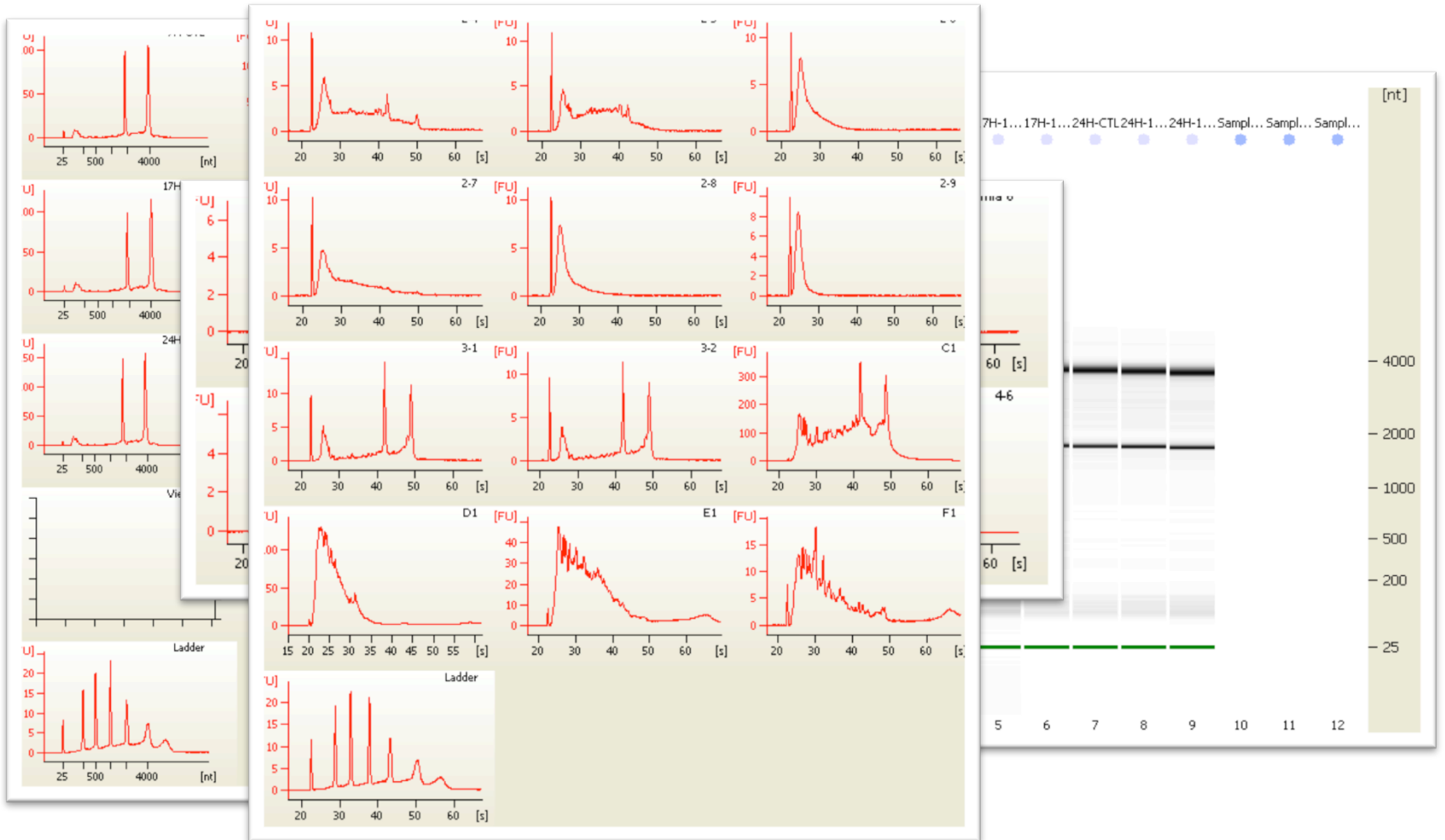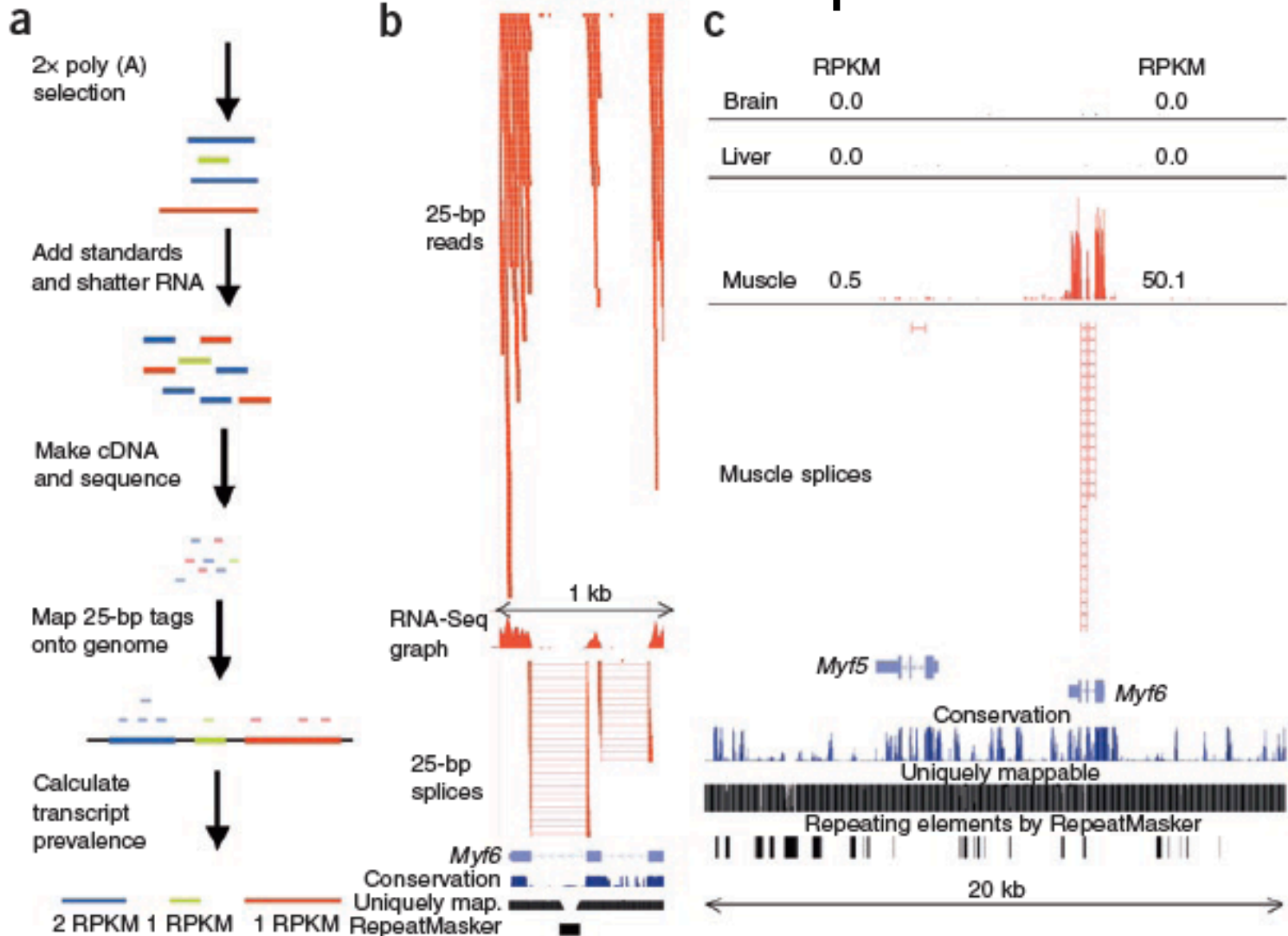# Sequencing RNA

