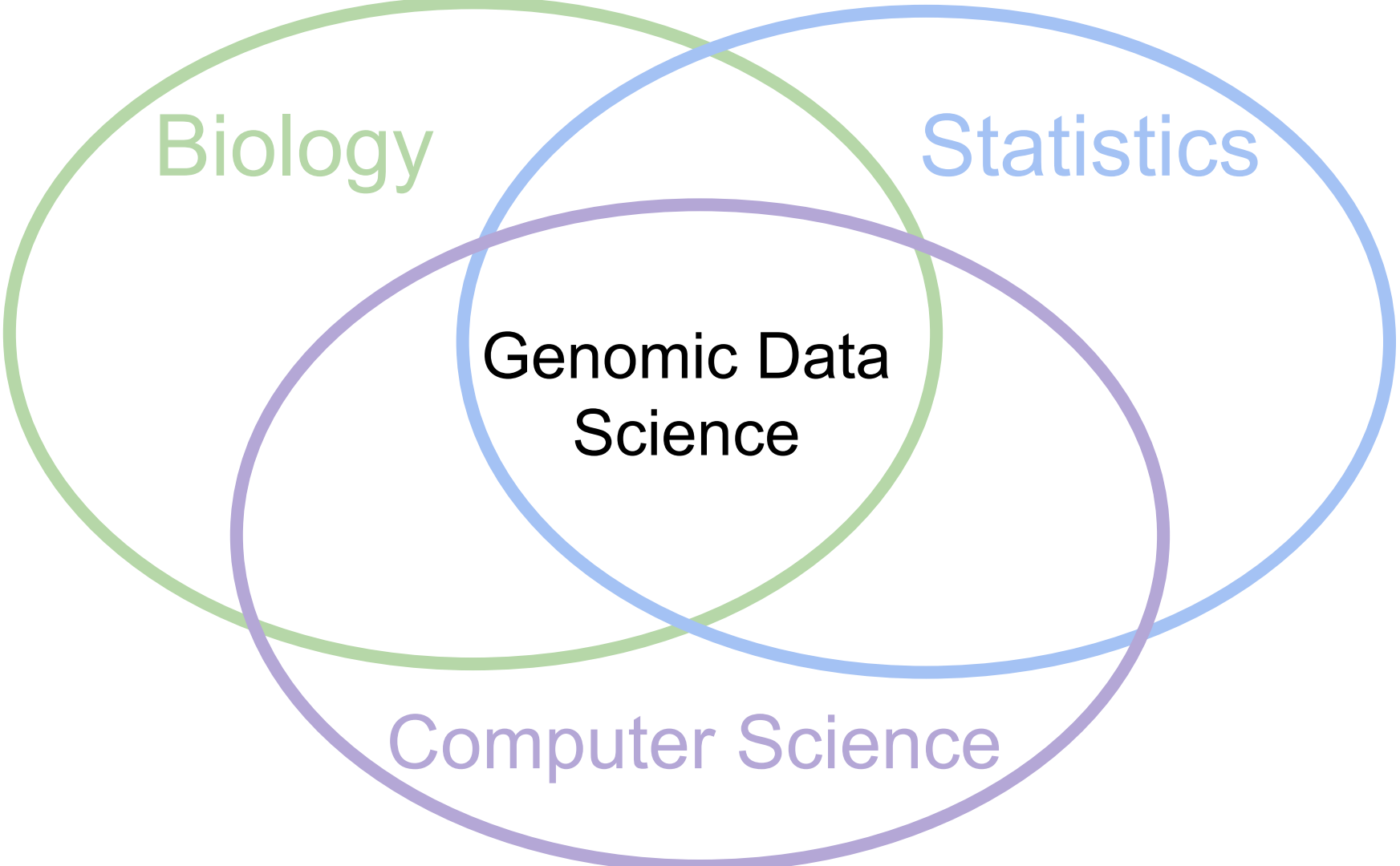
What is genomic data science?

