

Invloed van leeftijd op gen expressie van Ebf1 in muizen

F.J.A. Menninga ¹, I. Lottman ¹ S, Steller ¹, J.J Duiker ¹ | ¹Bio_informatica Hanzehogeschool

Samenvatting

In dit onderzoek wordt de invloed van leeftijd op de genexpressie van Ebf1 (Early B-Cell Factor 1) in muizen onderzocht. Ebf1 speelt een cruciale rol bij de celdifferentiatie. Verwacht wordt dat leeftijd een significante invloed heeft op de genexpressie. Om dit te onderzoeken, wordt het DNA van jonge en oude muizen geanalyseerd om te bepalen hoe de expressie van Ebf1 tussen beide groepen verschilt.

0.1 Setup *Chunk*

Aan het begin van je paper kan je een zogenaamde *setup chunk* toevoegen. Hierin kan je het gedrag van `knitr` configureren en de bibliotheken laden die je in je code gebruikt. Hieronder is een voorbeeld van zo'n *setup chunk*.

NB Gebruik altijd een naam voor iedere chunk. Dat maakt het *debuggen* van problemen bij het *knitten* van je RMarkdown document veel gemakkelijker.

1 Introductie

Je start altijd met de maatschappelijke drijfveer voor je onderzoek; waarom is het van belang dat dit onderzoek is uitgevoerd? Vervolgens bespreek je de achtergronden van je onderzoek. Wat is er al eerder onderzocht op dit vlak en wat mist er juist nog; welke speciale technieken heb je gebruikt? Refereer zorgvuldig naar bronnen die je hebt gebruikt - zie ook paragraaf ??.

Je eindigt de introductie met de doelstelling(en) van jouw onderzoek, en hoe je deze doelstellingen denkt te gaan verwezenlijken; de aanpak. Geef hier ook eventuele hypothesen.

De introductie bevat tussen de 400-800 woorden.

2 Materialen en Methoden

2.1 Materialen.

Tool	Referentie	Versie	Waarom
Featurecounts	https://academic.oup.com/bioinformatics/article/30/7/923/232889?searchresult=1		Featurecounts is een zeer efficiënt algemeen “read” samenvattingsprogramma dat mapped reads telt voor genomische kenmerken zoals genen, exonen, promotor, genlichamen, genomische bins en chromosomale locaties. het kan worden gebruikt om zowel RNA-seq als genomische DNA-seq leesbewerkingen te tellen
HISAT2	https://daehwankimlab.github.io/hisat2/		HISAT2 is een snelle en gevoelige aligner voor mapping NGS reads voor zowel DNA als RNA naar een enkele referentie genoom.
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/		FastQC wordt gebruikt om de kwaliteit te checken van de rauwe data, hier uit is te zien of de data gelijk te gebruiken is of dat deze moet worden getrimmed. De trimmer kan ook afgesteld worden op basis van de fastqc.
Fastqcr	https://rpkg.sdatanovia.com/fastqcr/index.html		
freebayes	https://github.com/freebayes/freebayes		freebayes is een haplotype gebaseerde gen variant detector, ontworpen om kleine polymorfismen te detecteren, SNP's, inserties en deleties in het bijzonder. Dit programma gebruikt .BAM bestanden met een Phred+33 encoding.
seqtk	https://github.com/lh3/seqtk		seqtk wordt gebruikt voor het proceseren van sequences in het FASTA of FASTQ formaat. het “seamlessly parses” beide FASTA en FASTQ welke dan ook optimaal compressed wordt door gzip
Trimmomatic	https://github.com/usadellab/Trimmomatic		Trimmomatic wordt gebruikt om de data op te schonen nadat deze uit FastQC komt. Deze haalt de slechte kwaliteit paren af van de streng waardoor een hoge kwaliteit RNA- of DNA-streng overblijft die gebruikt kan worden.

