

# Outline

## Riboprofiling pipeline

- Samples

- Trimming reads

- Removing contaminants

- Genome mapping STAR

- Comparing stats with references

## Quantification

- Correlation between data types

- Ratio CDS/UTR - riboprofiling VS RNAseq

- Ratio protein coding/long non coding RNA

# Outline

## Riboprofiling pipeline

- Samples

- Trimming reads

- Removing contaminants

- Genome mapping STAR

- Comparing stats with references

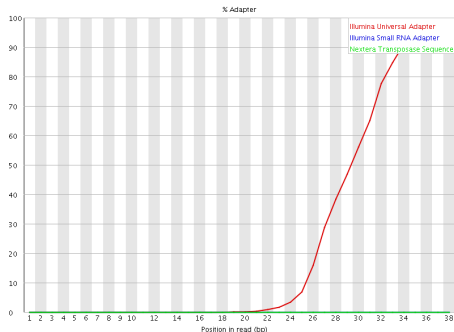
## Quantification

## Riboprofiling samples

	6h	18h	120h
Reads	42,130,444	45,982,977	44,077,266
Sequence length	50	50	50
%GC	59	55	56

# Trimming

## ✖ Adapter Content



- ▶ min length: 25
- ▶ min adapter alignment length: 5
- ▶ unclipped discarded
- ▶ first base discarded

# Trimming - Cutadapt

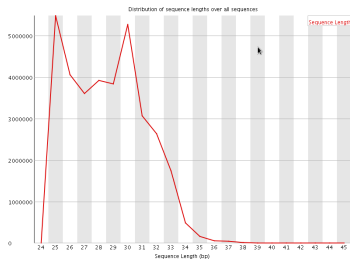
`-quality-cutoff=10` Trim low-quality bases from 3' ends of each read before adapter removal.

id	input	output	discarded TooShort	reads with Adapter
H006	42,130,444	34,439,444 (81.7%)	6,933,352 (16.5%)	41,271,051 (98.0%)
H018	45,982,977	39,441,134 (85.8%)	5,682,240 (12.4%)	44,907,869 (97.7%)
H120	44,077,266	32,501,482 (73.7%)	10,638,847 (24.1%)	42,985,050 (97.5%)

# Sequence length distribution after trimming - Cutadapt

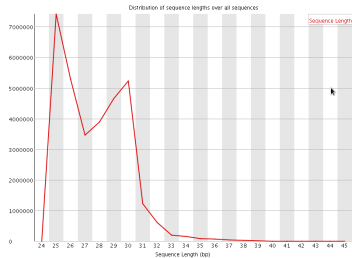
① Sequence Length Distribution

H006



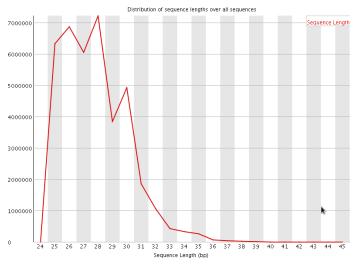
① Sequence Length Distribution

H120



① Sequence Length Distribution

H018



time	sequence length
H006	25-44
H018	25-44
H120	25-44

# Removing contaminants - rRNA

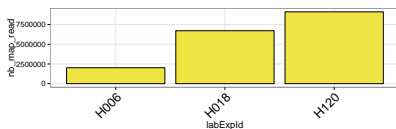
- ▶ STAR: without split mapping; max 10 multimaps; outFilterMatchNmin 16

id	reads processed	uniquely mapped	multiple loci	too many loci	discarded too short
H006	34,439,444	29,465,892 (85.56%)	45,281 (0.13%)	810 (0.00%)	14.31%
H018	39,441,134	12,559,738 (31.84%)	193,421 (0.49%)	4,945 (0.01%)	67.65%
H120	32,501,482	13,389,260 (41.20%)	119,345 (0.37%)	12,214 (0.04%)	58.39%

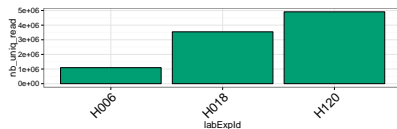
# Genome mapping STAR - max 10 multimaps

- ▶ Unaligned reads from rRNA mapping
- ▶ `-outFilterMatchNmin 16`
- ▶ **max 10 multimaps**

