Genomics Vl 8 bei Dr. Jung Gene-Set Analysen

Wir sind immer noch im Bereich Transkriptom und wir machen heute damit weiter.

Schauen wir uns an, wo wir letztes Mal stehen geblieben sind. Wir haben uns die differenzielle Genanalyse angeschaut. Die meisten dieser Experimente laufen als 2- Gruppen Experimente ab. Das heißt, wenn man zum Beispiel Patientendaten hat, dann hat man eine gesunde Kontrollgruppe und man vergleicht die Genexpression von kranken Patienten mit der Kontrollgruppe. Wenn man stattdessen Laborversuche mit Zelllinien macht, dann vergleicht man behandelte Zelllinien mit Nicht-behandelten Zelllinien. Man schaut sich die Expression von Tausenden Genen an. Das macht man entweder mit Microarrays oder RNA-seq. Man berechnet für jedes Gen ein statistisches Testergebnis. Dieses gibt einem Aufschluss darüber, ob die Expressionsänderung rein zufällig ist oder ob man sie als signifikant betrachten kann. Dazu hatten wir letztes Mal die p-Werte und FDR adjusted P-Werte besprochen. Um Die Expressionsänderung zu quantifizieren, hat man den Log Fold Change. Bei einem negativen Log Fold Change hat man eine Runterregulation eines Gens. Bei positiven log Fold Change Werten ist das Gen hochreguliert. Es ist wichtig, dass man sowohl die P-Werte für die Signifikanz und die log Fold Changes für die Quantifizierung beobachtet. Wir haben auch gelernt, dass wir bei multiplen Tests viele falsch positive Ergebnisse haben können. Dagegen adjustierte man den P-Wert, um die Anzahl Falsch-positiver Ergebnisse zu kontrollieren und zu reduzieren. Was machen wir jetzt aber mit den Genen, die man selektiert hat? Wir haben es am Ende der letzten VL schon mal angesprochen. Man schaut sich an, um welche Gene es sich handelt. Sind Sie an irgendwelchen Pathways beteiligt? Al das soll uns bei einer Interpretation helfen.

S.10.

Was wir brauchen ist eine Annotation der Gene. Wir kennen ja schon eine ganze Reihe an Datenbanken, in denen schon vielfach Gene annotiert vorliegen. Die klassische Annotation erfolgt über die so genannte "**GO-Annotation**". Go steht für Gene ontology. Es sind Schlüsselwörter mit denen Gene annotiert werden. Diese Schlüsselwörter geben dann zum Beispiel an, ob die **molekulare Funktion** eines Gens bekannt ist. Es zeigt auch an, ob der **biologische Prozess** und die **zelluläre Komponente** bekannt sind. Das sind die 3 GO-Kategorien. Daneben kann man Gene aber auch einem Pathway zuordnen. Es gibt also auch Datenbanken, wo Gene mit passenden Pathway-Angaben hinterlegt sind. Auf **uniprot** kann auch zum Beispiel Annotationen finden. Zu beachten ist, dass Gene mehrere GO-Annotationen haben können, weil Gene oft an mehreren Prozessen beteiligt sind. Es gibt GO-IDs, die auf Seiten wie Uniprot mit angegeben werden. Wenn ich sage, dass ein Gen mit mehreren Annotationen verknüpft sein kann, dann kann die einzelne GO-Annotation auch mit mehreren Genen verknüpft sein. Eine Annotation wie "antibacterial humoral response" kann also bei vielen Genen dabeistehen. Was bei der Analyse in zum Beispiel R die Annotation und Annotationsverknüpfungen erschwert ist, dass ich auf Seiten wie Uniprot viel Information pro Zeile finde, wenn ich mir die Annotationen anzeigen lasse. Am besten ist es für den Bioinformatiker, wenn pro Zeile nur eine Information enthalten ist. Wenn mehrere Informationen hintereinander geschaltet stehen, muss man die wichtige Information aus den Zeilen extrahieren. Es gibt noch viele andere Datenbanken. Wir werden in der Übung lernen, wie man direkt aus R eine Datenbankabfrage machen kann. So können Sie direkt über Packages auf Datenbanken zugreifen und direkt mit der Annotation beginnen. Gut, jetzt läuft es so ab, dass man seine Gene nach dem P-Wert sortiert und die Top-Gene annotiert man und schaut sich die molekulare Zugehörigkeit an. Oft hat man aber mehrere 100te signifikant differenziell exprimierte Gene und dann muss man alles ein bisschen zusammenfassen.

