

experimentelle Methoden der Bioinformatik

Inhaltsverzeichnis

1	Allgemein / Hintergrund	1
2	ChIP-Chip und ChIP-Seq	1
2.1	Ablauf	1
2.1.1	Crosslinking	1
2.1.2	Sonication	1
2.1.3	Immunoprecipitation (Selektion mittels Antikörper)	1
2.1.4	Reverse Immunoprecipitation	2
2.1.5	Reverse Cross Linking	2
2.1.6	Auswertung	2
2.2	Probleme/Fehler	2
2.3	Antikörper	3
3	Peak Calling	4
3.1	MACS	5
4	CLIP-Seq	6
4.1	ICLIP	6
5	PAR-CLIP	6
6	Protein-Protein-Interaktion	6
7	Tandem Affinity Purification (TAP)	6
7.1	Local clique merging algorithm (LCMA)	6
7.2	Clique Finding Algorithm (CFA)	6
8	RNA structure probing	6
8.1	chemical probing	6
9	X-ray crystallography	8
10	NMR spectroscopy	9

1 Allgemein / Hintergrund

Messung von Strukturen vs. Messung von Interaktionen

Motifsuche:

- Proteine (Transkriptionsfaktoren) haben Domäne die Nukleotidsequenzen erkennen
- Position weight matrix (PWM), position specific scoring matrix (PSSM)
- MEME zum erkennen von Sequenzen / Motifen

2 ChIP-Chip und ChIP-Seq

ChIP: **Ch**romatin-**I**mmuno**P**recipitation

Kein Single Cell Protocol -> es werden Zellpopulationen benötigt

Ziel: Man will feststellen an welcher Stelle Proteine binden

Quellen für Fehler / Ungenauigkeiten: Messung des Populationsmittelwerts

ChIP-Chip: Chromatin-Immunoprecipitation Chip

ChIP-Seq: Chromatin-Immunoprecipitation DNA-Sequencing

2.1 Ablauf

2.1.1 Crosslinking

Stabilisierung der Bindungen zwischen DNA und Protein

Geschieht reversibel zwischen DNA (**Chromatin**) und rekombinanten Proteinen

- Formaldehyd (CH₂O) vernetzt Base (B) mit Proteinen (P-NH₂) quer
- $\text{P-NH}_2 + \text{CH}_2\text{O} \rightleftharpoons \text{PN}=\text{CH}_2 + \text{NH}_2\text{-B} \rightleftharpoons \text{PNH-CH}_2\text{-NH-B}$
- Rekombinant: Biotechnologisch hergestellte Proteine aus genetisch veränderten Organismen

2.1.2 Sonication

Zerstören und Zerkleinern (fragmentieren) der Zellen, Zellbestandteile und DNA durch Ultraschall

(Vorher: Waschen der Zellen mit Protease Inhibitor, Lyse + homogenisieren)

- zeitkritisch → Länge bestimmt Grad der Zerkleinerung
- 200-1000 BP Fragmente im Idealfall

Ergebnis sind DNA Fragmente mit gebundenen Proteinen

2.1.3 Immunoprecipitation (Selektion mittels Antikörper)

- Antikörper (binden an Beads oder Membranen, Chip/in Gel) binden an rekombinante Proteine

oder Protein-TAG (kurze Aminosäuresequenz, markieren Protein)

- Aufreinigung:

- Zentrifugation des Präzipitats: Beads+(Protein-DNA) am Boden, Zellfragmente/Rest in Lösung
- Abkippen der Lösung
- Aufnehmen des Beadspellets in Puffer, erneut zentrifugieren (x-Mal)
- Manchmal noch
- DNase Verdau der DNA in Lösung
- Aufheben der DNA in Lösung, als total-Chromatin-Probe

2.1.4 Reverse Immunoprecipitation

Durch Aufreinigungsschritte sind Beads/Gel/Chip idealerweise frei von Zellfragmenten/ungebundener DNA.

Umkehren der IP mit Elutionspuffer → Antikörper von DNA+Proteine trennen

→ Salzgehalt und PH-Wert an Rückreaktion angepasst

2.1.5 Reverse Cross Linking

- Thermische Zerstörung der Bindung zw. Protein und DNA
- Salzgehalt des Buffer angepasst auf Rückreaktion - Proteinase K und RNase bauen Proteine und RNA ab (zur Aufreinigung)
- Extraktion der übrig gebliebenen DNA durch Zentrifuge

2.1.6 Auswertung

Chiphybridisierung

- Hybridisierung der DNA an Microarray
- Färbung der DNA
- Messung der Farbintensität

→ *mit dem ChIP Background kann ich nichts anfangen...* ←

Sequencing

Hochdurchsatzsequenzierung der aufgereinigten DNA.