Tool	Referentie	Versie	Waarom
bwa mem2	https://github.com/bwa-mem2/bwa-mem2?tab=readme-ov-file		Bwa mem2 wordt gebruikt om DNA en RNA reads te alignen tegen een gekozen referentie genoom.
samtools	https://www.htslib.org		samtools is een set van “utilities” dat alignments in de SAM, Bam en CRAM formatten kan manipuleren. het kan veranderen tussen de formats, sorteren, samenvoegen en indexen, ook kan het “reads” snel vinden in elke regio
R	https://www.r-project.org		R is de code taal die gebruikt wordt om alle statistieken testen te doen en tevens de visualisatie van de data die komt uit het onderzoek
R-studio	https://posit.co		R studio is het programma wat wordt gebruikt als IDE voor R
NCBI-GEO	https://www.ncbi.nlm.nih.gov/geo/		NCBI-GEO is gebruikt om het originele onderzoek te vinden waar dit onderzoek inspiratie vanaf neemt

Materialen en Methoden beschrijft verhalend wat je hebt gebruikt (data, tools) en wat je hebt gedaan (ontwikkelde methodes). ***Het is essentieel dat dit hoofdstuk je onderzoek reproduceerbaar en valideerbaar maakt.*** Verwijs bij de start van dit hoofdstuk naar je code repository (meestal je github link).

Neem referenties op!

De materialen zijn de meetinstrumenten die je hebt gebruikt, maar ook bijvoorbeeld datasets die je hebt gedownload.

Beschrijf de gebruikte software tools, alsmede de bibliotheken/plugins, met naam, versie, referentie en gebruiksdoel (in dit project). Dit kan eventueel in een tabel an als die lang is mag het een online bijlage zijn.

Beschrijf bestaande methodologieën met hun relevantie voor je project. Geef het doel, de toepassing en welke software en parameters er zijn gebruik. Voeg eventueel een flowchart toe. Beschrijf de gebruikte statistische methoden.

Beschrijf wat je zelf in het kader van dit onderzoek hebt ontwikkeld aan methodologieën. Geef de naam van scripts/programma's en waar deze (in je repo) te vinden zijn.

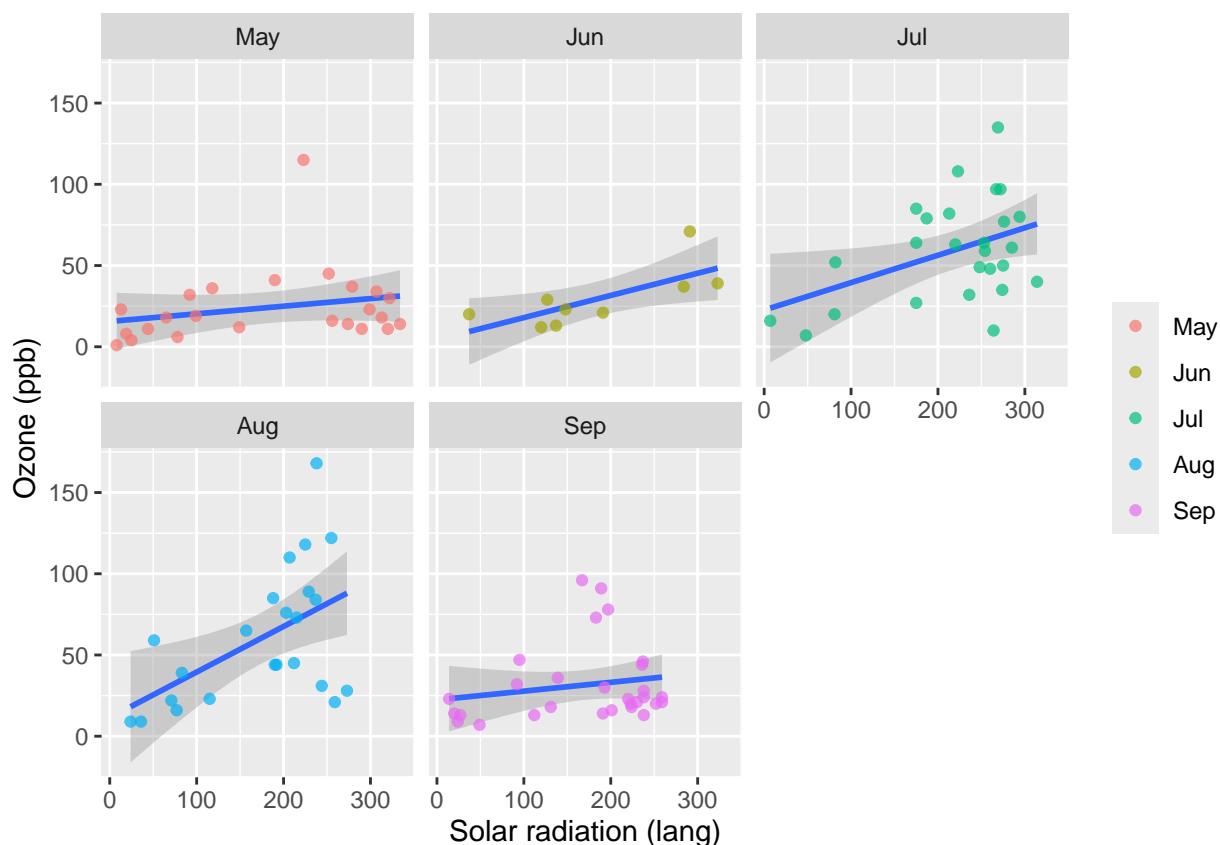
Deze sectie bevat tussen de 400-800 woorden.