S.11. Einige Datenbanken zu Pathways und GO-Annotationen von Genen.

www.uniprot.org.

reactome.org

www.genome.jp/kegg/

www.geneontology.org/

S.12.

Abbildung 1: Gene-Set-Analyse Tabelle.

- χ^2-test: $p < 0.01$
- → Gruppenfaktor ist mit Apoptose assoziiert

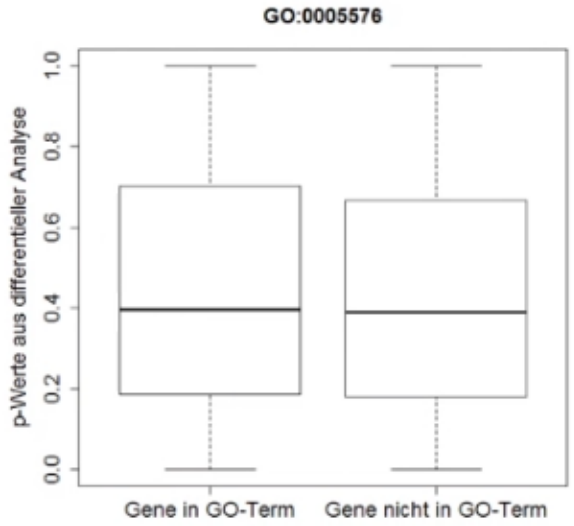
	Not differentially expressed genes	Differentially expressed genes	
Genes, not related to apoptosis	39600 (99.7%)	250 (83%)	39850
Genes, related to apoptosis	100 (0.3%)	50 (17%)	150
	39700	300	40.000

Ich stelle Ihnen hier eine ganz einfache Gene-Set Analyse vor. Diese beruht auch auf einem statistischen Test. Damit kann man überprüfen, ob die differenziell exprimierten Gene mit einer bestimmten Annotation verknüpft sind. Wir nehmen als Beispiel einfach den biologischen Prozess der Apoptose. Sagen wir mal, dass wir 40000 Gene haben, die wir unsere Analyse miteingeschlossen haben. Was wir in den Zeilen von Abbildung 1 sehen, ist die Information, ob ein Gen mit Apoptose assoziiert ist oder nicht. In Abbildung 1 sind 150 Gene mit Apoptose assoziiert und der Rest nicht. In den Spalten haben wir die Information, ob ein Gen als differentiell exprimiert selektiert wurden ist (über P-Wert und Log Fold Change als Kriterien). In Abbildung 1 wäre es jetzt so, dass 300 Gene differenziell exprimiert wurden und die restlichen Gene nicht. Dann kann man die Informationen aus Zeilen und Spalten miteinander verknüpfen und man kann sehen, dass unter den differenziell exprimierten Genen 17 % mit Apoptose assoziiert sind, während unter den nicht differenziell exprimierten Genen 0,3 % der Gene mit Apoptose assoziiert sind. Prozentual gesehen haben wir also mehr differenziell exprimierte Gene, die mit Apoptose assoziiert sind, als bei den Nicht-Differentiellen. Normalerweise würde man erwarten, dass, wenn ein biologischer Prozess mit einer Erkrankung nicht assoziiert ist, sich die Gene, die mit dem Prozess annotiert sind, gleichmäßig verteilen über die Tabelle. Das ist in Abbildung 1 nicht der Fall. Da ist es so, dass sich Überproportional viele Apoptose-Gene unter den differentiell exprimierten Genen befinden. Deshalb nennt man diese Analysen auch **Gene-Set-Enrichment-Analysen**. Das heißt, die Apoptose-

