UNIVERSITÄT LEIPZIG

Medizinische Fakultät

Genomweite mRNA-Expressions-profile der in vitro-Adipozyten-differenzierung

Andreas Kühnapfel[1,2], Kathrin Landgraf[1], Holger Kirsten[2], Elena Kempf[1], Markus Scholz[2] & Antje Körner[1]

Motivation

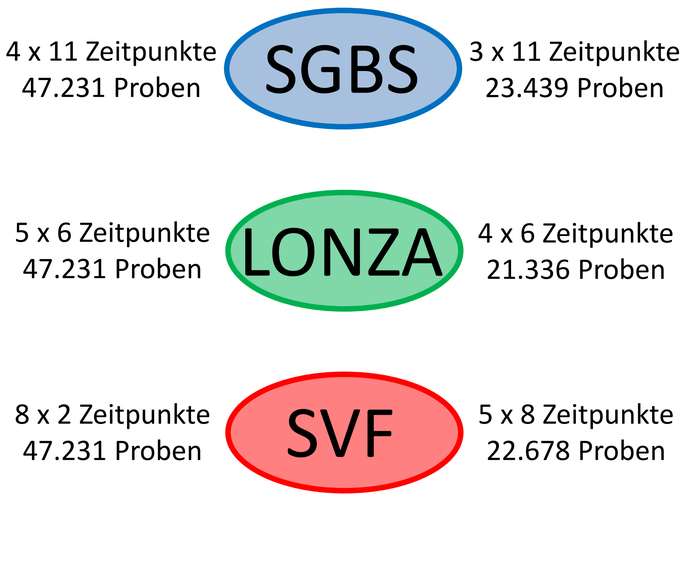
Adipositas entwickelt sich bereits in den ersten Lebensjahren und kann verschiedene metabolische und kardiovaskuläre Beeinträchtigungen nach sich ziehen. Welche Gene sind mit Adipositas und entsprechenden Parametern der Fettgewebsdysfunktion assoziiert und tragen zur Entwicklung klinischer Komorbiditäten bei? Diese Frage soll mittels differentieller Genexpressionsanalysen über den Zeitverlauf der Adipozytendifferenzierung beantwortet werden.

Datengrundlage bilden die Proben aus Simpson-Golabi-Behmel-Syndrom-Zellen („SGBS“, 4 Experimente x 11 Zeitpunkte), isolierte Prä-Adipozyten aus Erwachsenen („LONZA“, 5 x 6 Zeitpunkte) und aus dem Fettgewebe von Kindern isolierte Adipozytenvorläuferzellen („SVF“, 8 x 2 Zeitpunkte). Gemessen wurde jeweils vom Stadium der Prä-Adipozyten über verschiedene Stufen der Adipozytendifferenzierung bis hin zu reifen Adipozyten. Dazu wurden *Illumina HumanHT-12 v4 Expression BeadChip*-Arrays verwendet.

Die Analysen erfolgten nach der Durchführung diverser, standardisierter Präprozessierungsschritte und der Modellierung der Genexpression über den Zeitverlauf der Adipozytendifferenzierung. Sich anschließende Pathway-Analysen beziehen sich auf Gene Ontology und Kyoto Encyclopedia of Genes and Genomes. Damit sollen schließlich Aussagen zum Prozess der menschlichen Adipogenese getätigt werden.

Präprozessierung

Innerhalb der Präprozessierung wurden einerseits die Proben und auf der anderen Seite die Stichproben selbst gefiltert. Für das Filtern der Proben wurden niedrig exprimierte Proben herausgenommen sowie Proben mit einer schlechten Abbildung zum menschlichen Transkriptom. Stichproben wurden bezüglich der Anzahl der detektierten Proben, dem Mahalanobis-Abstand bezüglich diverser Qualitätsmerkmale der Kontroll-Proben und dem euklidischen Abstand innerhalb der Expressionswerte untersucht und gegebenenfalls exkludiert.



Analysen

Modelliert wurde die logarithmierte ($\log_2$) Genexpression in Abhängigkeit von der Zeit als quantitative Einflussgröße. Zur Berücksichtigung der individuellen Expressionsebenen wurde auf die Klasse der gemischten linearen Modelle – konkret auf das sogenannte „Fixed Effect“-Modell – zurückgegriffen. Damit soll primär die Frage beantwortet werden:

Welche Gene
werden über den Zeitverlauf
der Adipozytendifferenzierung
hoch- oder herunterreguliert?

Mathematisch betrachtet man dafür also das folgende Modell:

$$GE_{it} = (\beta_0 + b_i) + \beta_1 * t + \varepsilon_{it}$$

mit GE_{it} als Genexpression für Individuum i zum Zeitpunkt t, β_0 als individuen- & zeitunabhängiges (mittleres) Genexpressionsniveau, b_i als individuenspezifisches & zeitunabhängiges (mittleres) Genexpressionsniveau und β_1 als positiven/negativen Zeiteffekt auf die Genexpression. ε_{it} bezeichnet den Fehlerterm des Regressionsmodells.

