

Genekspressionsanalyse -prøv selv databehandlingen

Af: *Kresten Cæsar Torp, Aalborghus Gymnasium*

Genekspression-analyse (Gene Expression Analysis) er en metode til at sammenligne hvilke gener der udtrykkes under forskellige forhold. Et gens ekspression vil sige, hvor meget det udtrykkes, dvs. hvor meget RNA der dannes fra genet ved transkription.

Man kan fx undersøge:

- Udtrykkes nogle gener mere eller mindre hos personer med en bestemt lidelse, sammenlignet med personer en rask kontrolgruppe?
- Udtrykkes nogle gener mere eller mindre i bestemte væv i kroppen end i andre? Det kan fx være i forskellige dele af hjernen, eller i kræftvæv.
- Sker der en ændring i hvilke gener der udtrykkes, hvis man indtager en bestemt medicin eller udsættes for bestemte giftstoffer?

Undersøgelsen udføres ved at man sammenligner hvordan gener udtrykkes hos mennesker med en bestemt lidelse, sammenlignet med en kontrolgruppe uden lidelsen. Det gør man ved hjælp af en såkaldt microarray-analyse. Herefter undersøger man statistisk, hvilke gener der udtrykkes anderledes. Du kan læse om hvordan metoden og om microchips i filen: *Hvordan laver man en genekspressionsanalyse*.

Ved at undersøge de gener nærmere, som udtrykkes anderledes, kan man fx blive klogere på *årsagssammenhænge* der er mellem bestemte gener og en lidelse eller virkningen af et medicinsk præparat. Den information giver et genekspressionsstudie ikke i sig selv, men de giver spor at arbejde videre med.

1. Forklar, hvorfor man kan undersøge en gens ekspression ud fra at om der dannes RNA ved transkription fra genet.
2. Forklar, hvordan man med en microchip kan undersøge om gener udtrykkes mere i en patientgruppe end end hos en kontrolgruppe.

SRP-tip: Hvordan kan du anvende genekspressionsstudier?

I SRP i naturvidenskabelige fag er det en stor fordel at der indgår både teori og empiri. Empiri vil sige erfaring, og er i naturvidenskab stort set identisk med eksperimentelle data. Du kan få dem fra egne eksperimenter og undersøgelser eller andres. I SRP lægges vægt på metoder. Kan du lave databehandling eller dele af databehandlingen selv viser du både at du kan anvende kompetencer fra dit fag og du viser en metodisk forståelse.

Hvis du fx i SRP arbejder med en lidelse, vil du ofte kunne støde på genekspressionsstudier. Det kan være ved lidelser som alkoholisme, depression, diabetes eller kræft. Du vil også kunne støde på studier af virkningen af bestemte medicinske præparater eller behandlinger. Kan du skaffe data og vise dem på en relevant måde viser du dine kompetencer. Kan du ikke skaffe data fra præcist din problemstilling kan du evt. behandle data du kan skaffe for at demonstrere hvordan metoden fungerer.

I det følgende kan du arbejde dig igennem et datasæt som handler om diabetes.

Hvordan finder man data fra genekspressionsstudier?

Mange data fra genekspressionsstudier er frit tilgængelige på offentlige databaser. En del ligger på Gene Expression Omnibus-databasen (GEO) på NCBI.

Gå ind på databasen: <https://www.ncbi.nlm.nih.gov/geo/>

The screenshot shows the GEO homepage with the following sections:

- Getting Started**
 - Overview
 - FAQ
 - About GEO DataSets
 - About GEO Profiles
 - About GEO2R Analysis
 - How to Construct a Query
 - How to Download Data
- Tools**
 - Search for Studies at GEO DataSets
 - Search for Gene Expression at GEO Profiles
 - Search GEO Documentation
 - Analyze a Study with GEO2R
 - Studies with Genome Data Viewer Tracks
 - Programmatic Access
 - FTP Site
 - ENCODE Data Listings and Tracks
- Browse Content**
 - Repository Browser
 - DataSets: 4348
 - Series: 181426
 - Platforms: 24213
 - Samples: 5214178

Information for Submitters

Vælg "Analyze a study with GEO2R". Her kan man søge et emne, en sygdom eller organisme (fx Homo sapiens), og finde studier frem.

