State-of-the-art epigenomics using NGS-based BiSulfite Amplicon Sequencing (BSAS)

and

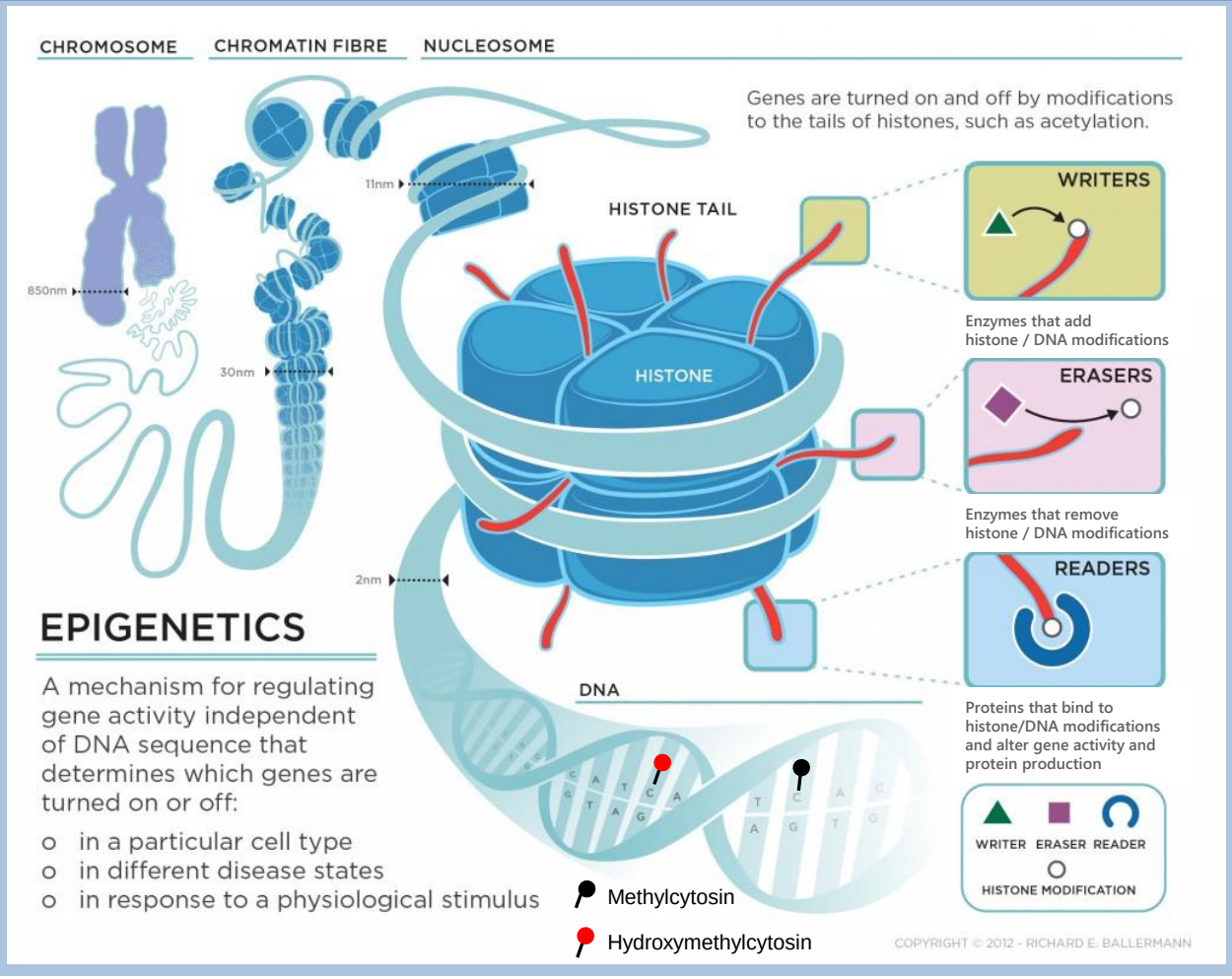
targeted Allele-Specific RNAseq

Matthias Linke, Institute of Human Genetics, Mainz

Agenda

- Introduction Epigenetics
- BiSulfite Amplicon Sequencing (BSAS)
  - Workflow
  - Analysis / Results
- BisPCR^2 – Alternative to BSAS
  - Workflow
  - Analysis / Results
- Allele-specific RNAseq
  - Background
  - Workflow
  - Analysis / Results

Epigenetics

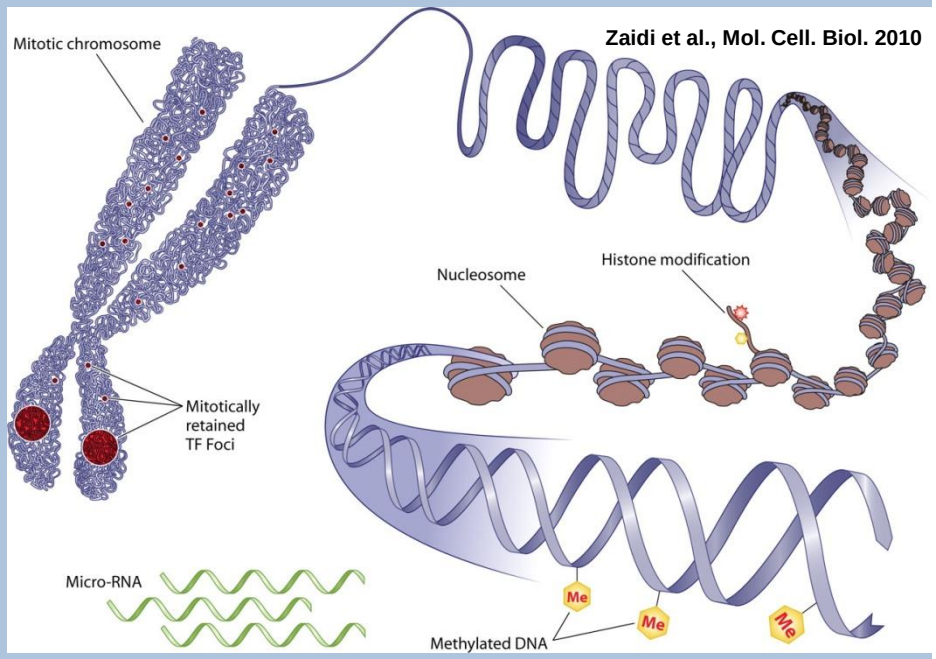


Modified after Richard E. Ballermann from resverlogix.com

Major biochemical mechanisms in neuroepigenetics

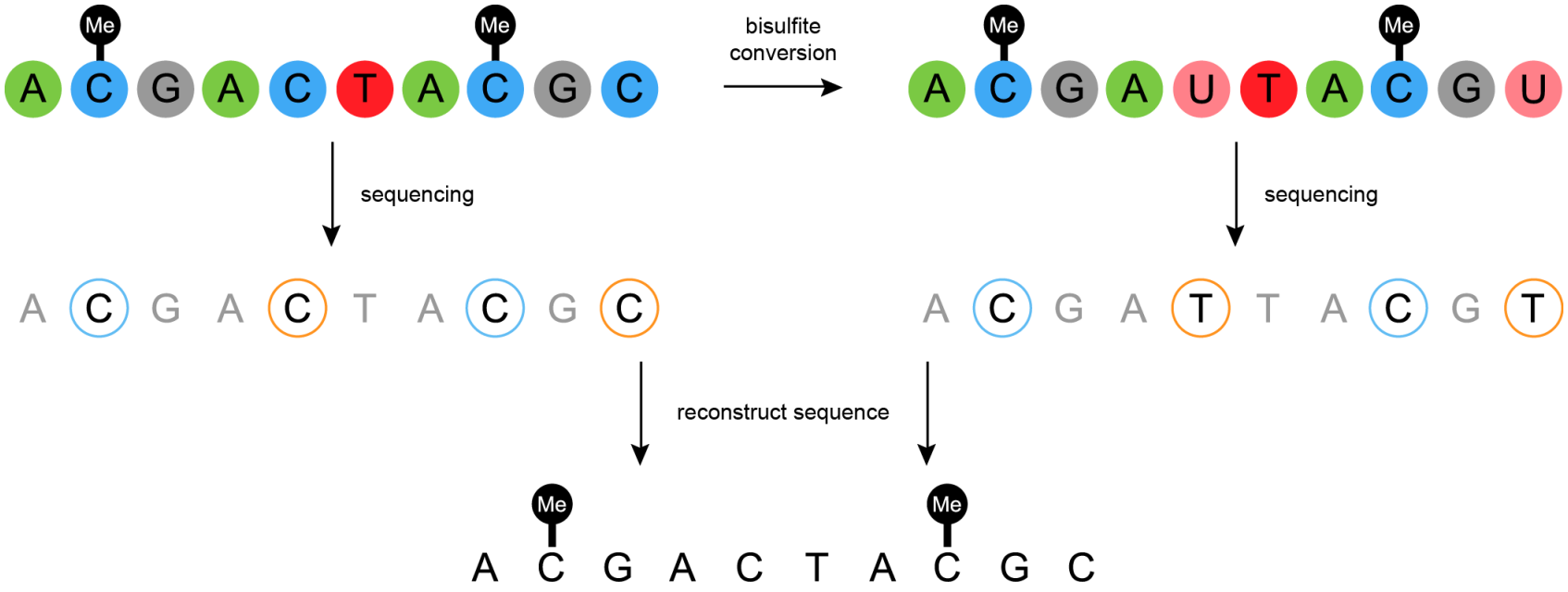
Covalent modification of DNA
DNA cytosine methylation
active cytosine demethylation
hydroxymethylcytosine formation
methylcytosine oxidation (5-formylcytosine, 5-carboxylcytosine)
Histone posttranslational modifications
lysine acetylation
lysine (mono/di/tri) and arginine (mono/di) methylation
serine/threonine phosphorylation
monoubiquitination
poly ADP-ribosylation
ATP-dependent chromatin remodeling (SWI-SNF)
Histone subunit exchange
H2A.Z
H3.3
RE1-silencing transcription factor (REST)/REST corepressor
(CoREST)/Sin3A system
Noncoding RNAs
piRNAs
microRNAs
small interfering RNAs (siRNAs)
small nuclear RNAs (snRNAs)
Line 1 retrotransposition
Prion protein-based mechanisms

Sweatt, Neuron 2013

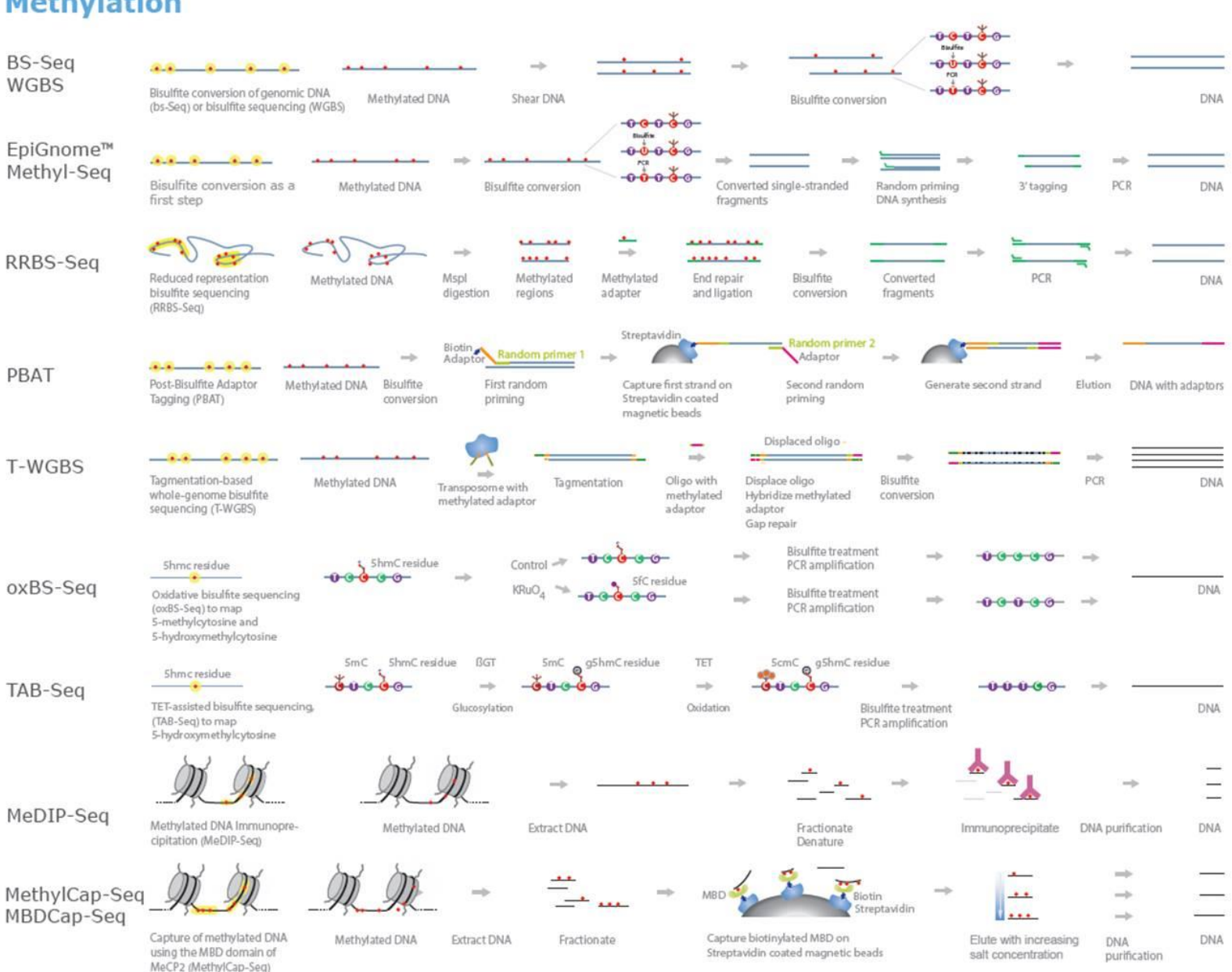


Zaidi et al., Mol. Cell. Biol. 2010

Bisulfite conversion



Methylation



Illumina MiSeq

Max throughput with MiSeq Reagent Kit v3: 25 Mio reads (15 Gb with 600 cycles)

Diagnostics:

BRCA1/2

BRCA +

HNPCC, FAP

Gene panels for:

- Hearing loss
- ID, mental disorders, ASD
- connective tissue disease

Hypothesis driven BS analysis – Why?

