

High Throughput QC

Quality Control at the Swedish National Genomics Infrastructure

Phil Ewels

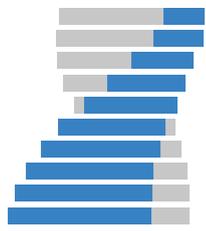
 @ewels

 @tallphil

Bioinfo-Core Workshop

2017-07-24

ISMB 2017, Prague



NGI stockholm

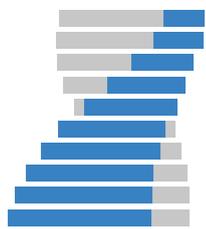


Ackred. nr 1850
Provning
ISO/IEC 17025

ISO accredited facility

- ✓ Library preparation
- ✓ Sequencing
- ✓ Bioinfo analysis

“This means that our services are subject to highly stringent quality control procedures, so that you can be sure that your data is of excellent quality.”



NGI stockholm



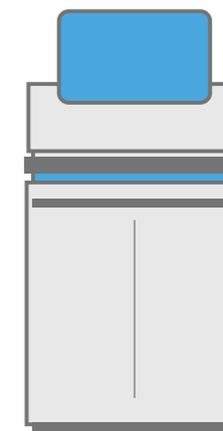
2 x MiSeq



5 x HiSeq 2500



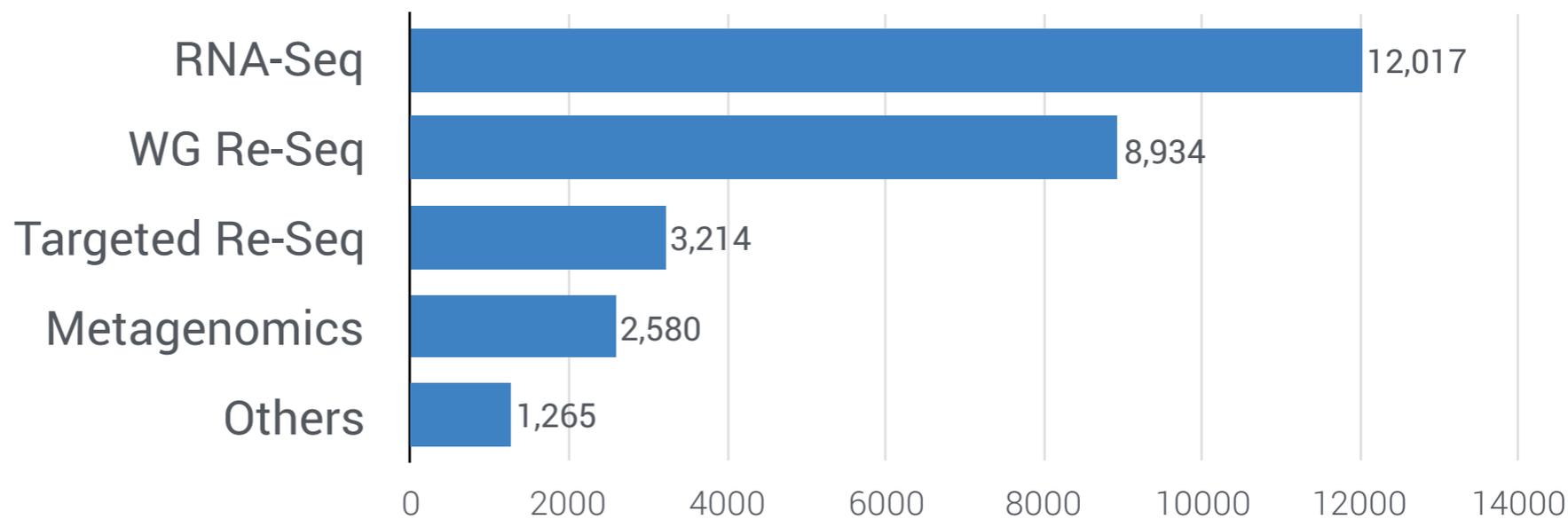
5 x HiSeq X10



NovaSeq

Number of