Ergebnisse

SGBS

Proben-ID	ø $\log_2$-Genexpression	Zeit	p-Wert (FDR)	Gene
ILMN_1797154	7,18	0,21	2,89E-16	AZGP1
ILMN_2091347	10,77	0,17	2,89E-16	IDH1
ILMN_2326273	8,33	0,39	6,49E-16	CHI3L2
ILMN_2301722	7,32	0,22	6,95E-16	PDE8B
ILMN_1814787	7,26	0,25	6,95E-16	ICA1

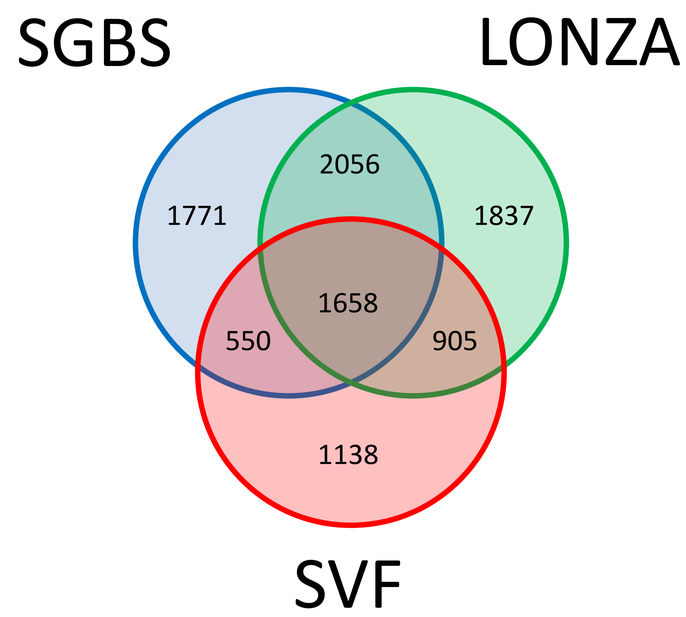
LONZA

Proben-ID	ø $\log_2$-Genexpression	Zeit	p-Wert (FDR)	Gene
ILMN_1651498	6,92	0,13	9,14E-14	GADD45G
ILMN_1740586	7,18	0,20	1,06E-13	PLA2G2A
ILMN_1797154	6,92	0,14	3,00E-13	AZGP1
ILMN_1778240	7,11	0,08	5,23E-13	GFOD1
ILMN_1694268	7,31	0,22	1,07E-12	HES6

SVF

Proben-ID	ø $\log_2$-Genexpression	Zeit	p-Wert (FDR)	Gene
ILMN_2083469	8,42	0,25	2,78E-06	IRS2
ILMN_2305407	7,29	0,17	2,78E-06	ZBTB16
ILMN_1689431	8,30	-0,34	2,94E-06	APCDD1L
ILMN_2411236	8,27	0,37	4,01E-06	NRCAM
ILMN_1722809	7,82	0,28	4,36E-06	NRCAM

Neben den nach Signifikanz geordneten Genlisten der drei Einzelanalysen bezüglich der SGBS-Zellen, der Prä-Adipozyten aus Erwachsenen und der isolierten Adipozytenvorläuferzellen aus Kindern, sind Überschneidungen der Genlisten von Interesse. Dazu wurden alle Konstellationen der als signifikant detektierten Gene der Einzelanalysen bezüglich ihrer Schnittmengen untersucht. Der Schwellwert für Signifikanz lag nach Kontrolle durch die False Discovery Rate bei 0,05.



Bezüglich der Gesamtschnittmenge von 1658 als signifikant detektierten Gene ergibt sich folgende Übersicht:

Gen	p-Wert (SGBS)	p-Wert (LONZA)	p-Wert(SVF)
AZGP1	2,89E-16	3,00E-13	0,011507176
IDH1	2,89E-16	7,93E-07	0,0003
CHI3L2	6,49E-16	3,83E-08	0,0123
PDE8B	6,95E-16	1,58E-06	0,0015
ICA1	6,95E-16	2,19E-08	0,0027
GHR	9,06E-16	2,50E-05	0,0005
OMD	2,12E-15	0,0234	0,0002
FAM89A	2,89E-15	7,23E-07	0,0013
APOD	8,09E-15	0,0007	4,02E-05
ACSS2	1,03E-14	2,52E-07	0,0012

Es wird davon ausgegangen, dass die identifizierten Kandidatengene in Prozesse der Adipozytendifferenzierung, der Lipolyse und des zellulären Energiestoffwechsels involviert sind. Damit könnten die detektierten Gene die Haupteinflussgrößen der Adipozytenhypertrophie darstellen. Dies muss mit Pathway-Analysen überprüft werden, um schließlich Zusammenhänge zur menschlichen Adipogenese aufdecken zu können. Als Ontologien werden dabei GO (Gene Ontology)[1] und KEGG (Kyoto Encyclopedia of Genes and Genomes)[2] verwendet.

[1] The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat Genet 25 (1), 25-9

[2] Kanehisa & Goto (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28 (1), 27-30

Andreas Kühnapfel
Institut für Medizinische Informatik, Statistik und Epidemiologie
Härtelstraße 16-18, 04107 Leipzig
0341 97 16346
andreas.kuehnapfel@imise.uni-leipzig.de

[1] Klinik und Poliklinik für Kinder- und Jugendmedizin, Medizinisches Forschungszentrum, Liebigstraße 21, 04103 Leipzig
[2] Institut für Medizinische Informatik, Statistik und Epidemiologie, Härtelstraße 16-18, 04107 Leipzig