→ DNA extrahieren → DNA fragmentieren → Primer an Fragmente → Sequenzierung

→ Herausrechnen der Primer (idealerweise kennt man sie) →

Quality control → Phred-score Berechnung (Güte der erkannten

Nukleobase) → Cutoff bei zu niedrigem Phred-score → Mapping des

sequenzierten Teilstücks auf Genom

2.2 Probleme/Fehler

Cross-Linking

FN: Protein an DNA gebunden, aber kein Cross-Linking

FP: Proteine, die sehr nahe an der DNA sind, aber ungebunden, werden

auch cross linked

Sonication

- Größe der Fragmente abhängig von Ultraschalleinsatz – zeitkritisch!
- Kürzere und längere Fragmente können Informationen enthalten

Immunoprecipitation

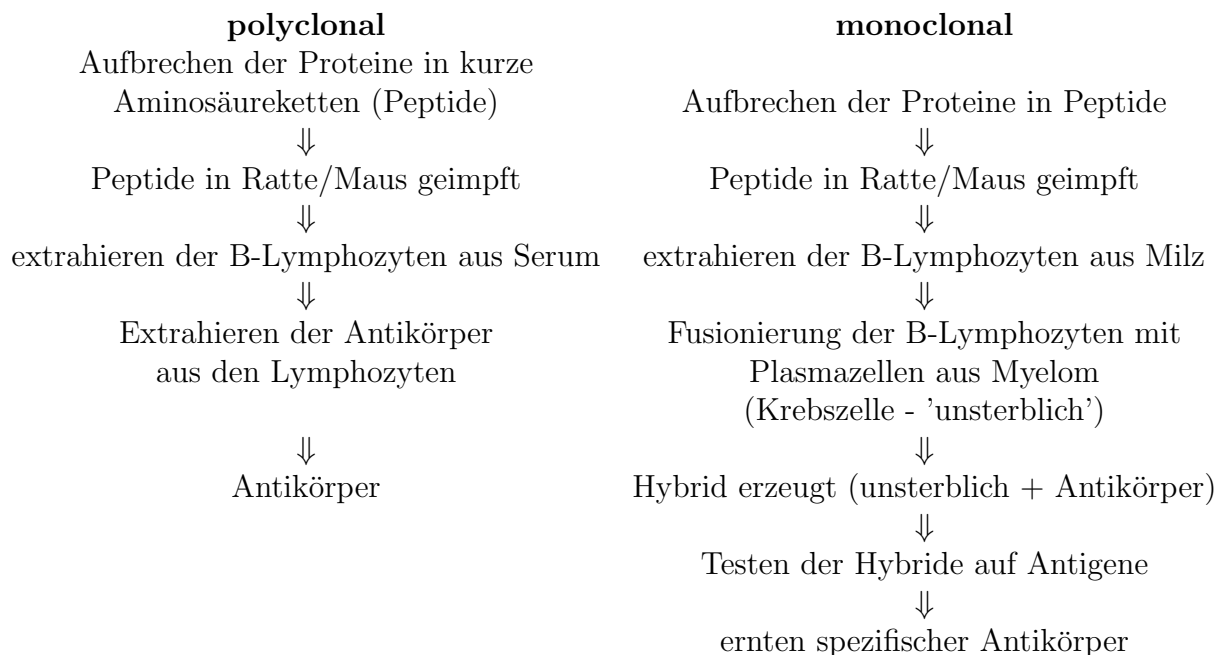
FP: Mangelnde Reinheit der rekombinanten Proteine; Spezifität der heterophilen Antikörper zu gering
Aufreinigung führt zu **FP** und **FN**

Chip

FN: Hybridisierung nicht effektiv genug

2.3 Antikörper

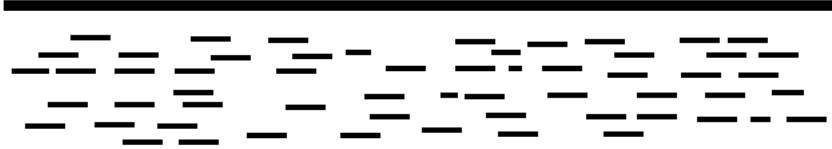
- Antikörper bindet spezifisch und sensitiv
- Antikörper sind fixiert an:
 - Beads
 - Chip (kein Microarray)
 - Gel
- Antikörper werden im Experiment erzeugt