LETTER
doi:10.1038/nature12433

Charting a dynamic DNA methylation landscape of the human genome

Michael J. Ziller[1,2,3], Hongcang Gu[1], Fabian Müller[3]†, Julie Donaghey[1,2,3], Linus T.-Y. Tsai[4], Oliver Kohlbacher[5], Philip L. De Jager[1,6], Evan D. Rosen[1,4], David A. Bennett[7], Bradley E. Bernstein[1,8], Andreas Gnirke[1] & Alexander Meissner[1,2,3]

studies. The results also highlight the general inefficiency of whole-genome bisulphite sequencing, as 70–80% of the sequencing reads across these data sets provided little or no relevant information about CpG methylation. To demonstrate further the utility of our DMR set, we use it to classify unknown samples and identify representative signature regions that recapitulate major DNA methylation dynamics. In summary, although in theory every CpG can change its methylation state, our results suggest that only a fraction does so as part of coordinated regulatory programs. Therefore, our selected DMRs can serve as a starting point to guide new, more effective reduced representation approaches to capture the most informative fraction of CpGs, as well as further pinpoint putative regulatory elements.

of power constraints. Extreme conditions *in vitro* or *in vivo* such as loss or misregulation of the maintenance methylation machinery will affect a larger subset including many intergenic CpGs that are generally static, but most of these additional CpGs are unlikely to overlap with functional elements such as TFBSs or enhancers. In combination with the fact that sequencing of WGBS libraries is very inefficient, as about 65% of all 101-bp reads in our set did not even contain any CpGs to begin with, this amounts to an approximate, combined loss of around 80% of sequencing depth on non-informative reads and static regions. Furthermore, once defined, it will probably be sufficient in most cases to profile only a representative subset of CpGs across a comprehensive set of DMRs using an array-based[28] or hybrid-capture-based[29] technology to recover representative dynamics and measure regulatory events. Using these results as a guiding principle, we expect further improved efficiencies in mapping DNA methylation and enhance its applicability as a marker for various regulatory dynamics in normal and disease phenotypes.

BiSulfite Amplicon Sequencing (BSAS)

Masser *et al. Epigenetics & Chromatin* 2013, **6**:33
http://www.epigeneticsandchromatin.com/content/6/1/33

Epigenetics & Chromatin

METHODOLOGY Open Access

Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing

Dustin R Masser[1], Arthur S Berg[2] and Willard M Freeman[1,3*]

Abstract

Background: The growing interest in the role of epigenetic modifications in human health and disease has led to the development of next-generation sequencing methods for whole genome analysis of DNA methylation patterns. However, many projects require targeted methylation analysis of specific genes or genomic regions. We have developed an approach, termed BiSulfite Amplicon Sequencing (BSAS), for hypothesis driven and focused absolute DNA methylation analysis. This approach is applicable both to targeted DNA methylation studies as well as to confirmation of genome-wide studies.

Results: BSAS uses PCR enrichment of targeted regions from bisulfite-converted DNA and transposome-mediated library construction for rapid generation of sequencing libraries from low (1 ng) sample input. Libraries are sequenced using the Illumina MiSeq benchtop sequencer. Generating high levels of sequencing depth (>1,000 ×) provides for quantitatively precise and accurate assessment of DNA methylation levels with base specificity. Dual indexing of sequencing libraries allows for simultaneous analysis of up to 96 samples. We demonstrate the superior quantitative accuracy of this approach as compared to existing Sanger sequencing methods.

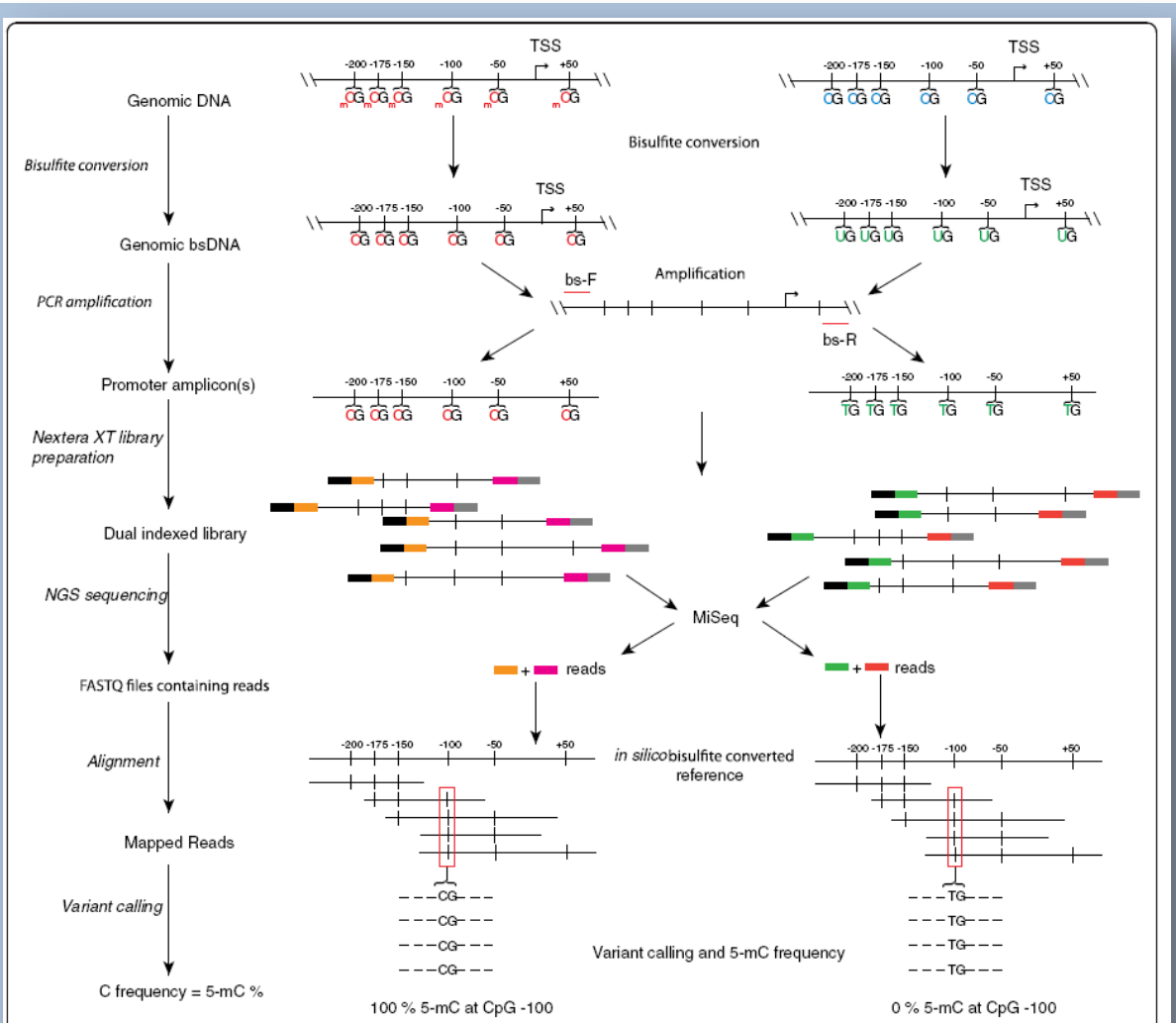
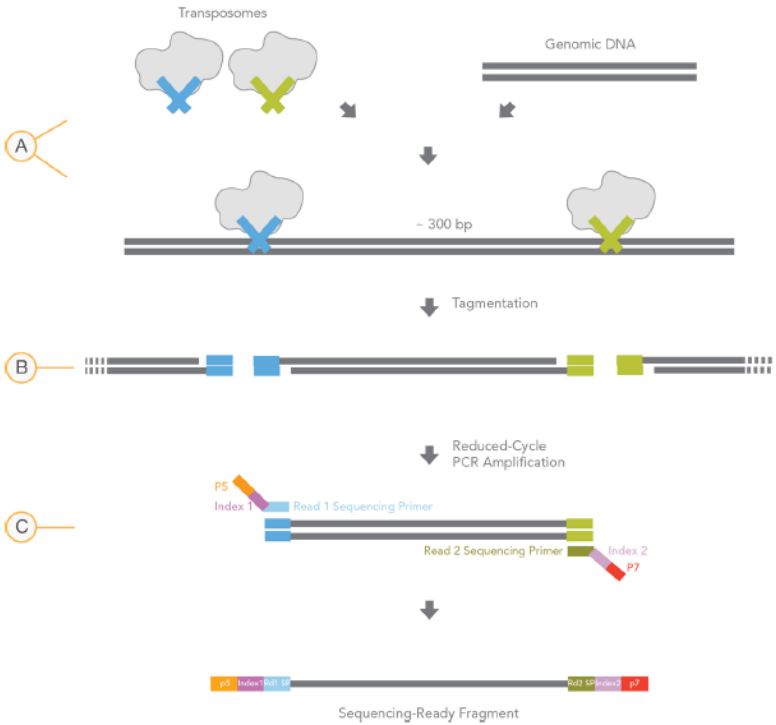


Figure 1 Bisulfite amplicon sequencing (BSAS) method schematic. Genomic DNA is bisulfite converted and subjected to bisulfite-specific PCR, using primers specific for bisulfite converted DNA (bs-F and R red lines). Amplicons are subjected to Nextera XT library preparation including dual indexing. Final libraries consist of a random insert of bisulfite converted, amplified DNA, capture probes (black and gray) and specific index sequences (orange, magenta, green, pink). These libraries are multiplexed and sequenced on the Illumina MiSeq. Demultiplexing separates the dual indexed reads from each sample (orange and magenta are one sample, green and pink represent the other sample). These reads are aligned to an *in silico* converted reference sequence, and variant calling is used to identify the percentage of 5-mC.

Masser *et al.*, 2013

How does the Nextera XT Assay Work?

The Nextera XT DNA Sample Preparation Kit uses an engineered transposome to simultaneously fragment and tag ("tagment") input DNA, adding unique adapter sequences in the process. A limited-cycle PCR reaction uses these adapter sequences to amplify the insert DNA. The PCR reaction also adds index sequences on both ends of the DNA, thus enabling dual-indexed sequencing of pooled libraries on any Illumina Sequencing System.



