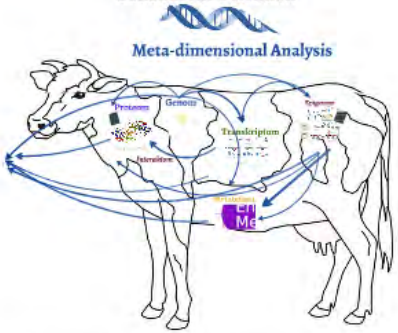
Konzepte und Ideen hinter GWAS Studien und der Bestimmung von QTL

Phänotypen basierte Selektion

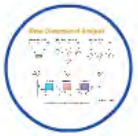
- Abhängig von Umweltfaktoren: z.B. Klima, Futterangebot (Art und Menge), Haltungsform,...
- Viele Merkmale
  - können erst später im Leben / nächsten Generationen erkannt werden
  - sind schwer - ethisch bedenklich oder kostenintensiv zu bestimmen

Genomische Selektion



- Zuchtwertschätzung und genetische Veranlagungen der Tiere können direkt aus den Erbinformationen abgelesen werden
- Schon im frühesten Stadium der Entwicklung
- Screen für tausende Merkmale gleichzeitig möglich

Ansätze zur Datenintegration



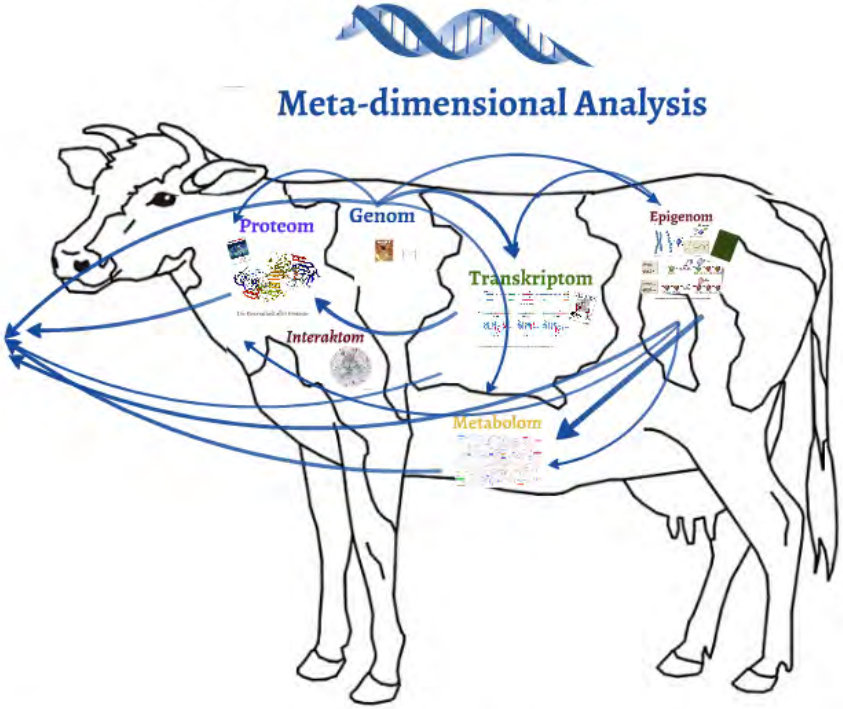
Herausforderungen

der Bestimmung von QTL

Phänotypen basierte Selektion

- Abhängig von Umweltfaktoren: z.B. Klima, Futterangebot (Art und Menge), Haltungsform,...
- Viele Merkmale
  - können erst später im Leben / nächsten Generationen erkannt werden
  - sind schwer - ethisch bedenklich oder kostenintensiv zu bestimmen

Genomische Selektion



- Zuchtwertschätzung und genetische Veranlagungen der Tiere können direkt aus den Erbinformationen abgelesen werden
- Schon im frühesten Stadium der Entwicklung
- Screen für tausende Merkmale gleichzeitig möglich

Herausforderungen

Probleme der MAS:

Individuelle Selektion: Marker nur in den schwersten Fällen auch der Grund für das Auftreten eines bestimmten Merkmals.

Falsch positiv durch Rekombination

Abhängig von externen Faktoren / Populationsstratifikation / Störvariablen

Nur eine Ebene

Genomweite Assoziationen

Diskrete Merkmale

In der Regel sind nur ein oder wenige Gene an der Ausprägung solcher Merkmale beteiligt.

Identifizierung der verantwortlichen Gene vergleichsweise einfach.

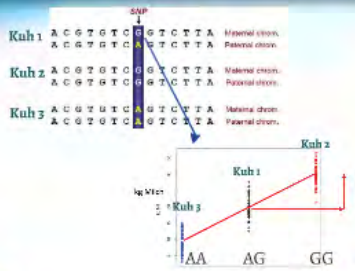
Quantitative Merkmale

Interaktionen einer größeren Zahl an Genen sowie Wechselwirkungen mit Umweltfaktoren .

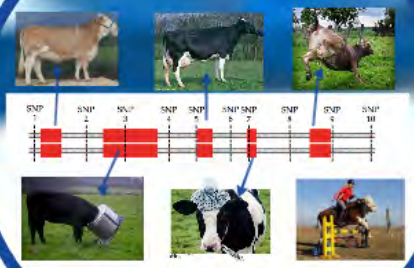
Identifizierung der beteiligten Gene ist deshalb methodisch sehr aufwändig.

=> Bestimmung von Quantitative Trait Loci (QTL)
Die Größe der QTL / Zahl der Kandidatengene ergibt sich dabei aus der Markerdichte.