3 Peak Calling

Sequenziertes Genom/RNA/DNA aus dem Experiment = viele, kurze Reads

→ naiv: Jedes Nukleotid, dass von Reads bedeckt ist = Gebunden



→ Problem: Viele FP, da kurze Reads mehrere Treffer haben können

→ Lösung: Cutoff für Anzahl der Reads auf Nukleotid

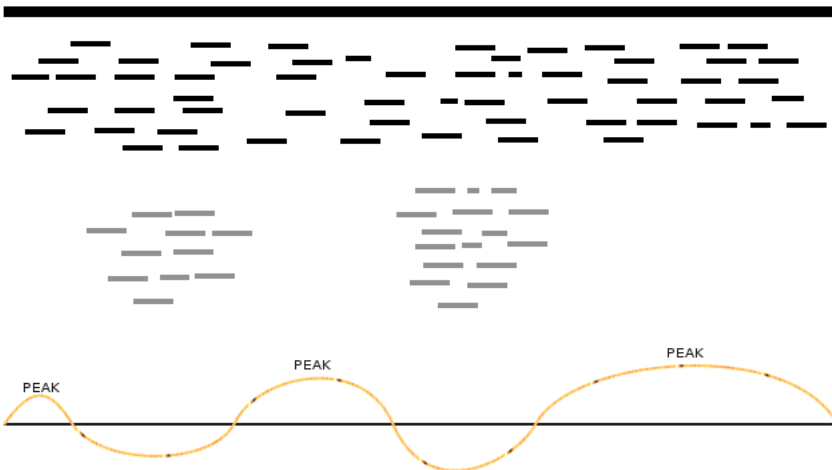
→ Problem: Manche Basen einfach zu binden = viele FP

So geht das nicht!

Lösung:

Enrichment: $\log \frac{Expression}{Background}$

naiv: Wenn Enrichment > Cutoff → Peak!



3.1 MACS

Model-based Analysis of ChIP-Seq (MACS)

1. Einteilen des Genoms in Bins (Eimer)

Window: 200 BP und Offset von 1/4 der window size

In Bins werden Reads eingeordnet

2. Zählen der Fragmente pro Bin, +/- Strang

→ Poisson verteilt!

$$P(x > k, \lambda) = \sum_{i=k}^{\infty} P\lambda(i) = 1 - \sum_{i=0}^{k-1} P\lambda(i) = 1 - \sum_{n=0}^{k-1} \frac{\lambda^n}{n!} e^{-\lambda}$$

λ =Mittelwert der read counts aus Background, k =read counts aus Experiment
read count signifikant größer Mittelwert → Peak!

Mittelwert kann abhängig von Menge der reads in Window sein:

$$\lambda = \max(\lambda_{\text{global}}, \lambda_{1000}, \lambda_{5000}, \lambda_{10000})$$

→ Window jeweils zentriert an Bin

3. p-Value Correction

Holm-Bonferroni

q-Value

4. Peakmerging

Wenn Abstand zwischen Peaks < Cutoff → Merge Peaks

(bei MACS 2xWindowSize)

Wo sind die Bindungsstellen?

Protein → RNA - **ChIP**: Regionen, mit denen das Protein assoziiert ist

DNA → RNA - **ChIRP**: Match von RNA auf sequenzierter DNA

Verfahren ähnlich zu CLIP

→ RNA cross-linking (UV o. formalin) → aufreinigen →

reverse cross-linking → Read → Match mit DNA

(Chromatin isolation by RNA purification)

Protein → RNA - **RIP**: RNA zu cDNA, hybridisieren mit Chip

→ RNA cross-linking (UV o. formalin) → aufreinigen →

reverse cross-linking → RNA in cDNA →

Hybridisierung auf Chip

(RNA immunoprecipitation protocol)

4 CLIP-Seq

4.1 ICLIP

5 PAR-CLIP

6 Protein-Protein-Interaktion

7 Tandem Affinity Purification (TAP)

7.1 Local clique merging algorithm (LCMA)

7.2 Clique Finding Algorithm (CFA)

8 RNA structure probing

8.1 chemical probing

MACS (Model-based Analysis of ChIP-Seq):

