## Statistical and Computational Methods in Neuroscience

Qiongling Peng Dongliang Chen Zijian Liang 2024-11-28



## Content

From Statistical Approaches to Data-Driven Neuroscience ------ Qiongling Peng

• Exploring Omics Data in Neuroscience: Techniques and Statistical Tools ------ Dongliang Chen

## From Statistical Approaches to Data-Driven Neuroscience

Qiongling Peng 2024-11-28

# What is Statistics ?

Statistics is a body of methods for making wise decisions in the face of uncertainty (Wallis & Roberts, 1962)

Statistics is a set of concepts, rules, and methods for (1) collecting data, (2) analyzing data, and (3) drawing conclusions from data (Iversen & Gergen 1997)

Statistics is a way of reasoning, along with a collection of tools and methods, designed to help us understand the world (De Veaux et al. 2006)



## The development of Statistics



**Bills of Mortality** 





Ronald Fisher proposed Analysis of Variance (ANOVA) and Maximum Likelihood Estimation



The integration of statistics and computational science has advanced into the fields of machine learning and neuroscience

21 Century

20 Century

18-19 Century

1662 John Graunt

### The Early Applications of Statistics in Neuroscience

#### RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX

BY D. H. HUBEL AND T. N. WIESEL From the Neurophysiolo jy Laboratory, Department of Pharmacology Harvard Medical School, Boston, Massachusetts, U.S.A.

(Received 31 July 1961)

What chiefly distinguishes cerebral cortex from other parts of the central nervous system is the great diversity of its cell types and interconnexions. It would be astonishing if such a structure did not profoundly modify the response patterns of fibres coming into it. In the cat's visual cortex, the receptive field arrangements of single cells suggest that there is indeed a degree of complexity far exceeding anything yet seen at lower levels in the visual system.











Using spike frequency quantification and analysis of variance (ANOVA), the receptive field structure of visual cortical neurons, including the spatial distribution of excitatory and inhibitory regions, was successfully revealed.

Hubel DH, Wiesel TN, J Physiol, 1959

From t-test to Mixed-Effects Models: The Evolution to Address the Complexity of Neuroscience Data



### **Intraclass Correlation Coefficient**

**(ICC):** Measures the degree of similarity or correlation between observations within the same group (e.g., neurons within the same animal).



Yu Z et al., *Neuron*, 2022



Table 2. p values for comparing pCREB immunoreactivity at eachtime point (24 h, 48 h, 72 h, and 1 week) after ketamine treatmentto the baseline (saline)

	Overall	24 h	48 h	72 h	1 week	
.inear model ANOVA)	1.2 × 10 <sup>-78</sup>	6.0 × 10 <sup>-38</sup>	6.8 × 10 <sup>-26</sup>	0.0291	1.1 × 10 <sup>-8</sup>	
ME	0.0029	0.0049	0.0164	0.5601	0.2525	

Type I Error: Overestimation of significance

Linear Model (LM): Ignores correlation

Linear Mixed-Effects Model (LME): Models within-group correlation

Yu Z et al., Neuron, 2022



How to Choose the Right Statistical Model for Complex Data

• Linear Models (LM): For simple data without clustering.

- ANOVA: Comparing group means with independent samples.
- LME: Handling hierarchical or nested data structures.
- **GLM/GLMM:** For non-continuous outcome variables.
- Random Effects: Capturing variability from nested structures like animals, cells, or cages.
- Key Decision Points:
- Is the outcome variable continuous?
- Are there nested or clustered structures?
- Are there repeated measures per subject or cell?

From Regression to Prediction: Bridging Descriptive Modeling and Predictive Power in Neuroscience



Bzdok D, Ioannidis JPA, *Trends Neurosci*, 2019



## Regression [ri-'gre-shən]

A statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).



Wilcox RR, Rousselet GA, Curr Protoc, 2023



The blessing and curse of dimensionality

Huys QJ et al., Nat Neurosci, 2016



This ROC curve compares the performance of classifiers trained on raw behavioral data versus derived parameters. The use of parameters derived from a mechanistic model significantly improves classification accuracy, as shown by the higher AUC for the parameter-based classifier.

Huys QJ et al., Nat Neurosci, 2016

Focused and Simplified: Feature Selection and Dimensionality Reduction in Neuroscience



Gomez-Marin A et al., Nat Neurosci, 2014



Dimitriadis SI et al., Neural Regen Res, 2018



Cunningham JP, Yu BM, Nat Neurosci, 2014



Cunningham JP, Yu BM, Nat Neurosci, 2014

## Take home message

- From Comparison of Differences to the Evolution of Predictive Models: Traditional statistical methods are suitable for exploring relationships between variables, while modern machine learning methods (such as LASSO and Random Forest) are better equipped to handle high-dimensional and complex data.
- Dimensionality Reduction is Key for High-Dimensional Neural Data Analysis: Through methods like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), we can simplify data structures and extract the most meaningful information.
- Feature Selection Enhances Model Performance: Algorithms like Random Forest not only classify data effectively but also identify critical features, providing more precise inputs for models.

## Exploring Omics Data in Neuroscience: Techniques and Statistical Tools

20241128

CDL

## Introduction : Why Omics Data?





### Omics analysis enhance the persuasiveness of our experiments



Veiner, Marcell et al. *Molecular ecology resources*,2022

## How to get genomic data?

## The process of obtaining genomic data



## The principle of Illumina sequencing process.



Mardis, Elaine R. Annual review of genomics and human genetics, 2008

The principle of PacBio sequencing process.