Genomische Marker



Quantitative Trait Loci



Assoziationskartierung

Diskrete Merkmale

In der Regel sind nur ein oder wenige Gene an der Ausprägung solcher Merkmale beteiligt.

Identifizierung der verantwortlichen Gene vergleichsweise einfach.

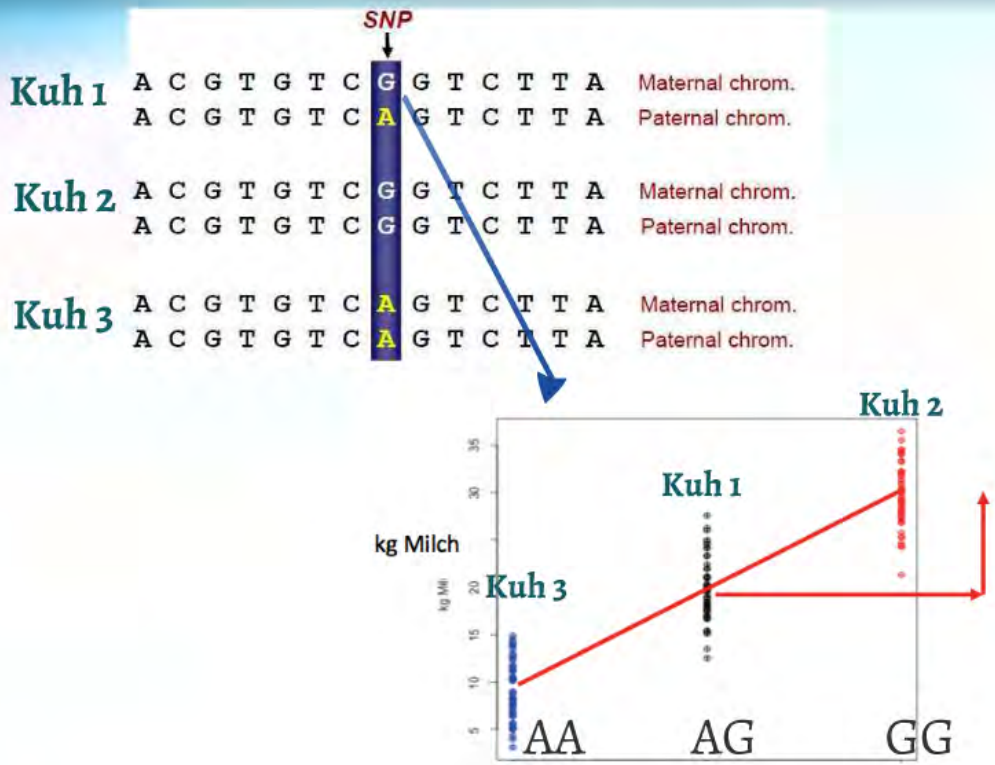
Quantitative Merkmale

Interaktionen einer größeren Zahl an Genen sowie Wechselwirkungen mit Umweltfaktoren .

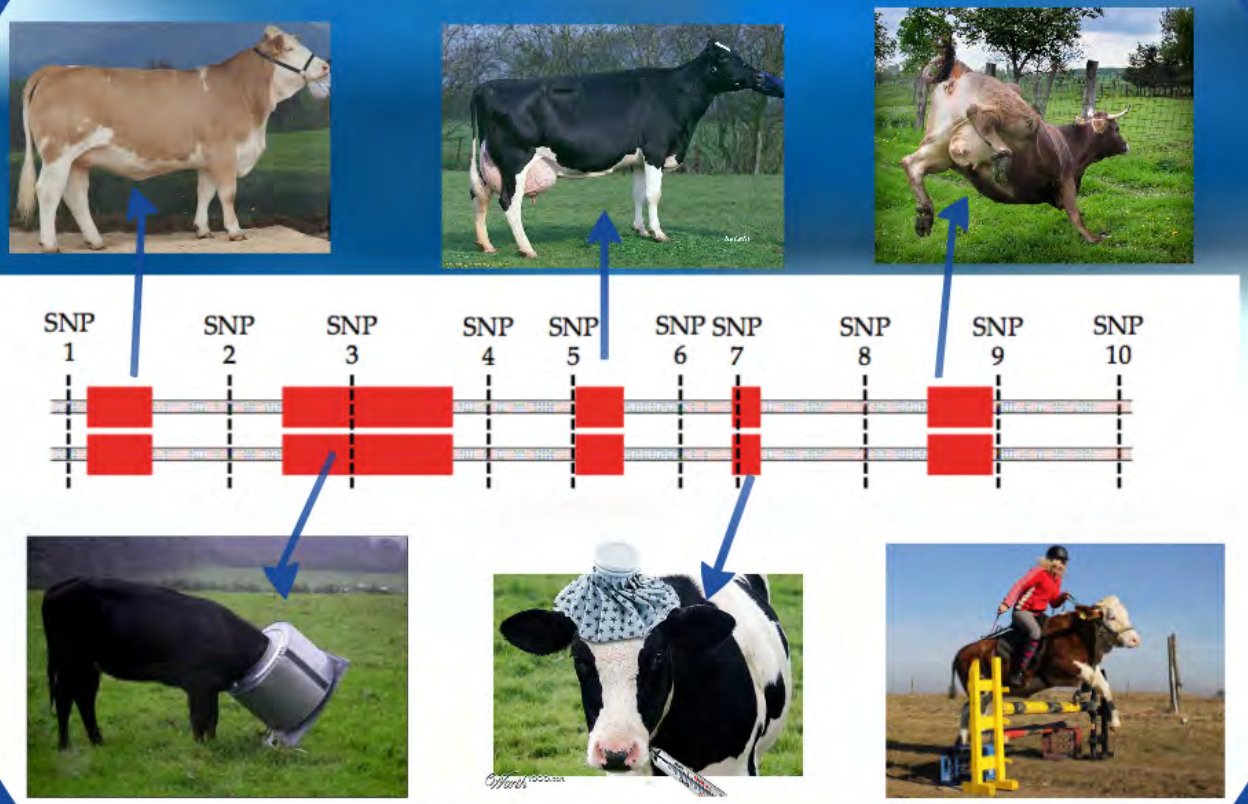
Identifizierung der beteiligten Gene ist deshalb methodisch sehr aufwändig.