assoziierten Gene in diesem Beispiel sind unter den differentiell exprimierten Genen angereichert (enriched). Es gibt im Prinzip zwei Tests in der Statistik, die solche 4-Felder-Tafeln analysieren und schauen, ob die Information aus den Spalten mit der Information aus den Spalten signifikant verknüpft sind oder ob es eine zufällige Anordnung ist. Hier in Abbildung 1 haben wir den **Chi-Quadrat-Test**. Damit würde man überprüfen, ob die 17 % signifikant verschieden sind von den 0,3 % in Abbildung 1. Hier ist der Unterschied ganz klar, weil 17 % viel größer ist als 0,3 % und das erkennt der Test mit einem P-Wert kleiner als 0,01. Nehmen wir mal an, wir hätten statt 0,3 % 17,1 % und die 17 % bleiben. 17,1 % zu 17 % ist zwar ein Unterschied, aber kein signifikanter Unterschied. Ein anderer Test, den man anwenden könnte, wäre der Fischers exakter Test. Das kennen Sie vielleicht aus der Statistikveranstaltung, aber bei so großen Zahlen ist es üblich, dass man den CHI-Quadrat-Test nutzt. Üblicherweise haben wir mehrere Tausende GO-Terme oder Pathways, mit denen wir unsere Pathways verknüpfen. Das heißt, der CHI-Quadrat-Test wird für alle 40.000 GO-Annotationen oder Pathways durchgeführt. Wir haben also wieder eine Multiple Testsituation. Insofern müsste ich die P-Werte, die dabei rauskommen, wieder adjustieren. Also bitte unterscheiden: Wir haben die differenzielle Genexpressionsanalyse, die sich auf die Einzelgene bezieht, und wir haben die Gen-Set-Analyse, die Teilmengen von Genen betrachtet und nicht die einzelnen Gene selbst. Beide sind aber multiple Testsituationen und beide haben P-Werte, die adjustiert werden müssen, um falsch positive Ergebnisse zu vermeiden.

S.13. T-Test und Wilcoxon-Rangsummen-Test.