Samples in 2016
1141 Gbp/day

1X Human Genome
every 4 minutes



QC with lots of samples

Good quality control at scale is essential

Always

Occasionally



Automation

Manual checks

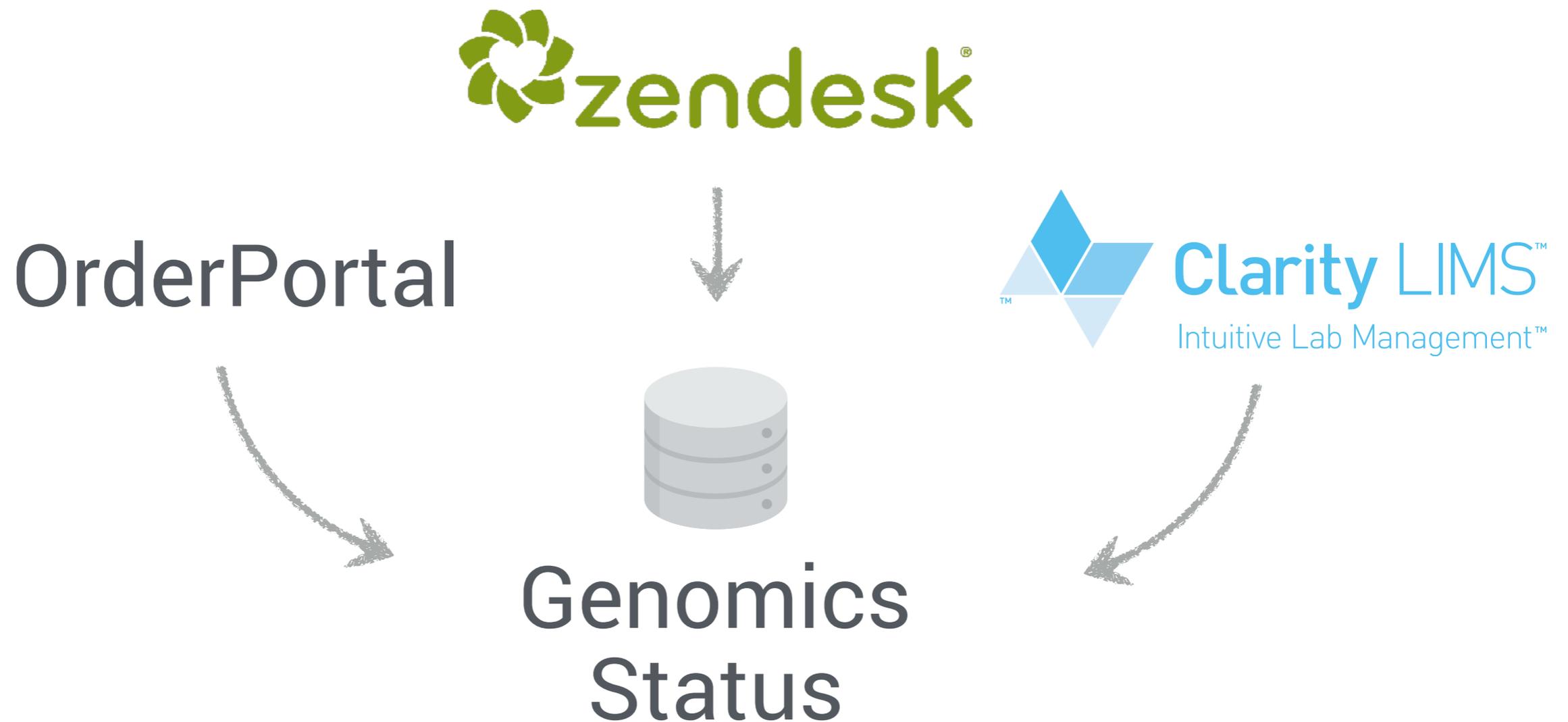
Validations

Visualisation
(spot outliers)

Looking for
trends

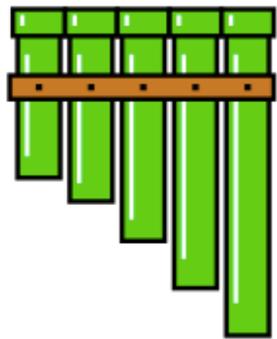
QC Automation

All sample metadata in a database



QC Automation

Automated workflows



PIPER

nextflow

QC Automation

Collection + visualisation of results



MultiQC

The logo for MultiQC features the text 'MultiQC' in a bold, black, sans-serif font. A magnifying glass icon is superimposed over the letter 'Q'. Below the text are three horizontal lines in red, green, and blue.