A Nextera XT transposome with adapters is combined with template DNA
B Tagmentation to fragment and add adapters
C Limited cycle PCR to add sequencing primer sequences and indicies

Costs: BiSulfite Amplicon Sequencing vs PyroSeq 1/2

	BSAS	PyroSeq	PyroSpecs to meet BSAS
Sequencing depth (sensitivity)	1000x (recomm.) easy to increase	100x difficult to increase (multiple tech. Replicates)	10 technical replicates to achieve BSAS coverage/sensitivity (1000x)
read length	> 300bp PCR-product - average: 400bp only limited by BS conversion (0,5-1kB)	< 140bp (new chemistry, algorithms)	> 3 assays to cover 400bp of target region
plex-level for 1 sample	Alexis: 23 assays	Not possible	23 seperate Pyros
Summary	**1** sample (23 assays à 400bp with 1000x coverage)	690 Pyro-Rxn > 7 plates à 96 wells	

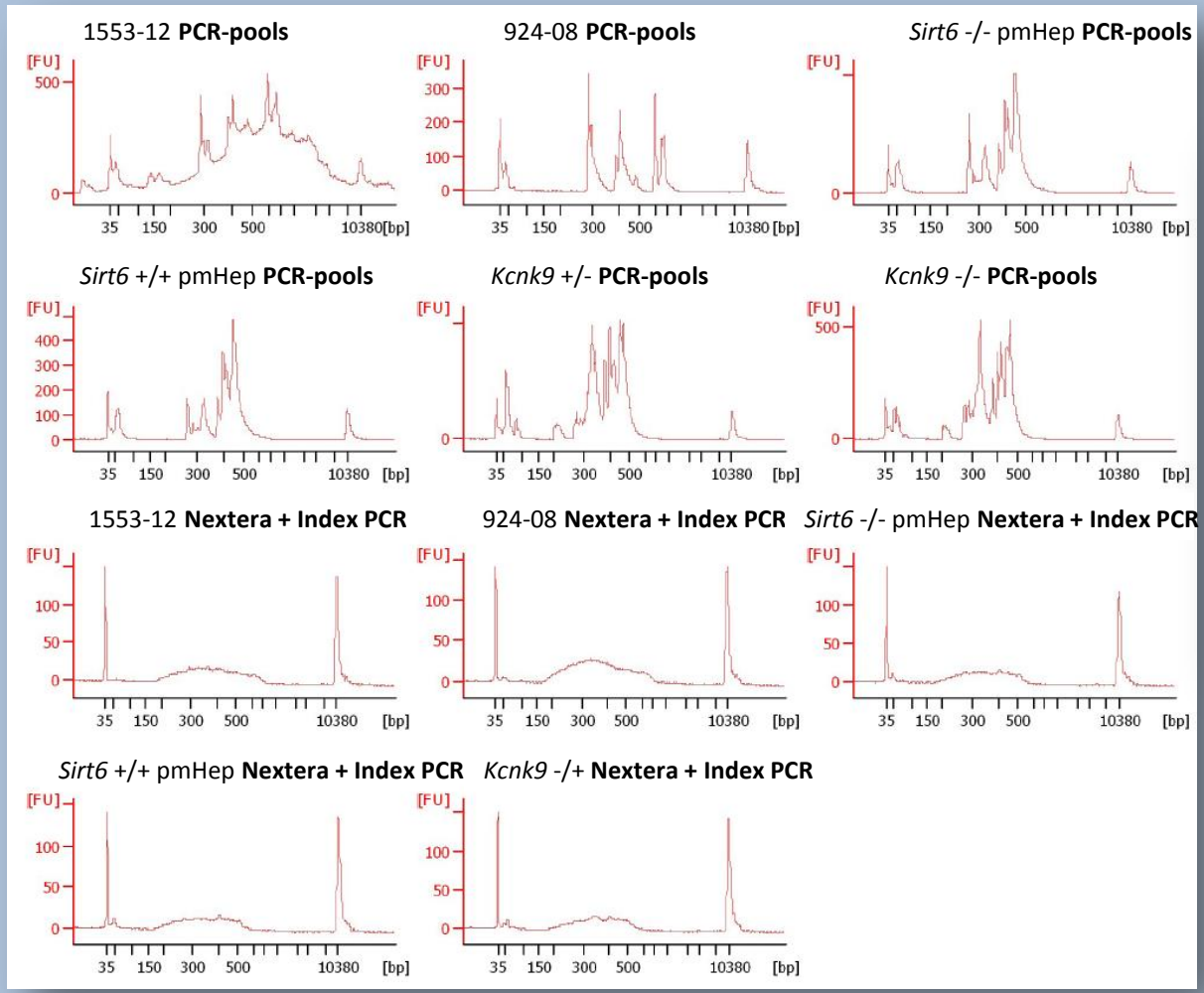
… and we got **12** samples in our 1st run = **8280** Pyro-Rxn (86 plates)

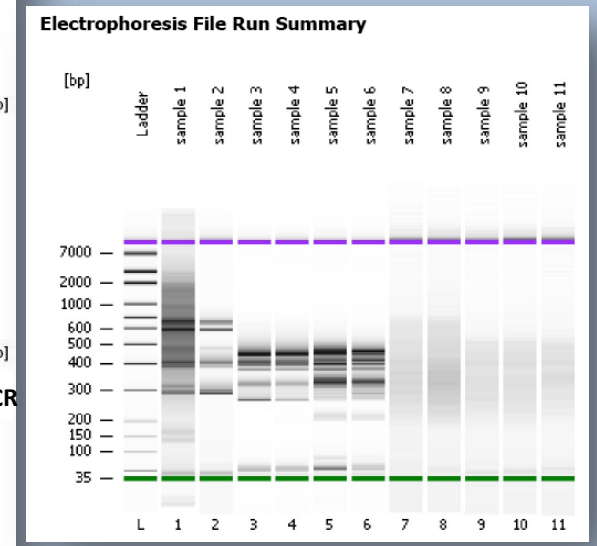
Costs: Bisulfite Amplicon Sequencing vs PyroSeq 2/2

without Qubit/BioAnalyzer reagents (BSAS) and Sepharose beads etc. (PyroSeq)

	PyroSeq	BSAS
PSQ 96 Pyrogold Reagent Kit 96rxn	451,34€	-
MiSeq Reagent Kit v2 (300cycle)	-	908,92 €
Nextera XT DNA Sample Prep Kit 24rxn	-	734,83€
Nextera XT Index Kit (24rxn)	-	249,90€
Summary (costs per run)	451,34€ (1 x 96-well plate)	1893,65€
extrapolation on same data output of 1xBSAS run	**38.815,24€** (451,34€ x 86 plates)	**1893,65€**

QC


1553-12 PCR-pools
924-08 PCR-pools
Sirt6 -/- pmHep PCR-pools
Sirt6 +/+ pmHep PCR-pools
Kcnk9 +/- PCR-pools
Kcnk9 -/- PCR-pools
1553-12 Nextera + Index PCR
924-08 Nextera + Index PCR
Sirt6 -/- pmHep Nextera + Index PCR
Sirt6 +/+ pmHep Nextera + Index PCR
Kcnk9 -/+ Nextera + Index PCR
[FU]
[bp]
35 150 300 500 10380


Electrophoresis File Run Summary
[bp]
Ladder
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6
sample 7
sample 8
sample 9
sample 10
sample 11
7000
2000
1000
600
500
400
300
200
150
100
35
L 1 2 3 4 5 6 7 8 9 10 11

Complete workflow

Wet lab

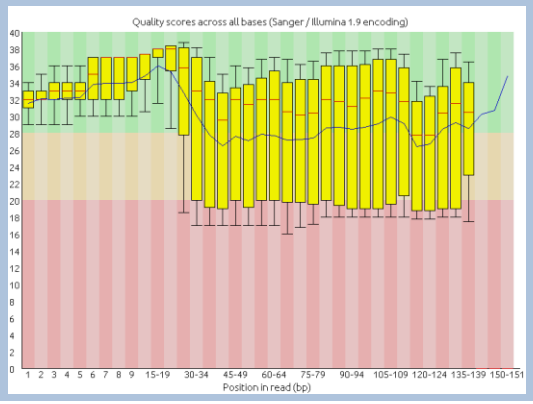
illumina

DNA isolation → Bisulfite conversion (Zymo) → Bisulfite PCR (FastStart; Roche) → Library preparation (Nextera XT) → Sequencing MiSeq V2 300 cycles