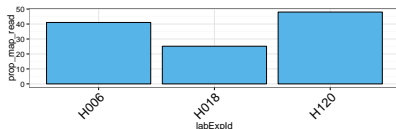
## Number of mapped reads



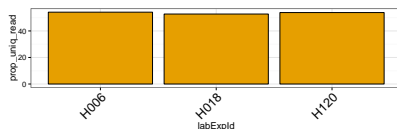
## Number of uniquely mapped reads



## Proportion of mapped reads



## Proportion of uniquely mapped reads

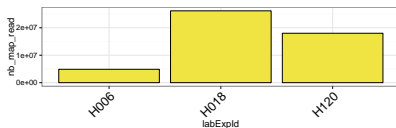




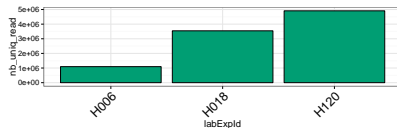
# Genome mapping STAR - max 100 multimaps

- ▶ Unaligned reads from rRNA mapping
- ▶ `-outFilterMatchNmin 16`
- ▶ **max 100 multimaps**

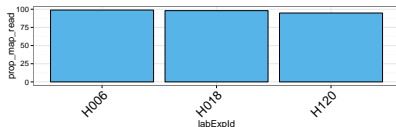
## Number of mapped reads



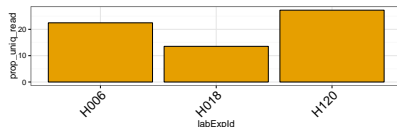
## Number of uniquely mapped reads



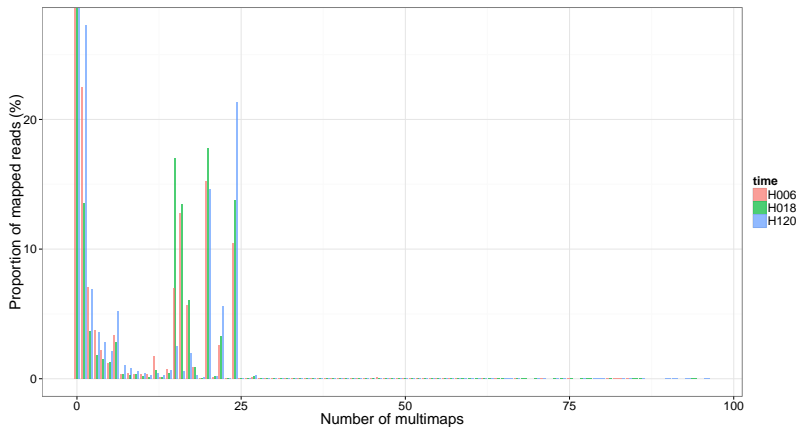
## Proportion of mapped reads



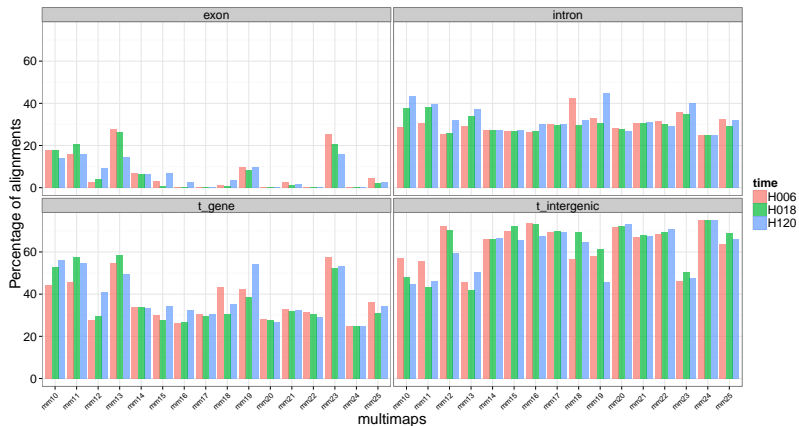
## Proportion of uniquely mapped reads



# Distribution of multimaps



# Distribution of multimaps



# Comparing stats with references

## Fatima's lab

	RPF1 Mock	RPF1 KD	RPF2 Mock	RPF2 KD	RPF3 Mock	RPF3 KD
Total Reads	110,347,659	86,886,294	70,701,122	61,970,268	187,463,074	147,577,976
size-selected (22-36)	105,447,994	80,382,820	65,538,346	60,464,127	159,938,111	134,116,969
After rRNA,tRNA filtering	57,452,964	40,704,687	8,788,516	8,790,692	45,542,640	29,858,389
Aligned (-rRNA,tRNA)	22,252,759	15,717,018	4,376,899	4,232,701	23,323,320	12,759,154
In annotated CDSs	16,520,263	11,497,323	2,451,068	2,742,875	17,831,699	10,158,765