RNA Applications

- mRNA Sequencing (RefSeq, RNASeq)
- microRNA Sequencing
- RNA-IP-Sequencing

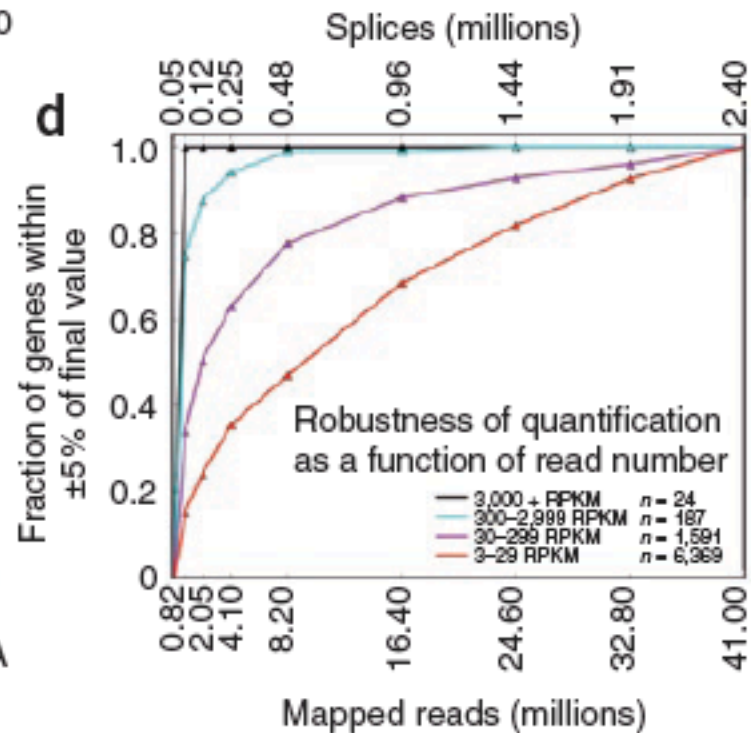# RNA Quality

mRNA-Seq

a
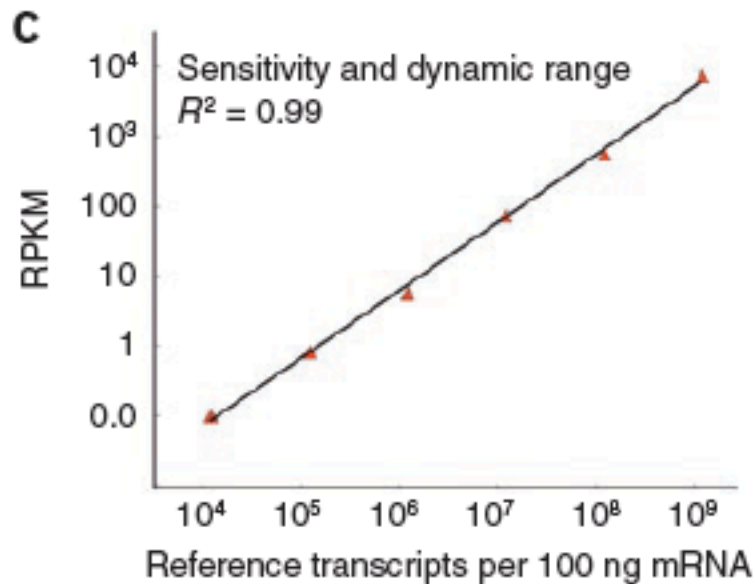
2× poly (A) selection

Add standards and shatter RNA

Make cDNA and sequence

Map 25-bp tags onto genome

Calculate transcript prevalence

2 RPKM 1 RPKM 1 RPKM

b

25-bp reads

1 kb

RNA-Seq graph

25-bp splices

Myf6
Conservation
Uniquely map.
RepeatMasker

c