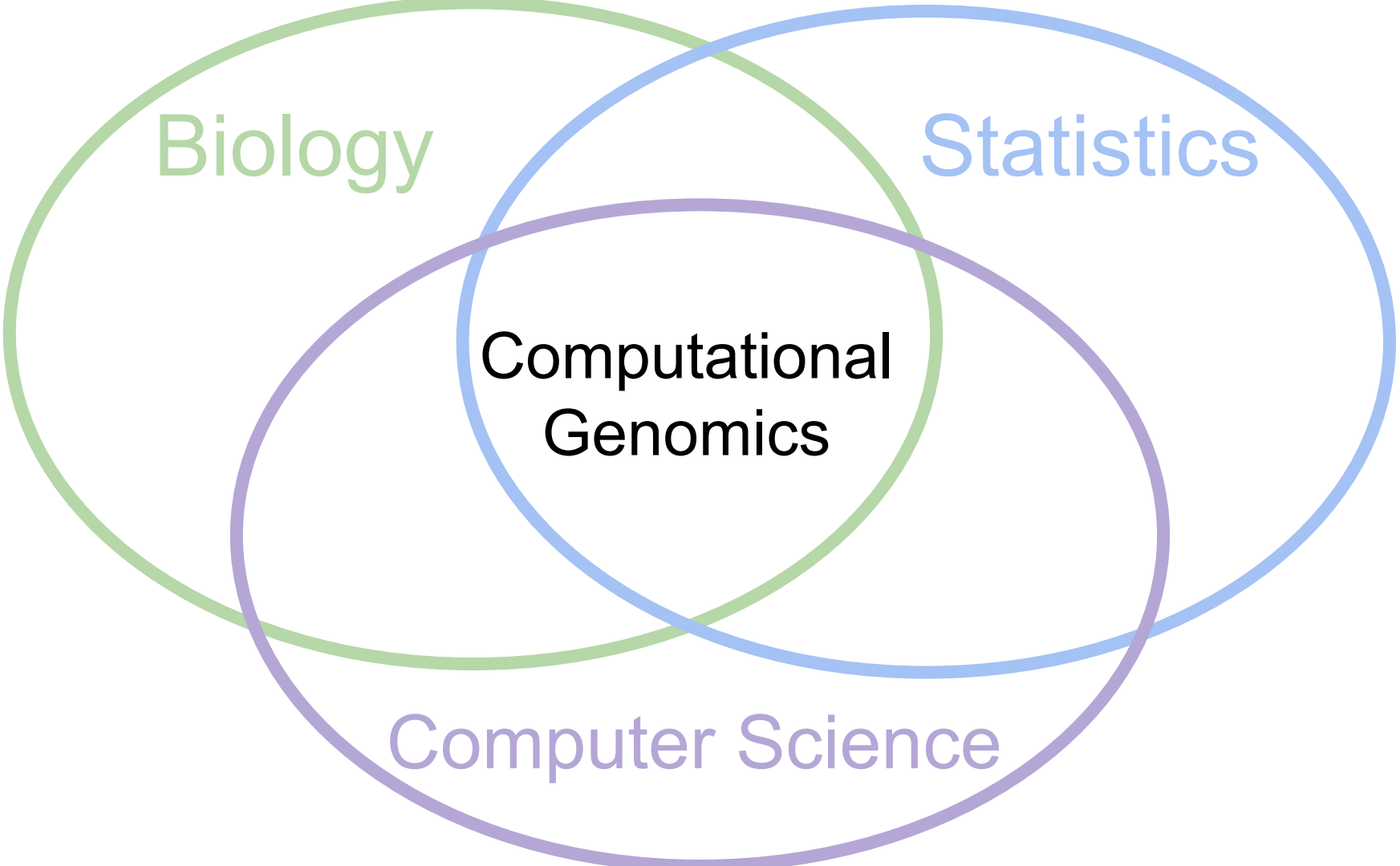
Steven Salzberg

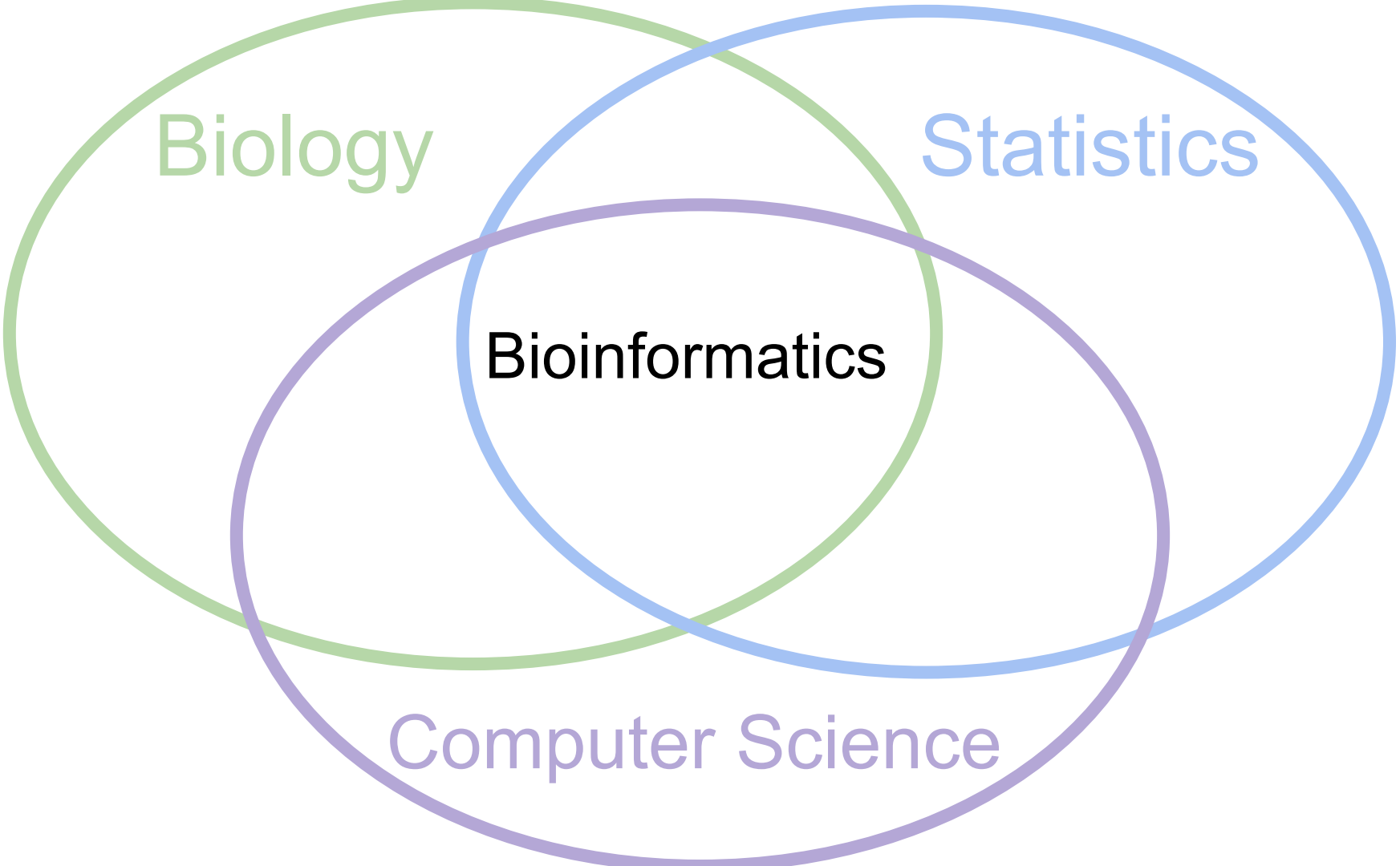
Intersection of three disciplines

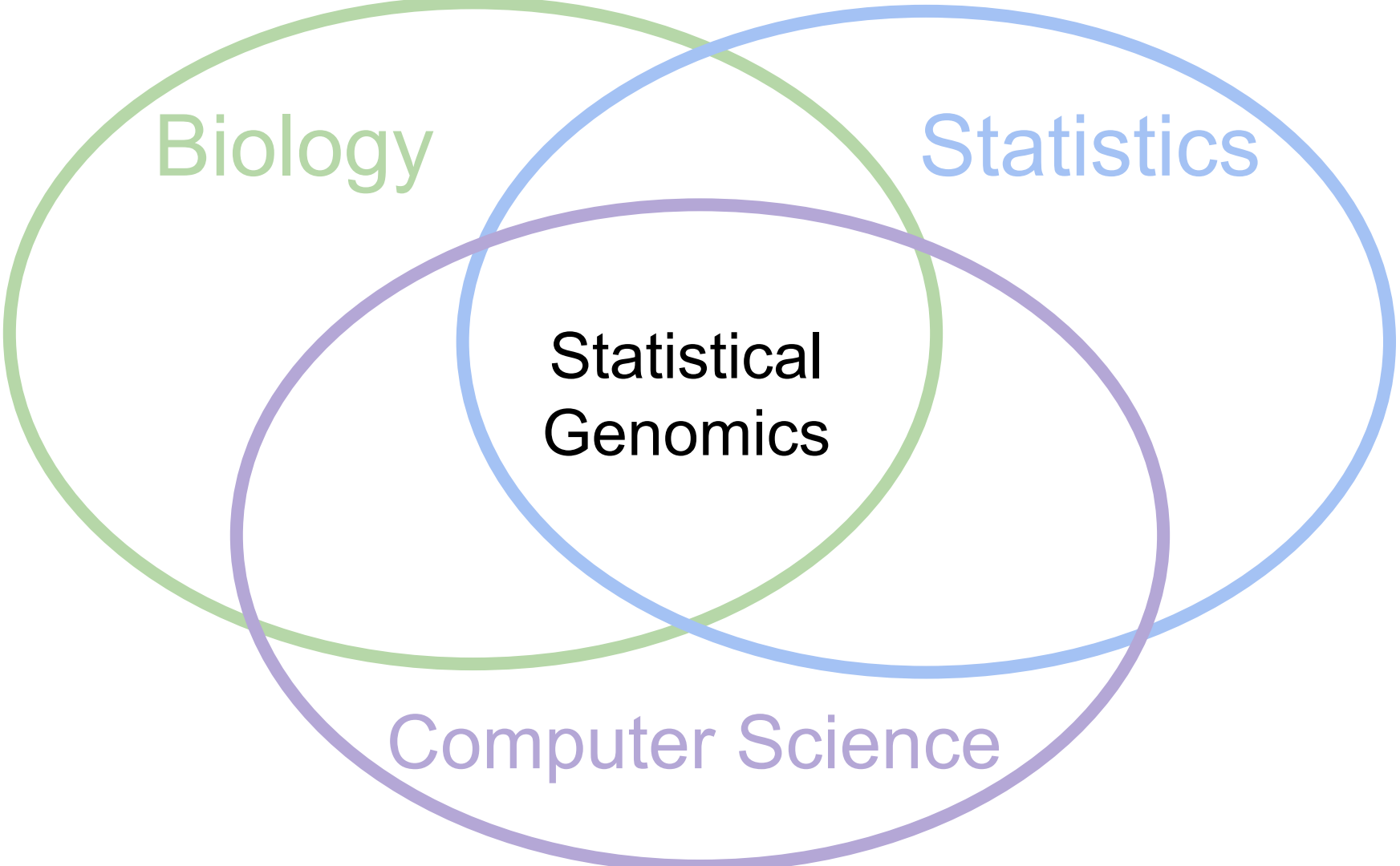
- biology
- statistics