## Current stats ERC

	6h	18h	120h
Reads	42,130,444	45,982,977	44,077,266
Size selected (25-44)	34,439,444	39,441,134	32,501,482
After rRNA filtering	4,928,271	26,687,975	18,992,877
Aligned (-rRNA, 10mm)	2,022,381 (41.04%)	6,718,578 (25.17%)	9,117,294 (48.00%)
Uniquely aligned (-rRNA, 10mm)	1,095,120 (54.15%)	3,543,776 (52.75%)	4910138 (53.86%)
Exonic mapping primary alignment	918,722	3,241,787	3,881,137
Exonic uniquely mapped	697,556	2,486,593	3,071,028

# Comparing stats with references

Genik et al. 2015, Genome Res.

→ How many initial reads and reads mapping to coding regions you had approximately?

This varies quite a bit based on the efficiency of rRNA depletion.

Even after using the oligo-depletion a large fraction goes to rRNAs.

Just to give you an idea, I picked one of the 100 ribosome profiling libraries we sequenced:

Reads	16,739,559
After trimming	14,866,539 (88%)
Mapped to rRNA	64.49%
Reads after rRNA removal	5,265,866
aligned uniquely to APPRIS transcriptome	3,355,633 (63.72%)
aligned > 1 times	722,762 (13.73%)
remaining	1,187,471
mapped to the genome but not the transcriptome	476,051

Current stats ERC

	6h	18h	120h
Reads	42,130,444	45,982,977	44,077,266
Size selected (25-44)	34,439,444	39,441,134	32,501,482
After rRNA filtering	4,928,271	26,687,975	18,992,877
Aligned (-rRNA, 10mm)	2,022,381 (41.04%)	6,718,578 (25.17%)	9,117,294 (48.00%)
Uniquely aligned (-rRNA, 10mm)	1,095,120 (54.15%)	3,543,776 (52.75%)	4910138 (53.86%)
Exonic mapping primary alignment	918,722	3,241,787	3,881,137
Exonic uniquely mapped	697,556	2,486,593	3,071,028

# Outline

Riboprofiling pipeline

## Quantification

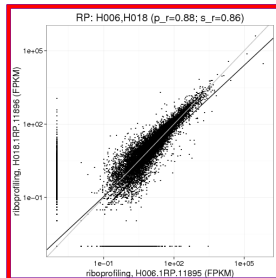
Correlation between data types

Ratio CDS/UTR - riboprofiling VS RNAseq

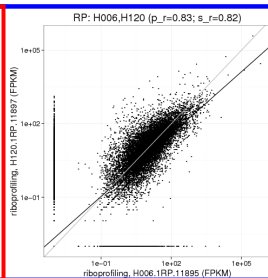
Ratio protein coding/long non coding RNA

# Correlation between RP samples (50,666 genes)

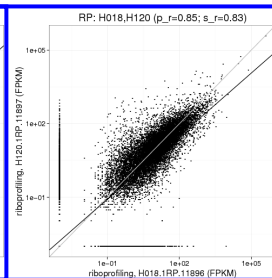
**H006 VS H018**



**H006 VS H120**



**H018 VS H120**



# H006 (6,109 genes)

## Spearman

id	H006.1RP	H006.2C	H006.2N	H006.2T	H006.3C	H006.3N	H006.3T	H006.2P
H006.1RP	1	0.879	0.759	0.864	0.853	0.809	0.885	0.454
H006.2C	0.879	1	0.812	0.925	0.973	0.894	0.956	0.407
H006.2N	0.759	0.812	1	0.895	0.774	0.91	0.9	0.331
H006.2T	0.864	0.925	0.895	1	0.903	0.922	0.97	0.398
H006.3C	0.853	0.973	0.774	0.903	1	0.921	0.923	0.38
H006.3N	0.809	0.894	0.91	0.922	0.921	1	0.927	0.344
H006.3T	0.885	0.956	0.9	0.97	0.923	0.927	1	0.422
H006.2P	0.454	0.407	0.331	0.398	0.38	0.344	0.422	1