Bioinformatics

Reference genome → FASTQ reads → FASTQC → Bismark → R script for visualization

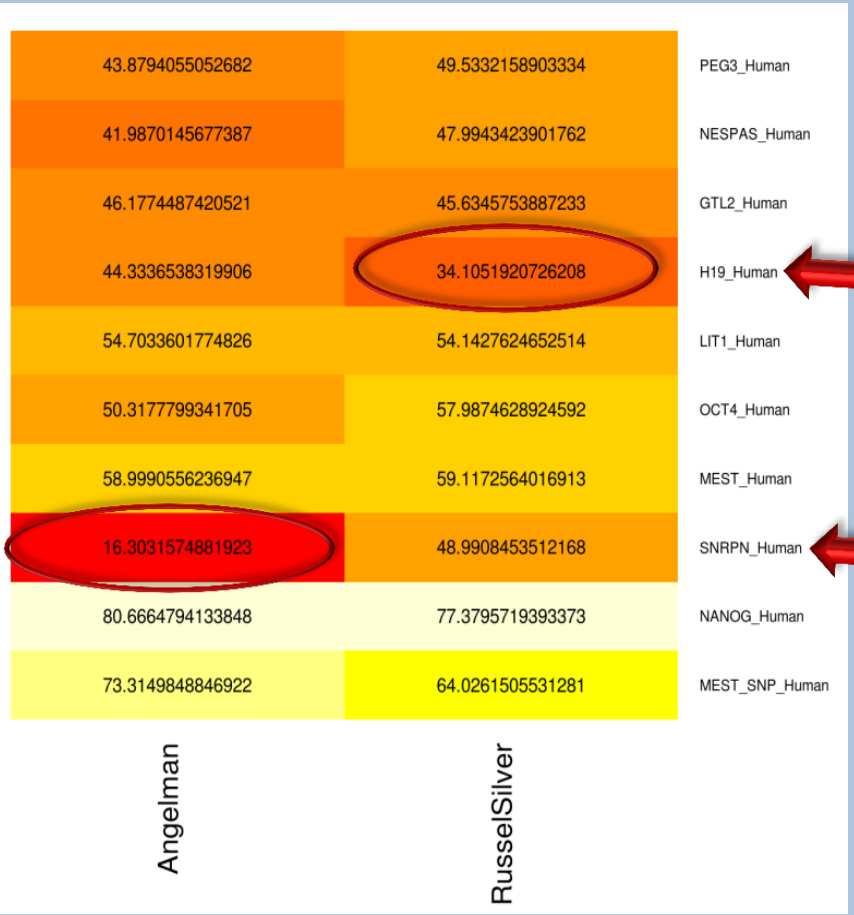
UCSC Genome Bioinformatics



Bismark

Babraham Bioinformatics

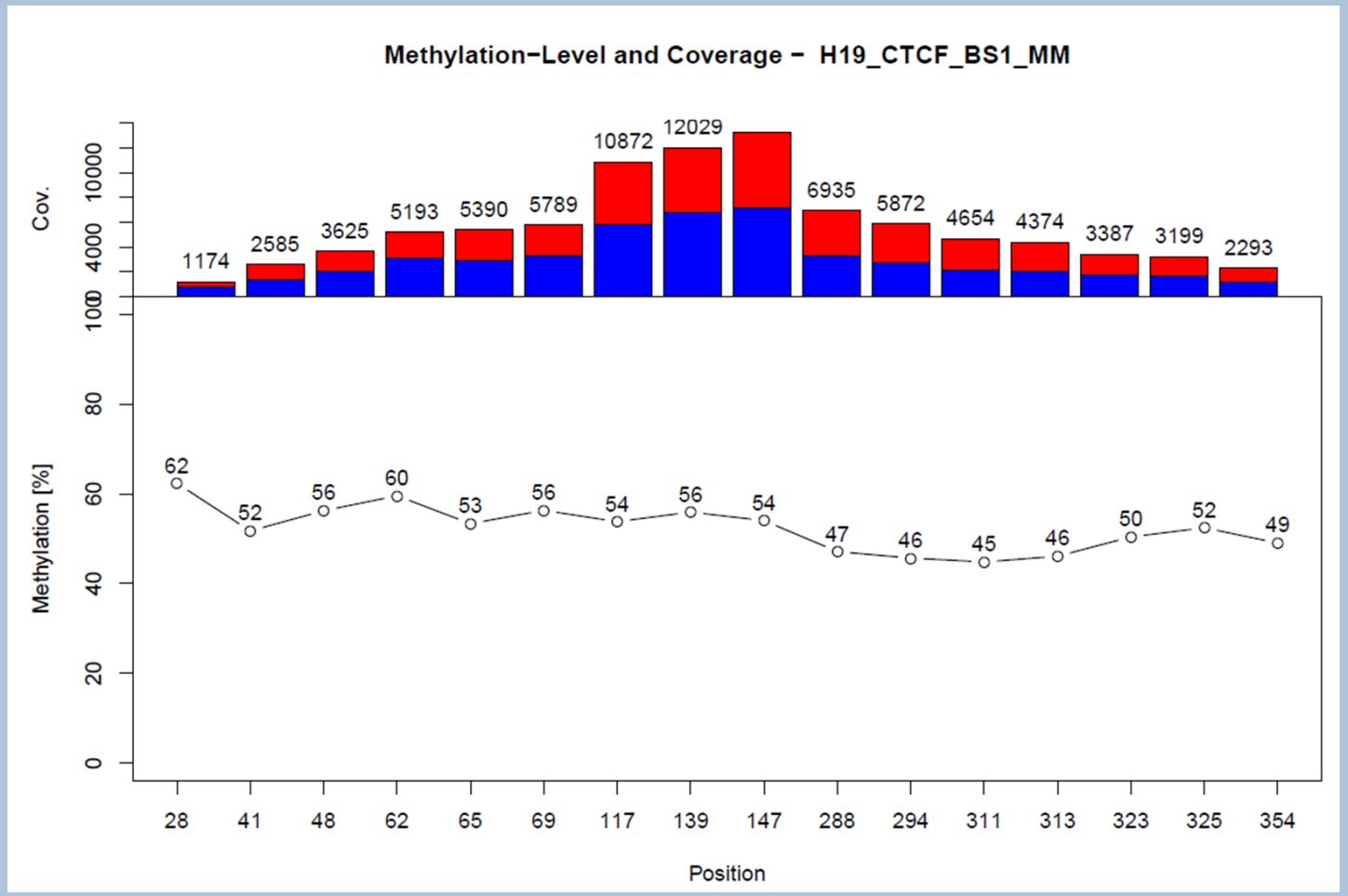
exemplary results



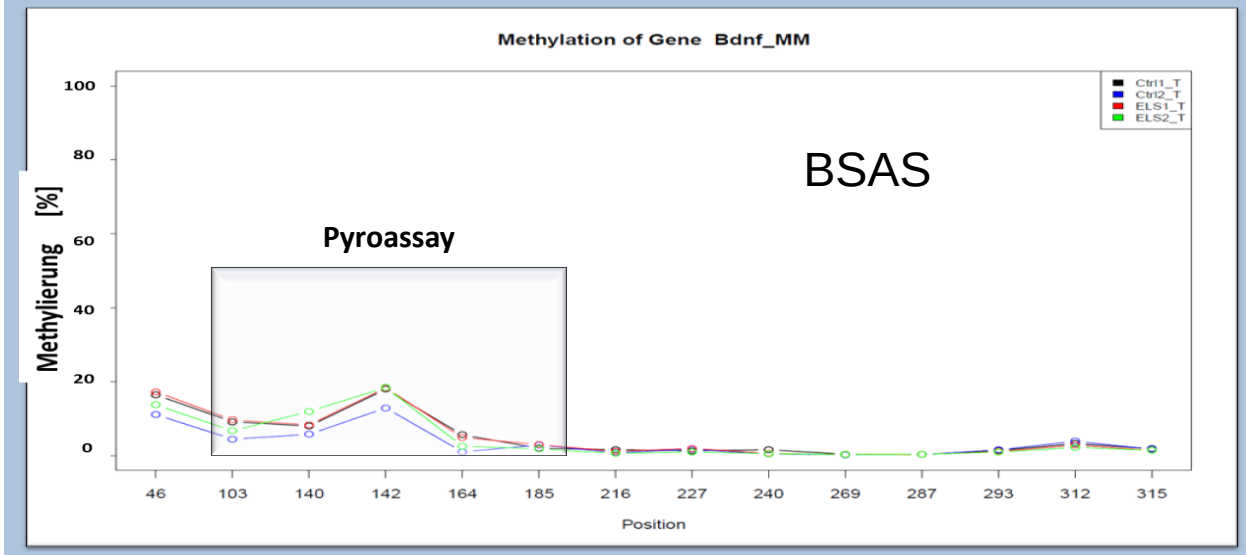
- Mean methylation of patients suffer from:
  - Angelman-Syndrom (IC *SNRPN* hypomethylated)
  - Silver-Russel-Syndrom (IC *H19* hypomethylated)

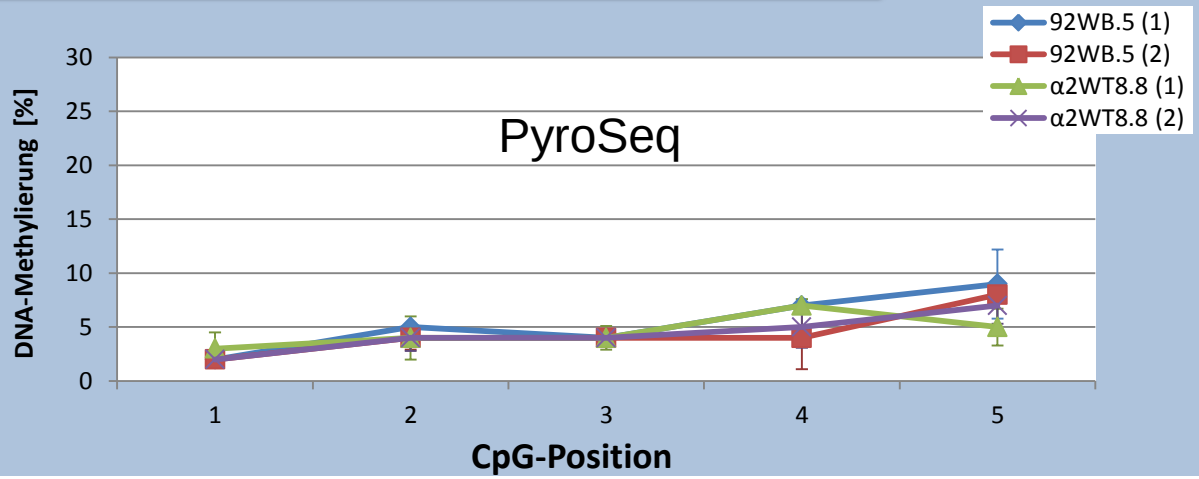
Pyrosequencing confirmed

exemplary results



exemplary results





BSAS: Conclusions

- Less costly than Pyrosequencing and 454 Pyrosequencing
- Lot more throughput than 454: **4,5Gb (15 Mio reads)** vs. **0,035Gb** (0,1 Mio Reads; Junior) or **0,7Gb** (1 Mio reads; FLX)
- Workflow relative simple and streamlined compared to Roche 454 Pyrosequencing (Würzburg, Essen, Saarbrücken with Amplicon BS Sequencing on Roche)
- reliable results on methylation level of imprinted genes in somatic tissue
- Critical to gain experience with Nextera XT also for other NGS applications (RNAseq etc.)

BisPCR² – Alternative to BSAS

Bernstein *et al. Epigenetics & Chromatin (2015) 8:27*
DOI 10.1186/s13072-015-0020-x

Epigenetics & Chromatin

METHODOLOGY Open Access

The BisPCR² method for targeted bisulfite sequencing

Diana L Bernstein, Vasumathi Kameswaran, John E Le Lay, Karyn L Sheaffer and Klaus H Kaestner*

Abstract

Background: DNA methylation has emerged as an important regulator of development and disease, necessitating the design of more efficient and cost-effective methods for detecting and quantifying this epigenetic modification. Next-generation sequencing (NGS) techniques offer single base resolution of CpG methylation levels with high statistical significance, but are also high cost if performed genome-wide. Here, we describe a simplified targeted bisulfite sequencing approach in which DNA sequencing libraries are prepared following sodium bisulfite conversion and two rounds of PCR for target enrichment and sample barcoding, termed BisPCR².

Results: We have applied the BisPCR² technique to validate differential methylation at several type 2 diabetes risk loci identified in genome-wide studies of human islets. We confirmed some previous findings while not others, in addition to identifying novel differentially methylated CpGs at these genes of interest, due to the much higher depth of sequencing coverage in BisPCR² compared to prior array-based approaches.

Conclusion: This study presents a robust, efficient, and cost-effective technique for targeted bisulfite NGS, and illustrates its utility by reanalysis of prior findings from genome-wide studies.

Keywords: Targeted bisulfite sequencing, DNA methylation, Next-generation sequencing

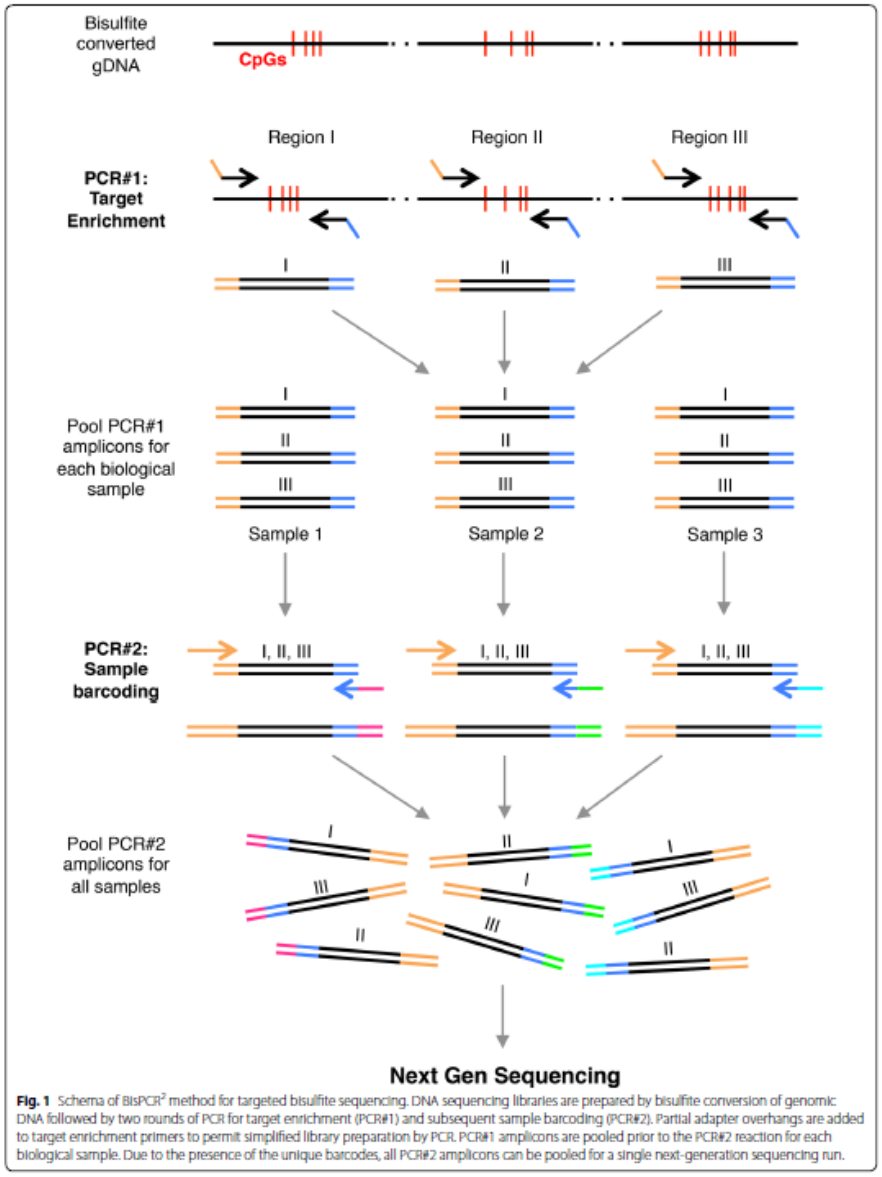
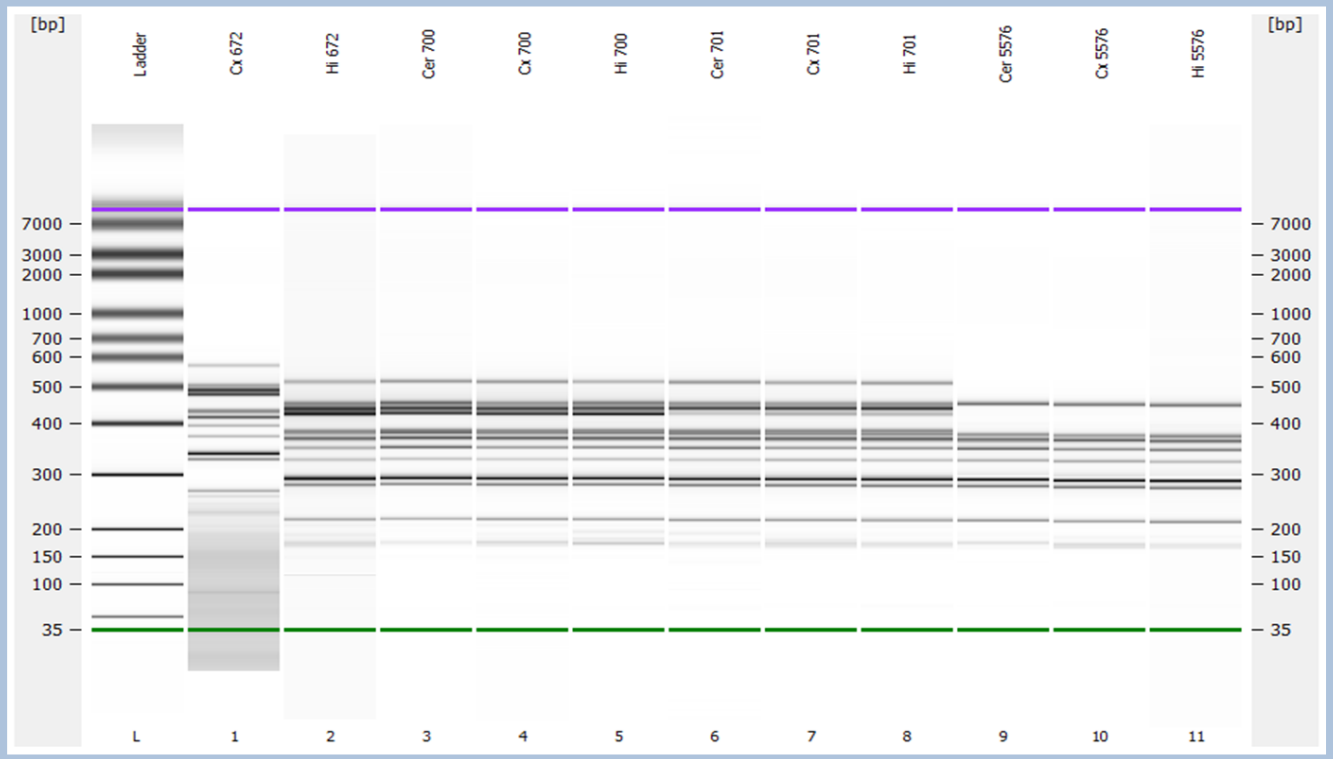


Fig. 1 Schema of BisPCR[2] method for targeted bisulfite sequencing. DNA sequencing libraries are prepared by bisulfite conversion of genomic DNA followed by two rounds of PCR for target enrichment (PCR#1) and subsequent sample barcoding (PCR#2). Partial adapter overhangs are added to target enrichment primers to permit simplified library preparation by PCR. PCR#1 amplicons are pooled prior to the PCR#2 reaction for each biological sample. Due to the presence of the unique barcodes, all PCR#2 amplicons can be pooled for a single next-generation sequencing run.

