

Project Proposal

Mark van de Streek

2023-02-24

Contents

Gen expressie van leukocyten bij mensen die roken.	1
Het onderzoek	1
De groep verdeling	2
Plan voor mijn eigen onderzoek	2
Software	2
Data	3

Gen expressie van leukocyten bij mensen die roken.

Het onderzoek

In het onderzoek is er gekeken naar onder meer gen expressie van leukocyten bij verschillende groepen mensen. De groepen mensen zijn:

- Vapers (mensen die een elektronische sigaret roken)
- Mensen die de traditionele sigaret roken
- Mensen die geen tabaksproducten gebruiken

De onderzoekers hebben vervolgens een aantal modelleringen uitgevoerd op de ruwe data die verkregen was van de personen. De modellen werden allemaal gemaakt in Rstudio met behulp van een aantal pakketten.

Om mensen te verkrijgen voor het onderzoek werd er om te beginnen een online vragenlijst gemaakt. Voor deze vragenlijst werd online reclame gemaakt op bijvoorbeeld, Reddit en Twitter. Aan de hand van de deze vragenlijst werden de mensen die geschikt leken voor het onderzoek uitgenodigd voor een gesprek. In dit gesprek werden nog een aantal vragen en checklists gemaakt om er zeker van te zijn dat de persoon geschikt is voor het onderzoek.

Uit de resultaten bleek dat de mensen die op dit moment Vapen, maar vroeger niet rookten, een significant andere genregulatie hebben ten opzichte van mensen uit de controle groep. Ook mitochondriale genen en immuunresponsogenen waren significant ontregeld.

Klik hier om het artikel te lezen die bij het onderzoek hoort.

De groep verdeling

In alle verschillende groepen waren een aantal mensen die voor het onderzoek werden gebruikt. Onderstaand de aantallen in een kleine tabel.

Table 1: Aantal mensen per groep

Vapers	Traditionele rokers	Niet-rokers
37	22	23

Er werd bij de deelnemers ook gekeken naar geslacht, leeftijd, etniciteit, rassen en ook naar mensen die kunnen lezen en schrijven in het Engels. Het onderzoek vond plaats in Los Angeles.

De opzet van het onderzoek was om de verschillen tussen alle drie de groepen duidelijk weer te geven. In vele eerdere onderzoeken werden vaak deelnemers gekozen die vaperen, maar vaak met een geschiedenis in andere rookgewoonten. Er namen dus mensen deel aan het onderzoek die eerder traditionele sigaretten rookten en op latere tijd zijn overgestapt naar elektronische sigaretten.

De groep die werd geclassificeerd als vapers in dit onderzoek heeft in de 6 maanden voor het onderzoek geen traditionele sigaret gebruikt. De onderzoekers hebben bewust voor een periode van ‘maar’ 6 maanden gekozen, omdat de elektronische sigaret nog niet zo heel lang bestaat. Door het kiezen van een periode van 6 maanden, zijn er meer mensen die deel konden nemen aan het onderzoek. Voor de traditionele rokers groep werd een periode van 1 jaar gebruikt voor deelname.

Met behulp van bioinformatische bevindingen zijn in dit onderzoek onderscheid gemaakt tussen gezonde volwassen vapers, met en zonder geschiedenis van roken, en ‘exclusieve’ sigaretten rokers. Ook is specifiek de expressie van genen in rokers (zowel e-sigaret als traditionele rokers) vergeleken met niet-rokers.

Plan voor mijn eigen onderzoek

Een bijzonder resultaat van het onderzoek wat ik eigenlijk helemaal niet wist is de beschadiging van de mitochondriën. De meeste DEG's werden gevonden in mitochondriale genen, die zijn sterk geassocieerd met mitochondriale aandoeningen. Zowel vaperen als roken is gekoppeld aan een verhoogde depolarisatie en beschadiging aan de mitochondriën. Omdat deze stap mij opviel, wil ik hier nog nader naar kijken. Op de afbeelding die zichtbaar is bij de resultaten van het onderzoek, staat een klein staafdiagram waar de verhoogde kans op aandoeningen worden vergeleken tussen vapers en rokers. Het verschil is niet extreem groot tussen beide groepen, maar bij traditionele rokers is het vaker wel (iets) groter. Dit figuur wil ik namaken om te kijken of ik het zelfde resultaat krijg met mijn anders bewerkte data.

Graag zou ik nog willen kijken of ik de genen van niek-rokers kan meenemen in de figuren, zodat het echt duidelijk wordt hoeveel genen nou ‘echt’ bijdragen aan de aandoeningen voor rokers. Uit de resultaten bleek dat het aantal DEG's bij traditionele rokers ongeveer ~ 7.4 keer hoger was dan bij vapers. Ik wil ook graag iets dieper kijken naar de DEG' in vapers, want in vele onderzoeken is er al gekeken naar normale rokers. Uiteraard is het nog niet zeker of dit überhaupt wel kan met de data die ik heb. Als ik een extra figuur wil maken, moet het natuurlijk wel echt iets toelichten.

