Quality Control and Target Validation in Sequencing

v1.0

Simon Andrews
Laura Biggins
Boo Virk

Babraham Institute

Babraham Bioinformatics

Babraham Bioinformatics

- Support service for bioinformatics
  - Academic – Babraham Institute
  - Commercial – Consultancy

- Support BI Sequencing Facility
  - HiSeq / MiSeq based sequencing service
  - Data Management / Processing / Analysis

Babraham Bioinformatics

Interests in QC

- Developed QC for in-house sequencing
- Developed QC packages
  - FastQC
  - BamQC
  - FastQ Screen

- Developed application specific QC
  - Bismark (bisulphite methylation)
  - HiCUP (Hi-C genome structure)

- Developed data visualisation QC
  - SeqMonk (generic sequencing visualisation / analysis)
    - RNA-Seq QC
    - Small RNA QC
    - Duplication QC

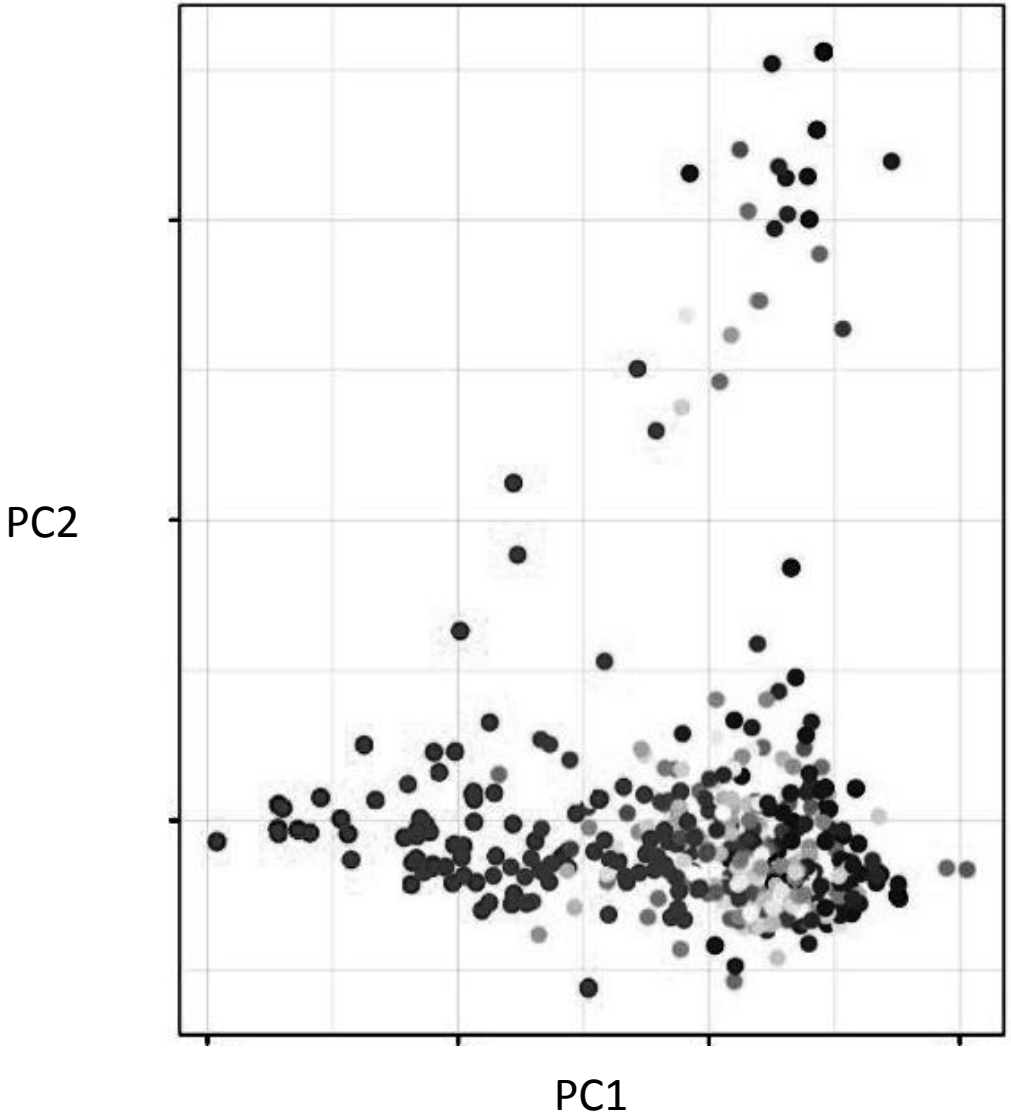
Babraham Bioinformatics

Areas for today

- How do sequencing experiments go wrong
  - Learn from mistakes of others
- How to construct good QC
  - What should you run
  - What should you look for
  - How should you interpret / act
- What software exists
  - Review of existing QC packages / use cases