QC Automation

Collection + visualisation of results



Clarity LIMS™
Intuitive Lab Management™



MultiQC



NATIONAL CTAC
ATC GENOMICS SGT
INFRASTRUCTURE



**Genomics
Status**

P7609, A.Lindstrand_17_02

NGI Portal: 10XPilot_16samples

Ongoing

Hide order dates on timeline

Key Project Information

| | |
|------------------|-----------------------|
| Type | Application |
| Application | WG re-seq |
| Sequencing Setup | HiSeq X, 2x150 |
| Library Method | DNA, Special, -, -, - |

Order Dates

| | |
|----------------------|------------|
| Order Received | 2017-02-17 |
| Contract Sent | 2017-02-22 |
| Plates Sent | 2017-02-21 |
| Contract Received | 2017-03-06 |
| Sample Info Received | 2017-03-08 |
| Samples Received | 2017-03-09 |

Project information

Samples

Running Notes

Order info & Contacts

| | |
|----------------------|--|
| Disposal Of Samples | Destruction one year after project completion |
| Bioinfo. Responsible | Francesco Vezzi |
| Lab Responsible | Joel Gruselius |
| Project Coordinator | Karin Gillner |
| Contact | Johanna.Lundin@ki.se |
| Affiliation | KS |
| Invoice Reference | array-WGS Johanna Lundin |

Project Timeline



Genomics Status

Project Co-ordinator Comment

chromium prep

Final Costs

project description Use with caution! ⓘ

over som ska helgenomsekvenseras med 10X Genomics minimum sample prep. Klinisk utveckling där vi kör prover ut med array och helgenom. Kontakt Max Käller.

Lab

| | |
|---------------------------|-------------|
| Sample type | Genomic DNA |
| Library type | |
| Number of samples ordered | 16 |
| Library | 10X |

| | |
|------------------------------|------------------------------|
| Best Practice Bioinformatics | No |
| Bioinfo Analysis | WG re-seq |
| Organism | |
| Reference Genome | Human (Homo sapiens, GRCh37) |

| | |
|---------------|-------|
| Low Diversity | False |
| Custom Primer | False |
| Lanes Ordered | 16 |

Samples

Sample Status

Progress summary

P7609, A.Lindstrand_17_02

NGI Portal: 10XPilot_16samples

Ongoing

Hide order dates on timeline

Key Project Information

| | |
|------------------|-----------------------|
| Type | Application |
| Application | WG re-seq |
| Sequencing Setup | HiSeq X, 2x150 |
| Library Method | DNA, Special, -, -, - |

Order Dates

| | |
|----------------------|------------|
| Order Received | 2017-02-17 |
| Contract Sent | 2017-02-22 |
| Plates Sent | 2017-02-21 |
| Contract Received | 2017-03-06 |
| Sample Info Received | 2017-03-08 |
| Samples Received | 2017-03-09 |

Project Timeline

| | | |
|------------------------|------------|-----|
| Days in Production | 72 | |
| Open date | 2017-03-09 | |
| QC Start Date | 2017-03-28 | |
| Queue Date | 2017-03-28 | JGR |
| Library Prep Start | 2017-04-10 | |
| QC Library Finished | 2017-04-10 | |
| Sequencing Start | 2017-04-11 | |
| All Samples Sequenced | 2017-04-11 | FV |
| All Raw Data Delivered | 2017-05-31 | |
| BP Analysis Completed | 2017-05-30 | FV |
| Close Date | - | |

Project Co-ordinator Comment

10X Chromium prep

Internal Costs

-

User project description **Use with caution!** ⓘ

16 prover som ska helgenomsekvenseras med 10X Genomics Chromium sample prep. Klinisk utveckling där vi kör prover parallellt med array och helgenom. Kontakt Max Käller.

Project information

Samples

Running Notes

User communication

Links

Open MultiQC report

Open bioinfo tab

Search Samples: Search...