1) Einteilen des Genoms in Bins

Window Size: typisch 200 bp & offset (ungefähr 0,25 windows size = 50 bp)

MACS empfiehlt Bin doppelt so groß wie Fragment

2) Zähle die Anzahl an hypothetischen Fragmenten pro Bin (=window)

Fragmente können in mehr als ein Bin fallen

CLIP Cross-linking & immunoprecipitation protocol - Ultraviolettes Licht für cross linking - UV cross linked nur RNA mit Proteinen - induziert UV Mutation der RNA - CIMS: cross-linking induced mutation sites

9 X-ray crystallography

Voraussetzung: regulären Kristall aus dem Protein



Bragg's Law: $n\lambda = 2d\sin(\Theta)$

X-ray crystallography diffraction:

X-ray \rightarrow Kristall \rightarrow Ablenkung

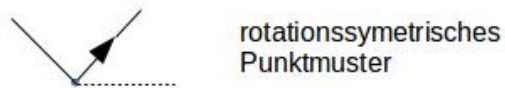
durch Atome \rightarrow Ablenkung wird durch einen Detektor gemessen

feste Wellenlänge λ , Winkel Θ variieren (Kristall rotieren) \rightarrow charakteristisches Diffraction pattern \rightarrow Amplitude ändert sich über den Winkel

$d_{hkl} = \frac{a_0}{\sqrt{h^2 + k^2 + l^2}}$ mit hkl=Laue-Index, a_0 = Gitterkonstante

oder:

Θ fest und λ variieren \rightarrow white x-ray

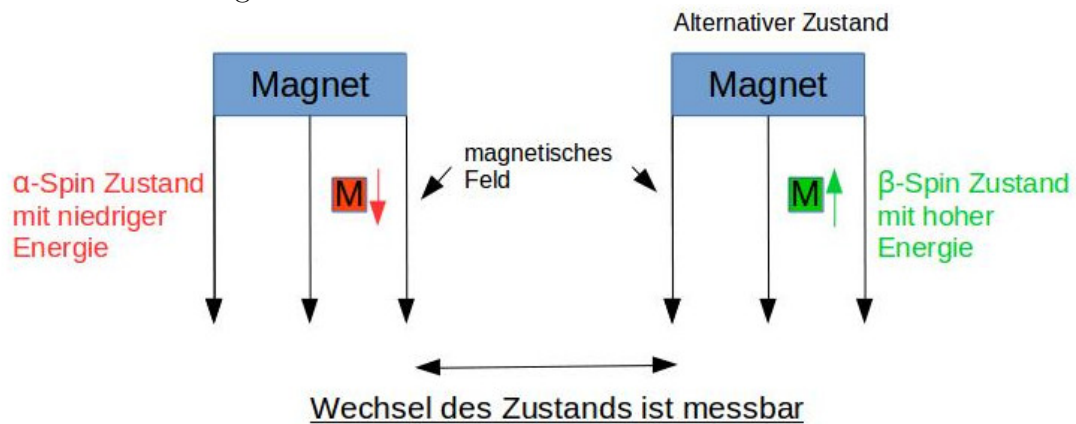


Kombinierte Information aus allen Messungen für verschiedene λ & Θ

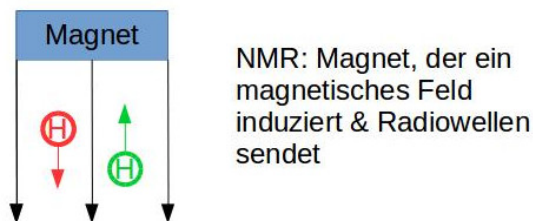
1. Backbone des Proteins ($COOH - NH_2$)
2. Bestimmung der Position der flexiblen Seitenketten der Aminosäuren
3. Verbesserung

10 NMR spectroscopy

NMR: nuclear magnetic resonance



Atome mit magnetischen Eigenschaften: H, Deuterium, N, C, Li, B, O



- ohne weitere äußere Einflüsse Atom in α - spin
- über Flips im Magnetfeld Ermittlung der Protein-Struktur

Spektren von H,C,N + Strukturformel der bekannten Aminosäure + Aminosäureketten
 → Wechselwirkungen zwischen den Gruppen herleiten → 3D Koordinaten berechnen