Babraham Bioinformatics

An example…

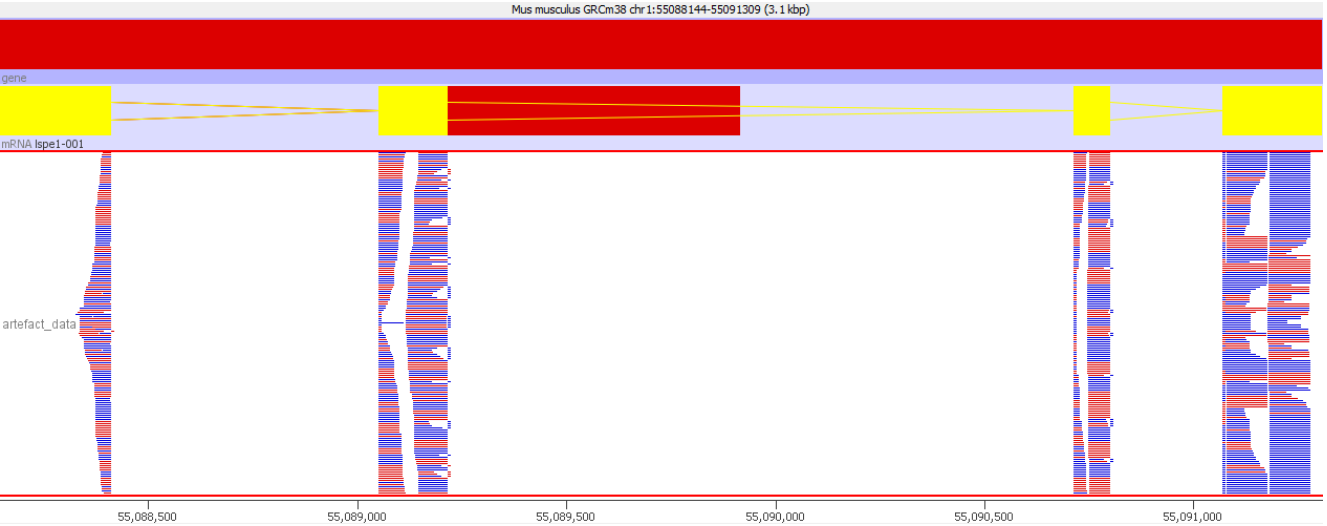


Babraham Bioinformatics

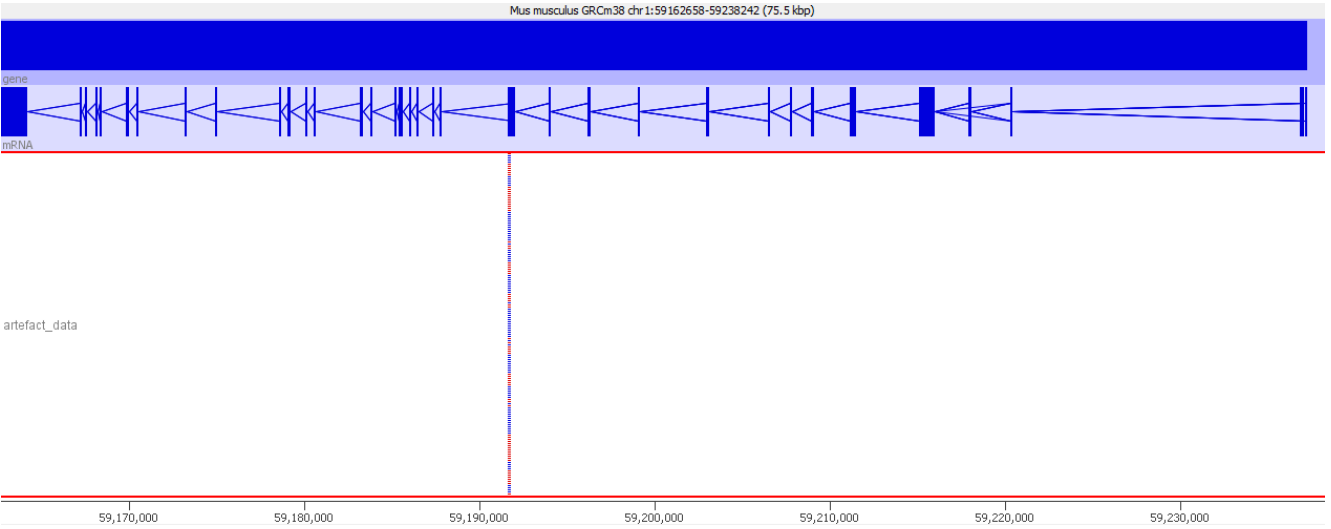
Genes for PC1 (85 total)

Gene	Description
Arhgef4	Rho guanine nucleotide exchange factor (GEF) 4
Cflar	CASP8 and FADD-like apoptosis regulator
Als2	amyotrophic lateral sclerosis 2 (juvenile) homolog (human)
Cxcr2	chemokine (C-X-C motif) receptor 2
Col4a3	collagen, type IV, alpha 3
Sag	retinal S-antigen
Gpr35	G protein-coupled receptor 35
Acmsd	amino carboxymuconate semialdehyde decarboxylase
Qsox1	quiescin Q6 sulfhydryl oxidase 1
9430070O13Rik	RIKEN cDNA 9430070O13 gene
Mrps14	mitochondrial ribosomal protein S14
Scyl3	SCY1-like 3 (S. cerevisiae)
Ildr2	immunoglobulin-like domain containing receptor 2
Atp1a2	ATPase, Na+/K+ transporting, alpha 2 polypeptide
Slamf8	SLAM family member 8
Wdr38	WD repeat domain 38
Exd1	exonuclease 3'-5' domain containing 1
Serf2	small EDRK-rich factor 2

Coverage of Raw Data



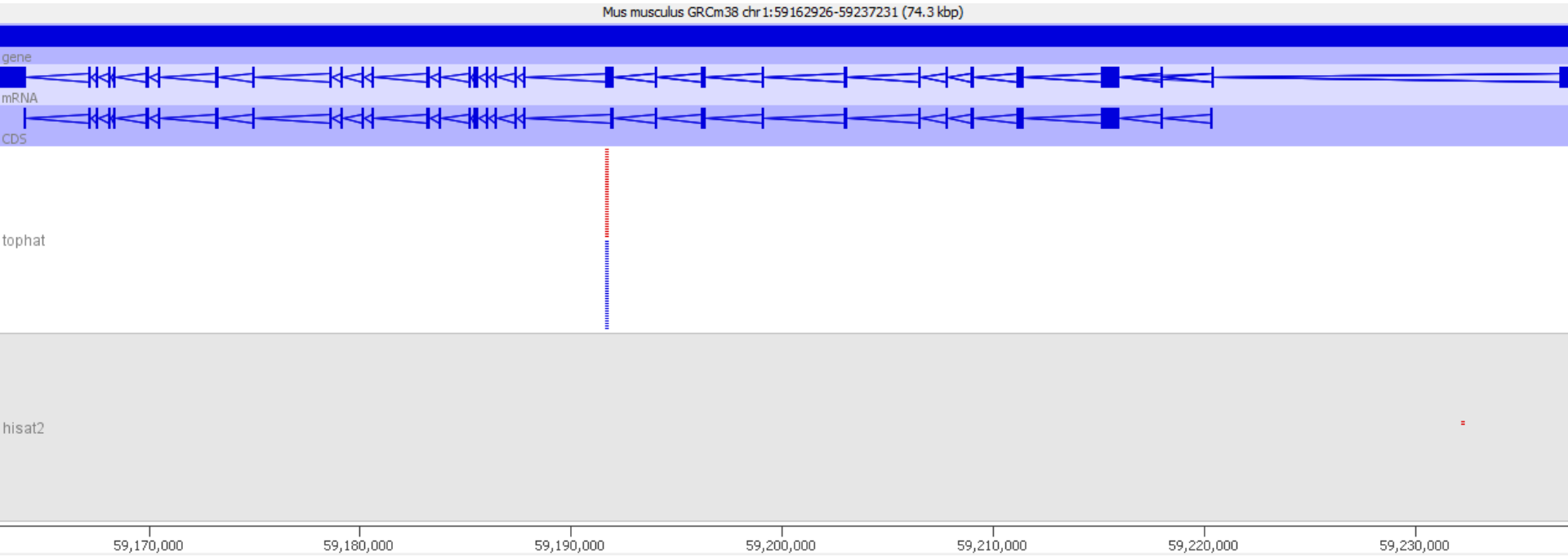
Normal Gene



PCA Gene

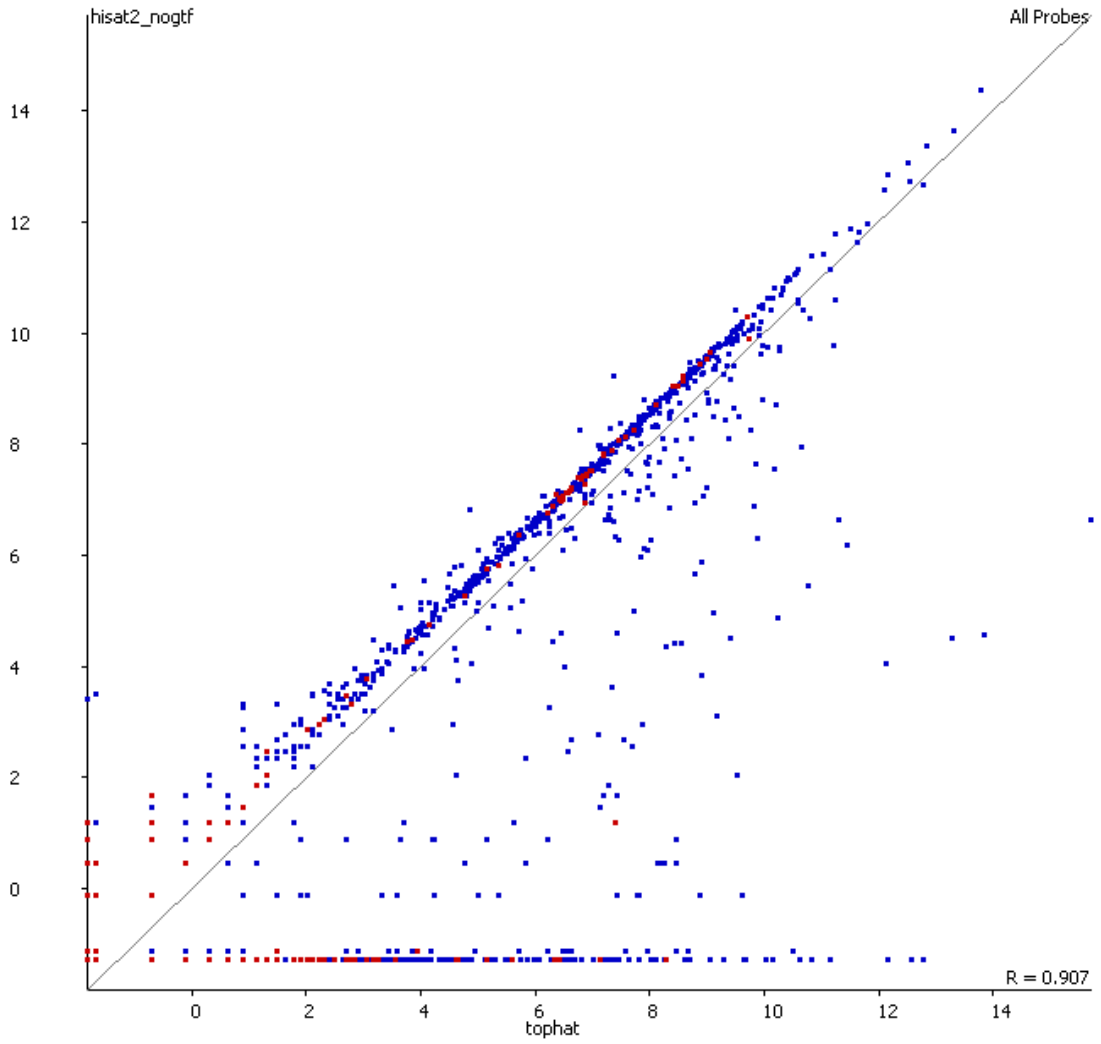
Babraham Bioinformatics

Using a different read mapper…

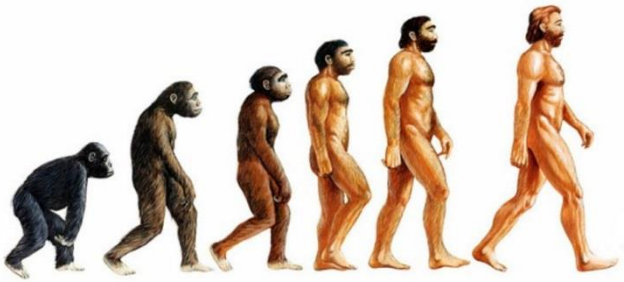


Babraham Bioinformatics

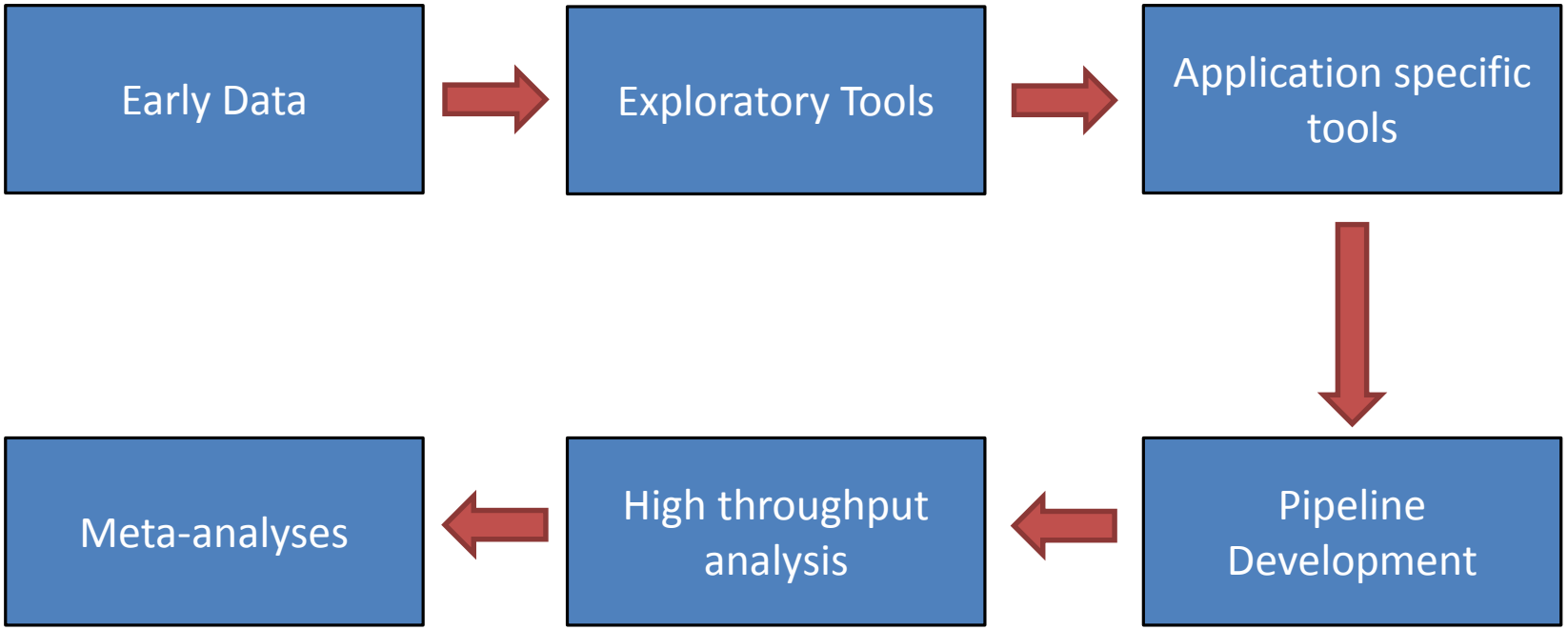
Using a different read mapper…



Babraham Bioinformatics



The Evolution of Sequencing Analysis



Babraham Bioinformatics

Good Lessons from exploration and tool development

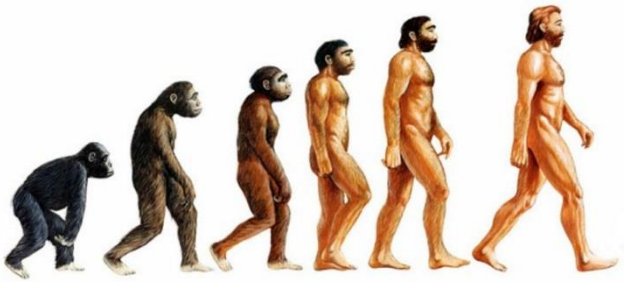
- General structure of the data
- Quantitation
- Points of commonality
  - Expectations
  - Reference points
  - Normalisation
- Statistics

Babraham Bioinformatics

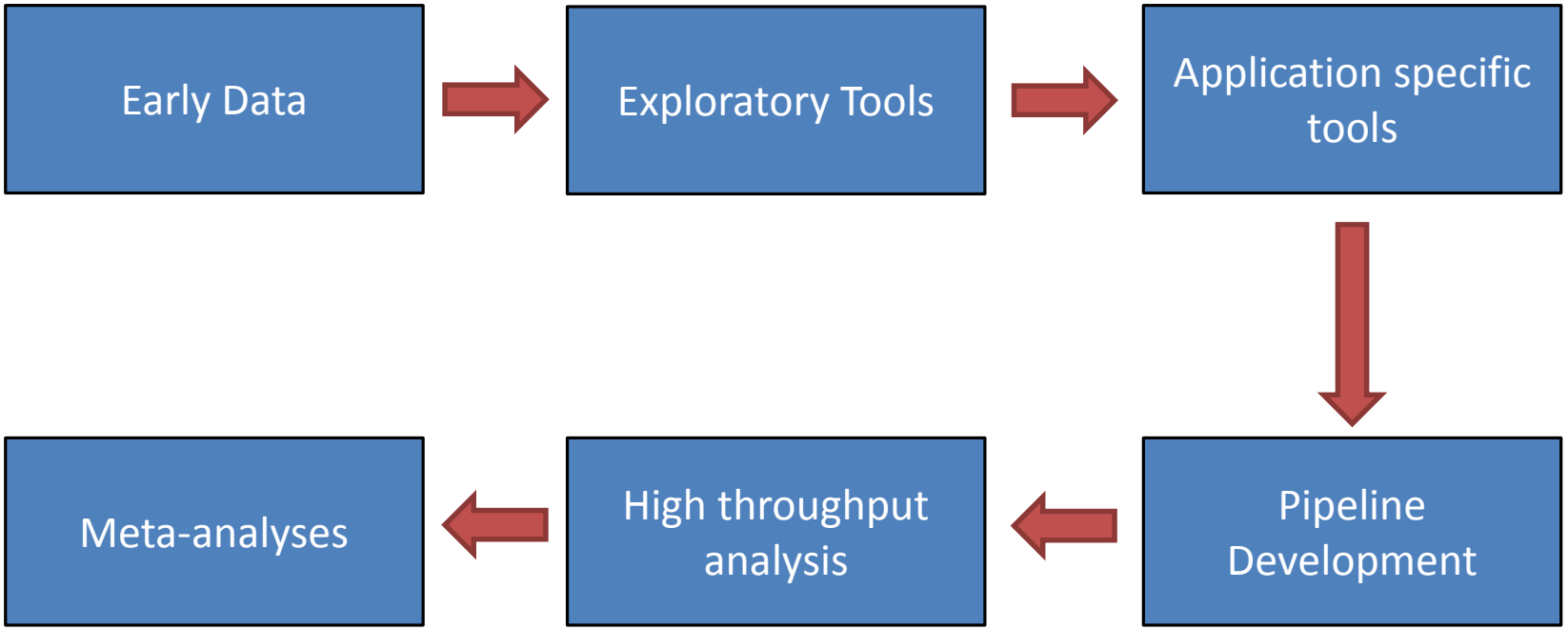
Bad Lessons from exploration and tool development

- Failure modes
  - Contamination
  - Library failures
- Artefacts
- Biases
- Mis-interpretations

Babraham Bioinformatics



The Evolution of Sequencing Analysis



Babraham Bioinformatics

Areas for today

- How do sequencing experiments go wrong
  - Learn from mistakes of others
- How to construct good QC
  - What should you run
  - What should you look for
  - How should you interpret / act
- What software exists
  - Review of existing QC packages / use cases

Babraham Bioinformatics