Bernstein *et al*. Epigenetics & Chromatin (2015) 8:27

QC

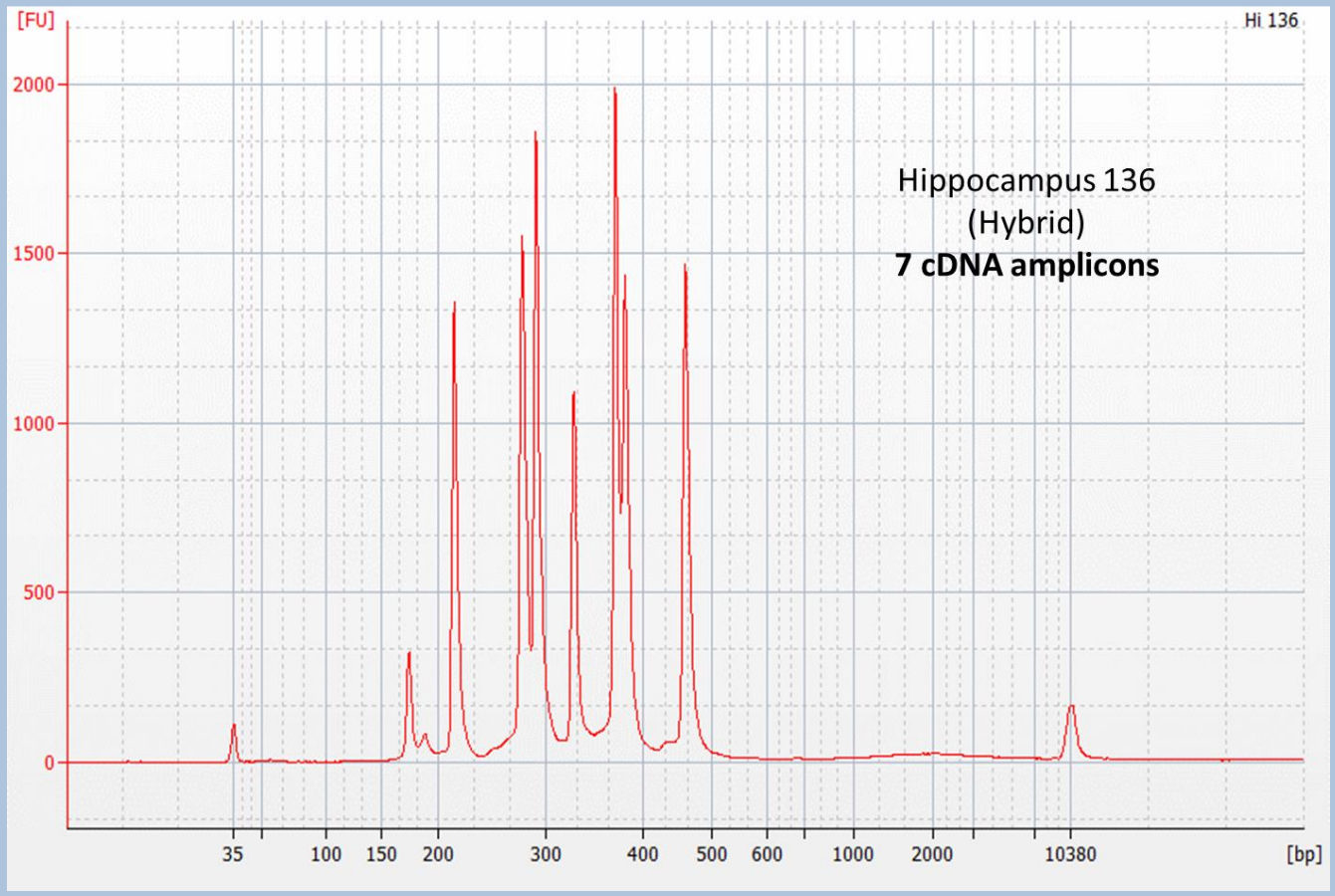
Bioanalyzer gel visualizing the amplicon fragments after pooling



Fragment length was determined by separating 1 ng of sample on an Agilent high sensitivity DNA assay using the 2100 Bioanalyzer (Agilent Technologies).

QC

Electropherograms

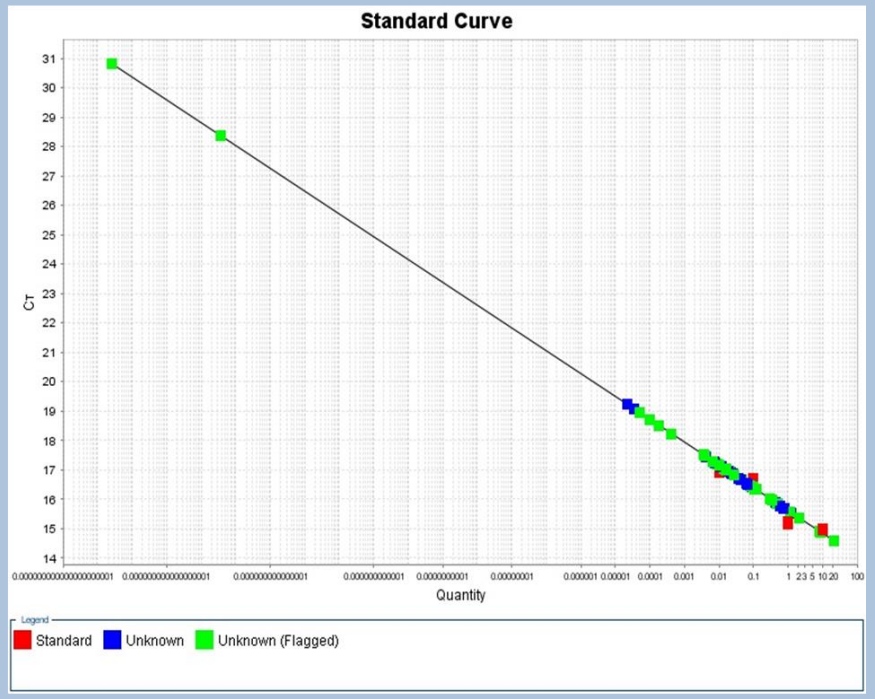


QC

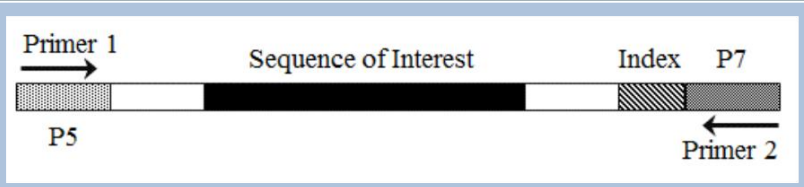
Molarity Quantification with qPCR
(Library Quantification Kit, Clontech Lab.)

Qubit measurement:
All amplicons are measured (100 %)

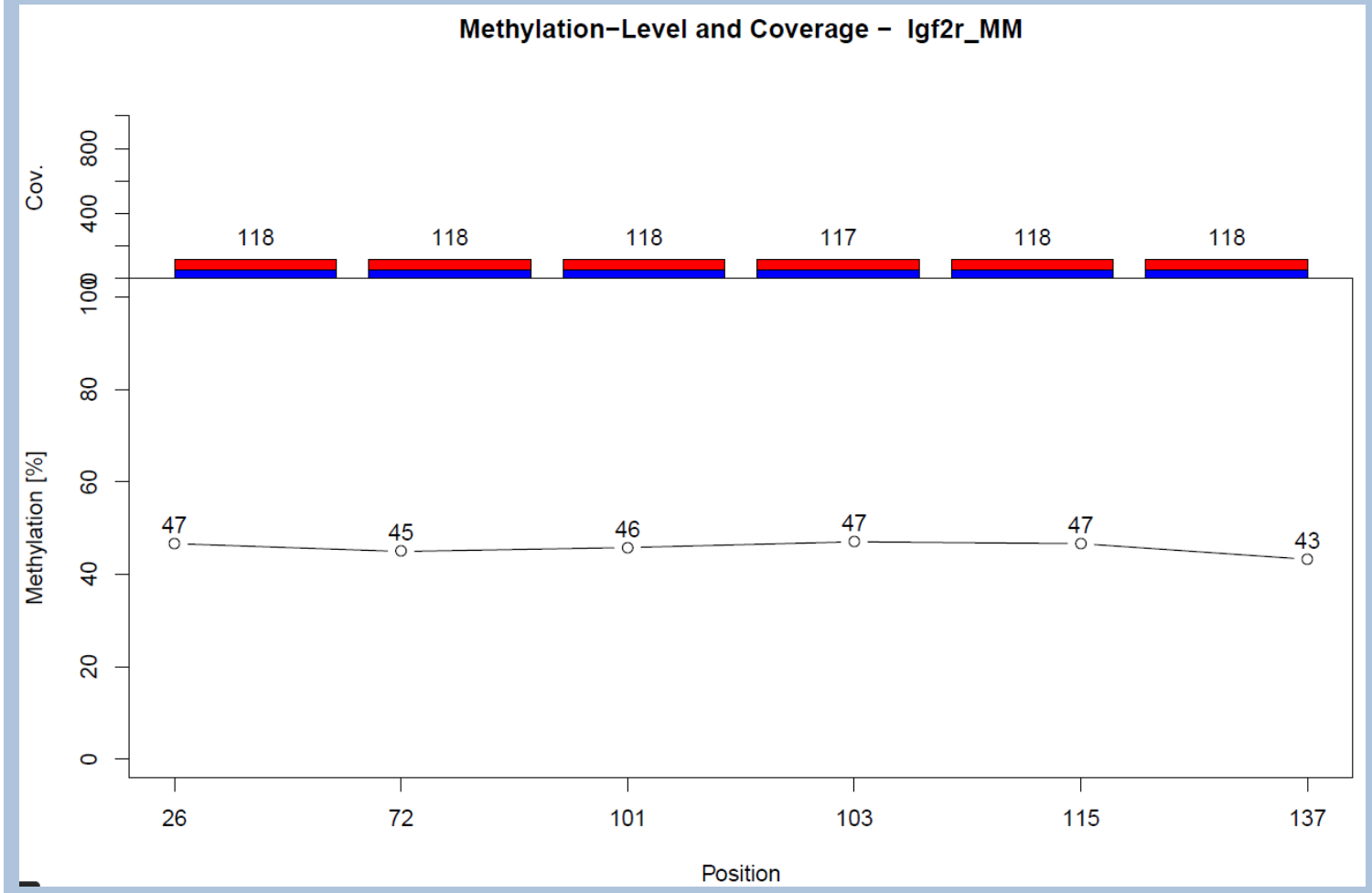
qPCR measurement:
Only amplicons with barcodes are measured
(Ø= 90 %)



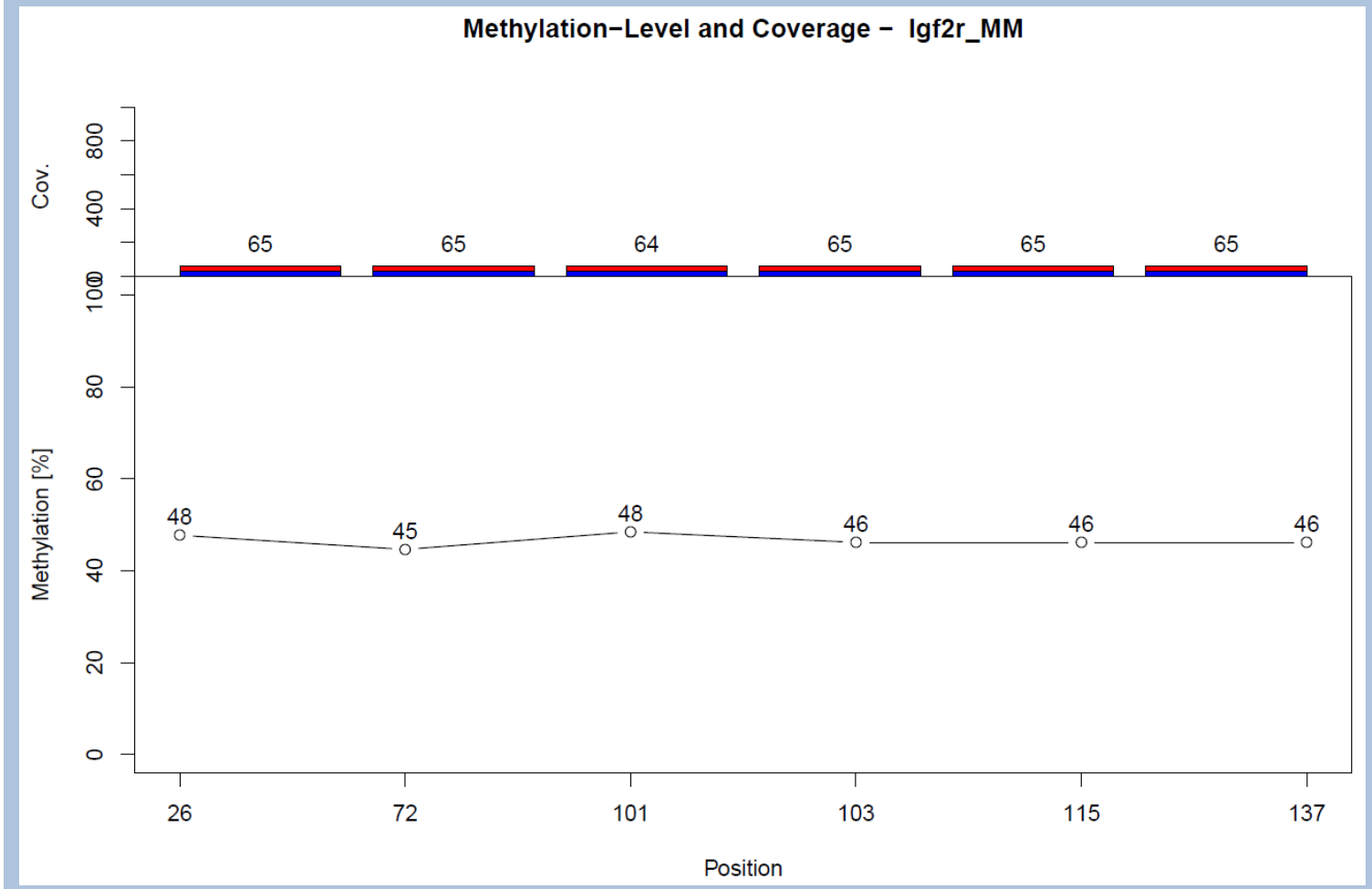
The primers are designed to anneal to the P5 and P7 regions at the ends of Illumina sequencing library fragments.