The screenshot shows the GEO2R analysis tool interface with the following elements:

- Header: NCBI, GEO Publications, FAQ, MIAME, Email GEO, Login
- Navigation: NCBI » GEO » GEO2R
- Text: Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full Instructions](#) [YouTube](#)
- Form: GEO accession
- Buttons: GEO2R, Options, Profile graph, R script
- Section: **Quick start**
 - Specify a GEO Series accession and a Platform if prompted.
 - Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
 - Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, source and characteristics) columns to help determine which Samples belong to which group.
 - Click 'Analyze' to perform the calculation with default settings.
 - You may change settings in the Options tab.
- Text: How to use
- Button:

Søg fx studier som beskæftiger sig med insulin:

NCBI DATASET BROWSER CURATED GEO Gene Expression Omnibus

Search for Search Clear Show All Advanced Search Page size 20

222 DataSet records Page 4 of 12

DataSet	Title	Organism(s)	Platform	Series	Samples
GDS4115	Diabetogenic 3A9 T-cell transfer model of type 1 diabetes: pancreatic islets	<i>Mus musculus</i>	GPL6246	GSE26147	7
GDS3850	Peroxisome proliferator-activated receptor- γ ligand treatment: adipose tissue, skeletal muscle, and liver	<i>Rattus norvegicus</i>	GPL341	GSE21329	54
GDS4000	Type 2 diabetic (TALLYHO/JngJ x C57BL6) F2 males: liver, pancreas, muscle and adipose tissue	<i>Mus musculus</i>	GPL1261	GSE24637	64
GDS3882	Type 2 diabetic pancreatic islets of Langerhans	<i>Homo sapiens</i>	GPL96	GSE25724	13
GDS3880	Prolonged exercise training effect on skeletal muscle from type 2 diabetes subjects	<i>Homo sapiens</i>	GPL570	GSE19420	42
GDS3807	Sebacic acid supplemented diet effect on db/db liver	<i>Mus musculus</i>	GPL6247	GSE19877	11
GDS4010	Effect of dietary obesity in fathers on islets of female offspring	<i>Rattus norvegicus</i>	GPL6244	GSE24533	9
GDS4239	Pandemic and seasonal influenza A H1N1 infection of differentiated type I-like alveolar epithelial cells in vitro	<i>Homo sapiens</i>	GPL570	GSE14503	7
GDS4011	Pancreatic islet-like cell clusters derived from T3 embryonic stem cells	<i>Mus musculus</i>	GPL1261	GSE24031	18

DataSet Record GDS3880: Expression Profiles Data Analysis Tools Sample Subsets

Title: Prolonged exercise training effect on skeletal muscle from type 2 diabetes subjects

Summary: Analysis of skeletal muscle biopsies from insulin-treated type 2 diabetes (T2D) subjects before/after exercise training, pre-diabetics and healthy controls. Results provide insight into the molecular basis of T2D, the role of mitochondrial dysfunction, and the effects of prolonged exercise training.

Organism: *Homo sapiens*

Cluster Analysis Download

Du kommer nu til en side, som uddyber det studie datasættet stammer fra. Scroll ned, og tryk på "Analyze with GEO2R"

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19420

Citation(s) Schoenderwoerd K, de Coo IF, Nicolay K, Prompers JJ, Smeets HJ, van Loon LL van Tienen FH, Praet SF, de Feyter HM, van den Broek NM et al. Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes. *J Clin Endocrinol Metab* 2012 Sep;97(9):3261-9. PMID: 22802091

Submission date Dec 10, 2009
Last update date Mar 25, 2019
Contact name Florence van Tienen
E-mail(s) florence.vantienen@gen.unimaas.nl
Organization name Maastricht University
Department Genetics and Cell biology
Street address Universiteitssingel 50
City Maastricht
ZIP/Postal code 6229ER
Country Netherlands

Platforms (1) GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array

Samples (42) GSM482933 BA26_IGT
 GSM482934 BE16_NGT
 GSM482935 BL21_IGT

Relations
 BioProject PRJNA120357

Analyze with GEO2R

Download family Format

SOFT formatted family file(s)	SOFT
MINIML formatted family file(s)	MINIML
Series Matrix File(s)	TXT

Du kommer nu til datasættet, hvor du skal udvælge de data, du skal anvende.