	RPKM	RPKM
Brain	0.0	0.0
Liver	0.0	0.0
Muscle	0.5	50.1

Muscle splices

Myf5

Myf6

Conservation

Uniquely mappable

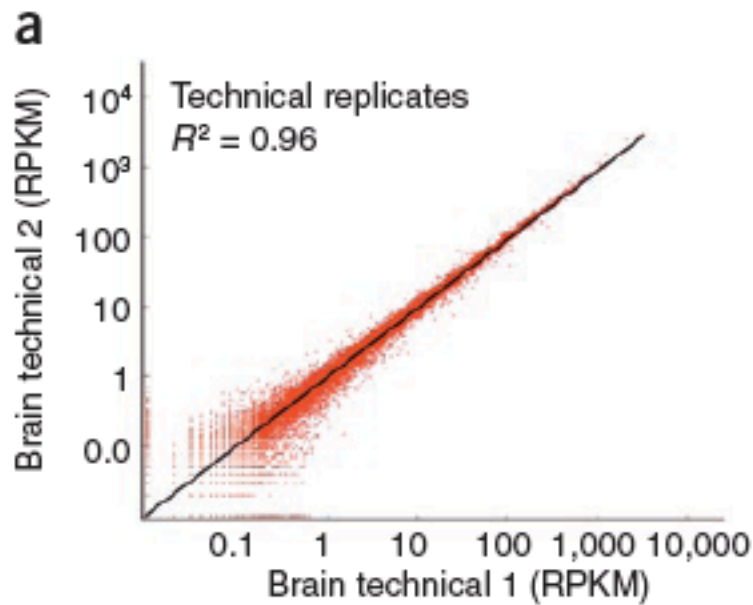Repeating elements by RepeatMasker
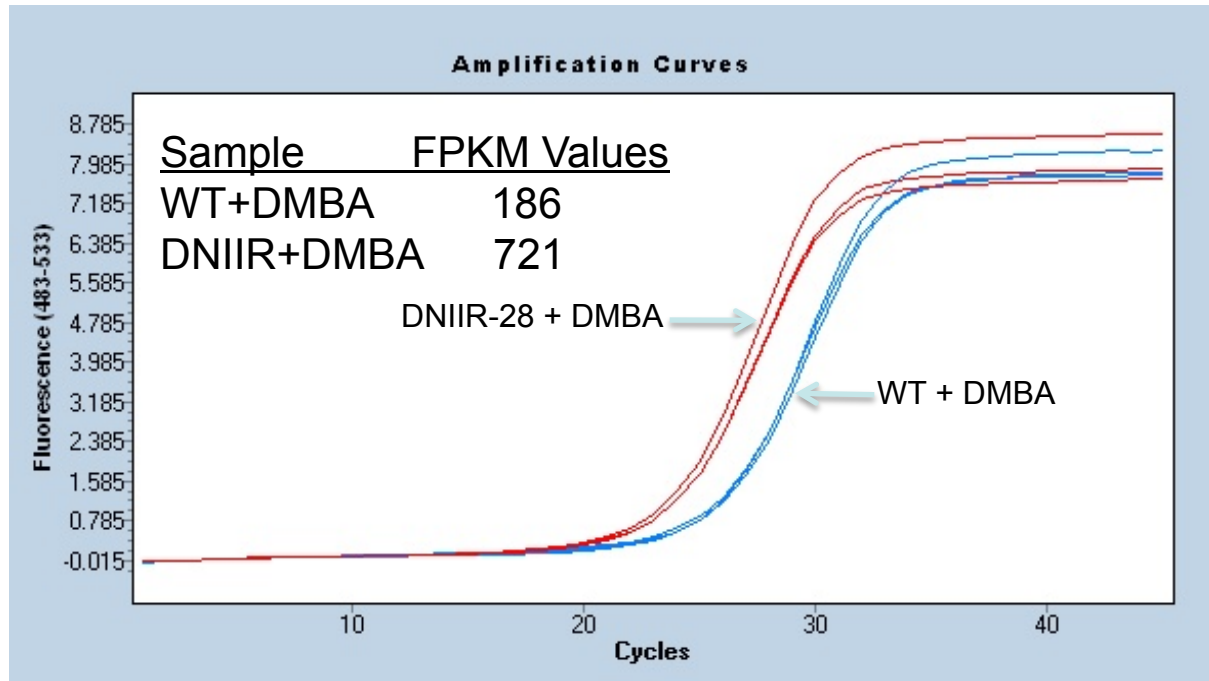
20 kb
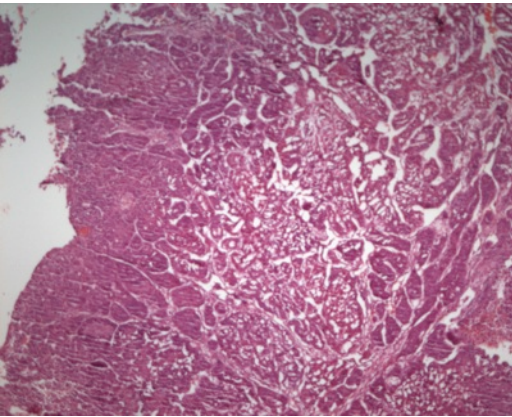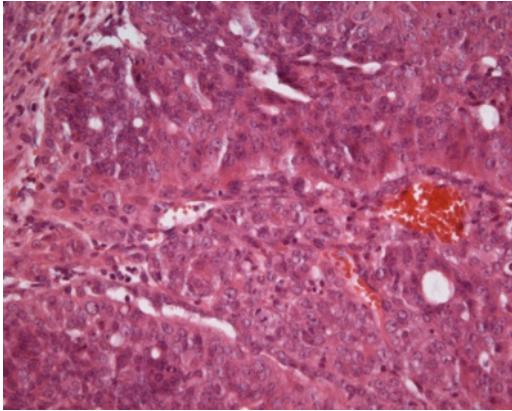
Mortazavi et al., Nature Methods 5:7:621 (2008)
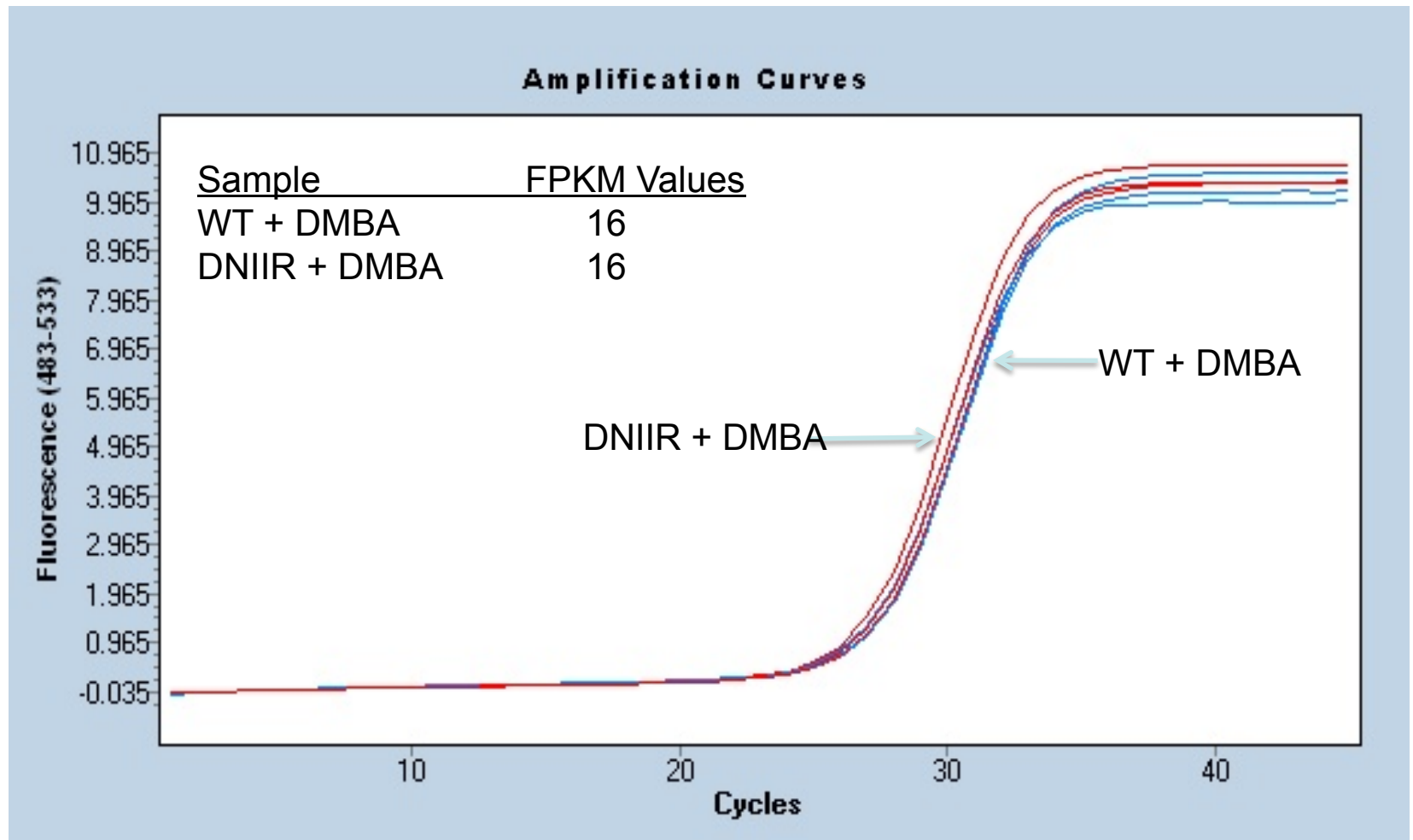
RPKM: reads per kilobase of exon model per million mapped reads