Software

Het trimmen, alignen en kwantificeren van de RNA-seq dat is uitgevoerd met **Partek Flow**, version 8.0.19.1027. voor het alignen werd referentie genoom hg38 gebruikt. Alle post-alignment processen, het kwantificeren en alle analysis zijn uitgevoerd met **R**. Voor het transformeren van RNA-seq dat werd **LimmaVoom** gebruikt. LimmaVoom is een pakket in R.

Data

In de data zit de volgende informatie: chromosome, start- en stop-positie, streng, gen symbool/gen naam, havana gen en natuurlijk de samples. Van elke sample is dus meer informatie op te zoeken dan simpelweg alleen de telwaarde. De havana gen kolom is een kolom die unieke waarden bevat die verwijzen naar een specifiek gen, dat op een specifieke locatie op het menselijk genoom ligt. De waarden zijn geannoteerd door het HAVANA-project, om genen, exonen, etc. op het menselijk genoom te identificeren.

Onderstaand een klein overzicht van de eerste paar regels van de data.

	Chromosome	Start	Stop	Strand	Gene.Symbol	gene_name	gene_type	havana_gene	BC.165_S3	BC.171_S7	BC.157_S28
1	4	67439992	67440119	+	5S_rRNA	5S_rRNA	rRNA	---	2.00000e+00	2.60193e+00	5.00000e+00
2	6	52995620	52995951	+	7SK	7SK	misc_RNA	---	4.66046e+02	7.38235e+01	3.57325e+02
3	19	58345178	58353500	-	A1BG	A1BG	protein_coding	OTTHUMG00000183507.2	3.24430e+02	9.64965e+01	2.58853e+02
4	19	58347751	58355184	+	A1BG-AS1	A1BG-AS1	antisense_RNA	OTTHUMG00000183508.1	2.86521e+02	8.19209e+01	1.81269e+02
5	10	50799409	50885676	-	A1CF	A1CF	protein_coding	OTTHUMG0000018240.5	0.00000e+00	0.00000e+00	0.00000e+00
6	12	9067664	9116230	-	A2M	A2M	protein_coding	OTTHUMG00000150267.7	1.33380e+02	4.10000e+01	1.29748e+02
7	12	9065177	9068061	+	A2M-AS1	A2M-AS1	antisense_RNA	OTTHUMG00000168289.2	2.26827e+01	0.00000e+00	1.11026e+01
8	12	8822472	8887002	+	A2ML1	A2ML1	protein_coding	OTTHUMG00000128499.8	1.00000e+00	7.00000e+00	1.20000e+01
9	12	8776219	8830948	-	A2ML1-AS1	A2ML1-AS1	antisense_RNA	OTTHUMG00000168273.1	4.04119e-36	0.00000e+00	0.00000e+00
10	12	9228533	9275818	-	A2MP1	A2MP1	transcribed_unprocessed_pseudogene	OTTHUMG00000168334.3	5.34553e+01	1.00000e+00	2.91391e+01
11	1	33306766	33321099	-	A3GALT2	A3GALT2	protein_coding	OTTHUMG00000004125.4	3.00000e+00	0.00000e+00	2.00000e+00
12	22	42692121	42721299	-	A4GALT	A4GALT	protein_coding	OTTHUMG00000150744.2	0.00000e+00	5.00000e+00	6.00000e+00
13	3	138123718	138132388	-	A4GNT	A4GNT	protein_coding	OTTHUMG00000159820.1	0.00000e+00	0.00000e+00	0.00000e+00
14	12	53307456	53324865	-	AAAS	AAAS	protein_coding	OTTHUMG00000169729.4	1.67300e+03	4.55000e+02	1.34700e+03
15	12	125065379	125143334	+	AACS	AACS	protein_coding	OTTHUMG00000168550.3	5.92648e+02	1.40000e+02	4.44591e+02
16	4	170060222	170091700	-	AADAT	AADAT	protein_coding	OTTHUMG00000160912.4	0.00000e+00	0.00000e+00	0.00000e+00
17	9	96639577	96655304	-	AAED1	AAED1	protein_coding	OTTHUMG00000020299.1	2.53274e+02	3.00000e+01	1.49406e+02
18	15	67201033	67255196	-	AAGAB	AAGAB	protein_coding	OTTHUMG00000172246.3	6.82997e+02	7.90000e+01	5.28999e+02
19	2	69457997	69674350	-	AAK1	AAK1	protein_coding	OTTHUMG00000129648.9	8.29397e+03	6.41710e+03	6.54465e+03
20	11	77821109	77918433	+	AAMDC	AAMDC	protein_coding	OTTHUMG00000166651.2	7.05037e+01	2.41169e+01	7.10807e+01
21	2	218264123	218270258	-	AAMP	AAMP	protein_coding	OTTHUMG00000155202.3	2.76099e+03	5.15000e+02	2.18398e+03
22	17	76453351	76470118	+	AANAT	AANAT	protein_coding	OTTHUMG00000180179.2	2.90000e+01	1.60000e+01	3.90000e+01
23	20	36236459	36270919	+	AAR2	AAR2	protein_coding	OTTHUMG00000032380.2	1.13300e+03	8.00000e+01	8.22996e+02
24	8	116938199	116944488	+	AARD	AARD	protein_coding	OTTHUMG00000164961.1	3.00000e+00	0.00000e+00	1.60000e+01

Figure 1: Overzicht van de data