# H018 (6,109 genes)

## Spearman

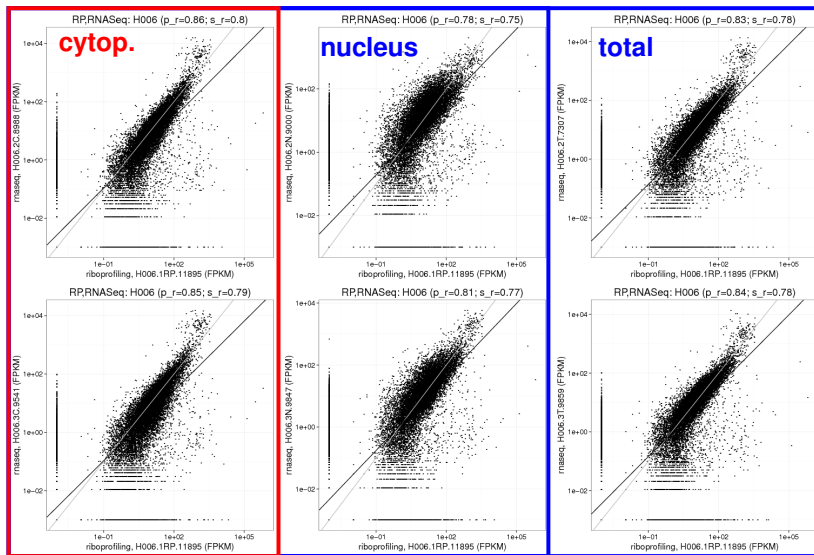
id	H018.1RP	H018.2C	H018.2N	H018.2T	H018.3C	H018.3N	H018.3T	H018.2P
H018.1RP	1	0.874	0.749	0.87	0.855	0.776	0.882	0.466
H018.2C	0.874	1	0.828	0.94	0.988	0.895	0.948	0.409
H018.2N	0.749	0.828	1	0.897	0.809	0.906	0.896	0.334
H018.2T	0.87	0.94	0.897	1	0.924	0.904	0.983	0.419
H018.3C	0.855	0.988	0.809	0.924	1	0.918	0.928	0.396
H018.3N	0.776	0.895	0.906	0.904	0.918	1	0.902	0.338
H018.3T	0.882	0.948	0.896	0.983	0.928	0.902	1	0.428
H018.2P	0.466	0.409	0.334	0.419	0.396	0.338	0.428	1

# H120 (6,109 genes)

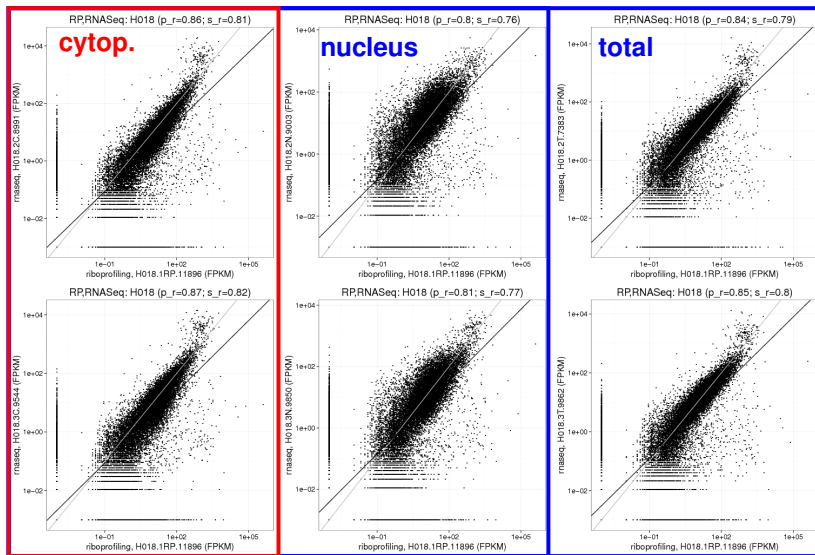
## Spearman

id	H120.1RP	H120.2C	H120.2N	H120.2T	H120.3C	H120.3N	H120.3T	H120.2P
H120.1RP	1	0.852	0.826	0.897	0.774	0.839	0.905	0.404
H120.2C	0.852	1	0.872	0.939	0.956	0.918	0.936	0.372
H120.2N	0.826	0.872	1	0.905	0.812	0.971	0.926	0.306
H120.2T	0.897	0.939	0.905	1	0.86	0.921	0.979	0.398
H120.3C	0.774	0.956	0.812	0.86	1	0.893	0.86	0.334
H120.3N	0.839	0.918	0.971	0.921	0.893	1	0.944	0.328
H120.3T	0.905	0.936	0.926	0.979	0.86	0.944	1	0.382
H120.2P	0.404	0.372	0.306	0.398	0.334	0.328	0.382	1

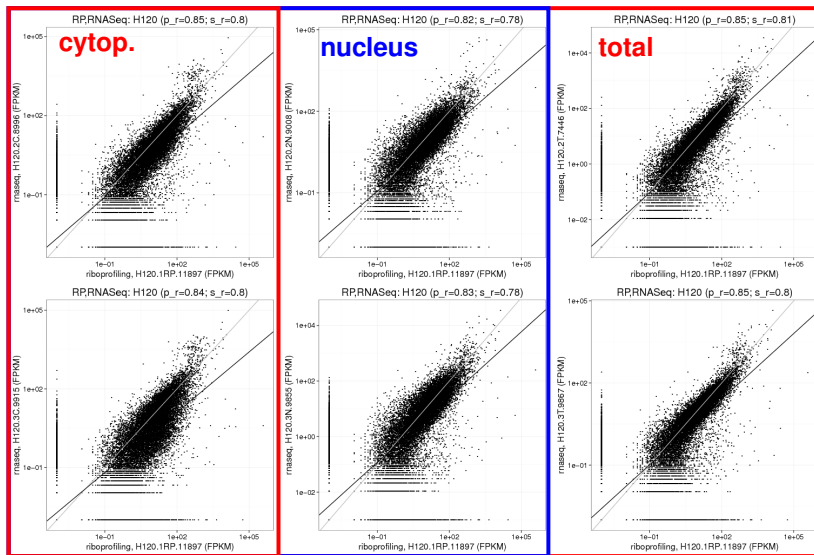
# Correlation: RP VS RNAseq - H006 (50,666 genes)



# Correlation: RP VS RNAseq - H018 (50,666 genes)



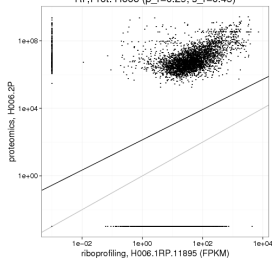
# Correlation: RP VS RNAseq - H120 (50,666 genes)



# Correlation: RP VS proteomics (6,109 genes)

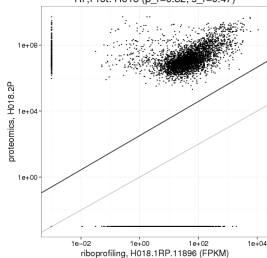
**H006**

RP,Prot: H006 ( $p_r=0.29$ ;  $s_r=0.45$ )



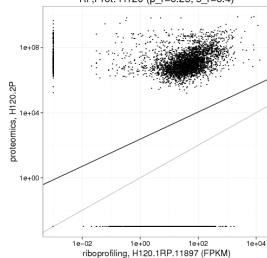
**H018**

RP,Prot: H018 ( $p_r=0.32$ ;  $s_r=0.47$ )

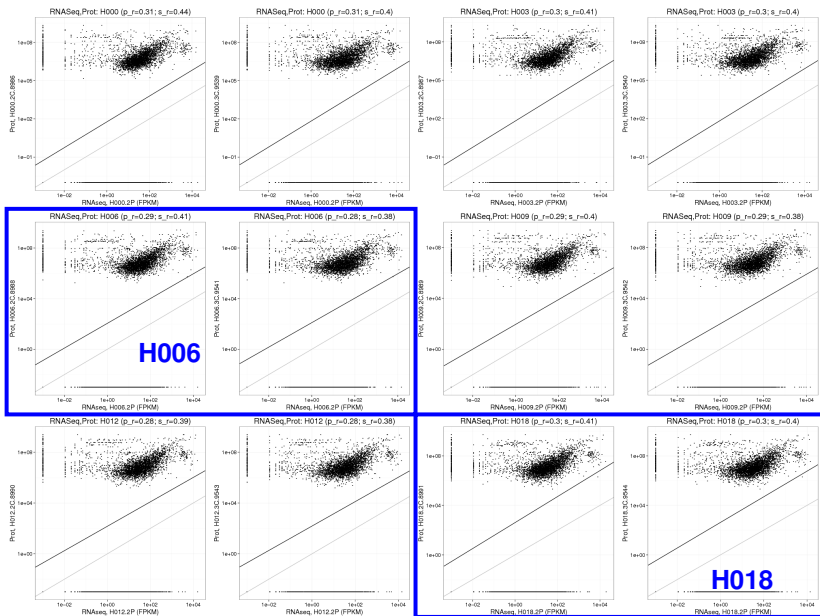


**H120**

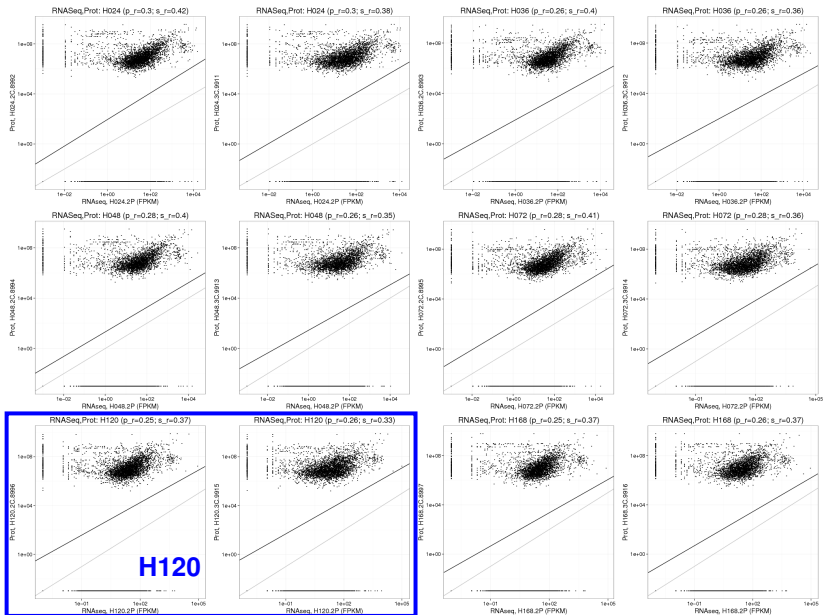
RP,Prot: H120 ( $p_r=0.25$ ;  $s_r=0.4$ )



# Correlation: RNAseq cytop. fraction VS Proteomics (6,109 genes)



# Correlation: RNAseq cytop. fraction VS Proteomics (6,109 genes)





# H006 (6,109 genes) - Spearman

## RP VS RNASeq

id	H006.1RP
H006.1RP	1
H000.2C	0.85
H000.3C	0.81
H003.2C	0.862
H003.3C	0.847
H006.2C	0.879
H006.3C	0.853
H009.2C	0.862
H009.3C	0.844
H012.2C	0.847
H012.3C	0.838
H018.2C	0.845
H018.3C	0.83
H024.2C	0.839
H024.3C	0.793
H036.2C	0.828
H036.3C	0.774
H048.2C	0.8
H048.3C	0.752
H072.2C	0.742
H072.3C	0.705
H120.2C	0.689
H120.3C	0.639
H168.2C	0.682
H168.3C	0.663

## RP VS Proteomics

id	H006.1RP
H006.1RP	1
H000.2P	0.469
H003.2P	0.447
H006.2P	0.454
H009.2P	0.462
H012.2P	0.466
H018.2P	0.484
H024.2P	0.484
H036.2P	0.448
H048.2P	0.448
H072.2P	0.41
H120.2P	0.335
H168.2P	0.329

## RNASeq VS Proteomics

id	H006.2C
H006.2C	1
H000.2P	0.421
H003.2P	0.403
H006.2P	0.407
H009.2P	0.418
H012.2P	0.418
H018.2P	0.436
H024.2P	0.443
H036.2P	0.406
H048.2P	0.405
H072.2P	0.364
H120.2P	0.284
H168.2P	0.277

id	H006.3C
H006.3C	1
H000.2P	0.397
H003.2P	0.377
H006.2P	0.38
H009.2P	0.391
H012.2P	0.392
H018.2P	0.412
H024.2P	0.421
H036.2P	0.386
H048.2P	0.384
H072.2P	0.35
H120.2P	0.282
H168.2P	0.276

# H018 (6,109 genes) - Spearman

## RP VS RNASeq

id	H018.1RP
H018.1RP	1
H000.2C	0.839
H000.3C	0.805
H003.2C	0.848
H003.3C	0.839
H006.2C	0.862
H006.3C	0.841
H009.2C	0.862
H009.3C	0.844
H012.2C	0.863
H012.3C	0.851
H018.2C	0.874
H018.3C	0.855
H024.2C	0.872
H024.3C	0.825
H036.2C	0.863
H036.3C	0.81
H048.2C	0.833
H048.3C	0.788
H072.2C	0.781
H072.3C	0.743
H120.2C	0.73
H120.3C	0.68
H168.2C	0.72
H168.3C	0.703

## RP VS Proteomics

id	H018.1RP
H018.1RP	1
H000.2P	0.448
H003.2P	0.426
H006.2P	0.431
H009.2P	0.44
H012.2P	0.44
H018.2P	0.466
H024.2P	0.474
H036.2P	0.446
H048.2P	0.454
H072.2P	0.42
H120.2P	0.352
H168.2P	0.348

## RNASeq VS Proteomics

id	H018.2C
H018.2C	1
H000.2P	0.391
H003.2P	0.372
H006.2P	0.372
H009.2P	0.385
H012.2P	0.383
H018.2P	0.409
H024.2P	0.424
H036.2P	0.401
H048.2P	0.405
H072.2P	0.376
H120.2P	0.31
H168.2P	0.302

id	H018.3C
H018.3C	1
H000.2P	0.379
H003.2P	0.359
H006.2P	0.358
H009.2P	0.37
H012.2P	0.371
H018.2P	0.396
H024.2P	0.411
H036.2P	0.388
H048.2P	0.389
H072.2P	0.364
H120.2P	0.303
H168.2P	0.296

# H120 (6,109 genes) - Spearman

## RP VS RNASeq

id	H120.1RP
H120.1RP	1
H000.2C	0.625
H000.3C	0.603
H003.2C	0.638
H003.3C	0.636
H006.2C	0.655
H006.3C	0.643
H009.2C	0.665
H009.3C	0.652
H012.2C	0.676
H012.3C	0.667
H018.2C	0.702
H018.3C	0.684
H024.2C	0.719
H024.3C	0.676
H036.2C	0.753
H036.3C	0.692
H048.2C	0.75
H048.3C	0.704
H072.2C	0.822
H072.3C	0.764
H120.2C	0.852
H120.3C	0.774
H168.2C	0.857
H168.3C	0.823

## RP VS Proteomics

id	H120.1RP
H120.1RP	1
H000.2P	0.292
H003.2P	0.267
H006.2P	0.266
H009.2P	0.279
H012.2P	0.278
H018.2P	0.309
H024.2P	0.328
H036.2P	0.323
H048.2P	0.354
H072.2P	0.401
H120.2P	0.404
H168.2P	0.412

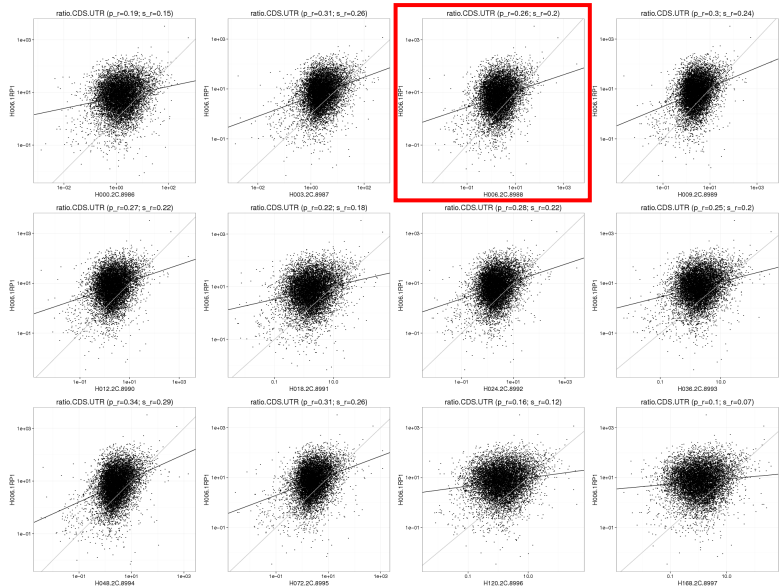
## RNASeq VS Proteomics

id	H120.2C
H120.2C	1
H000.2P	0.267
H003.2P	0.245
H006.2P	0.241
H009.2P	0.253
H012.2P	0.255
H018.2P	0.282
H024.2P	0.306
H036.2P	0.303
H048.2P	0.325
H072.2P	0.367
H120.2P	0.372
H168.2P	0.38

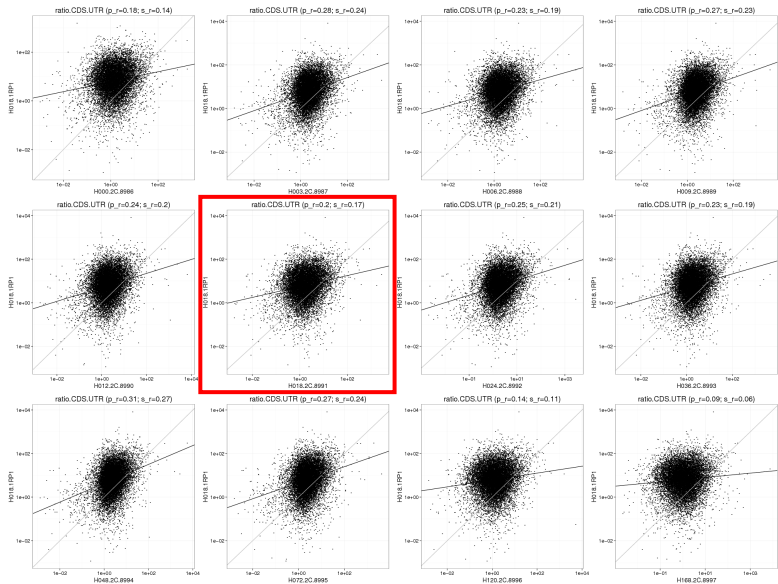
  

id	H120.3C
H120.3C	1
H000.2P	0.233
H003.2P	0.212
H006.2P	0.207
H009.2P	0.219
H012.2P	0.223
H018.2P	0.248
H024.2P	0.271
H036.2P	0.267
H048.2P	0.285
H072.2P	0.326
H120.2P	0.334
H168.2P	0.342

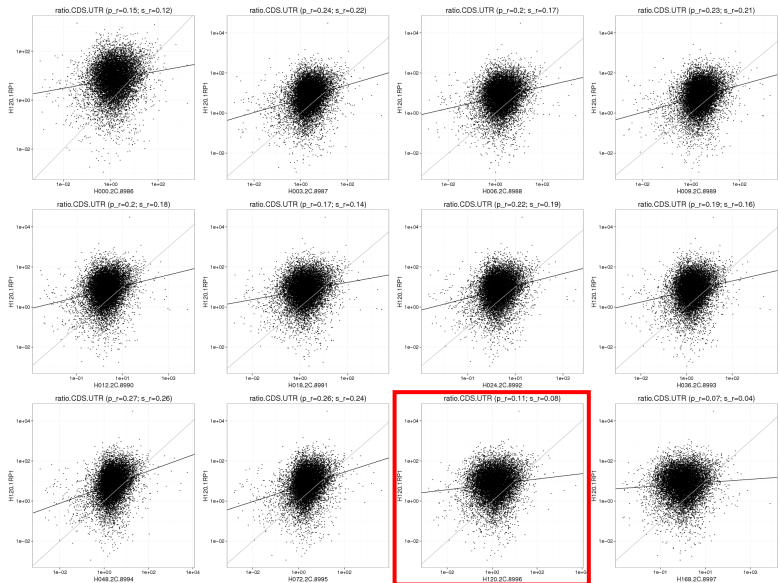
# Ratio CDS/UTR: H006.1RP VS all time points RNAseq cytoplasmic fraction



# Ratio CDS/UTR: H018.1RP VS all time points RNAseq cytoplasmic fraction



# Ratio CDS/UTR: H120.1RP VS all time points RNAseq cytoplasmic fraction



# Ratio protein coding/long non coding RNA

19815 PC  
15935 lncRNA

## Riboprofiling

	sum PC	sum lncRNA	ratio	mean PC	mean lncRNA	ratio
H006	1,002,966	821,368	1.22	50.617	51.545	0.98
H018	1,149,118	144,690	7.94	57.992	9.08	6.39
H120	1,072,901	94,444	11.36	54.146	5.9268	9.13

## RNAseq

	sum PC	sum lncRNA	ratio	mean PC	mean lncRNA	ratio
H006.2C	991,627	9,817.3	101.01	50.044	0.61609	81.23
H006.3C	1,035,817	9,777	105.94	52.274	0.61355	85.20
H018.2C	989,046	10,109	97.84	49.914	0.63441	78.68
H018.3C	954,880	9,249.3	103.24	48.19	0.58044	83.02
H120.2C	1,004,718	9,942.8	101.05	50.705	0.62396	81.26
H120.3C	1,020,129	7,199.2	41.70	51.483	0.45179	113.95