Eksempel: Diabetes, nedsat mitokondriefunktion og motion

Studiet undersøger, hvilken betydning motion har for nedsat mitokondriefunktion i skeletmuskler hos personer med diabetes 2.

Gå ind på datasættet, hvis du ikke har browsset dig frem som angivet ovenfor:

<https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE19420>

NCBI » GEO » GEO2R » GSE19420

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full instructions](#) [YouTube](#)

GEO accession: Set Skeletal muscle mitochondrial dysfunction is secondary to T2DM

▼ Samples Define groups Selected 0 out of 42 samples

Group	Accession	Title	Source name	Group	Age	Bmi
-	GSM482933	BA26_IGT	skeletal muscle vastus lateralis	IGT	67	33.44
-	GSM482934	BE16_NGT	skeletal muscle vastus lateralis	NGT	59	37.22
-	GSM482935	BL21_IGT	skeletal muscle vastus lateralis	IGT	48	36.30
-	GSM482936	CA07_T2DM_T0	skeletal muscle vastus lateralis	T2DM T0	64	29.03
-	GSM482937	CA07_T2DM_T52	skeletal muscle vastus lateralis	T2DM T52	64	29.03
-	GSM482938	DE25_NGT	skeletal muscle vastus lateralis	NGT	57	39.71
-	GSM482939	DO15_IGT	skeletal muscle vastus lateralis	IGT	58	30.16
-	GSM482940	EI03_T2DM_T0	skeletal muscle vastus lateralis	T2DM T0	67	31.23
-	GSM482941	EI03_T2DM_T52	skeletal muscle vastus lateralis	T2DM T52	67	31.23
-	GSM482942	EM04_T2DM_T0	skeletal muscle vastus lateralis	T2DM T0	66	31.00

Du kan læse mere om studiet ved at klikke på den blå titel ved siden af søgefeltet.

I dette tilfælde er der tale om resultater fra et antal forsøgspersoner, der alle er sammenlignelige på parametre som alder, BMO mm., men er fordelt på fire grupper:

- NGT: Kontrolpersoner uden Type 2 sukkersyge (Diabetes mellitus)
- IGT: Personer med begyndende type 2 sukkersyge
- T2DM T0: Personer med type 2 sukkersyge ved forsøgsstart
- T2DM T52: Personer med type 2 sukkersyge efter 52 dages træningsprogram

Først skal vi sammenligne personer med diabetes med kontrolgruppen uden. Vi skal altså have udvalgt data for de to grupper: NGT og T2DM T0.

NCBI » GEO » GEO2R » GSE19420

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full instructions](#) [YouTube](#)

GEO accession: Set Skeletal muscle mitochondrial dysfunction is secondary to T2DM

▼ Samples Define groups Selected 0 out of 42 samples

Enter a group name: List

Enter a group name

Cancel selection

NGT

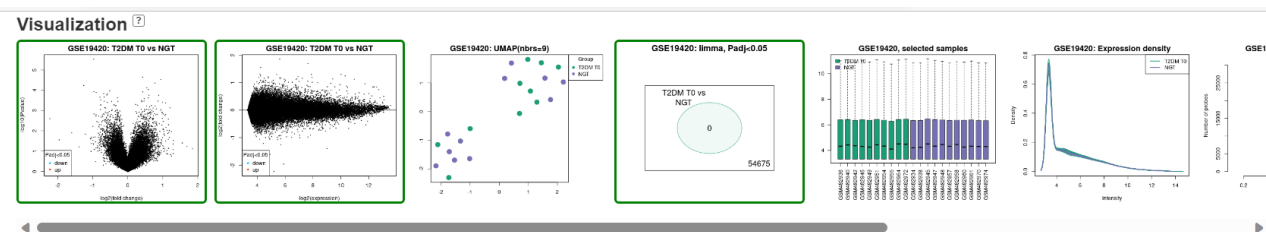
Group	Accession	Title	Source name	Group	Age	Bmi
-	GSM482933	BA26_IGT	skeletal muscle vastus lateralis	IGT	67	33.44
-	GSM482934	BE16_NGT	skeletal muscle vastus lateralis	NGT	59	37.22
-	GSM482935	BL21_IGT	skeletal muscle vastus lateralis	IGT	48	36.30
-	GSM482936	CA07_T2DM_T0	skeletal muscle vastus lateralis	T2DM T0	64	29.03
-	GSM482937	CA07_T2DM_T52	skeletal muscle vastus lateralis	T2DM T52	64	29.03
-	GSM482938	DE25_NGT	skeletal muscle vastus lateralis	NGT	57	39.71
-	GSM482939	DO15_IGT	skeletal muscle vastus lateralis	IGT	58	30.16
-	GSM482940	EI03_T2DM_T0	skeletal muscle vastus lateralis	T2DM T0	67	31.23

