu einem Doppelstrang verbinden kann
- M-Wert - Logarithmus des Verhältnisses von methyliertem Signal zu unmethyliertem Signal (alternativ zum Beta-Wert)
- Multidimensional Scaling (MDS) - Methode zur Dimensionsreduktion, bei der die paarweisen Distanzen zwischen Punkten erhalten bleibt
- Methylierung - Epigenetische Modifikation, bei der eine Methylgruppe an einem Cytosin angebracht wird
- Methylierungslevel - Relative Häufigkeit der Methylierung eines bestimmten CpGs in einer Probe (Menge von Zellen)
- Microarray - Chip mit Spots für unterschiedliche DNA Sequenzen, mit dem das Vorkommen der Sequenzen in einer Probe quantitativ bestimmt werden kann
- Polymerase-Kettenreaktion (PCR) - Experimentelles Verfahren zur Vermehrung zuvor selektierter DNA Sequenzen
- Pipeline - Abfolge von Bearbeitungsschritten von Daten mit verschiedenen Tools, um einen bestimmten Output aus den Daten zu erhalten
- Polymerase - Enzym, dass die Aneinanderkettung von Nukleinsäuren anhand einer Vorlage katalysiert
- Pool_ID - Zuordnung zu einer experimentellen Batch (z.B. Auftrennung nach Person oder Tag)
- Probe (eng. Sample) - Menge von Zellen (beispielsweise einer Zelllinie oder eines Gewebes), die zusammen experimentell verarbeitet werden
- Puffer - Stoffgemisch, dass den pH-Wert einer Lösung möglichst konstant hält
- Sample_Group - Gruppe/Zustand, der eine Probe zugeordnet wird (z.B. Tumor/Gesund)
- Sample_Name - Einzigartiger Name einer Probe (benötigt)
- Sample_Plate - (optional)
- Sample_Well - (optional)

- Sentrix_ID - einzigartige ID für jeden Infinium Methylation BeadChip (benötigt)
- Sentrix_Position - Position der Probe auf dem Beadchip (benötigt)
- Signalstärke (Bezug auf Fluoreszenz) - Menge des gemessenen Lichtsignals eines Spots auf einem Microarray eines Beadchips
- Sonde (eng. Probe) - einzelsträngiges DNA Fragment, das auf einem Microarray befestigt ist und dazu dient, mit einer komplementären DNA aus einer Probe zu binden
- Singular value decomposition (SVD) -
- Zentrifugation - Mechanische Separation von Stoffgemischen anhand ihrer Dichte durch hohe Zentrifugalkräfte (typischerweise in einer Zentrifuge)