3 Resultaten

Presenteer je resultaten in een logische volgorde. Beschrijf wat er te zien is en werk toe naar het beantwoorden van je doelstelling. Beschrijf indien mogelijk de logica van de keuze voor opeenvolgende experimenten. In elke paragraaf worden de feitelijke conclusies gegeven (bv ‘de vergelijking laat zien dat het gemiddelde van groep A significant afwijkt van het gemiddelde van groep B, met een p-waarde van ... Dit komt niet overeen met de in de literatuur (REF) beschreven waarden.’)

Gebruik zo veel mogelijk figuren om je informatie uit je resultaten over te brengen. Gebruik tabellen wanneer figuren minder geschikt zijn. In de tekst worden figuren en tabellen geïntroduceerd, besproken en de belangrijkste aspecten toegelicht.

Voorzie je figuren van een nummer en een beschrijvende titel. Zorg voor correcte as-labels (eenheid en grootte), legenda en bijschrift. Hier is een voorbeeldje.



Figuur 1: Ozon concentraties geplot tegen zonlicht intensiteit, vergeleken over verschillende maanden. Lineaire modellen (blauwe lijnen) en confidence intervals (grijze gebieden) zijn hieraan toegevoegd.

Geef tabellen bovenaan een titel en bijschrift die de inhoud beschrijft en onderaan voetnoten die kolomnamen of specifieke waarden verklaren.

Deze sectie bevat tussen de 600-1200 woorden.

4 Discussie en Conclusies

Formuleer je conclusie door eerst in te zoomen op je eigen data en daarna uit te zoomen. Zoom in door je resultaten samen te vatten. Zoom uit om de waarde van je werk te beoordelen, door je bijvoorbeeld de volgende vragen te stellen:

- Kunnen mijn resultaten gebruikt worden in het werkveld?
- Wat betekenen ze voor het werkveld?
- Zijn mijn data betrouwbaar?

Bespreek de resultaten zodanig dat je ze ter discussie stelt, wees kritisch. Vergelijk je resultaten met de literatuur of eerder ontwikkelde data. Geef aanbevelingen voor een vervolg en staaf je aanbevelingen door de impact op wetenschappelijk of maatschappelijk vlak te beschrijven.

Kom ten slotte altijd terug op de doelstelling (en hypothesen).

Deze sectie bevat tussen de 400-800 woorden.

5 Online bijlagen

Vaak zijn online bijlagen vele malen groter dan het eigenlijke artikel. Wees nooit bang om te veel aan bijlagen aan te bieden. Je kan hierbij denken aan

- de ruwe data
- de code voor dataverwerking
- de code voor analyse
- aanvullende figuren en tabellen

Natuurlijk is een git(hub) repo daar de beste plek voor! Zorg ervoor de je repo logisch is ingericht met goede Readme document(en). Ook de code zelf is waar nodig natuurlijk goed gedocumenteerd.

5.1 Wordcount

Voeg aan het eind een woord-telling in:

Method	koRpus	stringi
Word count	1129	1088
Character count	6972	7306
Sentence count	84	Not available
Reading time	5.6 minutes	5.4 minutes

6 Referenties

Een lijst van referenties wordt hier automagisch toegevoegd.