Abbildung 2: Boxplots der GO-assoziierten und nicht-assoziierten Gene

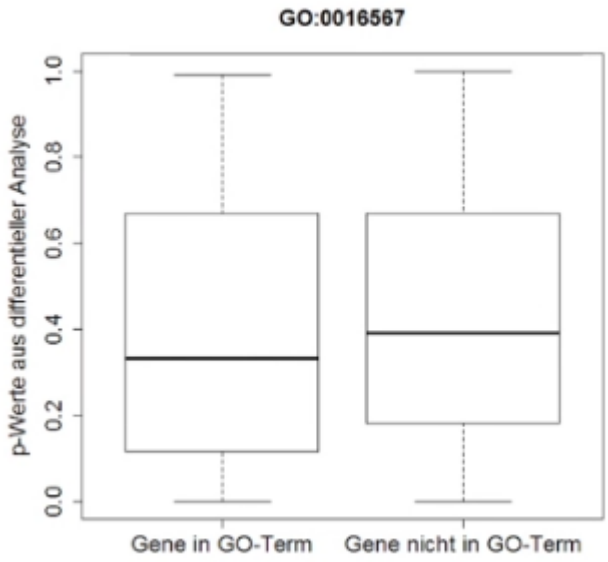


Es gibt wahrscheinlich ein paar Hundert Bioinformatik-Verfahren für Gen-Set-Analysen. Chi-Quadrat ist eine Möglichkeit. Ich stelle Ihnen noch eine weitere vor. Wir schauen uns dazu in Abbildung 2 diejenigen Gene an, die mit der GO-ID GO:0005576 verbunden sind. Dieser Term steht für die extrazelluläre Region. Es sind also Gene, die in der extrazellulären Region eine Rolle spielen. Jetzt schauen wir uns in Abbildung 2 die P-Werte (Y-Achse) an. Dazu schauen wir uns die P-Werte derjenigen Gene an, die mit dem GO-Term verbunden sind und die Gene, die nicht mit dem GO-Term verbunden sind. Auf der Y-Achse sind die P-Werte aufgetragen, die typischerweise zwischen 0 und 1 liegen. Die Verteilung der P-Werte sind als Box-plots dargestellt. Die sehen in Abbildung 2 sehr ähnlich aus und wir können nicht sagen, dass Gene, die in der extrazellulären Region etwas machen,

andere p-Werte haben als andere Gene. Insofern wäre die Schlussfolgerung, dass die beobachtete Erkrankung nichts mit der extrazellulären Region zu tun hat.

S.14. T-Test und Wilcoxon-Rangsummen-Test.

Abbildung 3: Box-Plot von GO:0016567 zum Vergleich mit Abbildung 2.



Schauen wir uns im Vergleich ein anderes Beispiel an. In Abbildung 3 habe ich zum Vergleich mit Abbildung 2 den GO-Term GO:0016567 in einem Box-Plot aufgetragen. Dieser Term steht für Protein Ubiquitinierung. In Abbildung 3 sehen wir, dass der Median der “Gene in GO-Term" etwas unter dem Median der “Gene nicht in GO-Term" steht. Wenn man jetzt den T-Test oder den Wilcoxon-Rangsummen-Test auf diese Daten anwendet (man vergleicht die Verteilungen der p-Werte zwischen den 2 Box-Plots mit den Tests), dann würde man ein signifikantes Ergebnis bekommen. Man kann dann schlussfolgern, dass die Protein Ubiquitinierung mit der untersuchten Erkrankung assoziiert ist. Das ist kein absoluter Beweis, aber ein erstes Indiz. Man würde jetzt mehrere Tests machen, um es zu überprüfen. Ich möchte Ihnen jetzt noch weitere Gene-Set-Analysen vorstellen. Die Analysen, die ich bis jetzt vorgestellt habe, fallen unter die **competitive Gene-Set-Analysen**. Das heißt, wir vergleichen immer die Menge an Genen, die mit einem Term assoziiert ist, und wir vergleichen es mit der Menge, die nicht damit assoziiert ist (Abbildung 1, 2 und 3).

s.15. Globaltests

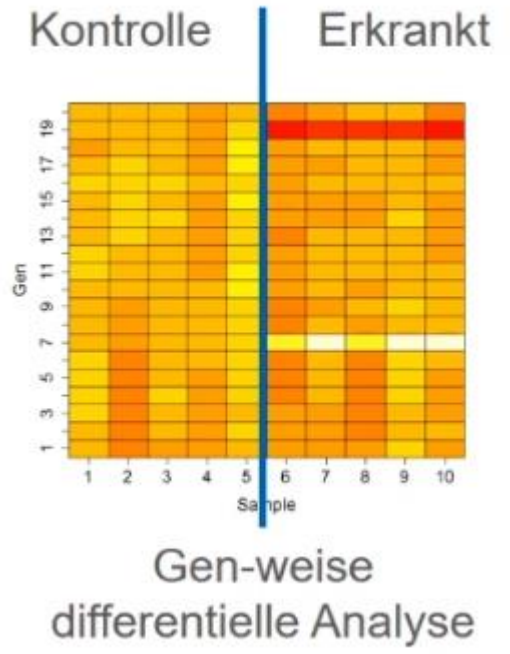
Dem gegenüber gibt es noch so genannte **Globaltests**. Das sind auch Gen-Set-Tests. Diese stellen aber nicht differenziell exprimierte und nicht-differentiell exprimierte Gene gegenüber, sondern Sie ziehen die Expressionsdaten selbst heran. Die Tests, die ich bisher gezeigt habe, basieren immer auf den P-Werten aus der differenziellen Genexpressionsanalyse. Diese Globaltests hingegen nutzen selbst nochmal die Genexpressionswerte. Das, was Sie in Grau in Abbildung 4 sehen, ist die Basis für competitive Gen-Set-Tests und das, was Sie in blau sehen in Abbildung 4, ist die Basis für die Globaltests. Der Globaltest nutzt also die Originaldatei aus den Spalten und Zeilen.