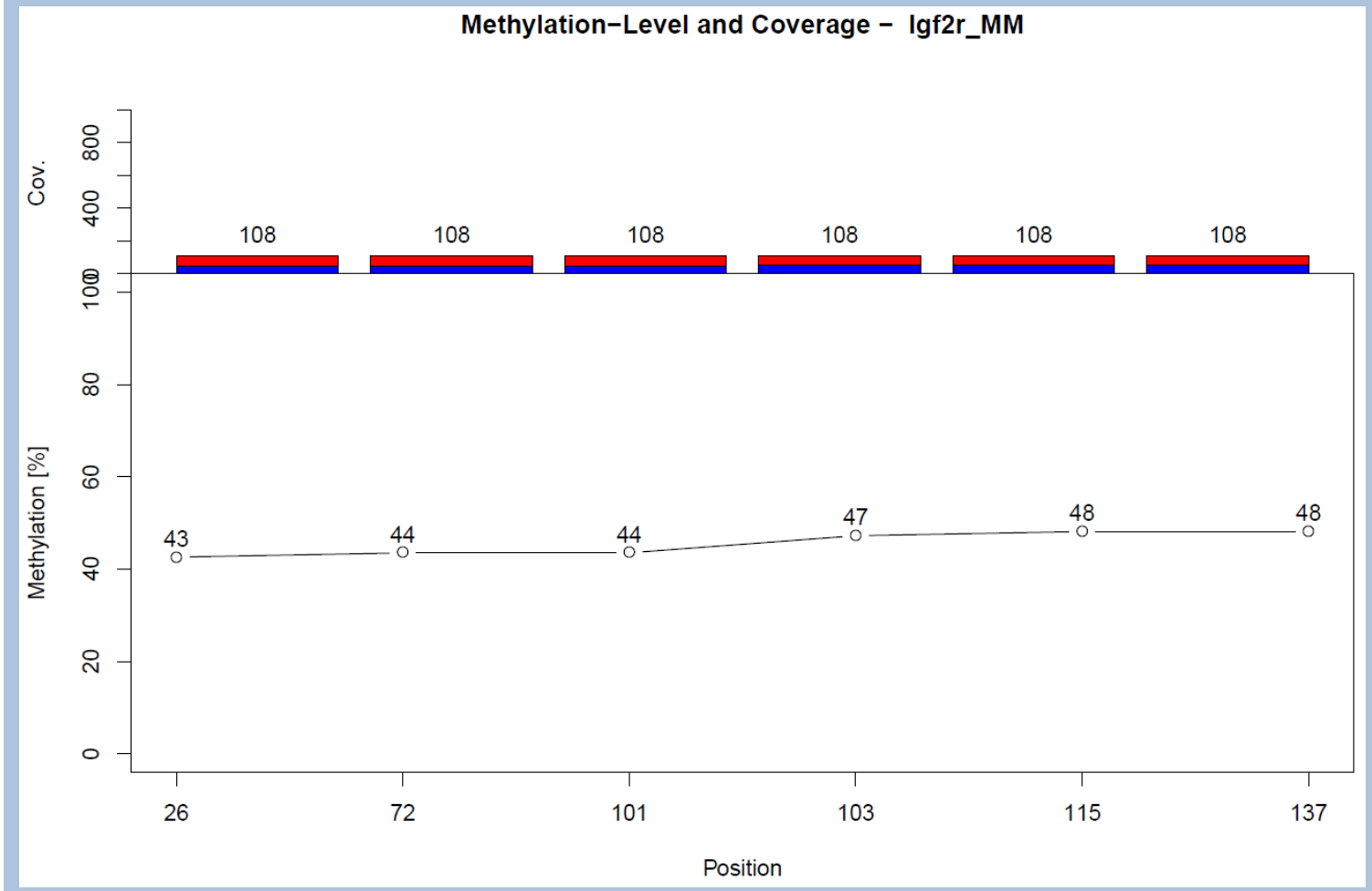
Murine Cerebellum DNA methylation *Igf2r*



Murine Cortex DNA methylation *Igf2r*



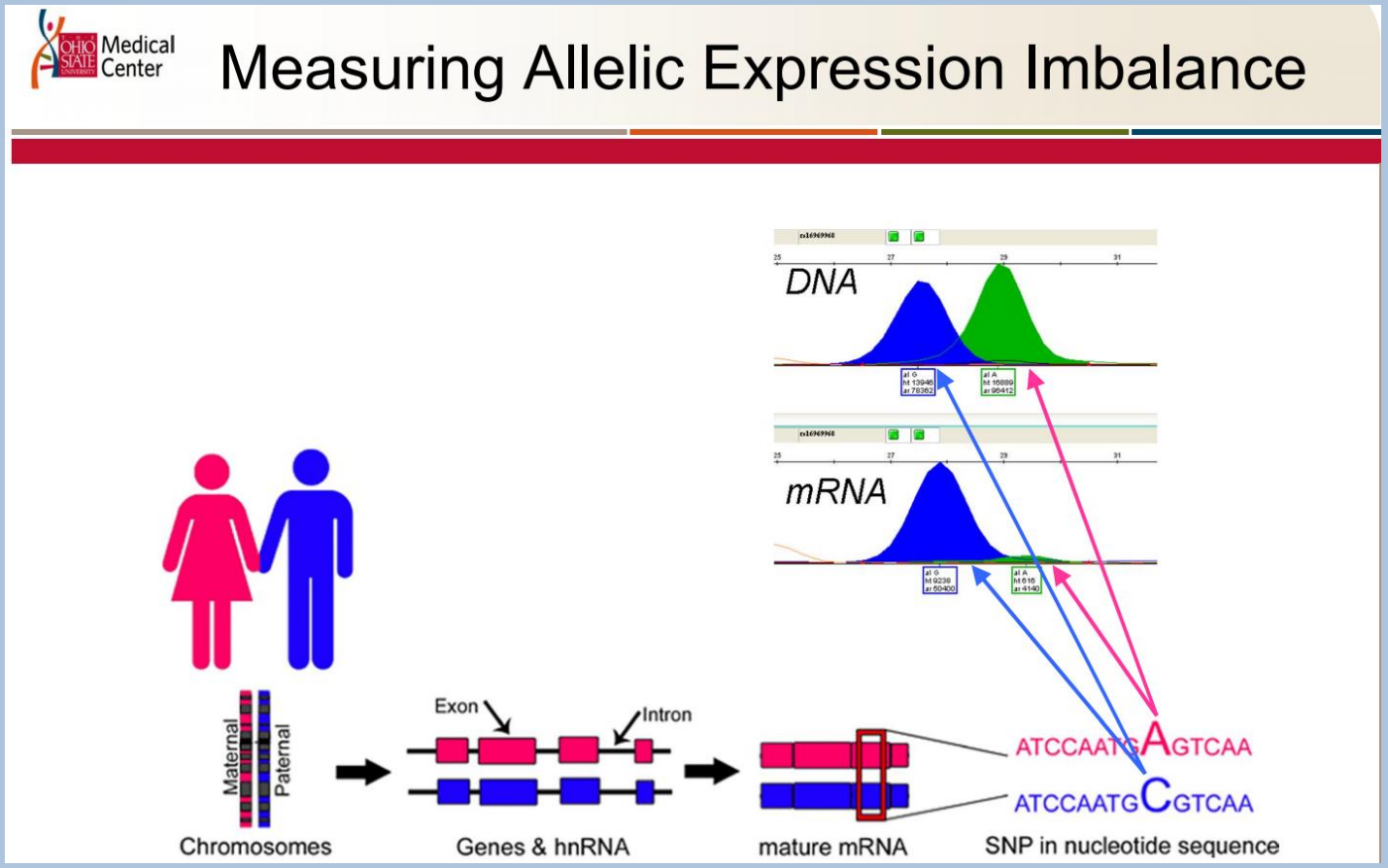
Murine Hippocampal DNA methylation *Igf2r*



BisPCR²: Conclusions

- Less costly than Pyrosequencing, 454 Pyrosequencing and BSAS
- Lot more throughput than 454: **4,5Gb (15 Mio reads)** vs. **0,035GB** (0,1 Mio Reads; Junior) or **0,7GB** (1 Mio reads; FLX)
- Workflow relative simple and streamlined compared to Roche 454 Pyrosequencing (Würzburg, Essen, Saarbrücken with Amplicon BS Sequencing on Roche)
- reliable results on methylation level of imprinted genes in somatic tissue
- Library preparation is less costly compared to BSAS by waiving Nextera XT
- **Epialleles could be reconstructed (no fragmentation of BS-PCR amplicons)**

Allele-specific RNAseq



Ryan M. Smith; Golden Helix Webinar October 27th, 2010

Allele-specific RNAseq

- ASE is a complex phenomenon caused by

a variety of mechanisms including:

  - imprinting
  - X-chromosome inactivation
  - modulating transcription factor binding affinity or
  - post-transcriptionally by changing splicing patterns

Areas with epigenetic mechanisms implicated in human nervous system function

Function or Disorder	Mechanism(s) Implicated
learning and memory	histone modifications, DNA methylation, piRNAs, miRNAs
maternal nurturing	histone modifications, DNA methylation
adult neurogenesis	histone modifications, DNA methylation
stress responses	histone modifications, DNA methylation
Alzheimer's disease	histone modifications, DNA methylation
Rett syndrome	MeCP2 methylcytosine binding
fragile X mental retardation	DNA methylation, miRNAs
schizophrenia	DNA and histone methylation, miRNAs
Rubinstein-Taybi syndrome	histone acetyltransferase deficiency
Prader-Willi/Angelman syndrome	genomic imprinting (DNA methylation)
depression and/or suicide	DNA methylation
bipolar disorder	histone modifications, DNA methylation, miRNAs
addiction and reward behavior	histone modifications, DNA methylation, miRNAs
PTSD	histone modifications, DNA methylation
ATR-X syndrome (α-thalassemia mental retardation)	SNF2 chromatin remodeling, H3.3
cognitive aging	histone modifications, DNA methylation
Coffin-Lowry syndrome	histone phosphorylation
Kleefstra syndrome	histone methylation
epilepsy	histone modifications, DNA methylation, miRNAs
autism	histone and DNA methylation? miRNAs?