=> Bestimmung von Quantitative Trait Loci (QTL)
Die Größe der QTL / Zahl der Kandidatengene ergibt sich dabei aus der Markerdichte.

Genomische Marker



Quantitative Trait Loci



Assoziationskartierung

Science
Livestock Decoded
AAAS
Probleme der MAS:
Falsch positiv durch Rekombination
Nur eine Ebene
A
C
G
T
C
A
T
G
C
A
G
T

Probleme der MAS:

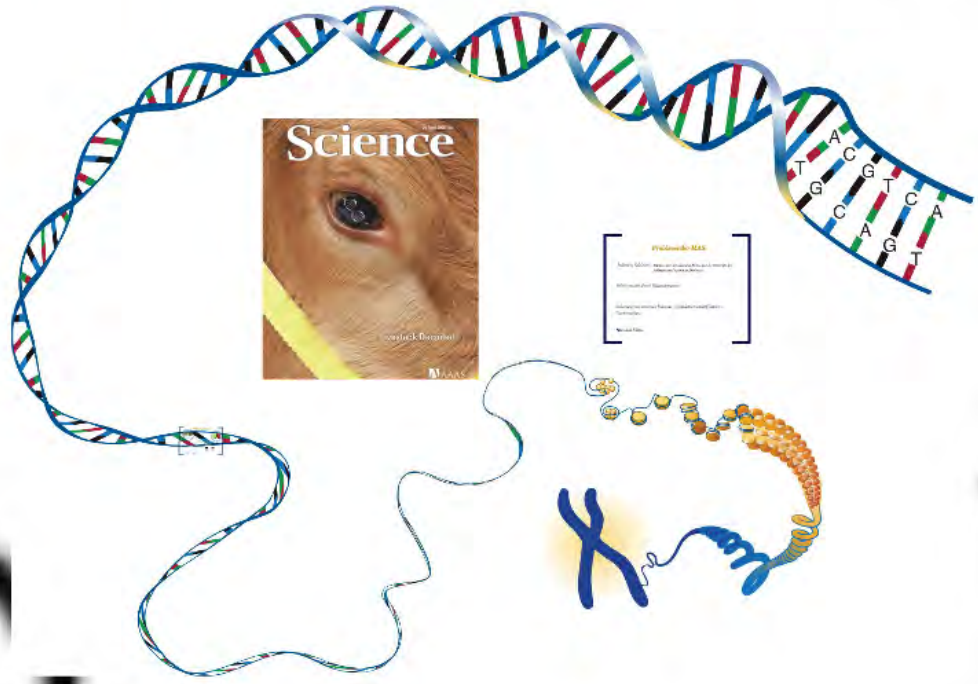
Indirekte Selektion: *Marker nur in den seltensten Fällen auch der Grund für das Auftreten eines bestimmten Merkmals.*