Abbildung 4: Was nutzen Globaltests und competitive Tests?

Gen	Kontrollgruppe			Erkrankte			p	log Fold Change
	1	…	n_1	1	…	n_2		
1	7.6	…	7.5	4.2	…	5.1	$2.7 * 10^{-10}$	-1.35
2	5.0	…	4.9	8.2	…	7.3	$4.7 * 10^{-09}$	-1.11
3	3.9	…	4.2	8.2	…	7.5	$9.4 * 10^{-09}$	+1.06
4	5.9	…	6.2	1.8	…	1.0	$1.1 * 10^{-08}$	-0.97
5	5.9	…	5.9	1.3	…	2.3	$1.9 * 10^{-09}$	-0.89
…	…	…	…	…	…	…	…	…
d	9.4	…	9.3	8.7	…	9.2	$9.9 * 10^{-01}$	0.01

So, wo ist denn der Nutzen von den so genannten Globaltests? Ich habe Ihnen mal eine Heatmap (farbliche Codierung einer Expressionsmatrix) dargestellt in Abbildung 5.

Abbildung 5: Heatmap.

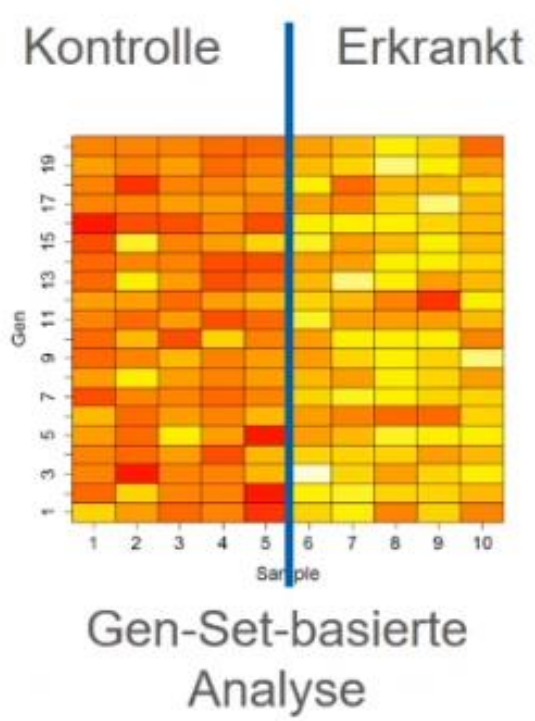


Starkes Hell-Gelb zeigt in Abbildung 5 eine starke Genexpression an und dunkel-Rot zeigt hier, dass keine Genexpression vorliegt. Wir betrachten in Abbildung 5 ein Genset, welches aus 20 Genen besteht (Y-Achse). Nehmen wir mal an, dass es sich um 20 Gene handelt, die alle zum selben Pathway gehören. Dann vergleicht man die Expression zwischen Kontrollgruppe und erkrankten Patienten. Es stehen in Abbildung 5 zwei Gene heraus: Gen 7 und Gen 19. Gen 7 ist bei den Erkrankten deutlich hochreguliert und Gen 19 ist bei den Erkrankten deutlich runterreguliert durch

die Erkrankung. Das würden wir ja bereits bei der differenziellen Genexpressionsanalyse herausfinden, wo wir Einzelgene betrachten.

S.16+17.

Abbildung 6: Heatmap 2.