Vælg "Define groups". Skriv de to grupper du ønsker ind i rullemenuen. Afslut hver gruppe med "enter".

Marker herefter for hver gruppe, de dataserier (forsøgspersoner) du vil have med. Du kan se, hvilken gruppe de tilhører i kolonne 5. Du kan anvende "ctrl"-knappen til at markere flere serier, og du kan anvende "shift" til at markere flere, der er listede under hinanden. Tryk på gruppen du vil have dem i, scroll længere ned og tilføj flere.

Samples		Define groups		Selected 9 out of 42 samples			
-	GSM482945			skeletal muscle vastus lateralis	NGT	59	31.89
-	GSM482946			skeletal muscle vastus lateralis	T2DM T0	54	35.15
-	GSM482947			skeletal muscle vastus lateralis	NGT	56	27.17
-	GSM482948	GR31_NGT		skeletal muscle vastus lateralis	NGT	49	33.59
-	GSM482949	HE08_T2DM_T0		skeletal muscle vastus lateralis	T2DM T0	66	34.11
-	GSM482950	HE08_T2DM_T52		skeletal muscle vastus lateralis	T2DM T52	66	34.11
-	GSM482951	HO02_T2DM_T0		skeletal muscle vastus lateralis	T2DM T0	51	32.65
-	GSM482952	HO02_T2DM_T52		skeletal muscle vastus lateralis	T2DM T52	51	32.65
-	GSM482953	HO07_IGT		skeletal muscle vastus lateralis	IGT	61	32.01

Scroll ned og tryk på knappen "Analyze". Du kan vende tilbage til at definere dine grupper ved at scrolle lidt op og vælge "Samples".