Ardui, Simon et al. Nucleic acids research ,2018

### Video introduction to PacBio sequencing



## The principle of Nanopore sequencing process.

## **Oxford Nanopore Sequencing Steps**



Video introduction to Nanopore sequencing



## The comparison of different generation of gene sequencing.

 Table 1. The comparison of different generation of gene sequencing.

	Reading Length (kb) N50	Estimated Cost per Gb (US \$)	Throughput per Flow Cell (Gb)	Read Accuracy (%)
Sanger(1st)	<1 kb	13,000 d	/	>99.9
Illumina(2nd)	0.075–0.15 <sup>a</sup>	50–63	16–30	>99.9
PacBio(3rd)	10–20 <sup>b</sup>	43–86	15–30	>99
ONT(4th)	10–60 <sup>c</sup>	21–42	50–100	87–98

X

AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC	CTCAGGACA/AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATACCTCAGGACA
+	+
AA- <fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<< td=""><td>:J-F&lt;7-A·AA-<fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<j-f<7-a< td=""></fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<j-f<7-a<></td></fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<<>	:J-F<7-A·AA- <fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<j-f<7-a< td=""></fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<j-f<7-a<>
ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA	@ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA
TATTTCAGTATTTAGTAGATATATTTTAAGAGGGGTATCCCTCTTTTTATTATCAATACTTATAATTGCAA	AGGGAATTTG:CTATTTCAGTATTTAGTAGATATATTTTAAGAGGGGTATCCCTCTTTTTATTATCAATACTTATAATTGCAAGGGAATTTG
•	+
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjj-<f-77fj< td=""><td>JFJ7&lt;<a\a<afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></a\a<afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a<></td></afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjj-<f-77fj<>	JFJ7< <a\a<afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></a\a<afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a<>
ST-E00310:147:HVT25CCXX:3:1101:14306:1872 1:N:0:TGAAGAGA	@ST-E00310:147:HVT25CCXX:3:1101:14306:1872 1:N:0:TGAAGAGA
IGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGTTTATCATCTACTGCCTTGACTGTATATGCAACC	TATAGAAC(TGTTCCGTCAGCTCAGAAGAGGTTTGATACTGCTAGAGGTTTATCATCTACTGGCTTGACTGTATATGCAACCATATAGAAC
	+
' \_&FF.T.T.T.&.T.TF&.T.T.F.TF.T.T.T_F&7_F&.TF.T&.TFFFFF.TF.T.T&F.T.T.T.T.T.T.T_<&7_<7F<7.T.T.TF_<<&	' 1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.
ST_F00310.147.HVT25CCXX.3.1101.6756.1907 1.N.O.TGAAGAGA	ARCH/ICOAR AITOCOCCOCICOCOCO IN/ INCICACIII IICICORICOCOCOC (A/ C/IC/COOI (CARACA/ICOA
TTACTTCCAAAGCTTTACCTTTATTCCGTCCAAAGTTCATTCCACCACCACCACCACCACCACCACCACCACCAC	GITETTETTETTETTECTA A GETTTA GETTTA TTECETCEA A GETTEA TTECA CEA CEA CEA CEA CEA CEA CEA CEA CEA
JIIACIIIGIICCAAAGCIIIAGCIIIAIICCGICCAAAGIICAIIGCACCACGIGCICCACCACCACCIG	STICTICAL STRUCTURE CARAGETTIAS CITIAS CONTRACTOR CONTRACT
- ST-E00310:147:HVT25CCXX:3:1101:13382:1819_1:N:0:TGAAGACA	т @ST-E00310:147:HVT25CCXX:3:1101:13382:1819_1:N:0:ТGAAGACA
AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC	CTCAGGACAJAGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAAATCCTTCATACCTCAGGACA
	+
' \\\	'. *.T_F/7_DDRTDDFF_FD7.TFD7FF/FF.T_/DF//DTD/_77_F7_//7FF77/D7/DF/_F/_F/_F/
ST_F00310.147.HUT25CCVV.3.1101.7162.1872 1.N.O.TGAAGAGA	AST_F00310.147.HUT25CCVV.3.1101.7162.1872 1.N.O.TGAAGAGA
TATTTCACTATTTACTACATATATTTAACACCCCTATCCCTCTTTTATCAATACTTATAATTAATCAATA	CCCANTTTC/CTATTTCACTATTTACTACTACTATATTTAACACCCCCTATTTTATTA
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff< td=""><td>nalyze this data?-FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<></td></afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff<>	nalyze this data?-FFFJ- <j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<>
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 9ST-E00310:147:HVT25CCXX:3:1101:14306:1872 : HOW to a</afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 	nalyze this data? 1-FFFJ- <j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<>
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 9ST-E00310:147:HVT25CCXX:3:1101:14306:1872 : GGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT(</afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 	nalyze this data? 1:N:0:TGAAGAGA TTTATCATCTACTGGCTTGACTGTATATGCAACCATATAGAAC
HACAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 05T-E00310:147:HVT25CCXX:3:1101:14306:1872: TGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGTTTAT(</jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 	nalyze this data? F-FFFJ- <j-fjj<<-<jj<jjj-<f-77fjfj7<<a TTTATCATCTACTGGCTTGACTGTATATGCAACCATATAGAAC</j-fjj<<-<jj<jjj-<f-77fjfj7<<a 
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : GGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGTTTAT( A-AFFJJJJAJJFAJJFJFJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a< td=""><td>nalyze this data? -FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a </j-fjj<<-<jj<jjj-<f-77fjfj7<<a </td></a7-<7f<7jjjf-<<a<></afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( 	nalyze this data? -FFFJ- <j-fjj<<-<jj<jjj-<f-77fjfj7<<a </j-fjj<<-<jj<jjj-<f-77fjfj7<<a 