Jetzt haben wir aber häufig Situationen wie in Abbildung 6. Hier sehen wir, dass nicht ein Einzelgen differenziell exprimiert wird wie in Abbildung 5. Es gibt in Abbildung 6 kein einzelnes Gen, welches heraussticht wie in Abbildung 5. In Abbildung 6 sticht kein einzelnes Gen an sich stark heraus, aber wir sehen, dass das Expressionsprofil für alle 20 Genen bei den Erkrankten erhöht ist (Viel helles Gelb bei den Erkrankten, was bei der Kontrollgruppe nicht der Fall ist). **Genau das soll ein Globaltest untersuchen!** Es soll sich nicht auf einzelne Gene fokussieren, sondern **er schaut insgesamt, ob das Muster bei den Erkrankten im Vergleich zu den Kontrollindividuen anders ist**. Ein Globaltest ist also wichtig, wenn kein einzelnes Gen für sich differenziell exprimiert ist, sondern wenn viele Gene, die zu einem Pathway beitragen, auch nur ein bisschen differenziell exprimiert werden und damit Unterschiede im Muster zur Kontrollgruppe aufweisen. Sowas kann ein Globaltest detektieren. Auch hier gibt es viele Tests. In der Software-Übung werden wir zwei kennenlernen.

s.18.

Hier habe ich ein Beispiel für ein Globaltest aus der Infektionsforschung. Es gab ja schonmal mehrere Pandemien und nicht unbedingt Pandemien, die den Menschen betreffen. Bei Nutztieren können zum Beispiel Pandemien auftreten. in diesem Beispiel ging es um den H1N1 Influenza Virus. Dieser Virus war 2009 für die Schweinegrippe mit verantwortlich. Dieser Virus führt auch zu einem Befall der Lunge. Die Autoren dieser Studie haben Genset-Tests verwendet, um zum Beispiel alle Gene zu analysieren, die in eigendeiner weise mit der Lunge assoziiert sind. Wir wissen ja bereits, dass Genexpression nicht statisch ist. In jedem Organ exprimieren wir andere Gene und so ist es auch in der Lunge. Wenn eine Erkrankung vorliegt, dann stellt sich natürlich auch die Frage, ob die Symptome auch auf die Genexpression der Lunge zurückzuführen sind. Das heißt, die Autoren haben

sich alle Gene genommen, bei denen man wusste, dass Sie in der Lunge eine Rolle spielen, und haben einen Globaltest gemacht. Sie haben Erkrankte Individuen mit Gesunden verglichen und gesehen, dass man bei den Lungen-Assoziierten Genen ein anderes Genexpressionsprofil beobachten kann zwischen den zwei Gruppen. Auch hier kann man nicht mit dem Auge erkennen, dass wir ein anders Expressionsprofil haben. Der statistische Test kann es aber erkennen. In Abbildung 6 konnte man die Muster klar mit den Augen erkennen, weil ich diese Heatmap selbst konstruiert habe. Wenn Sie es aber selbst irgendwann mal an echten Daten versuchen, werden Sie merken, dass es mit dem Augen erkennen und subjektiven Beurteilung nicht mehr funktioniert. Dafür brauchen wir Tests. Tests können das, was wir mit den Augen in der Heatmap nicht mehr erkennen, detektieren. Durch den Test weiß man dann, dass sich das Gesamtprofil dieser Lungen-assoziierten Gene signifikant verändert hat durch die Infektion.

s.19.

Ich habe den Begriff "competitive Test" schon gebracht. Dem gegenüber nennt man die Globaltests auch oft "**Self-contained Gen-Set Tests**". Self-contained weil Sie quasi nur die Gene mit einbeziehen, die mit einer bestimmten Annotation verknüpft sind, und nicht noch irgendwelche anderen Gene. Competitive Gen-Set-tests hingegen basieren auf der differenziellen Expressionsanalyse und ziehen alle Gene mit ein (d.h. alle Gene –egal, ob Sie zum Set gehören oder nicht). Üblicherweise sind die Competitive Tests deutlich sensitiver. Das heißt, Sie finden viel schneller einen Effekt. Die Globaltests hingegen wendet man dann an, wenn die competitive Tests nichts gefunden haben. Das heißt, die Globaltests können noch eher einen Unterschied aufdecken. Allerdings ist die Gefahr für falsch – positive Ergebnisse bei Globaltests höher.

s.20. Bootstrap-Verfahren.