Top differentially expressed genes

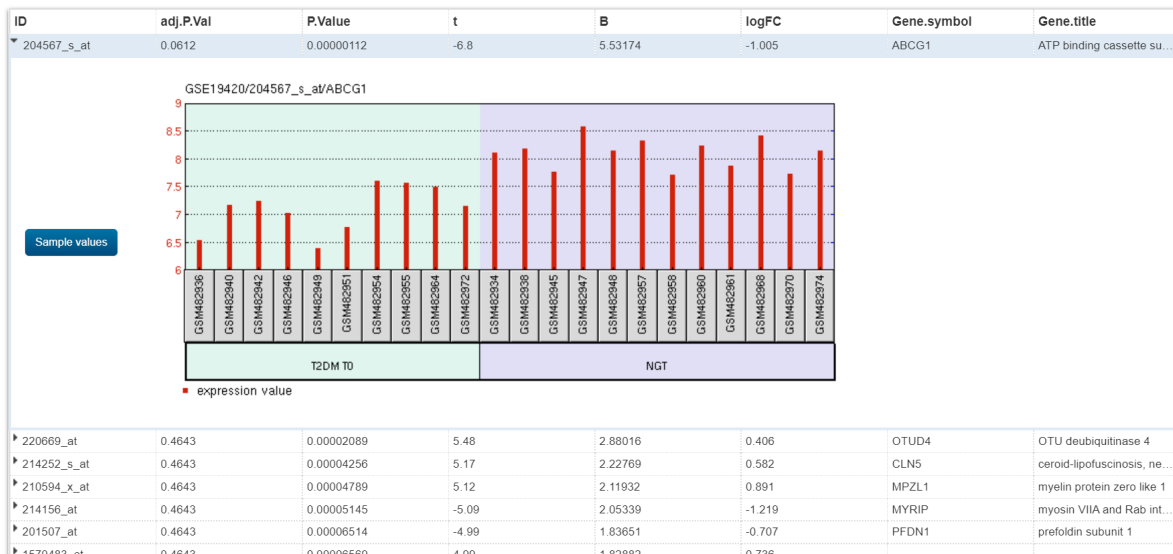
Download full table Select columns

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
204567_s_at	0.162	0.00000296	-6.46	4.6553	-0.975	ABCG1	ATP binding cassette s...
220669_at	0.548	0.00002271	5.52	2.8132	0.42	OTUD4	OTU deubiquitinase 4
210594_x_at	0.548	0.00003009	5.4	2.5565	0.939	MPZL1	myelin protein zero like 1
1570483_at	0.606	0.0000443	5.22	2.203	0.775		
214156_at	0.61	0.00005864	-5.1	1.9464	-1.258	MYRIP	myosin VIIA and Rab i...
214252_s_at	0.61	0.00007554	4.99	1.7142	0.542	CLN5	ceroid-lipofuscinosis, n...
212082_s_at	0.61	0.00009214	4.9	1.5319	0.588	MYL6	myosin light chain 6
225557_at	0.61	0.00010875	-4.83	1.3798	-0.52	CSRNP1	cysteine and serine ric...
202308_at	0.61	0.00011702	-4.79	1.3125	-0.971	SREBF1	sterol regulatory eleme...
204150_at	0.61	0.00012083	4.78	1.283	0.425	STAB1	stabilin 1
206303_s_at	0.61	0.00012972	-4.75	1.2178	-0.905	NUDT4P1//NUDT4P2...	nudix hydrolase 4 pse...
201507_at	0.61	0.00014672	-4.7	1.1046	-0.69	PFDN1	prefoldin subunit 1

I tabellen får du nu en liste over gener, som er over- eller underudtrykt hos gruppen af diabetikere, sammenlignet med kontrolgruppen.

- *t* angiver hvor meget middelværdien for de to datasæt afviger fra hinanden, set i forhold til spredningen for de to datasæt.
- *logFC* er logaritmen til hvor mange gange (fold) genet udtrykkes hyppigere hos den ene gruppe sammenlignet med den anden gruppe.
- *Gene.symbol* og *Gene.title* angiver hvilket gen der er tale om. *Gene.symbol* kan man benytte til at søge mere viden om hvilke gener der er påvirkede, og hvilke biokemiske funktioner de har. Gentitel kan i nogle tilfælde indeholde navne på proteiner, du kender fra undervisningen.

Du kan trykke på hvert gen for at se grafisk, hvor meget det er over- eller underudtrykt hos forsøgspersonerne:



Du kan også anvende diagrammerne for oven til at anskueliggøre resultaterne. Du kender sikkert allerede boxplots. Har du matematik eller informatik kan fx UMAP også være en måde at vise data, som du kan uddybe baggrunden for.



Prøv nu tilsvarende at sammenligne hvad der er sket med gruppen af diabetikere fra studiets dag 0 til dag 52, hvor de havde gennemført træningsprogrammet.

1. Hvilke gener udtrykkes anderledes efter træningsprogrammet? Fx ved at de ikke er i top 10 længere, eller at de er kommet i top 10.
2. Undersøg disse nærmere. Hvad koder de for i cellen?

3. Du kan også for sammenligningens skyld sammenligne diabetikere der har gennemført træningen med kontrolgruppen.

Data fra andre studier

- Genomisk associationsstudie for alkoholisme (AUD): [Multi-omics integration analysis identifies novel genes for alcoholism with potential overlap with neurodegenerative diseases \(nih.gov\)](#)
- Datasæt: [GEO Accession viewer \(nih.gov\)](#)
- Ændring i genekspression hos rotter ved alkoholpåvirkning: [Alcohol Dependence in Rats Is Associated with Global Changes in Gene Expression in the Central Amygdala \(nih.gov\)](#)
- Datasæt: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159136>