Falsch positiv durch Rekombination

Abhängig von externen Faktoren / Populationsstratifikation / Störvariablen

Nur eine Ebene

Genom



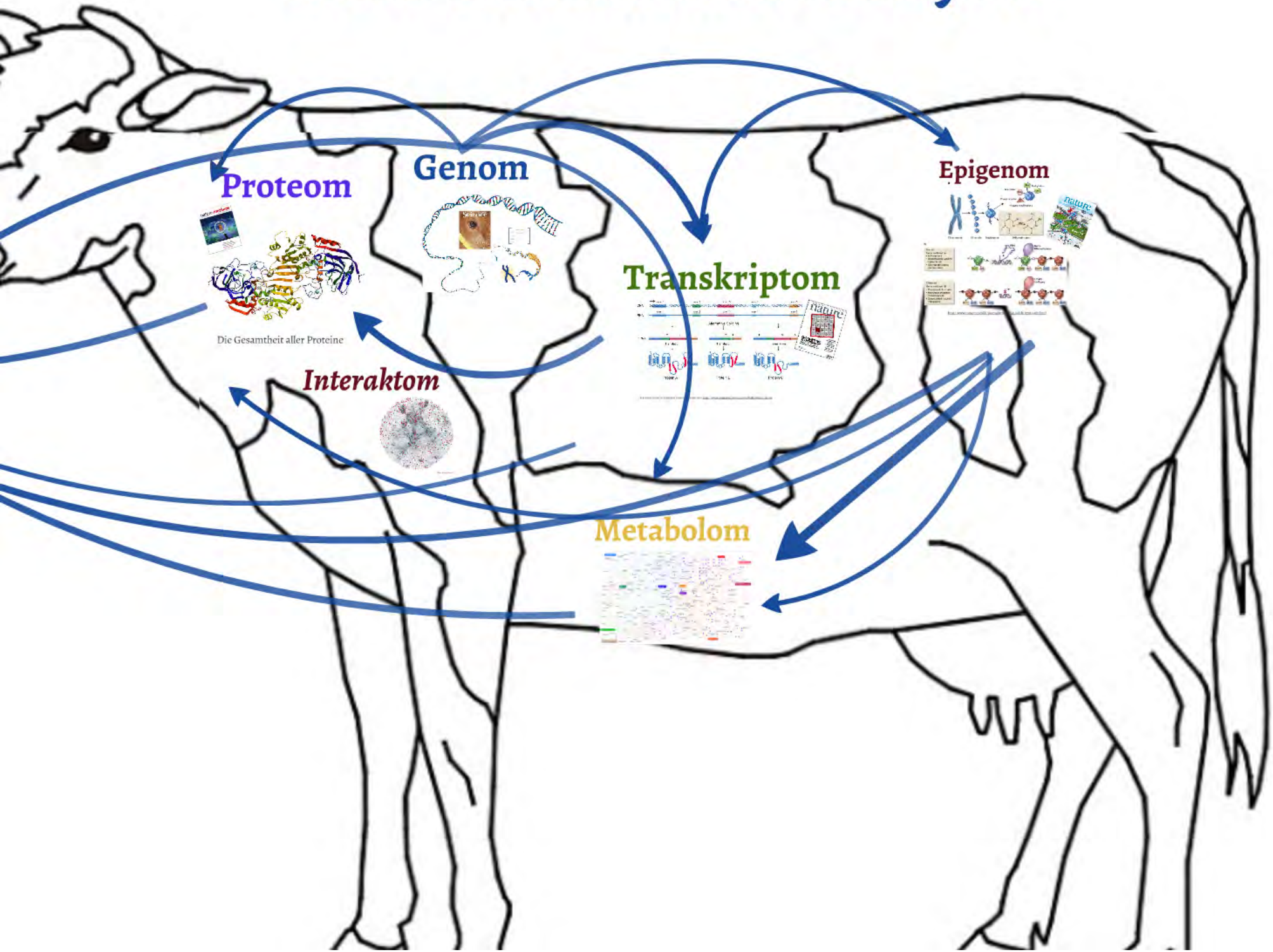
DNA

RNA

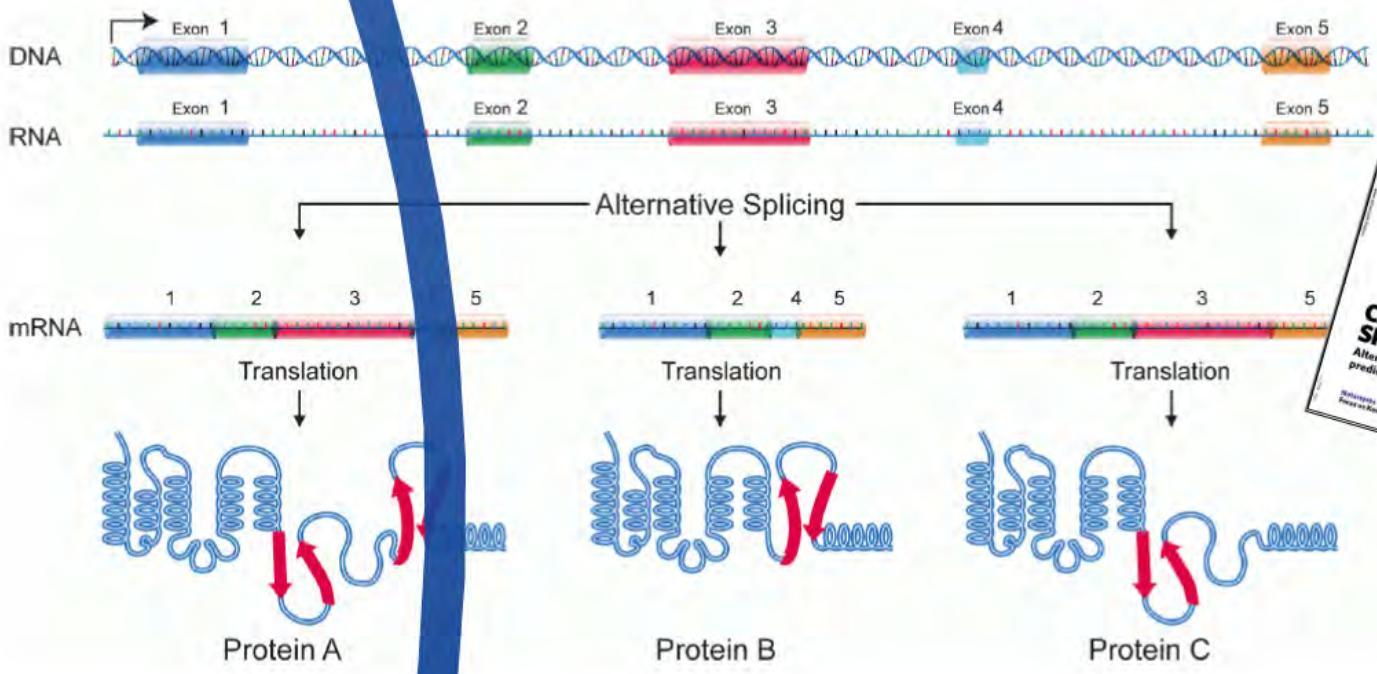