Copy table

Column Presets:

Default view

Reception Control

Library Validation

Reads Aggregation

RecCtrl Plates

Filter Columns

| Sample | Submitted Name | Sample runs | Status (manual) | Initial QC Status | Prep Start Date | Workset Setup | Library Concentration | Average Size bp | Prep Status | Library Well Location | Library Volume (ul) | Library Validation Initials |
|-----------|----------------|---|-----------------|-------------------|-----------------|---------------|-----------------------|-----------------|-------------|-----------------------|---------------------|-----------------------------|
| P7609_101 | 2017-00007 | 1_170411_AHH2Y3ALXX_CAGTACTG-AGTAGTCT-GCAGTAGA-TTCCCGAC | In Progress | Passed | 2017-04-10 | 24-267010 | 61.87 | 622 | Passed | A:1 | 16 | JGR |
| P7609_102 | 2017-00067 | 2_170411_AHH2Y3ALXX_GTGTATTA-TGTGCGGG-ACCATAAC-CAACGCCT | In Progress | Passed | 2017-04-10 | 24-267010 | 82.87 | 618 | Passed | B:1 | 16 | JGR |
| P7609_103 | 2017-00116 | 3_170411_AHH2Y3ALXX_TCAGCCGT-CAGAGGCC-GGTCAATA-ATCTTTAG | In Progress | Passed | 2017-04-10 | 24-267010 | 68.91 | 620 | Passed | C:1 | 16 | JGR |
| P7609_104 | 2017-00406 | 4_170411_AHH2Y3ALXX_ACATTACT-TTTGGGTA-CAGCCCAC-GGCAATGG | In Progress | Passed | 2017-04-10 | 24-267010 | 69.40 | 620 | Passed | D:1 | 16 | JGR |
| P7609_105 | 2017-00803 | 5_170411_AHH2Y3ALXX_AGGTATTG-CTCCTAGT-TCAAGGCC-GATGCCAA | In Progress | Passed | 2017-04-10 | 24-267010 | 68.24 | 635 | Passed | E:1 | 16 | JGR |

P7609, A.Lindstrand_17_02

Application **WGreseq** Francesco Vezzi

Open MultiQC report

Select view:

Save Changes

| Sample | | Multiplex QC | | | Quality | | Contamination | | WG Reseq | | | Data Delivered | |
|-----------|-----------|--------------|----------------------|--------------------|--------------|-----------------|---------------|--------------|---------------|----------|----------------------|----------------|---|
| Sample | Status | Fill row | Undetermined indexes | Unexpected indexes | Sample yield | PhiX error rate | %bases >= Q30 | fastq_screen | blast_wrapper | Coverage | Sequence duplication | Mapping rate | Data Delivered |
| P7609_101 | delivered | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | Pass | Pass | Pass | 2017-11-01 <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_102 | delivered | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | Pass | Warning | Pass | 2017-11-01 <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_103 | delivered | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | Pass | Pass | Pass | 2017-11-01 <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_104 | bp-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | Fail | Warning | Fail | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_105 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_106 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_107 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_108 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_109 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_110 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_111 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_112 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |
| P7609_113 | qc-done | → | Pass | Pass | Pass | Pass | Pass | Pass | N/A | ? | ? | ? | <input type="text" value=""/> <input type="button" value="x"/> <input type="button" value="↓"/> |

P1234: Test_NGI_Project

P1234: Test_NGI_Project

This is an example project. All identifying data has been removed.

Contact E-mail: phil.ewels@scilifelab.se
Application Type: RNA-seq
Sequencing Platform: HiSeq 2500 High Output V4
Sequencing Setup: 2x125
Reference Genome: hg19

ngi-rna/data

M Seqs

| Sample | Green | Purple | Blue | Orange | Yellow | M Seqs |
|------------|-------|--------|-------|--------|--------|--------|
| P1234_1003 | 64.7% | 21.7 | 11.0% | 72.3% | 50% | 33.7 |
| P1234_1004 | 55.2% | 17.0 | 13.2% | 73.4% | 51% | 31.2 |
| P1234_1005 | 53.0% | 17.7 | 15.9% | 75.8% | 52% | 33.8 |
| P1234_1006 | 52.7% | 16.1 | 14.1% | 73.8% | 52% | 30.8 |
| P1234_1007 | 33.0% | 7.0 | 32.0% | 80.5% | 52% | 21.8 |
| P1234_1008 | 27.5% | 4.3 | 44.2% | 79.1% | 50% | 16.7 |
| P1234_1009 | 52.3% | 10.5 | 20.9% | 64.2% | 46% | 20.5 |
| P1234_1010 | 44.2% | 4.9 | 24.9% | 79.2% | 48% | 11.4 |
| P1234_1011 | 42.2% | 11.6 | 27.1% | 45.1% | 50% | 28.0 |
| P1234_1012 | 40.6% | 10.6 | 27.1% | 45.2% | 51% | 26.6 |
| P1234_1013 | 60.8% | 35.6 | 14.0% | 51.6% | 46% | 59.0 |

Toolbox

- 🔍
- A
- 🔄
- ⬇️
- 🏠
- ?



QC with lots of samples

Always

Occasionally

Automation

Manual checks

Validations

Visualisation
(spot outliers)

Looking for
trends

Long term trends

coming soon...

Multi⊕QC



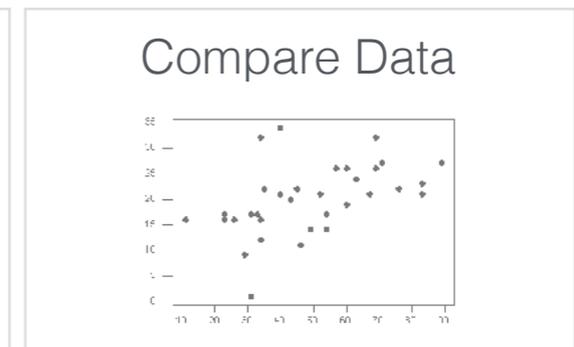
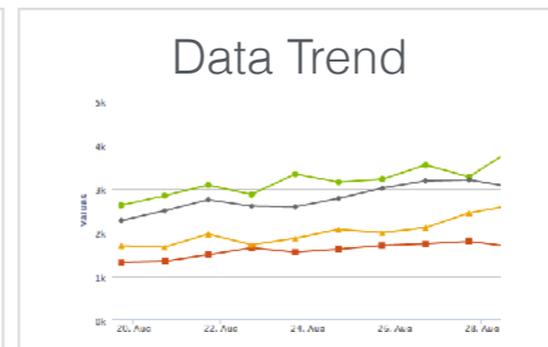
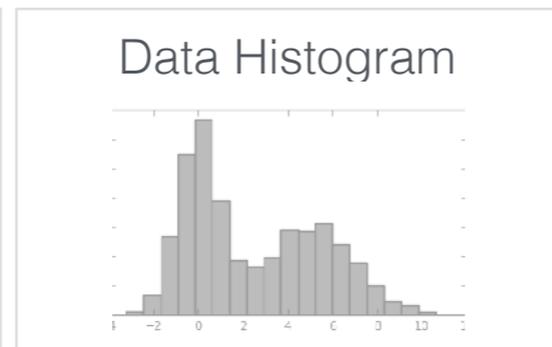
Mega⊕QC

Long term trends



Data available from
2253 samples, 738 reports

① Choose a plot type



② Select data

Long term trends



Data available from
2253 samples, 738 reports

① Choose a plot type Data Trend

② Select samples [837 samples]

 Date: 2017.01.01 - 2017.03.01

 Application Type: WGS

+ Add Filter

Metadata field:

Minimum:

Maximum

Long term trends

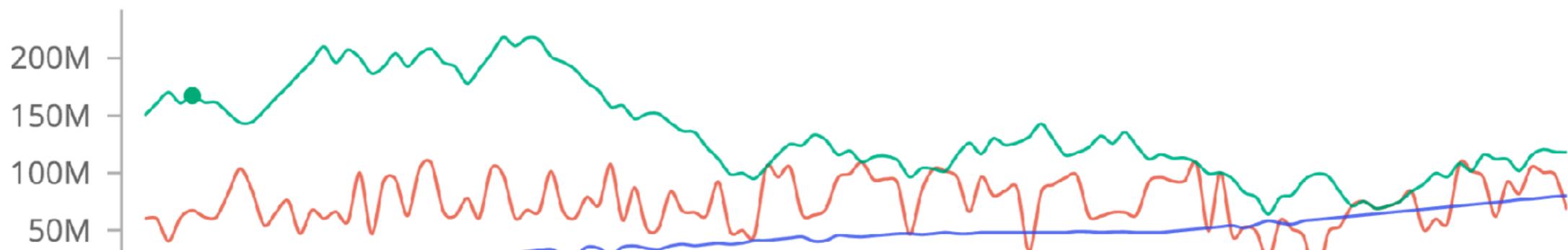
MegaQC

Data available from
2253 samples, 738 reports

① Choose a plot type Data Trend

② Select samples [837 samples]

③ Plot data 🗑 Picard: % Dups 🗑 Qualimap: > 30X [add data]



Validations

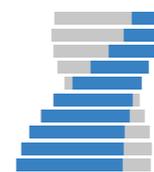
In-depth reproducibility checks

- **Each lab technique has a protocol**

- Changes require validation

- **Internal audits**

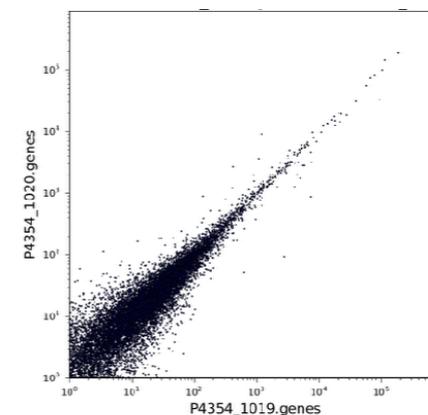
- Horizontal (one step)
- Vertical (one sample)



NGI

stockholm

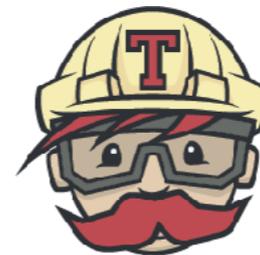
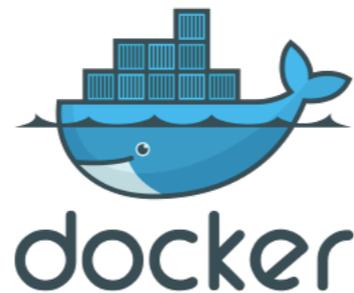
uppsala



r² = 0.991297

Analysis Validation

- Versioned pipeline releases
 - New versions require a validation test
 - Runs logged and reported
- Continuous integration tests



Travis CI

build **passing**

Analysis Validation



NGI-RNAseq: RNA-Seq Best Practice v1.1

Run Name: Test RNA Run: 1496849684

NGI-RNAseq execution completed successfully!

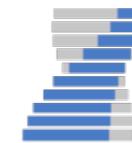
The workflow was completed at **Wed Jun 07 17:35:43 CEST 2017** (duration: 55.2s)

The command used to launch the workflow was as follows:

```
nextflow run ../main.nf -resume -name Test RNA Run: 1496849684 -profile testing --gtf
./test_data/ngi-rna_test_set/genes.gtf --bed12 ./test_data/ngi-rna_test_set/genes.bed --
hisat2_index ./test_data/ngi-rna_test_set/r64/ --aligner hisat2 --singleEnd --reads
./test_data/ngi-rna_test_set/*.fastq.gz
```

Pipeline Configuration:

| | |
|----------------|---|
| Run Name | Test RNA Run: 1496849684 |
| Reads | ./test_data/ngi-rna_test_set/*.fastq.gz |
| Data Type | Single-End |
| Strandedness | None |
| Genome | false |
| Aligner | HISAT2 |
| HISAT2 Index | ./test_data/ngi-rna_test_set/r64/ |
| GTF Annotation | ./test_data/ngi-rna_test_set/genes.gtf |
| BED Annotation | ./test_data/ngi-rna_test_set/genes.bed |
| Current home | /Users/philewels |
| Current user | philewels |
| Current path | /Users/philewels/GitHub/NGI-RNAseq/tests |
| Working dir | /Users/philewels/GitHub/NGI-RNAseq/tests/work |



NGI-RNAseq



NGI-smRNAseq*



NGI-MethylSeq*



NGI-ChIPseq*

Conclusion

Phil Ewels

 phil.ewels@scilifelab.se

 [ewels](#)

 [tallphil](#)

<http://opensource.scilifelab.se>

<http://multiqc.info>

<http://github.com/SciLifeLab>

Acknowledgements

Max Käller

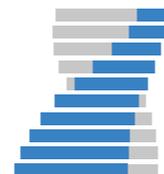
Rickard Hammarén

Denis Moreno

Francesco Vezzi

NGI Stockholm Genomics
Applications Development Group

SciLifeLab



NGI stockholm