- computer science


Biology
Statistics
Genomic Data Science
Computer Science

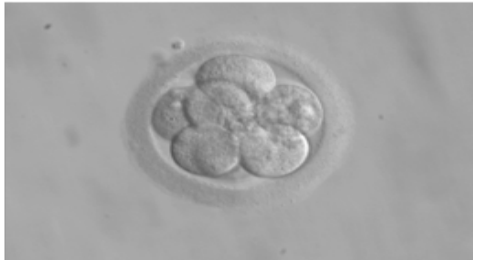
Genomic data science has multiple names


Biology
Statistics
Computational
Genomics
Computer Science


Biology
Statistics
Bioinformatics
Computer Science


Biology
Statistics
Statistical
Genomics
Computer Science

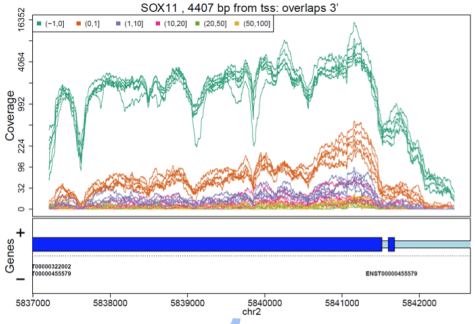
What do genomic data scientists do?

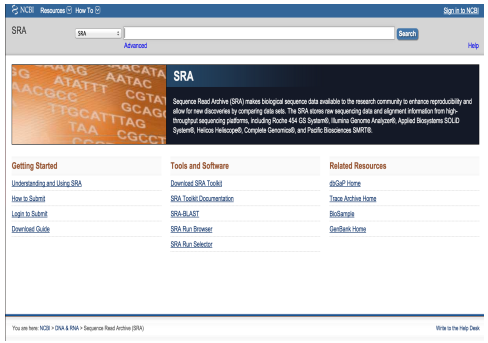


TACAAAATCA GATCGATCG

AGATCAGTC

TACAAAATCAGATCGATCGAGATCAGTC



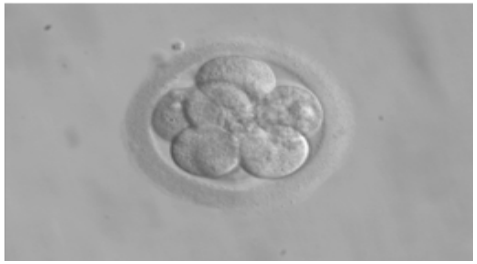


$$g(Y_{ij}) = \alpha(l_j) + \beta(l_j)X_i + \sum_{k=1}^{K} \gamma_k(l_j)W_{ik} + \varepsilon_{ij}$$

Galaxy

Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS



experimental design

TACAAAATCA GATCGATCG

AGATCAGTC

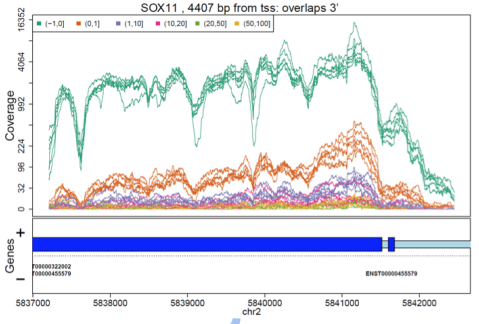
TACAAAATCAGATCGATCGAGATCAGTC

alignment & assembly

TACAAAATCA GATCGATCG

AGATCAGTC

TACAAAATCAGATCGATCGAGATCAGTC



preprocessing & normalization

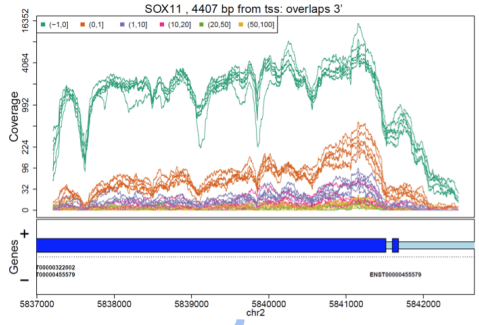
statistics & machine learning

Galaxy

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

TACAAAATCA GATCGATCG
AGATCAGTC
TACAAAATCAGATCGATCGAGATCAGTC

$$g(Y_{ij}) = \alpha(l_j) + \beta(l_j)X_i + \sum_{k=1}^{K} \gamma_k(l_j)W_{ik} + \varepsilon_{ij}$$



software development

Galaxy

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

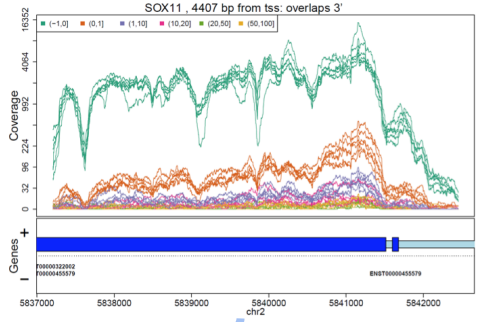
$$g(Y_{ij}) = \alpha(t_j) + \beta(t_j)X_i + \sum_{k=1}^{K}\gamma_k(t_j)W_{ik} + \varepsilon_{ij}$$

population genomics

Galaxy

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

$$g(Y_{ij}) = \alpha(l_j) + \beta(l_j)X_i + \sum_{k=1}^{K} \gamma_k(l_j)W_{ik} + \varepsilon_{ij}$$



integrative genomics

Galaxy

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

$$g(Y_{ij}) = \alpha(l_j) + \beta(l_j)X_i + \sum_{k=1}^{K} \gamma_k(l_j)W_{ik} + \varepsilon_{ij}$$