mRN

ne

Meta-dimensional Analysis

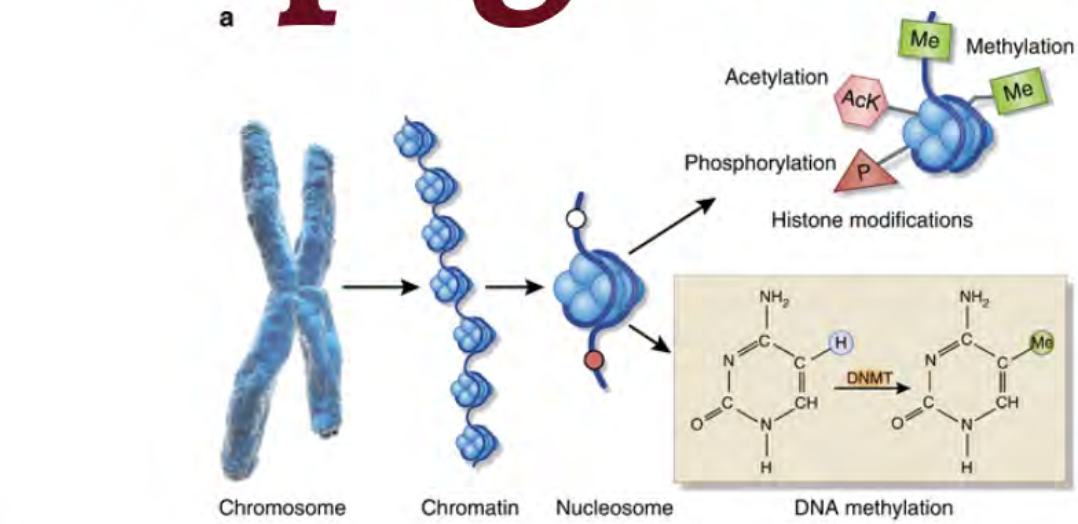


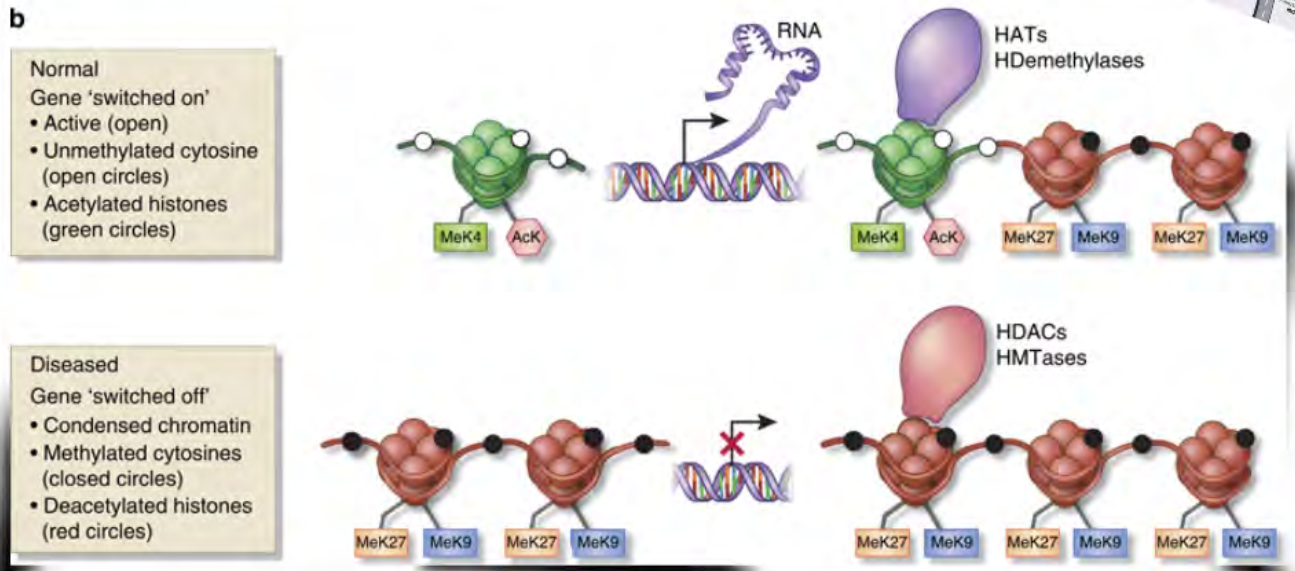
Transkriptom



National Human Genome Resea Institute (http://www.genome.gov/Images/EdKit/bio2j_large