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( ST-E00310:147:HVT25CCXX:3:1101:14306:1872 : GGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA</a7-<7f<7jjjf-<<a </afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( 	nalyze this data?
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff ST-E00310:147:HVT25CCXX:3:1101:14306:1872 : GGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG</a7-<7f<7jjjf-<<a </afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 	nalyze this data?
HAAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff ST-E00310:147:HVT25CCXX:3:1101:14306:1872 : GGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG H</a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 	haaja7F <jaa-affjjjjajjfajjfjfjjjj-fa7-fajfjajfffffjfjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja @ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA</jaa-affjjjjajjfajjfjfjjjj-fa7-fajfjajfffffjfjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja 
ACAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( BST-E00310:147:HVT25CCXX:3:1101:14306:1872: CGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA</a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( 	nalyze this data?
ACAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( BST-E00310:147:HVT25CCXX:3:1101:14306:1872: CGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC</a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( 	nalyze this data?
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( BST-E00310:147:HVT25CCXX:3:1101:14306:1872: GTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJFJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATACC AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATACC</a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( 	haaja7F <ja\a-affjjjjaajjfajjfjjjjj-fa7-fajfjajffffjfj-<j-fjj<-<-7f<7jjj-<-<<a * * * * * * * * * * * * * * * * * *</ja\a-affjjjjaajjfajjfjjjjj-fa7-fajfjajffffjfj-<j-fjj<-<-7f<7jjj-<-<<a 
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( BST-E00310:147:HVT25CCXX:3:1101:14306:1872: GTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC A-AFFJJAAFF-FA7JFA7FF<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<< td=""><td><pre>nalyze this data?-FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<a 1872 1:N:0:TGAAGAGA TTTATCATCTACTGGCTTGACTGTATATGCAACCATATAGAAC * haaja7F<ja(a-affjjjjjajjfajjfjfjjjjj-fa7-fajfjajfffffjfjjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja @ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GGTTCTTCAT(GTTACTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTGGTTCTTCAT @ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA CCTCAGGACAJAGATTCGGATTGGTCAACAATGTTTAAGTGATATTCCACCAAGGAGAGAGA</ja(a-affjjjjjajjfajjfjfjjjjj-fa7-fajfjajfffffjfjjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja </j-fjj<<-<jj<jjj-<f-77fjfj7<a </pre></td></ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f<<></a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( 	<pre>nalyze this data?-FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<a 1872 1:N:0:TGAAGAGA TTTATCATCTACTGGCTTGACTGTATATGCAACCATATAGAAC * haaja7F<ja(a-affjjjjjajjfajjfjfjjjjj-fa7-fajfjajfffffjfjjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja @ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GGTTCTTCAT(GTTACTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTGGTTCTTCAT @ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA CCTCAGGACAJAGATTCGGATTGGTCAACAATGTTTAAGTGATATTCCACCAAGGAGAGAGA</ja(a-affjjjjjajjfajjfjfjjjjj-fa7-fajfjajfffffjfjjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja </j-fjj<<-<jj<jjj-<f-77fjfj7<a </pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff BST-E00310:147:HVT25CCXX:3:1101:14306:1872: CGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJFJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCCG BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC A-<fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< BST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA</fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< </a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 	<pre>nalyze this data?-FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<a 1872 1:N:0:TGAAGAGA TTTATCATCTACTGGCTTGACTGTATATGCAACCATATAGAAC * AAAJA7F<ja:a-affjjjjajjfajjfjfjjjj-fa7-fajfjajfffffjfjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja @ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GGTTCTTCAT(GTTACTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTGGTTCTTCAT @ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA CCTCAGGACAAGATTCGGATTGGTCAACAATGTTTAAGTGATATTCCACCAAGGAGAGAGA</ja:a-affjjjjajjfajjfjfjjjj-fa7-fajfjajfffffjfjjafjjjjjjj-<a7-<7f<7jjjf-<<aaaja7f<ja </j-fjj<<-<jj<jjj-<f-77fjfj7<a </pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffg BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : GTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGTTTAT( A-AFFJJJJJAJJFAJJFJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCCCG BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC A-<fjaaff-fa7jfa7ff<ffj-<af<<aja-<-77-f7-<<7ff77<a7<af<-f< BST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA</fjaaff-fa7jfa7ff<ffj-<af<<aja-<-77-f7-<<7ff77<a7<af<-f< </a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffg 	<pre>nalyze this data?="FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<a "bt-e00310:147:hvt25ccxx:3:1101:6756:1907="" "ctccaggacajagatcggattggctcaacaatgtttaagtggatattccaaccaa<="" 1:n:0:tgaagaca"="" td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<a></pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff; BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : HOW to a A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCCG BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC HA-<fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< BST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA CTATTTCAGTATTTAGTAGATATTTTAAGAGGGGTATCCCTCTTTTATTATCAATACTTATAATTGCAA</fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< </a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff; 	<pre>nalyze this data?=FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<a<></pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff; BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : CGTTCCGTCAGCTCAGAAGAGTTTGATACTGCTAGAGGGTTTAT( A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGAGAGA</a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff; 	<pre>haaja7F<jaa-affjjjjjajjfajjjfjfjjjj-fa7-fajfjajfffffjfjjafjjjjjjj-<a7-<7f<7jjjjf-<<aaaja7f<ja @ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA @ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA @ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGAGA CCTCAGGACAIAGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCCACCAAGGAGATGAGAAGACTGAAATCCTTCATACCTCAGGACA # CJ-F&lt;7-AAA-<fjaaff-fa7jfa7ff<ffj-<af<<aja<77-f7-<<7ff77<a7<af<-f<j-f<7-a @ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA CCTCAGGACAIAGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATACCTCAGGACA # CJ-F&lt;7-AAA-<fjaaff-fa7jfa7ff<ffj-<af<<aja<77-f7-<<7ff77<a7<af<-f<j-f<7-a @ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA AGGGAATTTG:CTATTTCAGTATTTTAGTAGATATTTTAAGAGGGGTATCCTCTTTTTATTATCAATACTTATAATTGCAAGGGGAATTTG + FJ77&lt;<aaa<afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjjj-<f-77fjfj7<<a< pre=""></aaa<afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjjj-<f-77fjfj7<<a<></fjaaff-fa7jfa7ff<ffj-<af<<aja<77-f7-<<7ff77<a7<af<-f<j-f<7-a </fjaaff-fa7jfa7ff<ffj-<af<<aja<77-f7-<<7ff77<a7<af<-f<j-f<7-a </jaa-affjjjjjajjfajjjfjfjjjj-fa7-fajfjajfffffjfjjafjjjjjjj-<a7-<7f<7jjjjf-<<aaaja7f<ja </pre>
A <afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGAGA BST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGAGAGA</a7-<7f<7jjjf-<<a </afa-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, 	<pre>haaja7f<jaia-affjjjjjajjfajjfjjjjjjjjjjjjjajfa7-fajfjajjjjjffffjfjajjjjjjj-<a7-<7f<7jjjjf-<<aaaja7f<jaia-affjjjjjajjfajjfjjjjjjjjjjjjjajfa7-fajfjjjajfffffjfjjjaffjjjjjjjj-<a7-<7f<7jjjjf-<<aaaja7f<jaia- %ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA %ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA %ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA %CTCCAGGACAjaGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATACCTCAGGACA %ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA %CTCCAGGACAjaGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATACCTCAGGACA %ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA %GGGAATTTG:CTATTTCAGTATTTTAGTAGATATTTTAAGAGGGGTATCCCTCTTTTTATTATCAATACTTATAATTGCAAGGGGAATTTG %ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA %GGGAATTTG:CTATTTCAGTATTTTAGTAGATATTTTAAGAGGGGTATCCCTCTTTTTATTATCAATACTTATAATTGCAAGGGGAATTTG %ST-E00310:147:HVT25CCXX:3:1101:14306:1872 1:N:0:TGAAGAGA</jaia-affjjjjjajjfajjfjjjjjjjjjjjjjajfa7-fajfjajjjjjffffjfjajjjjjjj-<a7-<7f<7jjjjf-<<aaaja7f<jaia-affjjjjjajjfajjfjjjjjjjjjjjjjajfa7-fajfjjjajfffffjfjjjaffjjjjjjjj-<a7-<7f<7jjjjf-<<aaaja7f<jaia- </pre>
AAAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff; BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : HOW to a A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTACTTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCTG ACAAFFJJAJAFF-FA7JFA7FF<ffj-<af<<aja-<-77-f7-<<-7ff77<a7<af<-f< BST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC ACAAFF-FA7JFA7FF<ffj-<af<<aja-<-77-f7-<<-7ff77<a7<af<-f< BST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA CTATTTCAGTATTTAGTAGATATATTTAAGAGGGGTATCCCTCTTTTTATTACAATACTTATAATTGCAA ACAFA-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-ffj-<j-fjj<<-<jj<jjjj-<f-77fj BST-E00310:147:HVT25CCXX:3:1101:14306:1872 1:N:0:TGAAGAGA</jj-a-fjjf-aafjj<-fj-<faajf-7<f-ffj-<j-fjj<<-<jj<jjjj-<f-77fj </ffj-<af<<aja-<-77-f7-<<-7ff77<a7<af<-f< </ffj-<af<<aja-<-77-f7-<<-7ff77<a7<af<-f< </a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff; 	<pre>nalyze this data?-FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<></pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff BST-E00310:147:HVT25CCXX:3:1101:14306:1872 GTTCCGTCAGCTCAGAAGAGGTTTGATACTGCTAGAGGGTTTAT A-AFFJJJJAJJFAJJFJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJ-<a7-<7f<7jjjf-<a BST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA GTTCCGTCAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCCCCG ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGAGA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC AGATTCGGATTGGTCAACAATGTTTAAGTGGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC AGATTCGGATTGGTCAACAATGTTTAAGTGGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC AGATTCGGATTGGTCAACAATGTTTAAGTGGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC ACAFA-<jjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA CTATTTCAGTATTTAGTAGATATTTTAAGAGGGGGTATCCCTCTTTTTATTATCAATACTTATAATTGCAA ACAFA-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjjj-<f-77fj ST-E00310:147:HVT25CCXX:3:1101:14306:1872 1:N:0:TGAAGAGA GTTCCGTCAGGACGAAGAGTTGGTACTGCTAGGAGGGTTTATCATCATCTGGCTTGACTGTATATGCAACC</jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjjj-<f-77fj </jjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< </a7-<7f<7jjjf-<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff 	<pre>nalyze this data?=FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<></pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, ST-E00310:147:HVT25CCXX:3:1101:14306:1872: How to al A-AFFJJJJAJJFAJJFJFJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJ-<a7-<7f<7jjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC A-ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGACA AGATTCAGTATTTAGTAGATATTTTAAGAGGGTATCCCTCTTTTTATTATCAATAATTGCAA ACAFA-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<jj<jjjjjj-<f-77fj ST-E00310:147:HVT25CCXX:3:1101:14306:1872 1:N:0:TGAAGAGA CTATTCAGTATTAGTAGATATTTGATACTGCTAGAGGTTTATCATCATCAGGCTGACTGTATATGCAAC A-AFFJJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a< td=""><td><pre>malyze this data?=FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<></pre></td></a7-<7f<7jjjf-<<a<></jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<jj<jjjjjj-<f-77fj </a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, 	<pre>malyze this data?=FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<></pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, ST-E00310:147:HVT25CCXX:3:1101:14306:1872 : How to al A-AFFJJJJAJJFAJJFJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJ-<a7-<7f<7jjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTCCACCAAGGAGATGAGAAGACTGAAATCCTTCATAC </a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, 	<pre>nalyze this data?=FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<<a<></pre>
ACAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff, 	<pre>nalyze this data?-FFFJ-<j-fjj<<-<jj<jjj-<f-77fjfj7<a< td=""></j-fjj<<-<jj<jjj-<f-77fjfj7<a<></pre>
AAFA- <jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( BST-E00310:147:HVT25CCXX:3:1101:14306:1872 : HOW to al A-AFFJJJJAJJFAJJFJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA ST-E00310:147:HVT25CCXX:3:1101:13382:1819 1:N:0:TGAAGACA AGATTCGGATTGGTCAACAATGTTTAAGTGATATTTCAACCAAGGAGATGAGAAGACTGAAATCCTTCATAC A-AFFJAAFF-FA7JFA7FF<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA CATTCAGTATTTAGTAGATATATTTTAAGAGGGGTATCCCTCTTTTTATTATCAATACTTATAATTGCAA AGATCGGATTGGTCAACAATGTTTAAGTGAAAFTCTTCACCAAGGAGAGAGAGAGAGAGACTGAAATCCTTCATAC A- AA-<fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< ST-E00310:147:HVT25CCXX:3:1101:7162:1872 1:N:0:TGAAGAGA CTATTTCAGTATTTAGTAGATATATTTTAAGAGGGGGTATCCCTCTTTTTATTATCAATACTTATAATTGCAA CAFA-<jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjjjjjjj-<f-77fj ST-E00310:147:HVT25CCXX:3:1101:14306:1872 1:N:0:TGAAGAGA CTATTCCGTCCGAAGAGTTTGATACTGCTAGAGGGTTTATCATCTACTGGCTTGACTGTATATGCAACC CA-AFFJJJJJAJJFAJJFJJJJJJ-FA7-FAJFJAJFFFFFJFJJAFJJJJJJJ-<a7-<7f<7jjjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:16756:1907 1:N:0:TGAAGAGA CTATTTGTTCCAAAGCTTTAGCTTTATTCCGTCCAAAGTTCATTGCACCACGTGCTCCACCACCACCACCCCG CA-AFFJJJJJAJJFAJJFJJJFJJJJ-FA7-FAJFJAJFFFFFJFJAFJJJJJJJJ-<a7-<7f<7jjjjf-<<a ST-E00310:147:HVT25CCXX:3:1101:6756:1907 1:N:0:TGAAGAGA</a7-<7f<7jjjjf-<<a </a7-<7f<7jjjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fffj-<j-fjj<<-<jj<jjjjjjjj-<f-77fj </fjaaff-fa7jfa7ff<ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< </ffj-<af<<aja<-77-f7-<<7ff77<a7<af<-f< </a7-<7f<7jjjf-<<a </jj-a-fjjf-aafjj<-fj-<faajf-7<f-fff( 	<pre>nalyze this data?FFFJ-<j-fjj<<-<jj<jj-<f-77fjfj7<a< td=""></j-fjj<<-<jj<jj-<f-77fjfj7<a<></pre>