Abbildung 7: Bootstrap-Formel.

$$p_{bootstrap} = \frac{\#(p^* \leq p)}{B}$$

Damit komme ich nochmal zu einem allgemeinen Verfahren, welches wir bereits bei phylogenetischen Analysen kennengelernt haben. Bei **Gene-Set-Analysen verwendet man sehr häufig Bootstrap-Verfahren**. Bootstrap-Verfahren dienen dazu die Unsicherheit in einer Datenanalyse zu ermitteln. Beim Bootstrap-Verfahren zieht man Stichproben aus der eigentlichen Stichprobe. Wenn ich sage, dass mein Experiment eine Stichprobe ist aus einer größeren Population, dann nutze ich die Daten, die ich aus dieser Stichprobe gezogen habe, und ziehe aus den Daten, die ich schon vorliegen habe, mehrere neue Stichproben. Die Stichprobe zieht also aus sich selbst heraus neue Stichproben. In der phylogenetischen Analyse haben wir es kennengelernt, als man aus dem multiplen Sequenzalignment immer wieder Positionen rausgezogen hat. Daraus hatten wir dann einen Konsensusbaum berechnet, wo markiert wurde, wie oft eine bestimmte Verzweigung gefunden wurde. Bei diesen Gen-Set-Verfahren wendet man das jetzt auch an. Es gibt in Gen-Set-Analysen oft große Unklarheiten, ob die statistischen Verteilungsannahmen erfüllt sind. Wenn Sie nicht erfüllt sind, kann es dazu kommen, dass man verzerrte P-Werte hat. Damit wäre die Analyse nicht vertrauenswürdig. Eine Lösung für dieses Problem besteht darin, dass man die Analyse oft wiederholt, indem man immer wieder Stichproben aus der eigentlichen Stichprobe zieht. Man analysiert also immer wieder Teilmengen. Bleiben wir mal bei GO-Termen. Wenn ich GO-Terme

analysiere, dann kriege ich für jeden Go-Term auch einen P-Wert. Wenn ich jetzt aber pro GO-Term 100 Bootstraps analysiere, dann bekomme ich pro GO-Term 100 P-Werte raus. Schauen wir uns in Abbildung 7 die Formel an. Im Zähler haben wir zuerst **P***. P ohne Sternchen wäre der originale P-Wert für den Go-Term, wenn ich alle Daten analysiere. P* wäre, wenn ich 100 Bootstraps ziehe, 100 P-Werte, die aus Teilmengen der Gesamtdaten entstanden sind. Daraus kann ich einen neuen P-Wert – den **P-bootstrap-Wert**- berechnen. **B** ist die Anzahl an Bootstrap-Samples. In unserem Fall haben wir 100 Bootstraps pro GO-Term gezogen. Deshalb ist B=100. Wir kriegen also auch 100-mal P*. **#** steht für die Anzahl und wir schauen uns an, wie häufig die bootstrap-Samples (P*) kleiner sind als der Originale P-Wert. Die teile ich dann nochmal durch die Anzahl der Samples (B) und das ergibt meinen neuen P-Bootstrap-Wert. Bei dem weiß man, dass er nicht so anfällig ist für unerfüllte Verteilungsannahmen.

S.21-22. Globaltests zur Multi-Omics-Analysen.