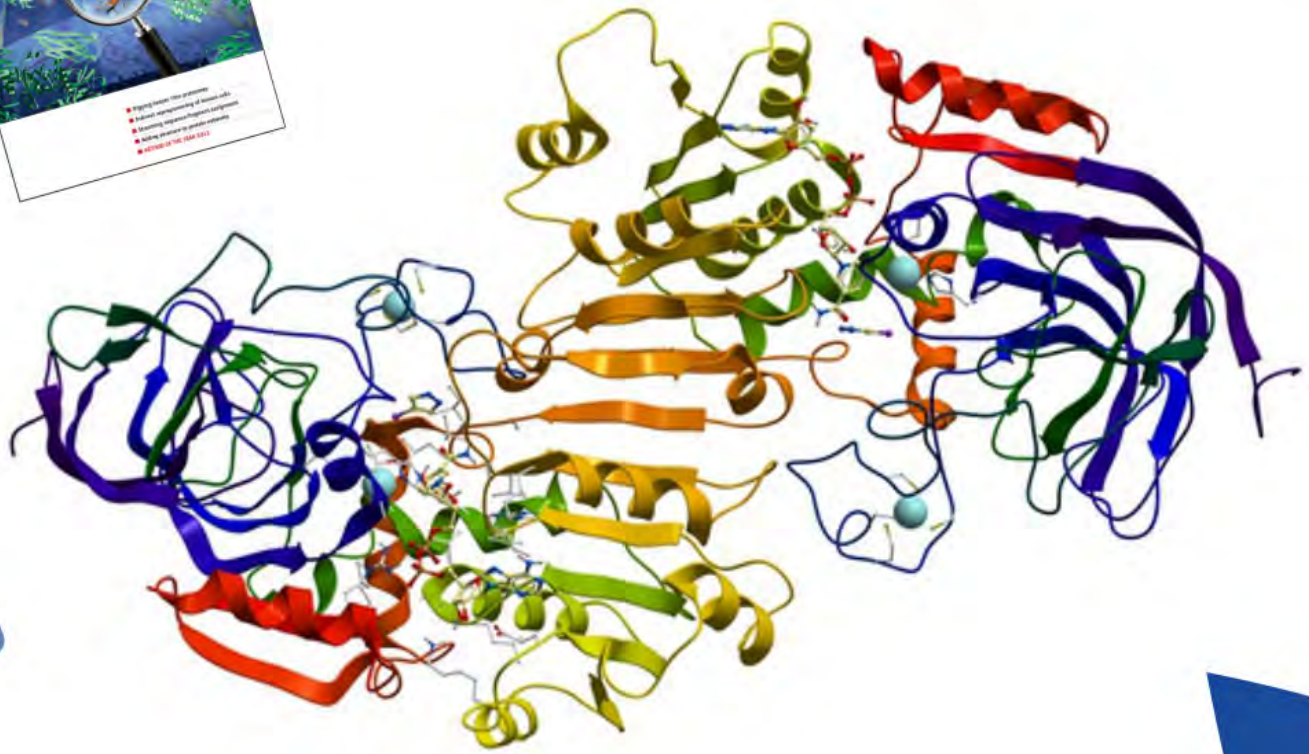
Epigenom





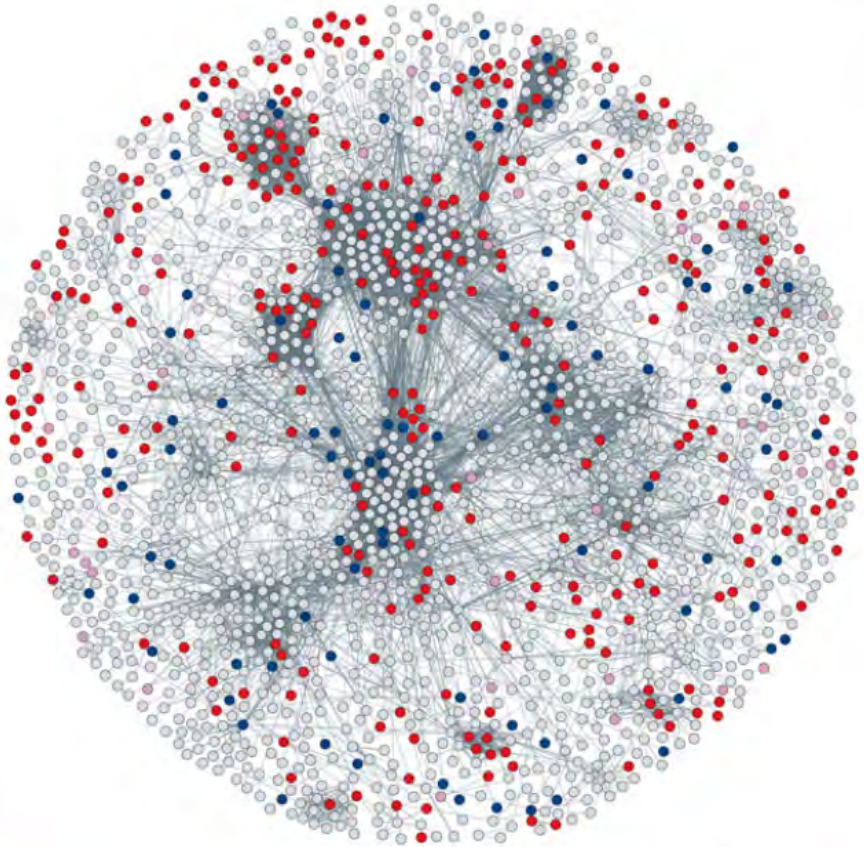
http://www.nature.com/ki/journal/v79/[illegible]/fig_tab/ki2010335f2.html