### **Overview of data processing step**



0.00

0.05 0.10 0.15 -Log10(Adjust.P)

Shi, Song et al. Journal of orthopaedic surgery and research, 2021.

## **Quality control**

## *Report* **Report**

## Summary

Basic Statistics Per base sequence quality Per tile sequence quality Per sequence quality scores Per base sequence content Per sequence GC content Per base N content Sequence Length Distribution Sequence Duplication Levels Overrepresented sequences Adapter Content

CSDN @YHC BI

### **Basic Statistics**

Q10-0.1

O20-0.01

Measure	Value	
Filename	1y2_2. fq. gz	文件名
File type	Conventional base calls	文件类型
Encoding	Sanger / Illumina 1.9	测序平台、编码版本号
Total Sequences	34194529	
Total Bases	5.1 Gbp	
Sequences flagged as poor quality	0	
Sequence length	150	测序长度150bp
%GC	53	GC含量



### Alignment



#### Table S3. Overview of mapping of RNA-seq reads

Sample name	Reads Passed Pipeline Casava 1.8	Total reads for mapping after depletion of rRNA reads	Mapped reads	Mapped reads (%)	Uniquely mapped reads	Uniquely mapped reads (%)
S1-BLO	14,979,030±567,503	14,645,020±726,712	10,413,447±1,038,711	71.01±4.36	8,642,167±687,620	58.96±2.05
S1-BLO1	14,331,274	13,808,859	9,318,562	67.48	7,860,447	56.92
S1-BLO2	15,388,745	15,001,973	11,384,966	75.89	9,153,474	61.02
S1-BLO3	15,217,072	15,124,229	10,536,813	69.67	8,912,579	58.93
S1-KID	14,348,868±304,077	13,956,939±199,609	9,985,279±365,562	71.53±1.62	8,196,252±24,544	58.73±1
S1-KID1	14,538,519	14,179,524	10,406,972	73.39	8,172,559	57.64
S1-KID2	14,509,948	13,793,822	9,758,102	70.74	8,221,567	59.6
S1-KID3	13,998,138	13,897,472	9,790,763	70.45	8,194,630	58.96
S2-BLO	16,243,503±3,578,949	14,163,395±1,630,064	9,379,959±990,424	66.31±2.89	7,848,435±831,241	55.48±2.48
S2-BLO1	20,289,683	16,045,356	10,439,342	65.06	8,734,231	54.43
S2-BLO2	14,948,561	13,194,586	8,477,144	64.25	7,085,403	53.7
S2-BLO3	13,492,265	13,250,242	9,223,390	69.61	7,725,672	58.31

### 转录组常用软件 基因组常用软件

Hisat2

• STAR

•

- Bowtie2
  - BWA

El-Sharkawy, Islam et al. Journal of experimental botany, 2015.

### **Differential Expression Analysis: Find key genes**

![](_page_34_Figure_1.jpeg)

常用的工具为

- DESeq2
- edgeR

![](_page_34_Figure_5.jpeg)

El-Sharkawy, Islam et al. Journal of experimental botany,2015.

Kingston, Belinda et al. Nature communications ,2021.

## **Enrichment Analysis : GO or KEGG analysis**

biological systems, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.	Your gene IDs here
The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and s a foundation for computational analysis of large-scale molecular biology and genetics.	
experiments in biomedical research.	biological process
	Homo sapiel ~ Examples Launch >
Search GO term or Gene Product in AmiGO	Hint: can use UniProt ID/AC, Gene Name, Gene Symbols

Intology (aspect)		dsx	doublesex	DNA-binding	FlyBase	Drosophila	IDA
(30)	р			transcription		melanogaster	
	F			activator			
				activity, KNA			
<b>a</b> (3)	U.			II-specific			
tology (aspect)		dsx	doublesex	nucleus	FlyBase	Drosophila	IC
ganism						melanogaster	
pe .							
idence							
30 plass (including_tregulatest)							
(37) balogical_proces	6	O dex	dsx doublesex	regulation of	EluBaca	Droconhila	IDA ter
(19) developmental pr	00936	U USA		transcription	melshor	molopopting	
(17) reproductive proc	225			transcription		melanogaster	
(14) anatomical struct	ure development			by RNA			
🧧 (14) multicellular orga	nismal process			polymerase II			
anatomical entity							
(13) anatomical struct	urb						
(13) biological entity							
(13) connected anotor	recal structure	dsx	doublesex	female	FlyBase Drosophila melanogast	Drosophila	TAS er
(13) developmental pr reproduction	ocess involved in		somatic sex determination	somatic sex		melanogaster	
🚦 (13) material anatomic	sal entity			determination			
(12) biological regulati	lón.						
🚦 (12) reputation of biolo	ogical process						
Feo (11)							

![](_page_35_Picture_3.jpeg)

Entry	Dme1_CG11094 CDS T00030
Symbol	dsx
Name	(RefSeq) doublesex, isoform D
Organism	dme Drosophila melanogaster (fruit fly)
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: DSX_dimer DM Motif
Other DBs	NCBI-GeneID: 40940 NCBI-ProteinID: NP_001262352 FlyBase: FBgn0000504 UniProt: P23023 A0A084KFS4
Structure	PDB PDBj
Position	3R Genome browser
AA seq	549 aa AA seq DB search MVSEENWNSDTMSDSDMIDSKNDVCGGASSSSGSSISPRTPPNCARCRNHGLKITLKGHK RYCKFRYCTCEKCRLTADRQRVMALQTALRRAQAQDEQRALHMHEVPPANPAATTLLSHH HHVAAPAHVHAHHVHAHHAHGHHSHHGHVLHHQQAAAAAAAPSAPASHLGGSSTAASS IHGHAHAHHVHMAAAAASVAQHQHQSHPHSHHHHHQNHHQHPHQQPATQTALRSPPHSD HGGSVGPATSSSGGGAPSSSNAAAATSSNGSSGGGGGGGGGGGGGGGGGGGGGGGGG

### Visualize the enrichment analysis results.

![](_page_36_Figure_1.jpeg)

Chen, Ke et al. *Molecular medicine reports*,2017.