Sweatt, Neuron 2013

Clinical characteristics of Prader-Willi syndrome (PWS)

Rare neurodevelopmental genetic disorder (1/20 000):

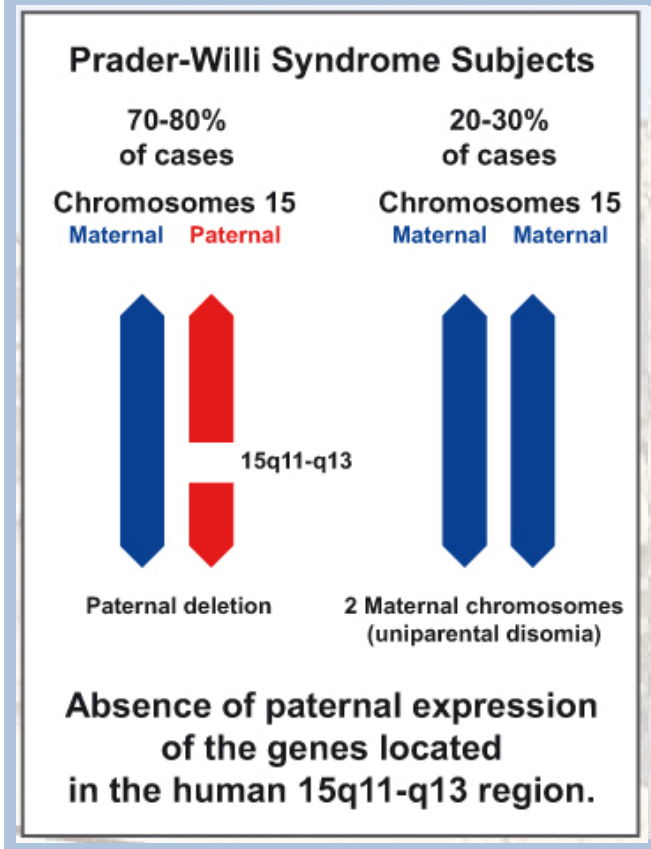
- Global developmental delay
- Behavioural & respiratory problems
- Variability in symptoms' intensity

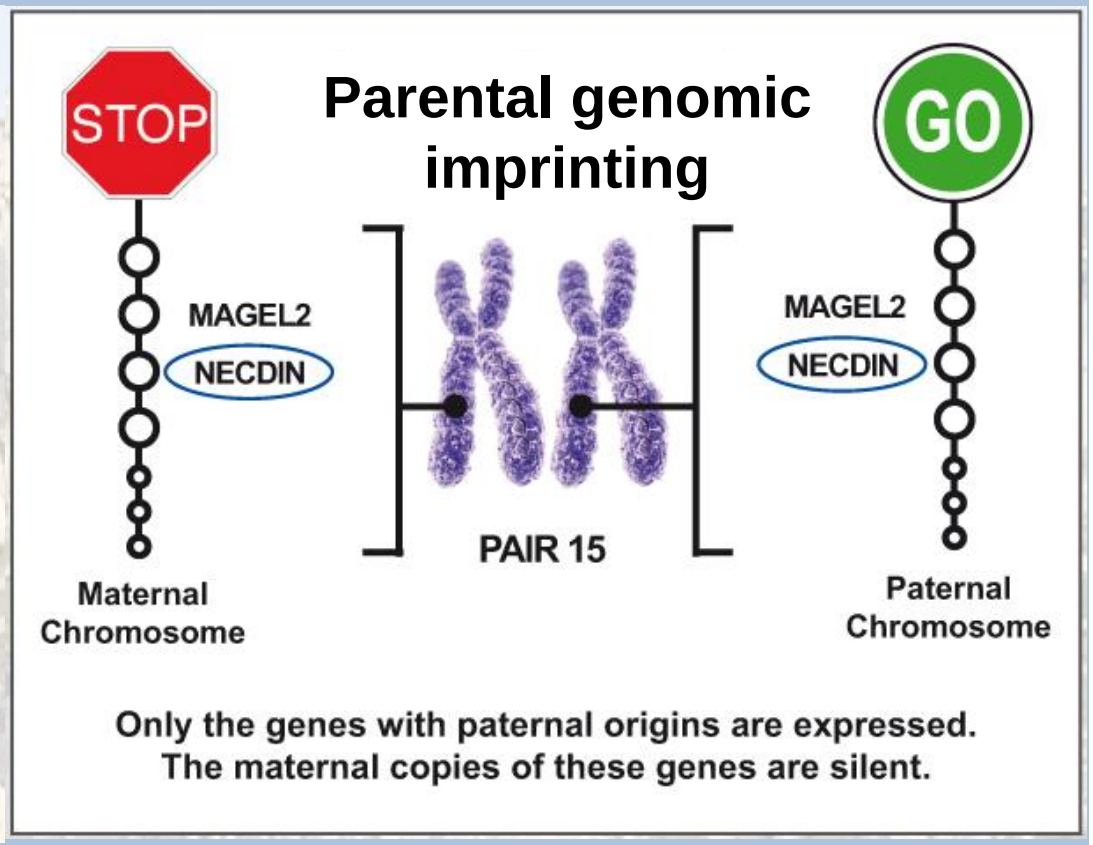
Hyperphagia

Severe neonatal hypotonia

- reduced fetal activity and reduced suckling at newborn age due to hypotonia
- weight gain between the ages of one and four due to hyperphagia
- short stature, hypogonadism
- distinctive facial features
- learning difficulties, mild intellectual disability
- Diabetes and cardiovascular disease at adult age

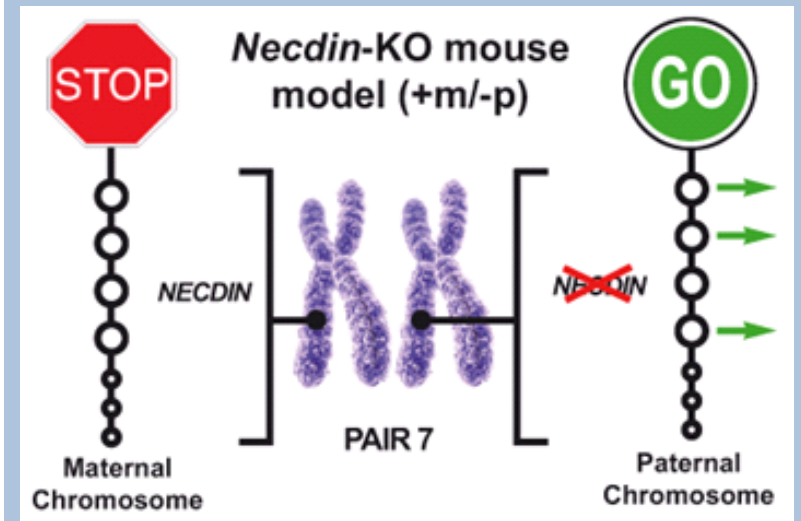
PWS is caused by loss of paternal gene expression on chromosome 15q11-q13





***Necdin (NDN)* gene:** maternally inherited allele normally silenced
paternally inherited allele normally expressed → deletion → PWS

Necdin deletion KO mouse model (+m/-p)



The maternal allele of the *Necdin* gene has been shown to be silent.

Phenotype of *Necdin* +m/-p mutant mice :

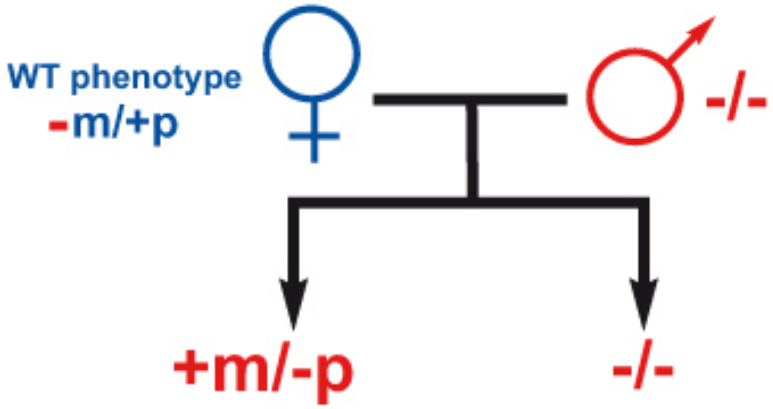
- Respiratory distress
- Growth retardation
- Sensory / motor deficits
- HIGH variability in symptoms' severity

Question:

Does this inter-individual variability in severity among *Ndn+m/−p* mice result from a "stochastic" activation of the putatively silent maternal allele of *Ndn*?

Rieusset et al., 2013, PLoS Genet

Survival difference between +m/-p and -/- *Necdin* KO mice



survival

n P0	57	56
n P1	51	46

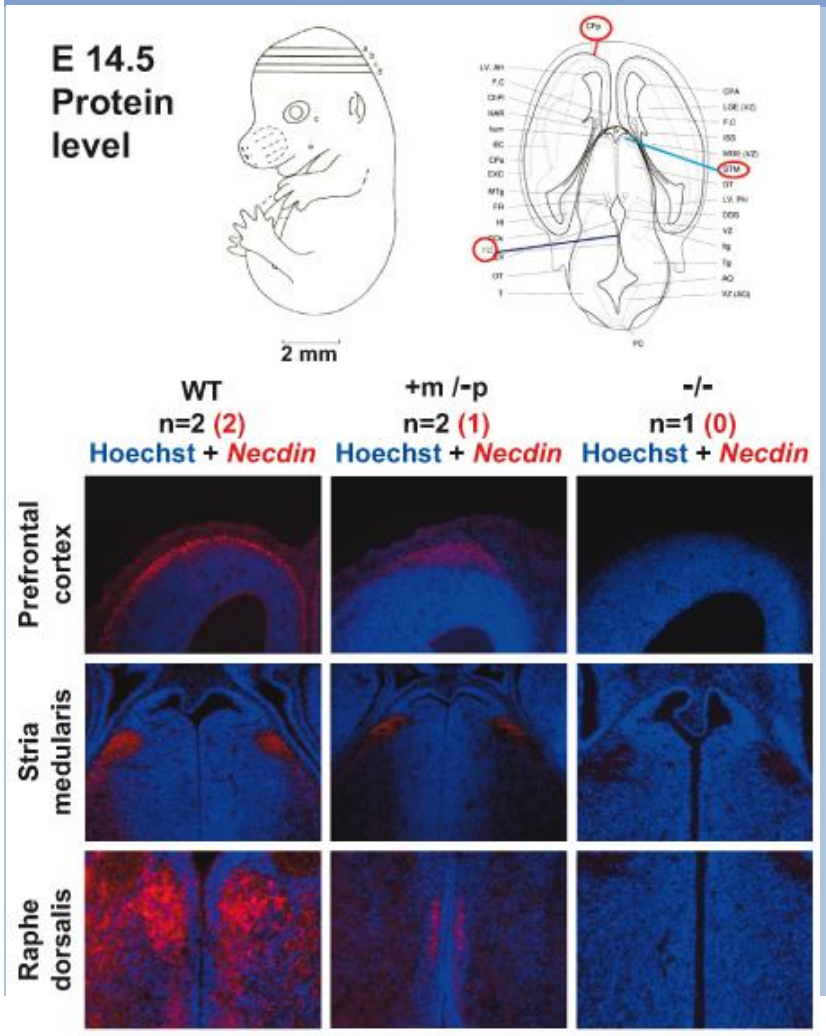
% of death

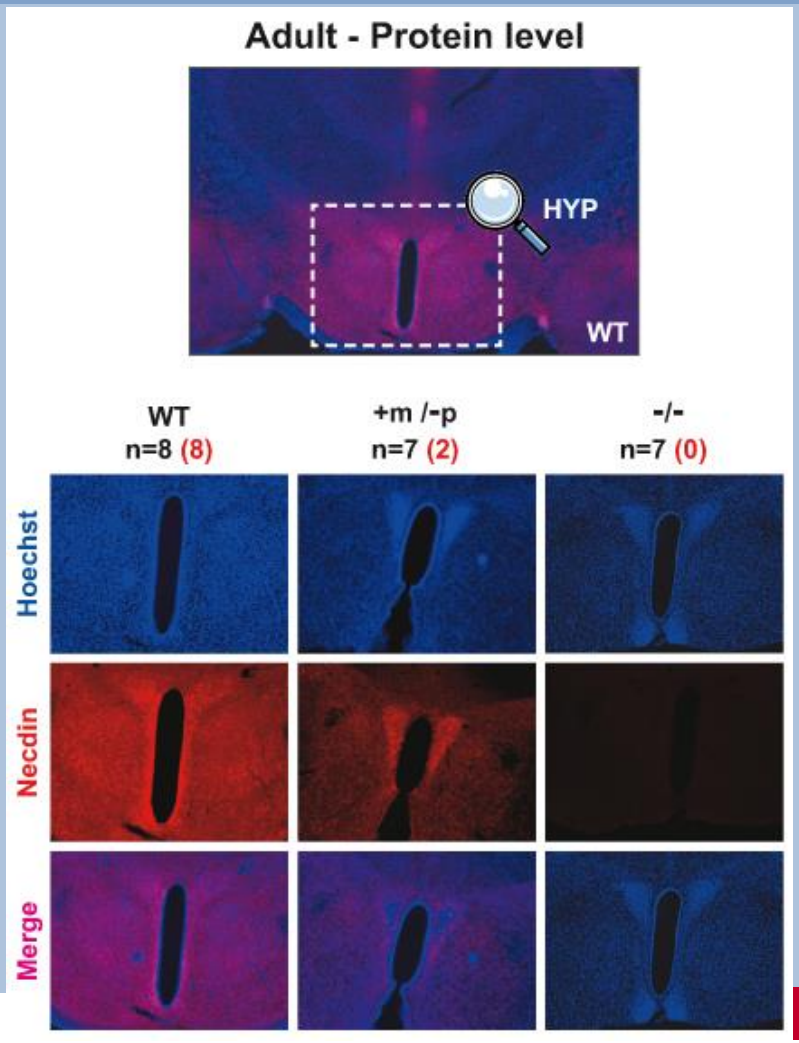
P0-P1	10%	20%

Maternal allele of the *Necdin* gene could be involved in the survival of +m/-p mutant mice.