FPKM: fragment of reads per kilobase of exon model per million mapped reads (usually 25bp fragments).
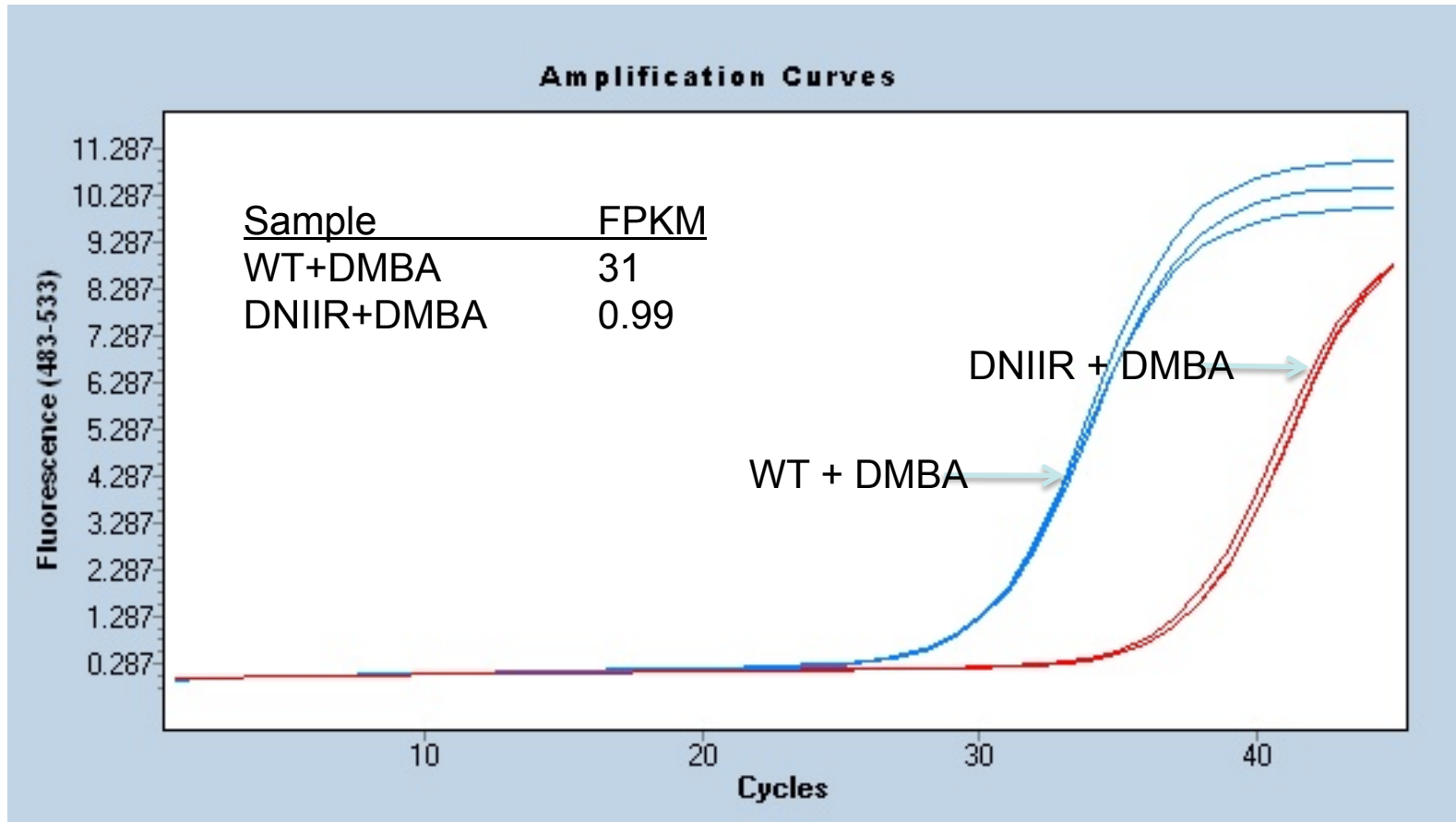
a Technical replicates
$R^2 = 0.96$