GO term

### Differentially expressed genes in chronic social isolated drosophila.

е

![](_page_37_Figure_1.jpeg)

d

![](_page_37_Figure_3.jpeg)

Li, Wanhe et al. Nature, 2021.

## Take home message

- Principles of the three sequencing technologies.
  - Illumina, PacBio, Nanopore
- The process of sequencing data processing
  - Quality control, Alignment, Differential Analysis, Enrichment

![](_page_38_Picture_5.jpeg)

![](_page_38_Figure_6.jpeg)

## Advanced Computational Techniques in Omics and Behavioral Data

Joshua Liang 2024-11-28

## Unlocking Complexity in Neuroscience with Advanced Analytics

![](_page_40_Figure_1.jpeg)

### **Non-Negative Matrix Factorization (NMF)**

$$A_{m_X n} = W_{m_X k} H_{k_X n}$$

![](_page_41_Figure_2.jpeg)

where,

- A -> Original Input Matrix (Linear combination of W & H)
- W -> Feature Matrix
- H -> Coefficient Matrix (Weights associated with W)
- k -> Low rank approximation of A (k ≤ min(m,n))

![](_page_42_Figure_0.jpeg)

NMF Reveals Expression Patterns in Olfactory Neuroblastoma

Yang J et al., Nat Cancer. 2024

![](_page_43_Figure_0.jpeg)

Connecting Brain Features with Clinical Insights via NMF

Han S et al., Nat Commun, 2023

## **Manifold Fitting**

![](_page_44_Figure_1.jpeg)

![](_page_44_Picture_2.jpeg)

Yao Z et al., *arXiv*, 2023

![](_page_45_Picture_0.jpeg)

![](_page_46_Figure_0.jpeg)

Yao Z et al., Proc Natl Acad Sci USA, 2024

![](_page_47_Figure_0.jpeg)

Yao Z et al., Proc Natl Acad Sci USA, 2024

### Mendelian Randomization (MR)

![](_page_48_Figure_1.jpeg)

Sanderson E et al., Nat Rev Methods Primers, 2022

a An RCT to test whether lowering CRP lowers SBP

![](_page_49_Figure_1.jpeg)

**b** An MR study to test whether lowering CRP lowers SBP

![](_page_49_Figure_3.jpeg)

Sanderson E et al., Nat Rev Methods Primers, 2022

![](_page_50_Figure_0.jpeg)

Mu C et al., Nat Hum Behav, 2024

![](_page_51_Figure_0.jpeg)

![](_page_51_Figure_1.jpeg)

Mu C et al., Nat Hum Behav, 2024

### Machine Learning & Deep Learning

![](_page_52_Figure_1.jpeg)

![](_page_53_Figure_0.jpeg)

Wittek N et al., Behav Res Methods, 2023

![](_page_54_Figure_0.jpeg)

## **Reinforcement Learning in ML**

![](_page_55_Figure_1.jpeg)

b

**Trial timeline** 

Choice

Outcome

#### **Possible outcomes**

Aversive sound, no reward

No reward, no aversive sound Reward, no aversive sound

![](_page_56_Figure_3.jpeg)

Reward and aversive sound

![](_page_56_Picture_5.jpeg)

![](_page_56_Picture_6.jpeg)

![](_page_56_Figure_7.jpeg)

![](_page_56_Figure_8.jpeg)

Yamamori Y et al., Elife, 2023

Fixation

Options

а

![](_page_57_Figure_0.jpeg)

Yamamori Y et al., Elife, 2023

## **Deep Neural Network**

![](_page_58_Figure_1.jpeg)

Convolutional Neural Network with Multiple Layers

![](_page_59_Figure_0.jpeg)

Danyal A et al., Nature, 2024

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 28, 28, 3)]	0
conv2d (Conv2D)	(None, 28, 28, 32)	896
batch_normalization (	BatchNo (None, 28, 28,	32) 128
conv2d_1 (Conv2D)	(None, 14, 14, 32	) 9248
batch_normalization_	1 (Batch (None, 14, 14,	32) 128
dropout (Dropout)	(None, 14, 14, 32)	0
conv2d_2 (Conv2D)	(None, 14, 14, 32	) 9248
batch_normalization_	2 (Batch (None, 14, 14,	32) 128
dropout_1 (Dropout)	(None, 14, 14, 32)	0
conv2d_3 (Conv2D)	(None, 14, 14, 32	9248
batch_normalization_	3 (Batch (None, 14, 14,	32) 128
dropout_2 (Dropout)	(None, 14, 14, 32)	0
conv2d_4 (Conv2D)	(None, 7, 7, 32)	9248
batch_normalization_	4 (Batch (None, 7, 7, 32	2) 128
dropout_3 (Dropout)	(None, 7, 7, 32)	0
conv2d_5 (Conv2D)	(None, 7, 7, 32)	9248
batch_normalization_	5 (Batch (None, 7, 7, 32	2) 128
dropout_4 (Dropout)	(None, 7, 7, 32)	0
conv2d_6 (Conv2D)	(None, 4, 4, 32)	9248
batch_normalization_	6 (Batch (None, 4, 4, 32	2) 128
dropout_5 (Dropout)	(None, 4, 4, 32)	0
conv2d_7 (Conv2D)	(None, 4, 4, 32)	9248
batch_normalization_	7 (Batch (None, 4, 4, 32	2) 128
dropout_6 (Dropout)	(None, 4, 4, 32)	0
conv2d_8 (Conv2D)	(None, 4, 4, 64)	18496
batch_normalization_	8 (Batch (None, 4, 4, 64	4) 256
dropout_7 (Dropout)	(None, 4, 4, 64)	0
conv2d_9 (Conv2D)	(None, 4, 4, 64)	36928
batch_normalization_	9 (Batch (None, 4, 4, 64	4) 256
dropout_8 (Dropout)	(None, 4, 4, 64)	0
dense (Dense)	(None, 4, 4, 1024)	66560
batch_normalization_	10 (Batc (None, 4, 4, 10	024) 4096
dropout_9 (Dropout)	(None, 4, 4, 1024)	0
conv2d_10 (Conv2D)	(None, 4, 4, 64)	589888