Kommen wir zum letzten Punkt der Gene-Set-Analysen. Wir haben ja noch eine eigene Vl zum Thema Multi-Omics-Analysen. Ich will aber hier ein bisschen voraus greifen, weil man Gen-Set-Tests (also Globaltests) auch anwenden kann, wenn man Multi-Omics-Situationen hat. Multi-Omics bedeutet, dass man Informationen auf verschiedenen Omics-Ebenen gewinnen kann (Transkriptom,Genom,Proteom). Das Gute daran ist, dass man so den Gesamtablauf abbilden kann.- also den Weg vom Genom zur RNA und durch die RNA kann ein Protein aufgebaut werden. So kann man das gesamte biologische System verstehen. Andere Omnics-Ebenen bedeutet, dass man zum Beispiel miRNA anschaut. Wir können uns auch Methylierungen als Änderung auf DNA-Ebene anschauen.

s.23.

Ich möchte hier einfach die Situation schildern. Wenn man in einem Experiment auf mRNA-Ebene (Transkriptom) die Genexpression misst (mit Array oder RNA-seq.), dann kriege ich einen Datensatz. Mit diesem Datensatz kann ich die Expression von microRNAs messen. Eine microRNA hat Auswirkung auf die Regulation mehrerer mRNAs. Man spricht dabei von so genannten **Target-Sets**. Man nennt es so, weil eine microRNA die Expression von mehreren Genen oder mRNAs (Target-Sets) inhibieren kann. Aber wie kann ich das jetzt gemeinsam auswerten? Wir haben ja schon ein bisschen Handwerkszeug an der Hand. Wir wissen wie man differenzielle Genexpressionsanalysen vornimmt. Wir kennen auch Globaltests. Fangen wir mal bei der microRNA an. Wir haben 7.000 microRNAs gemessen zwischen Kontrollgruppe und erkrankten Individuen. Jetzt können wir eine ganz einfache differenzielle Genexpressionsanalyse durchführen an den microRNAs. Wir berechnen P-Werte und Log Fold Changes für die microRNAs. Dann wissen wir, welche microRNAs signifikant hoch oder runterreguliert sind. Was machen wir auf mRNA-Ebene? Wenn wir ein Target-Set haben –also ein Set an mRNAs-, dann würden wir einen Globaltest nehmen. Ich kriege also einen P-Wert heraus für irgendeine microRNA und dann weiß ich aus Datenbanken, was die Target-mRNAs der microRNA sind. Ich wende auf diese Target-mRNAs einen Globaltest an. Ich kriege auch hier einen P-Wert heraus und ich verknüpfe dann beide P-Werte.

s.24.

Wenn bestimmte Methylierungsstellen verändert werden, dann wird das Gen möglicherweise nicht mehr abgelesen. Üblicherweise findet man mehrere Methylierungen vor einem Gen und man kann diesen Methylierungsgrad mittlerweile messen über Methyl-Sequenzierungen. Dann weiß ich, welche Methylierungsstellen zu welchem Gen und zu welcher mRNA gehören. Auch hier würde ich über einen Globaltest schauen, ob die Menge an Methylierungsstellen sich unter Erkrankten und

Gesunden unterscheidet. Ich kriege einen P-Wert. Dann mache ich eine differenzielle Genexpressionsanalyse für das Einzelgen. Dann kann ich beide Omnics-Ebenen – Transkriptom und Methylierung- miteinander verbinden.

s.25.

Wir haben gelernt, dass eine mRNA nicht unbedingt für ein Protein steht, sondern aus einem Gen kann durch Splicingvarianten auch mehrere Proteine gemacht werden. Man kann also ein Gen mit mehreren Proteinen assoziieren. Die mRNA kann ich hier über differenzielle Genexpressionsanalysen testen. Die Proteine teste ich über einen Globaltest und kann die Informationen dann verknüpfen.

s.26-27.

Es gibt mehrere Methoden, um die P-Werte zu verknüpfen. Nach der Verknüpfung kriegt man einen neuen P-Wert, der einem sagt, ob es eine signifikante Korrelation zwischen den Omnics-Ebenen gibt und ob es einen Effekt gibt zwischen Erkrankten und Gesunden.