Proteom



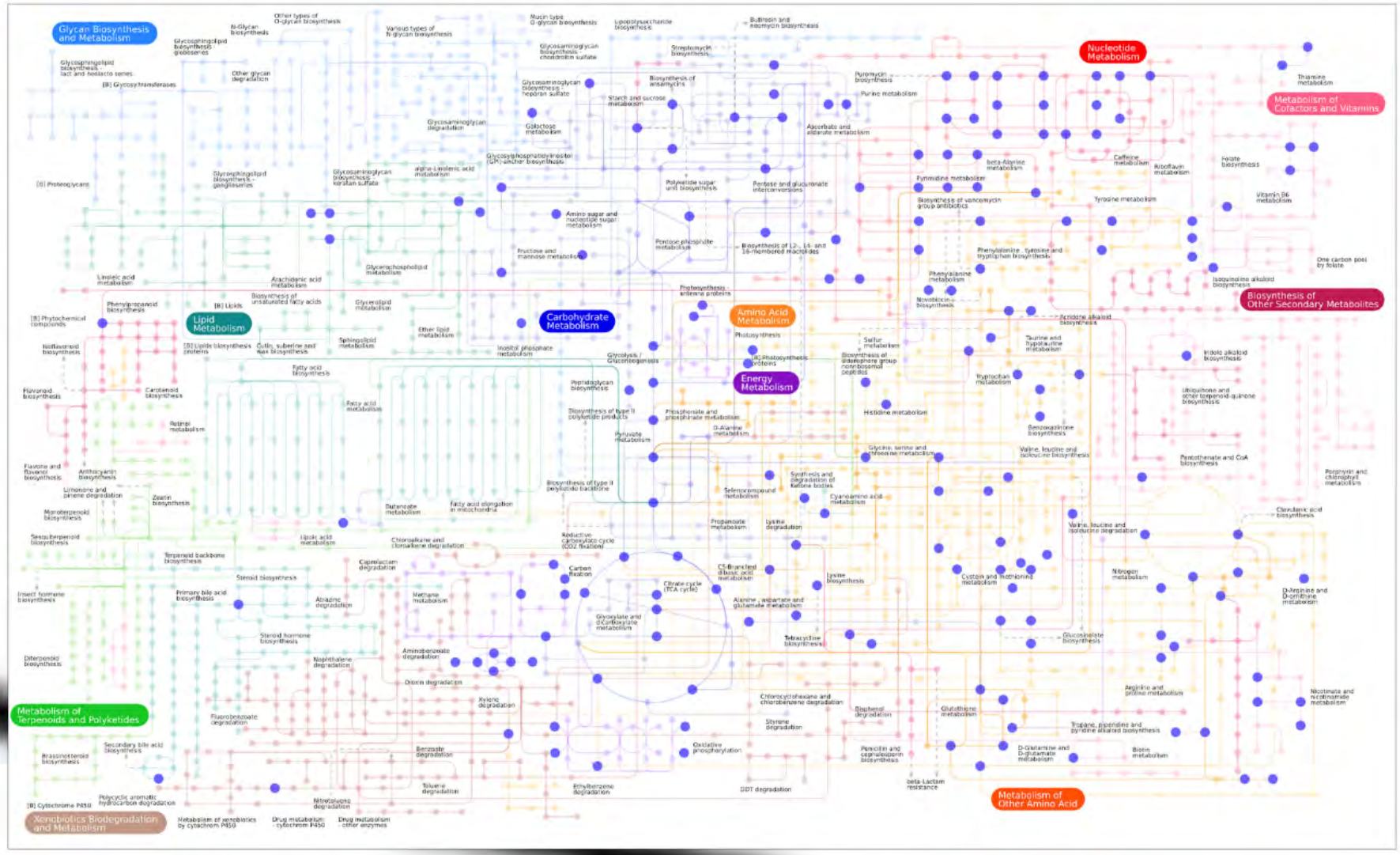
Die Gesamtheit aller Proteine

Interaktom



Nature Reviews | Genetics

Metabolom


Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid

Ansätze zur Datenintegration

Multi-Staged Analysis

Dreiecks Strategie:

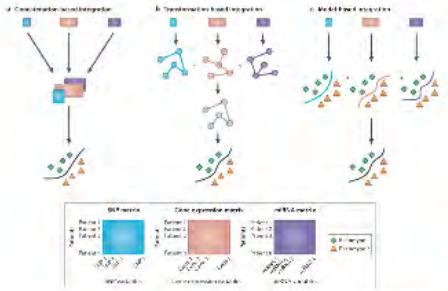
Schritt 1: *Filtere SNPs anhand von phenotypischen Assoziationen oder anhand von Annotationen*

Schritt 2: *Teste signifikante SNPs aus Schritt 1 mit einem anderen "Omics" Level --> e_QTL, m_QTL, met_QTL, p_QTL,....*

Schritt 3: *Teste Daten aus Schritt 2 auf Korrelation mit dem Phänotyp*

Annahmen: Integration kann hierarchisch vorgenommen werden
Schwächen: Durch Vor-Filterung geht viel ungekannte Information verloren

Meta-Dimensional Analysis



Nature Reviews | Genetics

http://www.nature.com/nrg/journal/v14/n1/pdf/nrg3351.pdf

Multi-Staged Analysis

Dreiecks Strategie:

Schritt 1: ***Filtere SNPs anhand von phenotypischen Assoziationen oder anhand von Annotationen***