batch_normalization_	11 (Batc (None, 4, 4, 64)	256
dropout_10 (Dropout)	(None, 4, 4, 64)	0
conv2d_11 (Conv2D)	(None, 4, 4, 128)	73856
batch_normalization_	12 (Batc (None, 4, 4, 128	3) 512
dropout_11 (Dropout)	(None, 4, 4, 128)	0
conv2d_12 (Conv2D)	(None, 4, 4, 128)	147584
batch_normalization_	13 (Batc (None, 4, 4, 128	3) 512
dropout_12 (Dropout)	(None, 4, 4, 128)	0
conv2d_13 (Conv2D)	(None, 2, 2, 128)	147584
batch_normalization_	14 (Batc (None, 2, 2, 128	3) 512
dropout_13 (Dropout)	(None, 2, 2, 128)	0
conv2d_14 (Conv2D)	(None, 2, 2, 128)	147584
batch_normalization_	15 (Batc (None, 2, 2, 128	3) 512
conv2d_15 (Conv2D)	(None, 2, 2, 128)	147584
batch_normalization_	16 (Batc (None, 2, 2, 128	3) 512
dropout_14 (Dropout)	(None, 2, 2, 128)	0
conv2d_16 (Conv2D)	(None, 2, 2, 128)	147584
batch_normalization_	17 (Batc (None, 2, 2, 128	3) 512
max_pooling2d (MaxF	Pooling2D) (None, 1, 1, 1	28) 0
flatten (Flatten)	(None, 128) 0	
dropout_15 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 2048)	264192
batch_normalization_	18 (Batc (None, 2048)	8192
dropout_16 (Dropout)	(None, 2048)	0
dense_2 (Dense)	(None, 1024)	2098176
batch_normalization_	19 (Batc (None, 1024)	4096
dropout_17 (Dropout)	(None, 1024)	0
dense_3 (Dense)	(None, 512)	524800
batch_normalization_2	20 (Batc (None, 512)	2048
dropout_18 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 10)	5130
Total params: 4,504,8 Trainable params: 4,4	74 93,226 • 11 648	

• **Convolutional Layers:** Extract audio spectrogram patterns.

• **Batch Normalization**: Stabilizes the training process by normalizing the output of each layer, accelerating convergence and improving performance.

• **Pooling Layers:** Reduce the dimensionality of spectrogram data while retaining critical features.

• **Dense Layers:** Combine learned features for classification.

• **Dropout Layers:** Regularize the model to avoid overfitting.

# Transformer

## **Attention Is All You Need**

$$a = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

![](_page_61_Figure_3.jpeg)

![](_page_62_Figure_0.jpeg)

Multi-Head Attention allows the model to "look" at the input from multiple perspectives simultaneously, capturing diverse relationships within the data for a deeper understanding.

![](_page_63_Figure_0.jpeg)

Szałata A et al., Nat Methods, 2024

![](_page_64_Figure_0.jpeg)

Szałata A et al., Nat Methods, 2024

## Take home message

➤ Transformers use attention mechanisms to analyze gene interactions and integrate multimodal data, advancing tasks like cell annotation and regulatory network inference.

NMF extracts latent features from high dimensional data, while manifold fitting captures nonlinear structures, enhancing insights into complex biological systems.

MR uses genetic variation to infer causal links between traits and diseases, aiding the study of brain networks and mental disorders.

## What can we learn from this report?

1. The diversity of statistical approaches in neuroscience data analysis: Beyond differential analysis, statistical methods encompass regression prediction, dimensionality reduction, and causal inference. These approaches enable us to uncover complex relationships and biological mechanisms within data.

## 2. Standardized sequencing data processing steps:

Common sequencing data processing pipelines include quality control, alignment, differential analysis, dimensionality reduction, and functional/network inference. These steps ensure data reliability and provide a solid foundation for biological interpretation.

## 3. The potential of deep learning in neurobiology:

Deep learning methods, such as the Transformer architecture, are driving innovations in neurobiology through their powerful feature extraction capabilities and flexibility. Embracing breakthroughs like "Attention Is All You Need" will further unravel the complexity of neural and gene regulatory networks.