Brain technical 2 (RPKM) vs Brain technical 1 (RPKM)

b

Exons (93%)
Intergenic (3%)
Introns (4%)

c Sensitivity and dynamic range
$R^2 = 0.99$

RPKM vs Reference transcripts per 100 ng mRNA

d Splices (millions)
0.05 0.12 0.25 0.48 0.96 1.44 1.91 2.40

Fraction of genes within ±5% of final value

Robustness of quantification as a function of read number

3,000 + RPKM $n = 24$
300–2,999 RPKM $n = 187$
30–299 RPKM $n = 1,591$
3–29 RPKM $n = 6,369$

Mapped reads (millions)
0.82 2.05 4.10 8.20 16.40 24.60 32.80 41.00

Mortazavi et al., Nature Methods 5:7:621 (2008)

Keratin 8

Sample	FPKM Values
WT+DMBA	186
DNIIR+DMBA	721

DNIIR-28 + DMBA

WT + DMBA

Lipase Maturation Factor 1 (Lmf1)

Amplification Curves

Sample	FPKM Values
WT + DMBA	16
DNIIR + DMBA	16

Lysophosphatidic acid receptor 3

Amplification Curves

Sample	FPKM
WT+DMBA	31
DNIIR+DMBA	0.99

DNIIR + DMBA

WT + DMBA

Insulin like growth factor binding protein 3 (Igfbp3)