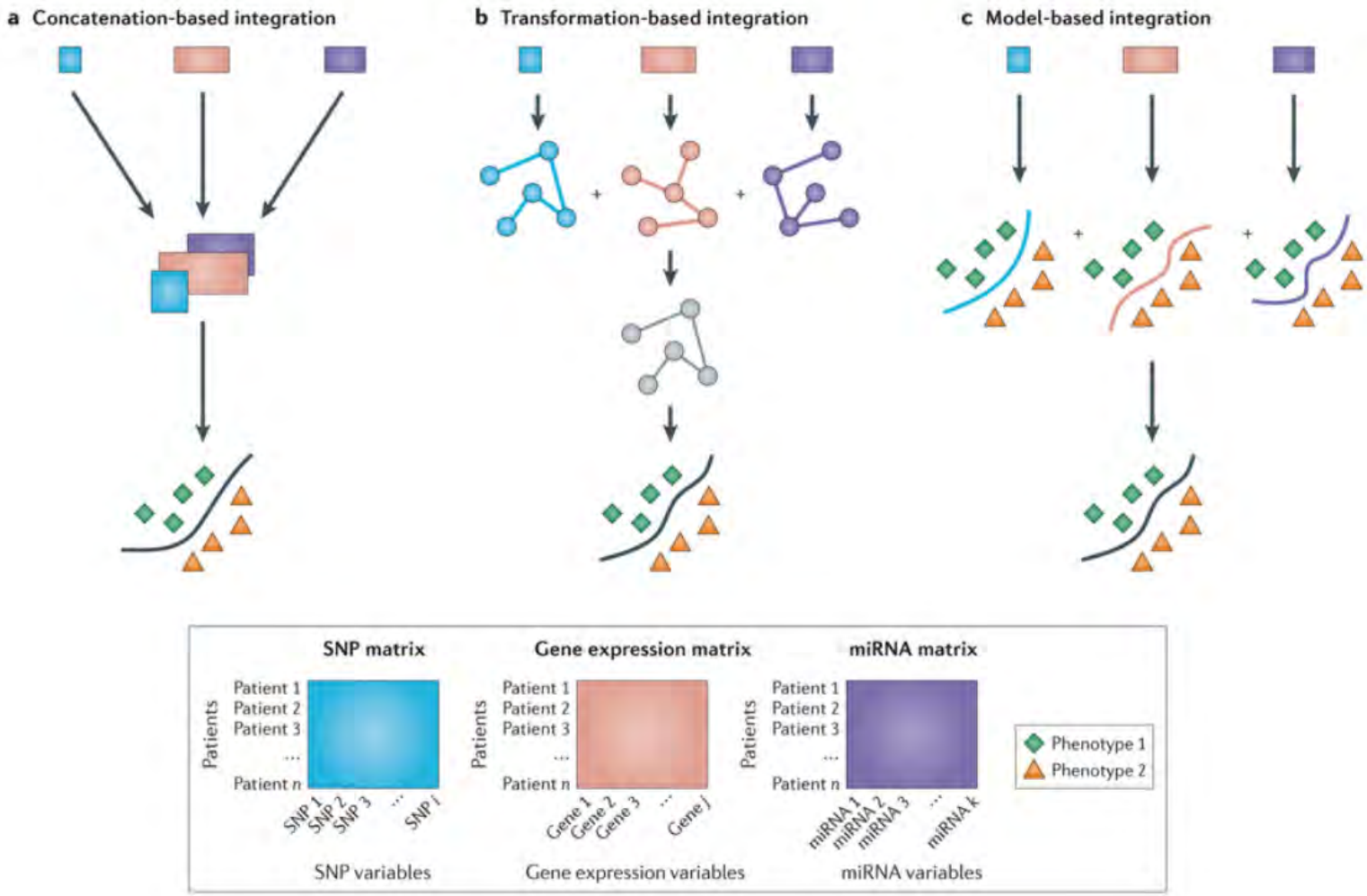
Schritt 2: ***Teste signifikante SNPs aus Schritt 1 mit einem anderen "Omics" Level --> e_QTL, m_QTL, met_QTL, p_QTL,....***

Schritt 3: ***Teste Daten aus Schritt 2 auf Korrelation mit dem Phänotyp***

Annahmen: Integration kann hierarchisch vorgenommen werden

Schwächen: Durch Vor-Filterung geht viel ungekannte Information verloren

Meta-Dimensional Analysis



Nature Reviews | Genetics

http://www.nature.com/nrg/journal/v16/n2/pdf/nrg3868.pdf

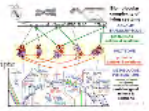
Herausforderungen

Verarbeitung und Qualitätssicherung der Rohdaten

Kausalität vs Korrelation vs Drittvariableneffekte

Grössenverhältnisse (Skalierung)
Dimensionalität

Dimensionen:
1 Mrd. Bp.
5 Mio. SNPs
23K Gene
10^{5}-10^{6} Transkripte
10^{5} Proteine



Datentypen und Technologien
Sequenzen, Positionen
Expressionslevel,
Netzwerke,
Methylierungszustand,
Annotationen
Phenot. Merkmale

Zeitliche / externe Abhängigkeiten

BIG Data Problem

Beispiel: SNP-SNP-Interaction (5 Mio SNP)

Paarweiser Vergleich: $1.45 \cdot 10^{13}$ Modelle
≈ 1400 Stunden

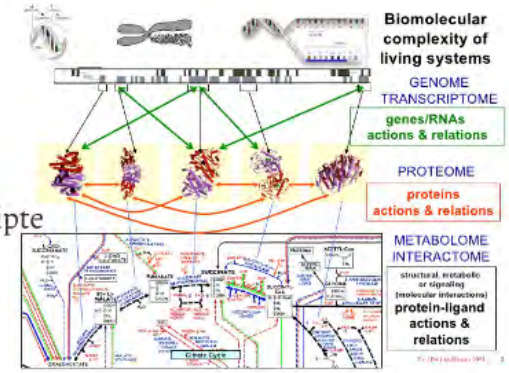
3 Variablen: $2.09 \cdot 10^{19}$ Modelle
≈ $5.7 \cdot 10^{9}$ Stunden

Verarbeitung und Qualitätssicherung der Rohdaten

Kausalität vs Korrelation vs Drittvariableneffekte

Grössenverhältnisse (Skalierung)
Dimensionalität

Dimensionen:
3 Mrd. Bp,
5 Mio. SNPs
23K Gene,
10^5-10^6 Transkripte
10^5 Proteine



Datentypen und Technologien

Sequenzen, Positionen
Expressionslevel,
Netzwerke,
Metylierungszustand,
Annotationen
Phenot. Merkmale

Zeitliche / externe Abhängigkeiten

BIG Data Problem

Beispiel: SNP-SNP-Interaction (5 Mio SNP)

Paarweiser Vergleich:

3 Variablen: