RNA-Seq Analysis of Tumor Resistance in Knockout (KT) vs Wild-Type (WT) Mice

- Lohit Marla

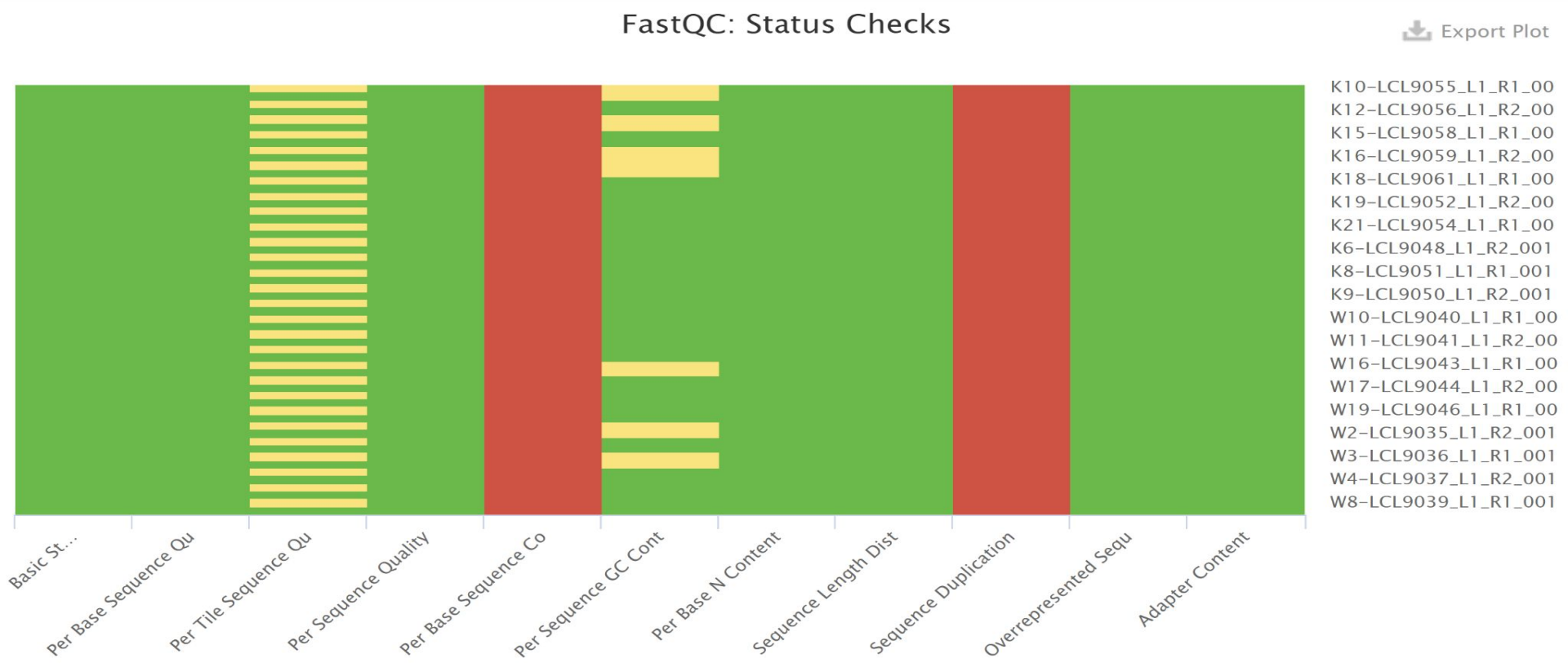
Agenda

1. Steps Accomplished
2. Data Quality Reports
3. Results
4. Current Progress
5. Achievements

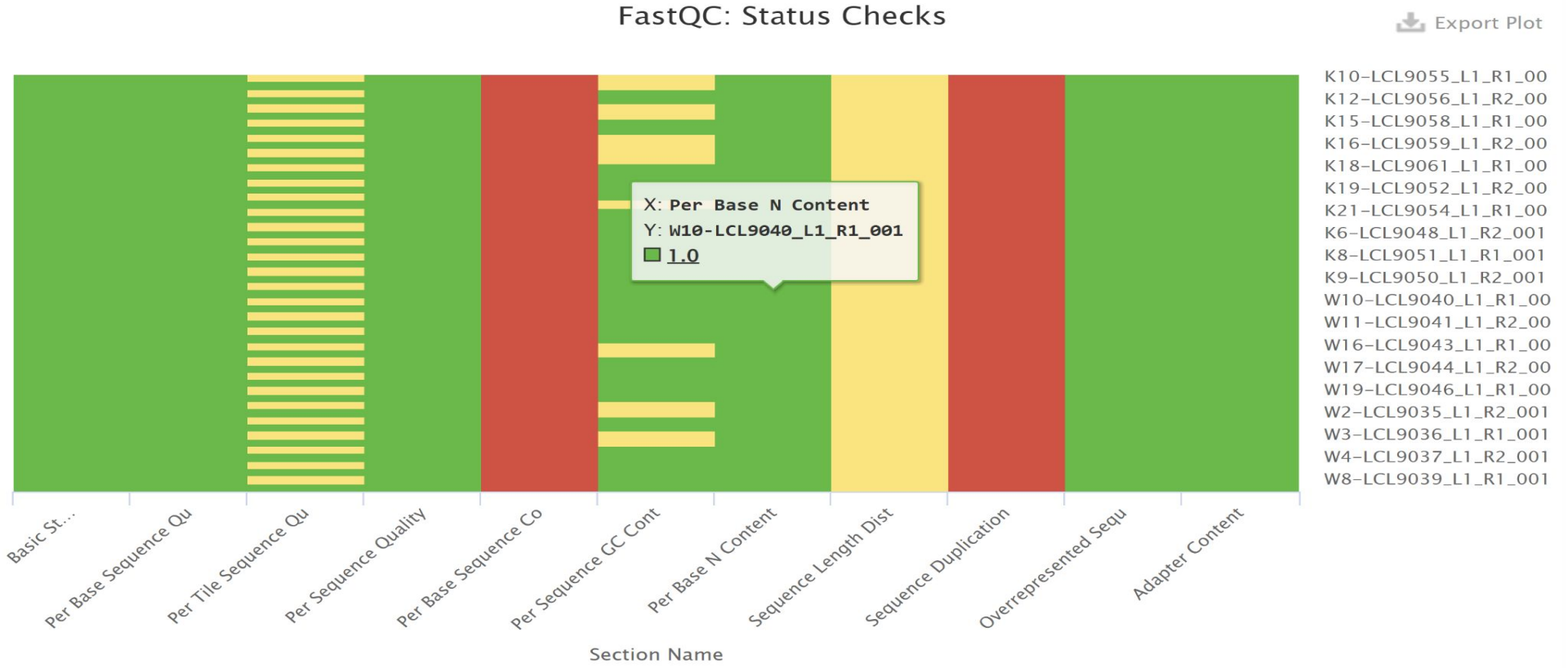
Steps Accomplished - High Level

1. Transferred 757 GB Data to HPC Clusters
2. Unzipping and FastQC, MultiQc reports
3. Trimming using multiple techniques
4. Verifying with data quality reports
5. Read Alignment using STAR
6. Collected statistics of Bam using Samtools
7. Generated featurecounts by GRCm39
8. Loaded the data into Rstudio
9. Processed the raw counts for analysis
10. Applied Deseq2
11. Filtered significant genes
12. Visualized the genes
13. Performed the Gene Ontology
14. Applied Kegg Pathway Analysis

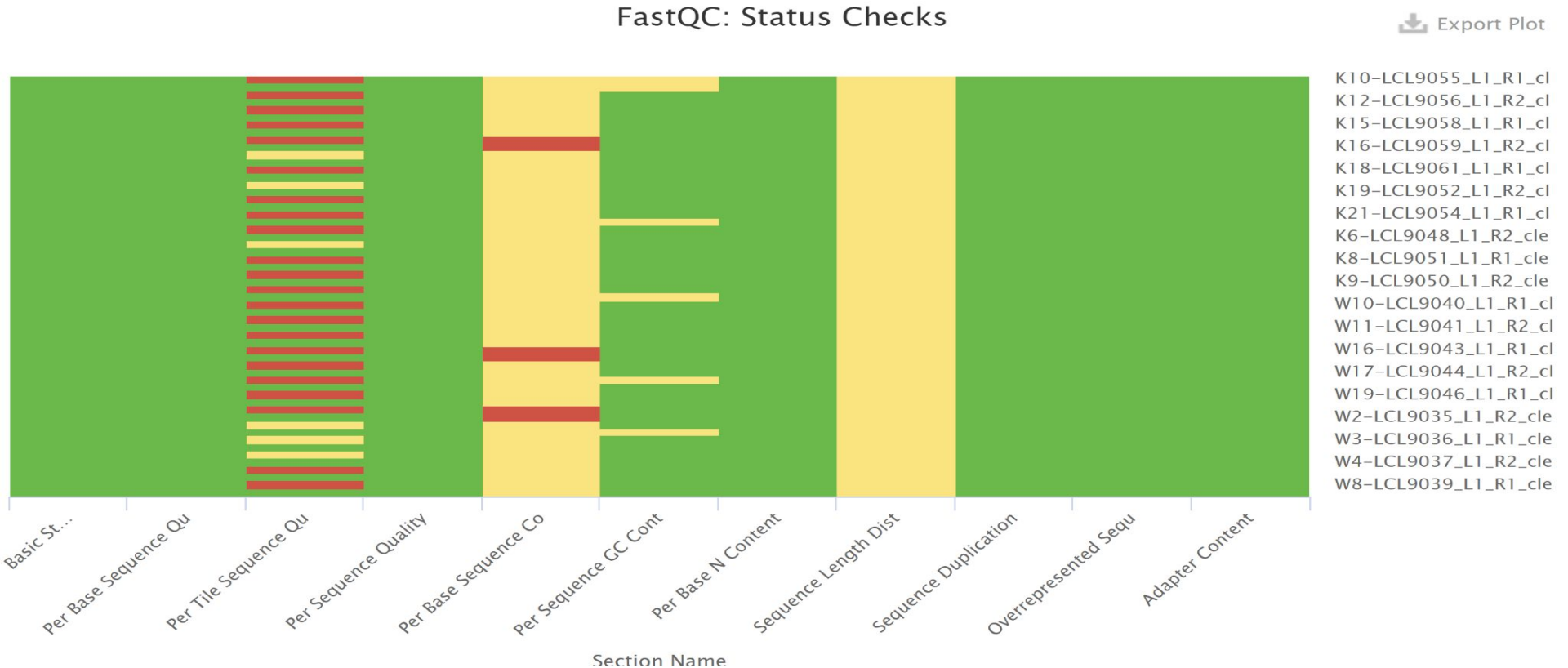
Data Quality Checks - Raw Data



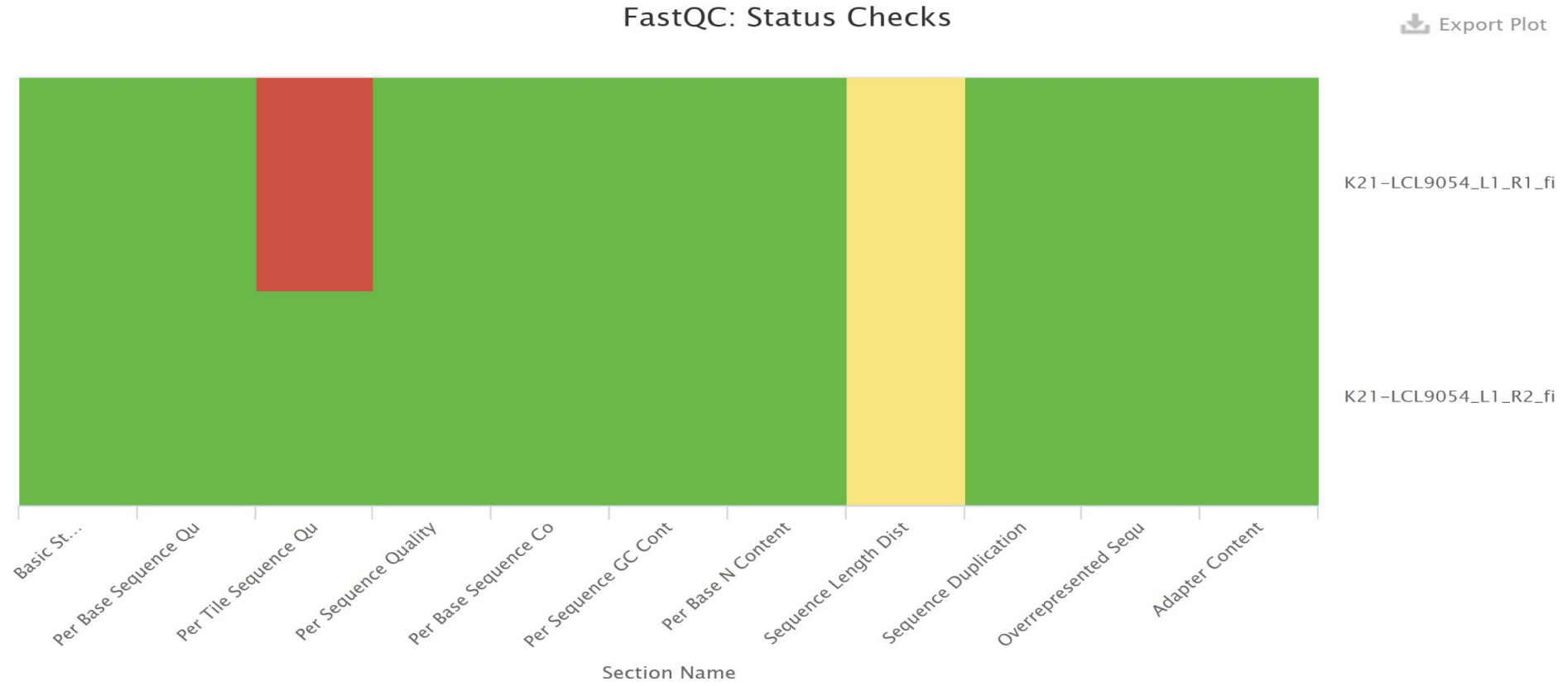
Data Quality Checks - Dedup Reads



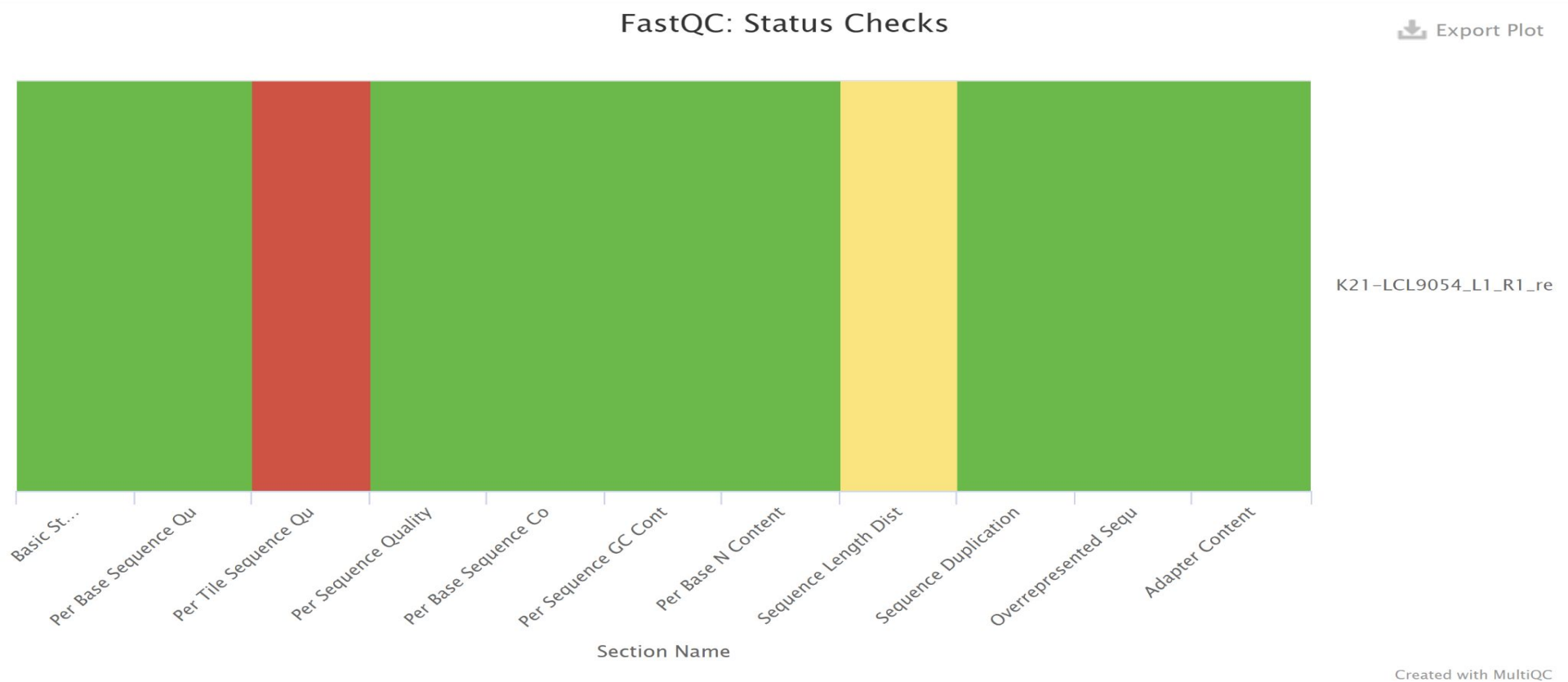
Data Quality Checks - Trimmomatic Min Len - 80%



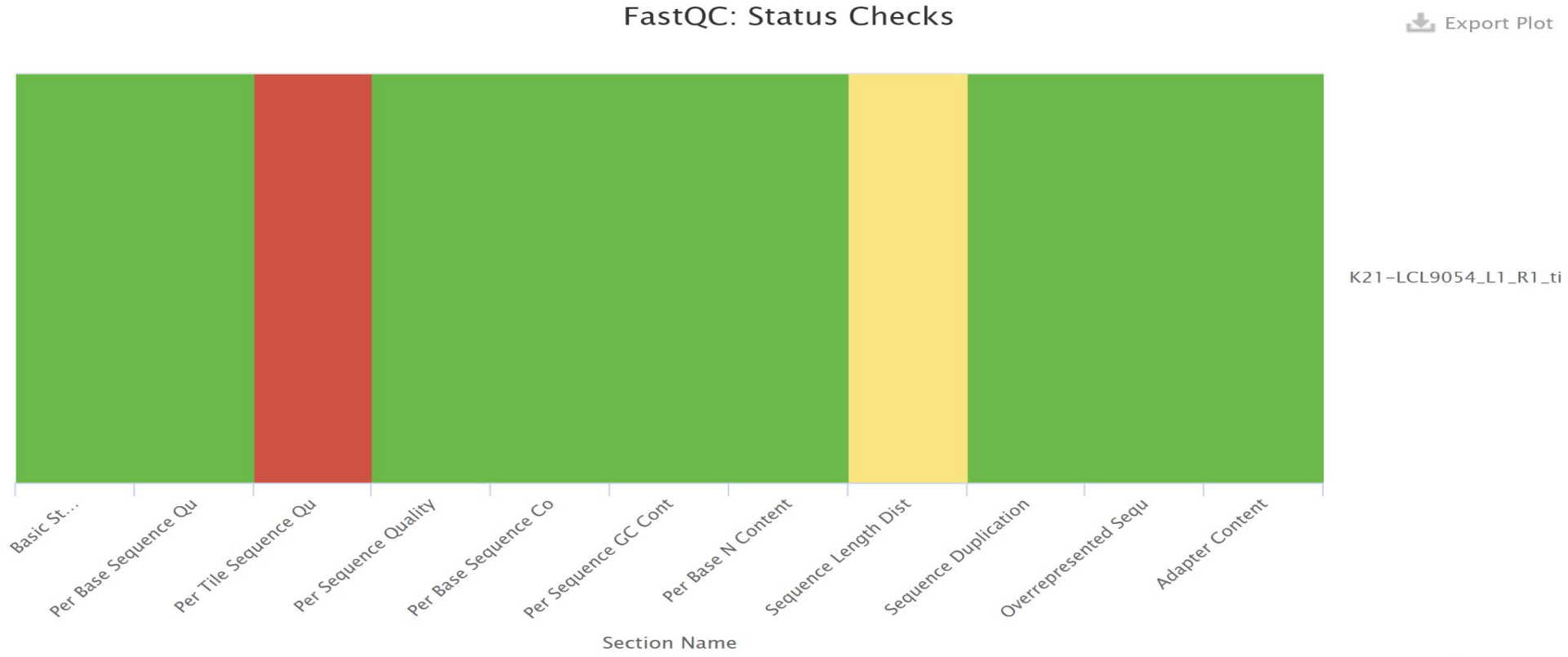
Data Quality Check - Filtered Reads



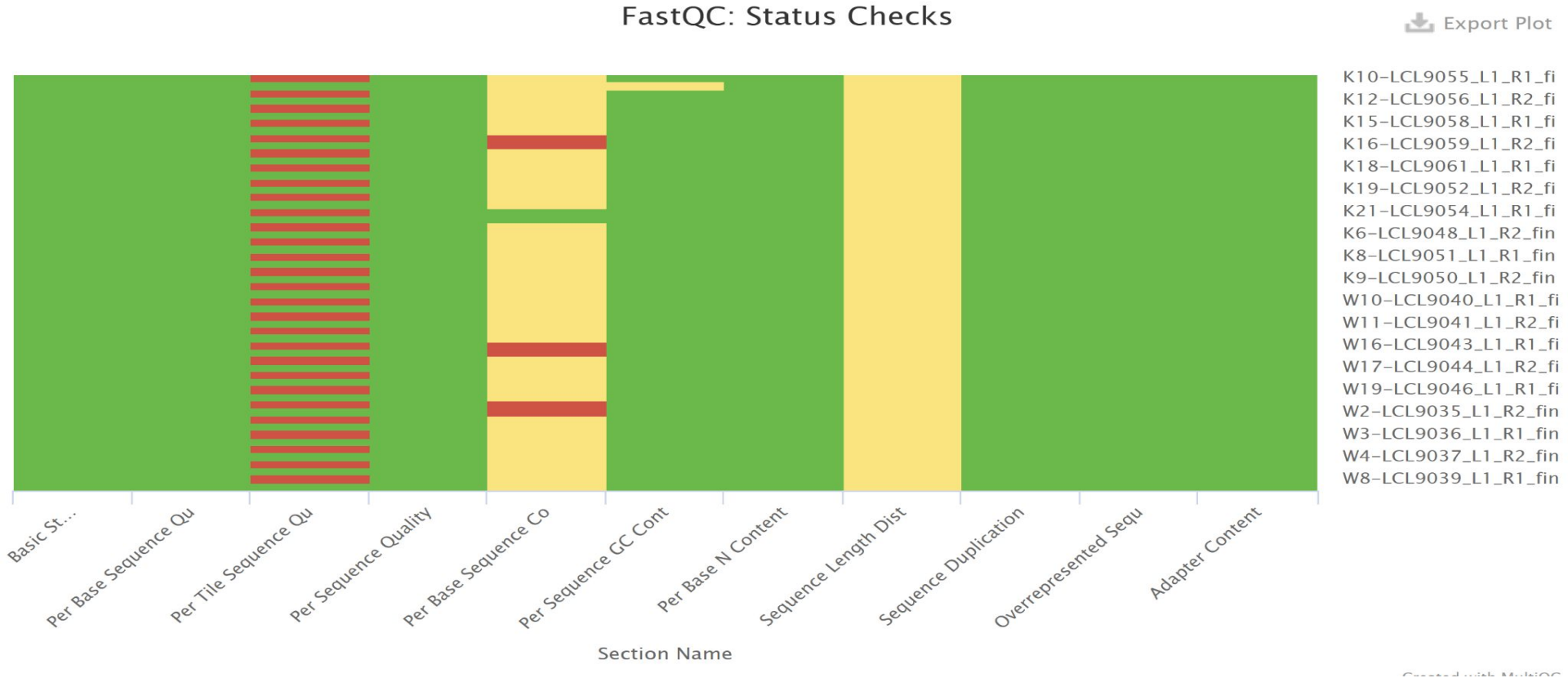
Data Quality Check - Recalibrated Reads



Data Quality Check - Tile Filtered



Data Quality Checks - Final



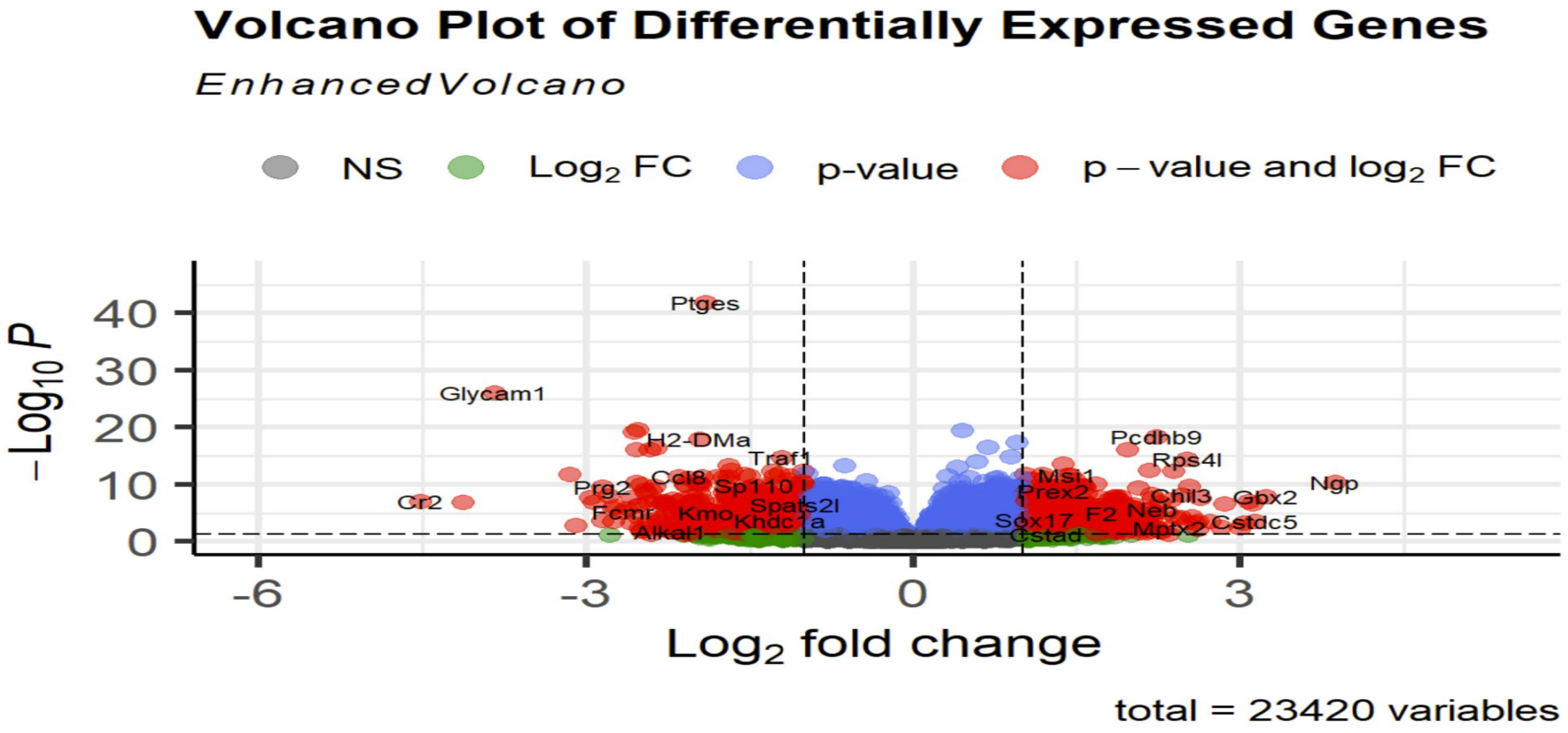
Data Quality Check - Samtools

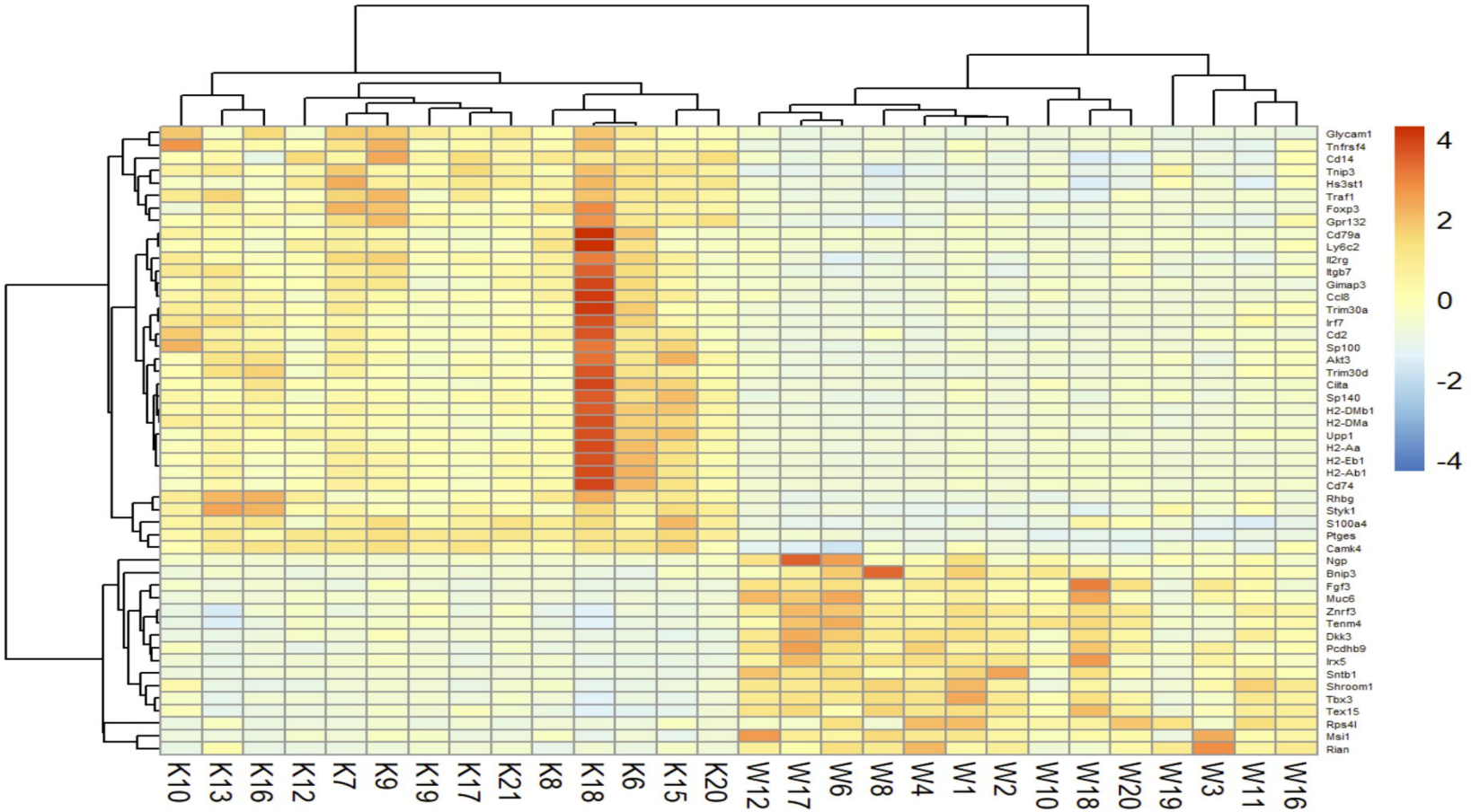
Samtools is a suite of programs for interacting with high-throughput sequencing data. *DOI: 10.1093/bioinformatics/btp352.*

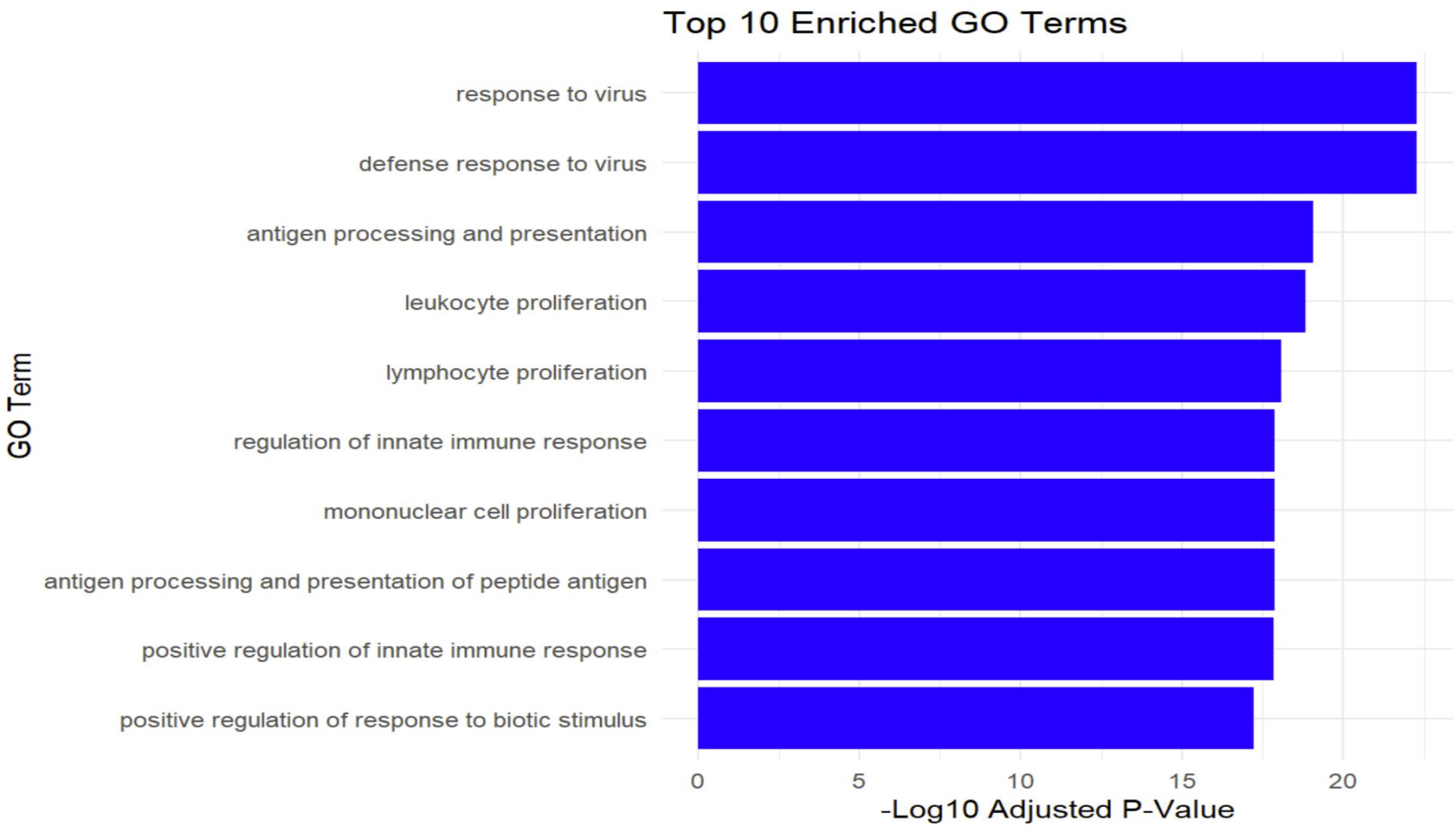
Samtools Flagstat

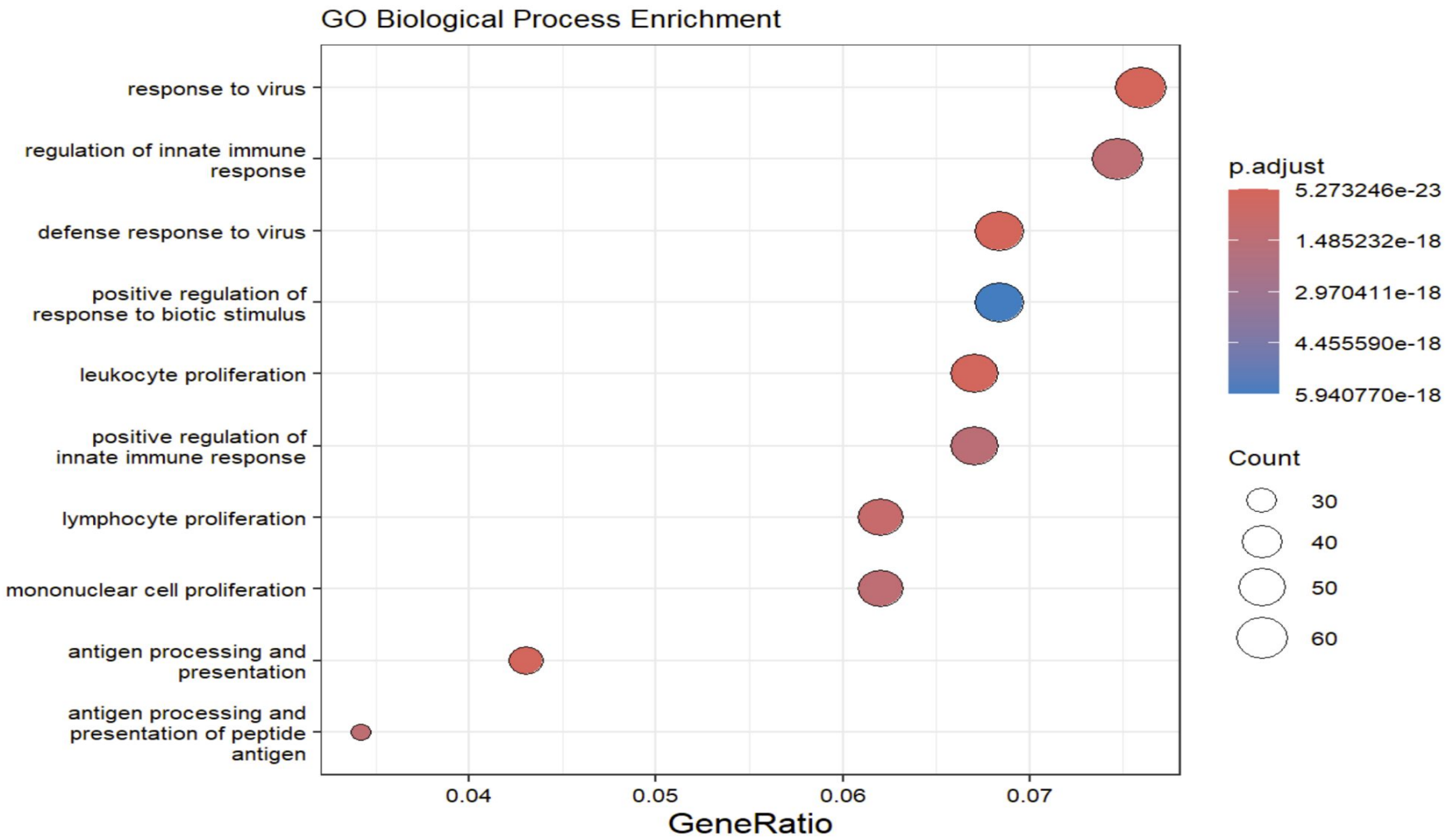
This module parses the output from `samtools flagstat`. All numbers in millions.

Results


Volcano Plot of Differentially Expressed Genes
EnhancedVolcano
NS
Log2 FC
p-value
p – value and log2 FC
−Log10 P
40
30
20
10
0
-6
-3
0
3
Log2 fold change
Ptges
Glycam1
H2-DMa
Traf1
Ccl8
Sp110
Prg2
Cr2
Fcmr
Kmo
Spats2l
Khdc1a
Alkal1
Pcdhb9
Rps4l
Msi1
Prex2
Chil3
Gbx2
Ngp
Sox17
F2
Neb
Cstad
Mptx2
Cstdc5
total = 23420 variables


Glycam1
Tnfrsf4
Cd14
Tnip3
Hs3st1
Traf1
Foxp3
Gpr132
Cd79a
Ly6c2
Il2rg
Itgb7
Gimap3
Ccl8
Trim30a
Irf7
Cd2
Sp100
Akt3
Trim30d
Cita
Sp140
H2-DMb1
H2-DMa
Upp1
H2-Aa
H2-Eb1
H2-Ab1
Cd74
Rhbg
Styk1
S100a4
Ptges
Camk4
Ngp
Bnip3
Fgf3
Muc6
Znrf3
Tenm4
Dkk3
Pcdhb9
Irx5
Sntb1
Shroom1
Tbx3
Tex15
Rps4l
Msi1
Rian
4
2
0
-2
-4
K10
K13
K16
K12
K7
K9
K19
K17
K21
K8
K18
K6
K15
K20
W12
W17
W6
W8
W4
W1
W2
W10
W18
W20
W19
W3
W11
W16


Top 10 Enriched GO Terms
response to virus
defense response to virus
antigen processing and presentation
leukocyte proliferation
lymphocyte proliferation
regulation of innate immune response
mononuclear cell proliferation
antigen processing and presentation of peptide antigen
positive regulation of innate immune response
positive regulation of response to biotic stimulus
GO Term
0
5
10
15
20
-Log10 Adjusted P-Value


GO Biological Process Enrichment
response to virus
regulation of innate immune response
defense response to virus
positive regulation of response to biotic stimulus
leukocyte proliferation
positive regulation of innate immune response
lymphocyte proliferation
mononuclear cell proliferation
antigen processing and presentation
antigen processing and presentation of peptide antigen
0.04
0.05
0.06
0.07
GeneRatio
p.adjust
5.273246e-23
1.485232e-18
2.970411e-18
4.455590e-18
5.940770e-18
Count
30
40
50
60

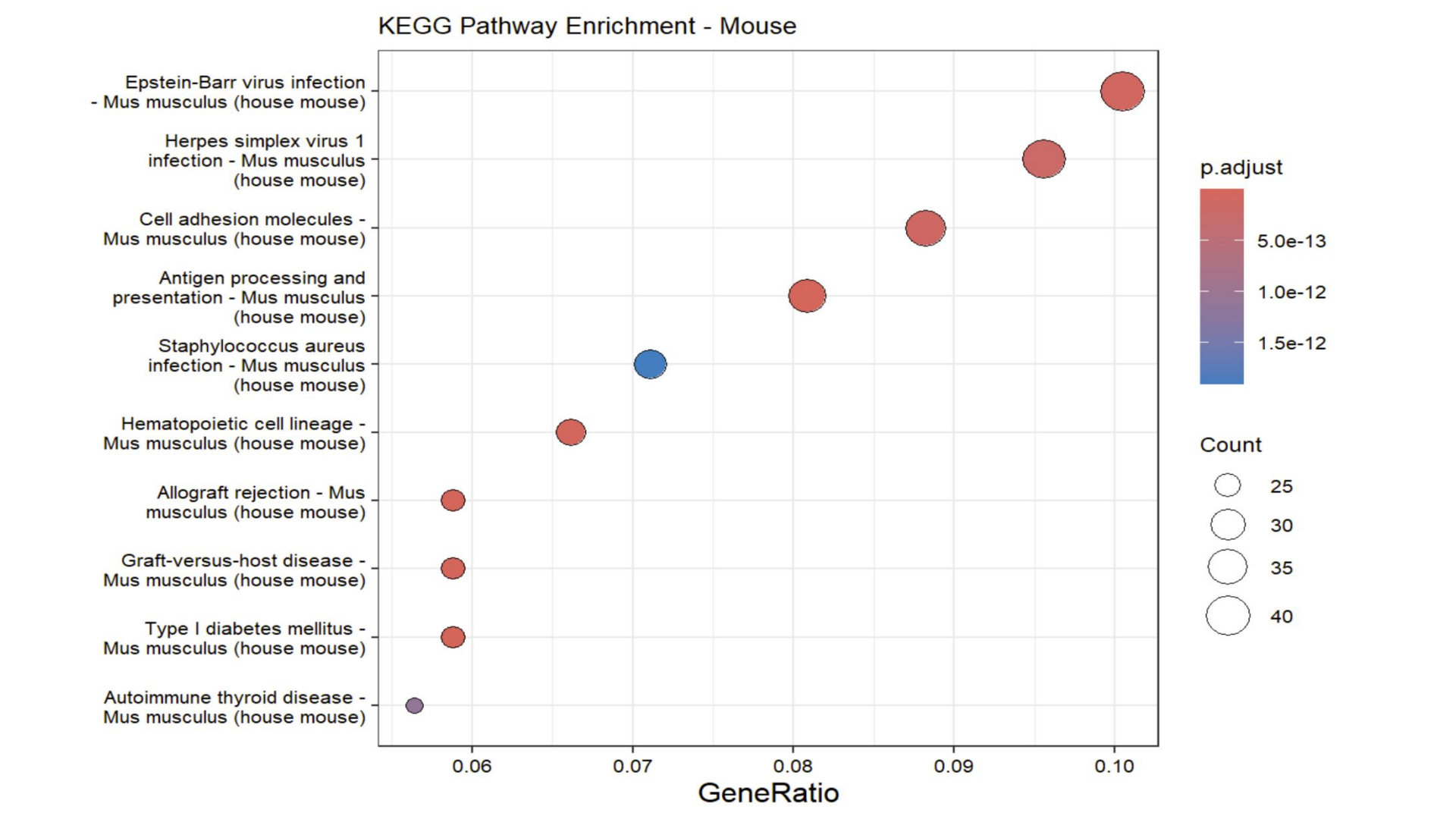
Gene Ontology (GO) Analysis

Key findings:

- "GO analysis revealed a **strong immune system activation signature**.
- The top enriched biological processes included:
  - **Defense response to virus**
  - **Antigen processing and presentation**
  - **Leukocyte and lymphocyte proliferation**
  - **Regulation of innate immune response**
  - **Leukocyte cell-cell adhesion**"

Interpretation:

- "These results suggest that the **KO gene plays a role in immune regulation**.
- The upregulation of antigen presentation and leukocyte activity indicates **a heightened immune response**, possibly due to **autoimmune activation or infection susceptibility**."


KEGG Pathway Enrichment - Mouse
Epstein-Barr virus infection - Mus musculus (house mouse)
Herpes simplex virus 1 infection - Mus musculus (house mouse)
Cell adhesion molecules - Mus musculus (house mouse)
Antigen processing and presentation - Mus musculus (house mouse)
Staphylococcus aureus infection - Mus musculus (house mouse)
Hematopoietic cell lineage - Mus musculus (house mouse)
Allograft rejection - Mus musculus (house mouse)
Graft-versus-host disease - Mus musculus (house mouse)
Type I diabetes mellitus - Mus musculus (house mouse)
Autoimmune thyroid disease - Mus musculus (house mouse)
0.06
0.07
0.08
0.09
0.10
GeneRatio
p.adjust
5.0e-13
1.0e-12
1.5e-12
Count
25
30
35
40

KEGG Pathway Enrichment Analysis

1. **Epstein-Barr Virus Infection (Top Pathway)**

   - "This pathway is highly enriched, suggesting that the **KO gene may impact the immune response to viral infections**."
   - "This could indicate an altered susceptibility to **latent viral infections**."

2. **Herpes Simplex Virus Infection**

   - "Similar to Epstein-Barr virus, this pathway suggests that the **KO gene may play a role in antiviral defense mechanisms**."
   - "Changes in this pathway may **influence susceptibility or immune response to herpesviruses**."

3. **Cell Adhesion Molecules (CAMs)**

   - "This pathway is critical for **immune cell communication and migration**."
   - "The enrichment of this pathway suggests that the KO might be affecting **immune cell adhesion, trafficking, or signaling**."

4. **Antigen Processing and Presentation**

   - "This is a key pathway in **adaptive immunity**, affecting how the immune system recognizes and responds to foreign antigens."
   - "If the KO gene disrupts antigen presentation, it could impact **T-cell activation and immune surveillance**."

5. **Staphylococcus aureus Infection**

   - "This indicates a potential change in **bacterial immune defense mechanisms**."
   - "The KO may alter how the host immune system recognizes and clears bacterial infections."
     -

Why KO Mice Are Protected from Cancer?

Based on pathway enrichment findings, we can hypothesize that KO mice have:

- **Stronger immune surveillance** via **enhanced antigen presentation and T-cell activation**
- **Increased immune system activation**, reducing tumor immune escape
- **Reduced viral oncogenesis**, preventing infection-driven cancers
- **A pro-inflammatory immune environment**, making it harder for tumors to establish
- **A potential autoimmunity-tumor trade-off**, where increased immune activation protects against cancer

Current Progress

1. Revisit the data quality checks and improve the score on fastqc reports.
2. Rerun the pipeline once the data quality checks are well established.
3. Resume the analysis on the data for finding insights.
4. Analyse the data on different scenarios such as:
    a. With genes starting with “Ig”.
    b. WT small tissue samples VS WT large tissue samples.
    c. KO small tissue samples VS KO large tissue samples.
    d. WT small tissue samples VS KO small tissue samples.
    e. WT large tissue samples VS KO large tissue samples.

Achievements

- Globus Transfer to transfer 757 GB data in minutes from hours(scp, rsync).
- End to End pipeline development from reading data to visualising insights.
- On average 60% mapping unique read with Genome (yet to improve more).
- Reduced fastqc generation on all files from 1.5 hours to under 26 mins.
    - Utilised parallel module to achieve this feet on the HPC clusters

Thank You