Amplification Curves

Sample	FPKM
WT + DMBA	36
DNIIR + DMBA	458

DNIIR + DMBA

WT + DMBA
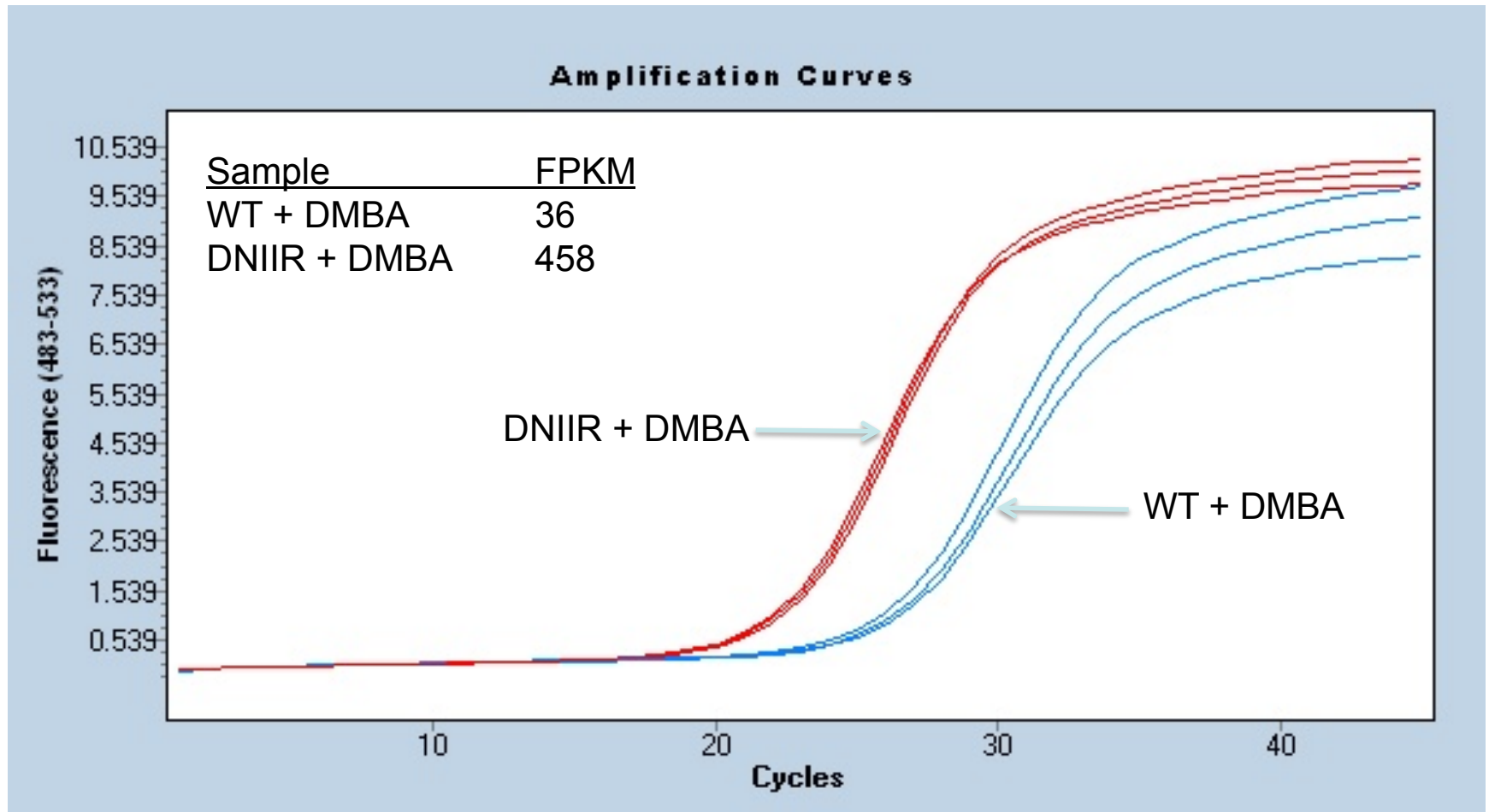
Ncapg: Non-SMC condensin I complex, subunit G
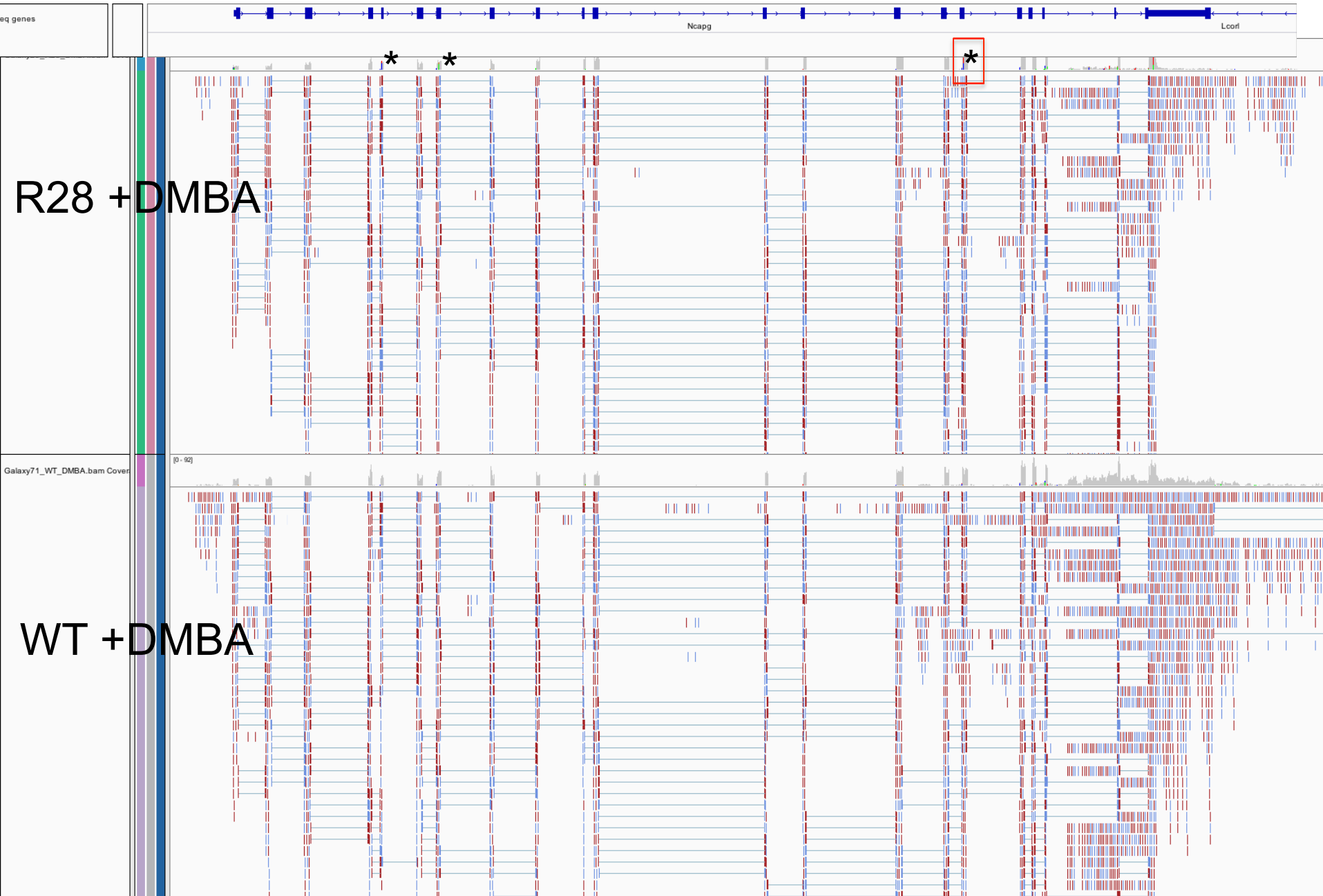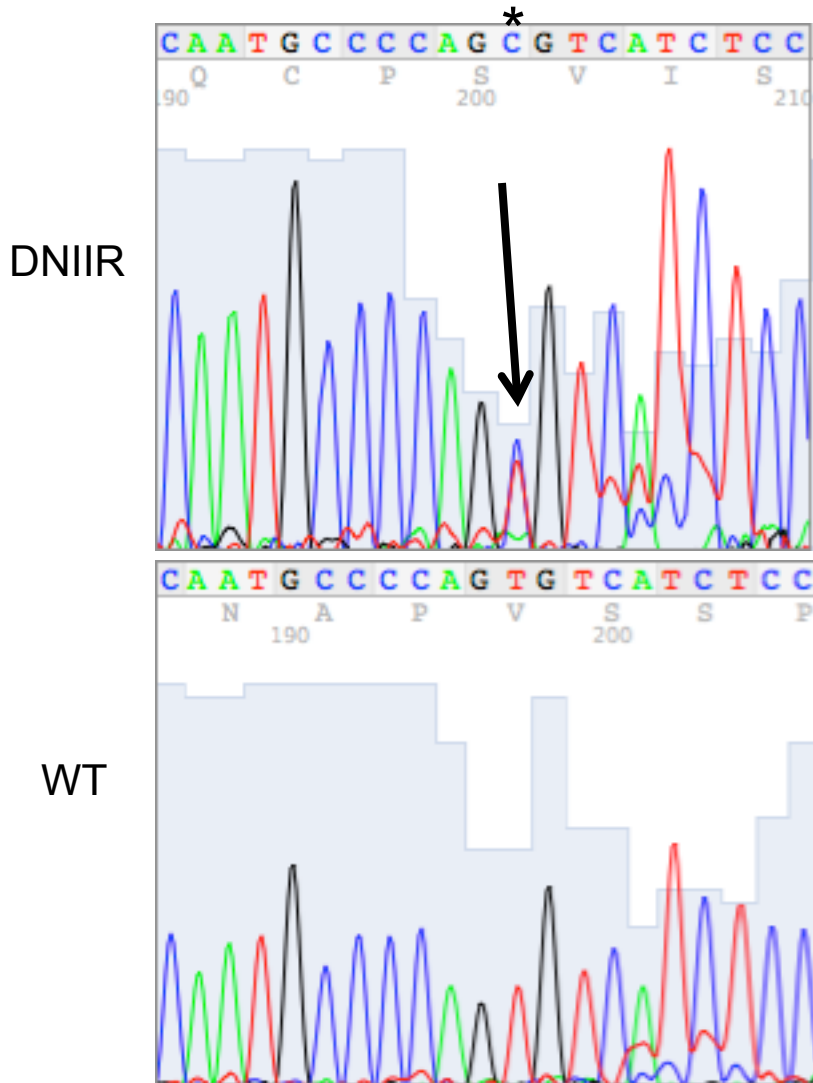
Exon 16 of Ncapg
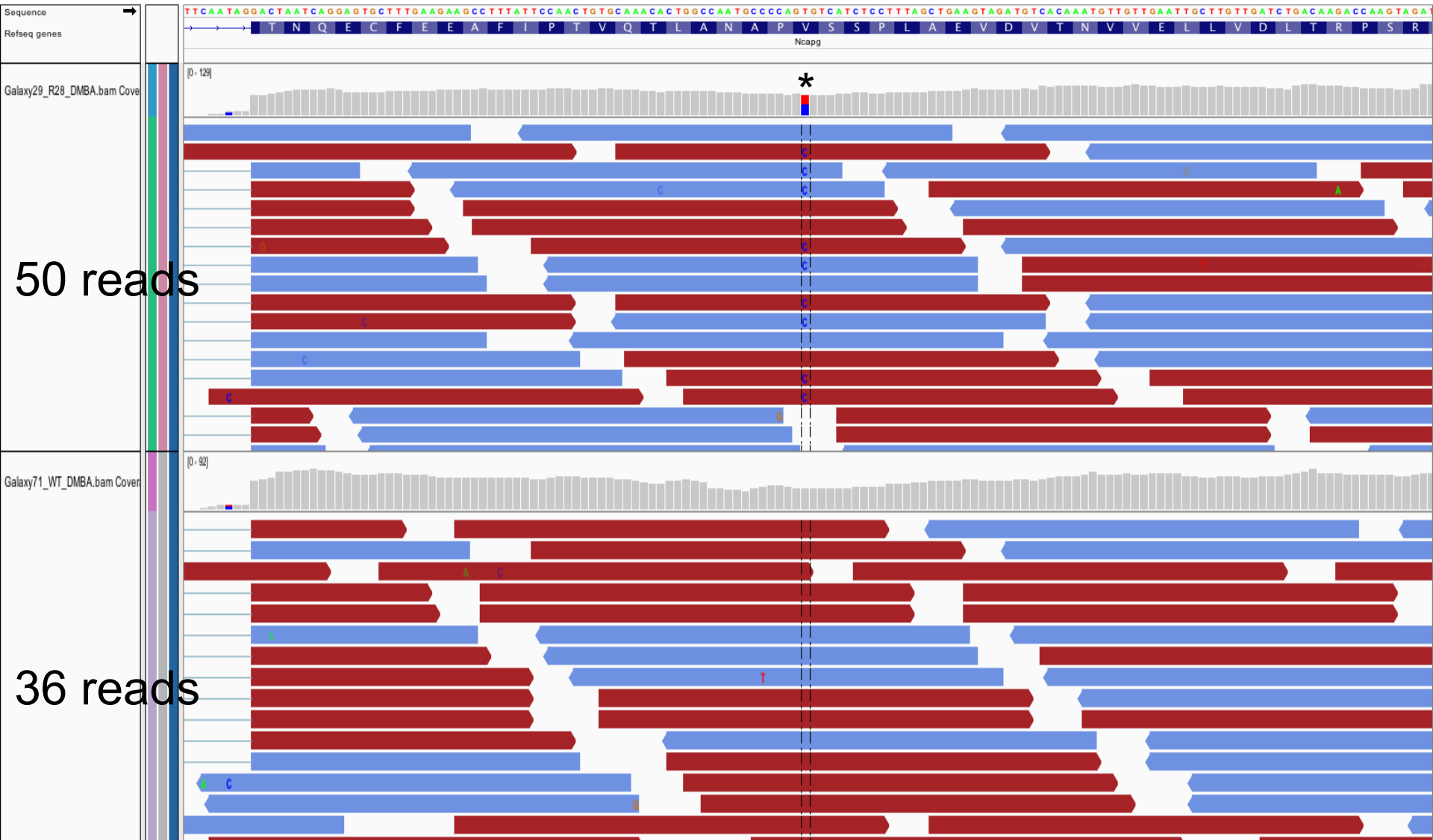
T-C mutation resulting in a Val-Ala change in the protein

Sequence Confirmation of Ncapg mutation
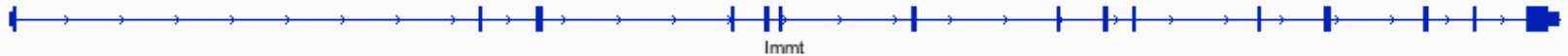

DNIIR

WT

T>C mutation resulting in an Ala>Val change at position aa784 in the protein. The other mutations were a polymorphic T>C change at aa242 and an A>G change at aa347 resulting in a non-synonymous change from Arg>Lys.
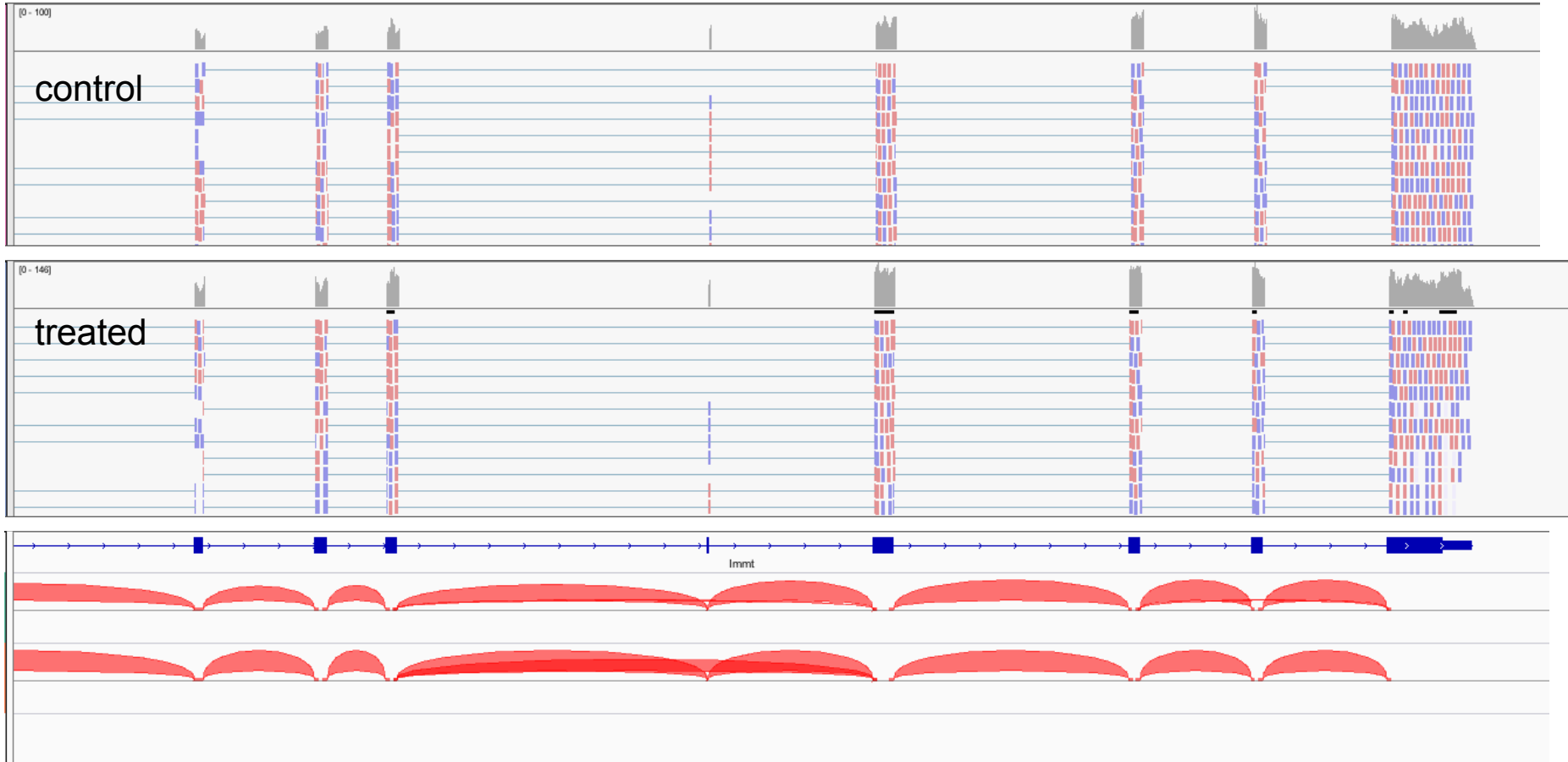
Alternative Exon Usage

Alternative splicing

Summary

- Several different platforms exist utilizing different technologies.

- Generate between 500 million to 600 Billion bases of sequence information per run.

- Several applications including Whole genome sequencing, Targeted genomic seq., ChIP-Seq and mRNA-Seq, among others.

- Data files are very large ≥1Tb of information.

- Personalized medicine via genome sequencing is not far away.