Rieusset et al., 2013, PLoS Genet

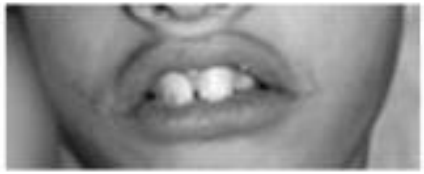
Maternal Necdin expression (loss of silencing) from E12 to adult stage in +m/-p mice

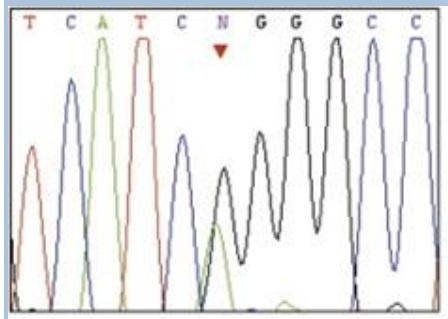




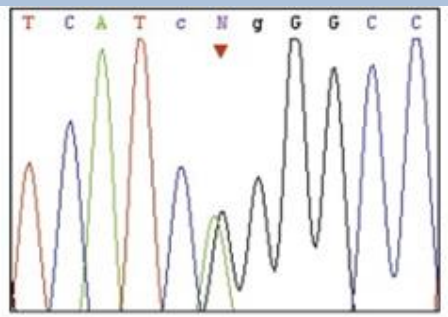
Rieusset et al., 2013, PLoS Genet

KCNK9 mutation accounts for the „Birk-Barel mental retardation dysmorphism syndrome“





Mother, obligatory carrier (5)



Affected (20)

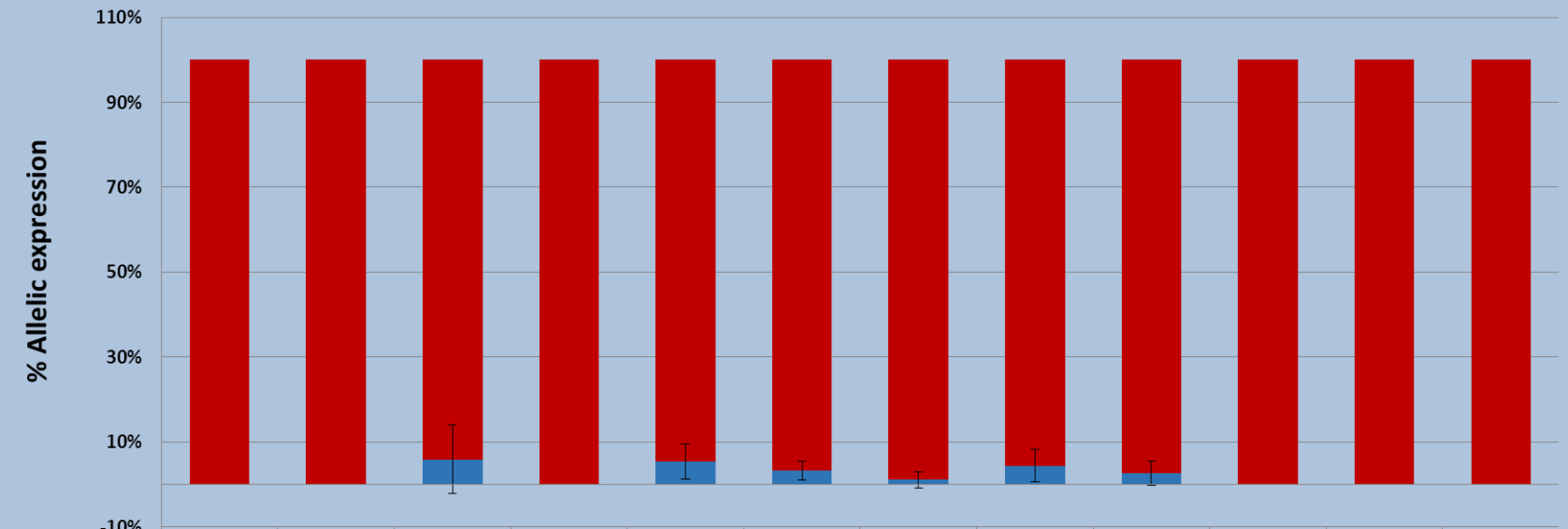
maternally inherited
→ paternally silenced gene

Clinical symptoms:
- **mental Retardation**
- **hypotonia**
- **elongated face**

***KCNK9*-mutation c.770G>A (p.Gly236Arg) prevents channel current in KCNK9-Homodimers as well as in Heterodimers with KCNK3**

Minor expression of the paternal *Kcnk9* allele in some brain areas

Allele-specific expression in wild-type-(B6^{mat} x Cast^{pat})F1-mice



	DG n=8	Hipp. n=11	OB n=3	Midbrain n=10	Cortex n=9	Cereb. n=7	Hypoth. n=3	Striatum n=3	OT n=3	Medulla n=3	Thalamus n=2	Pons n=2
Maternal (C)	100,0%	100,0%	94,1%	100,0%	94,6%	96,7%	98,9%	95,6%	97,4%	100,0%	100,0%	100,0%
Paternal (T)	0,0%	0,0%	5,9%	0,0%	5,4%	3,3%	1,1%	4,4%	2,7%	0,0%	0,0%	0,0%

AS-RNAseq - workflow

Wet lab

illumina

RNA isolation (C57/B6J x Cast/Ei) → cDNA synthesis → PCR 1 Gene-specific (Exon-bridging) → PCR 2 sample barcoding → Sequencing MiSeq V2 300 cycles

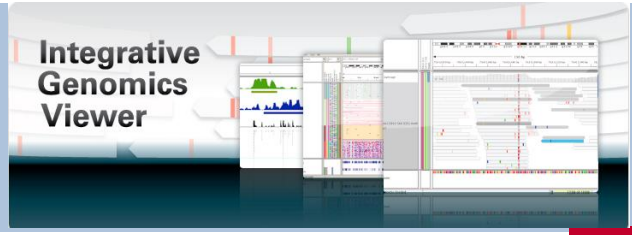
Bioinformatics

(artificial) Reference transcriptome → FASTQ reads → FASTQC → ALEA → Visualization IGV

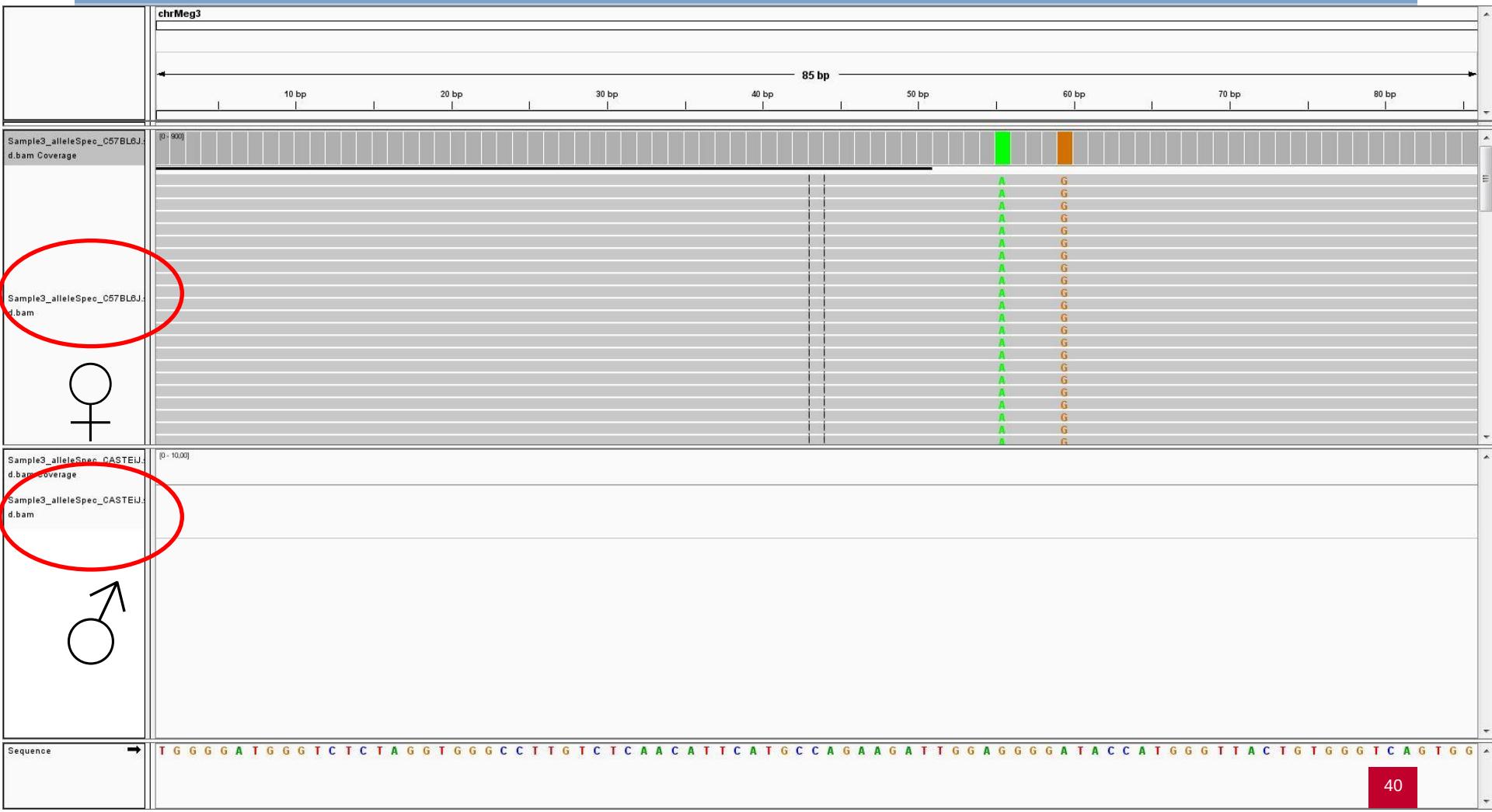
UCSC Genome Bioinformatics

ALEA

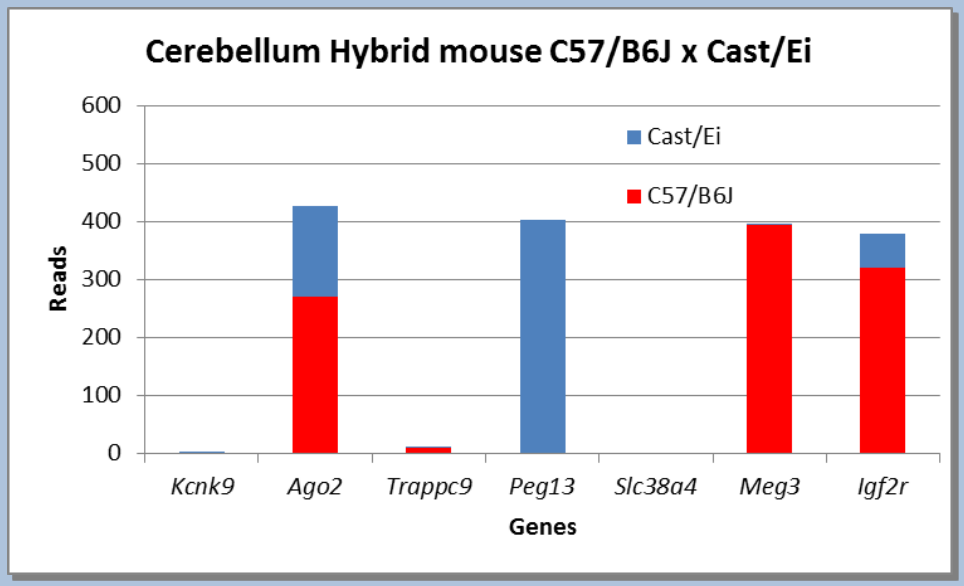
ALEA is a computational toolbox for allele-specific (AS) epigenomics analysis.

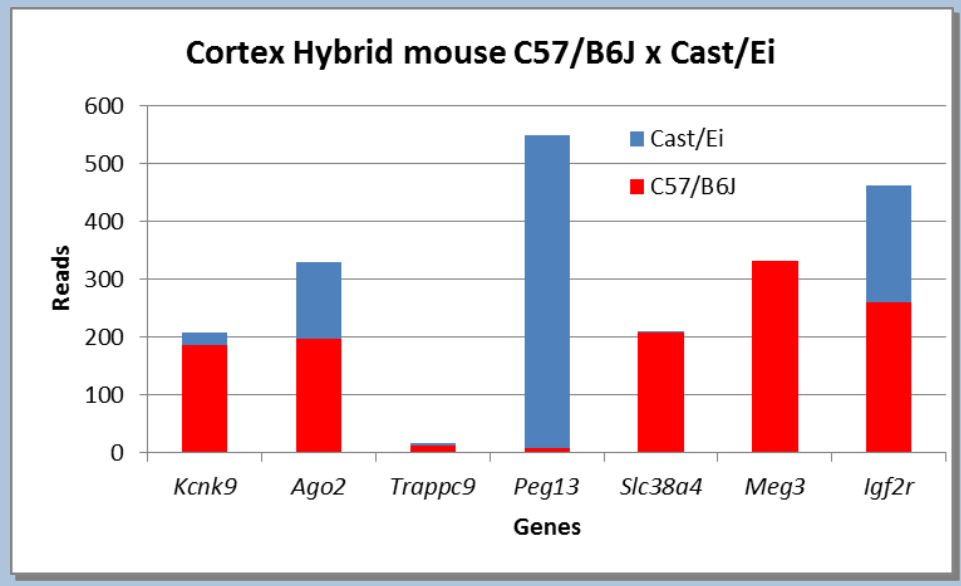


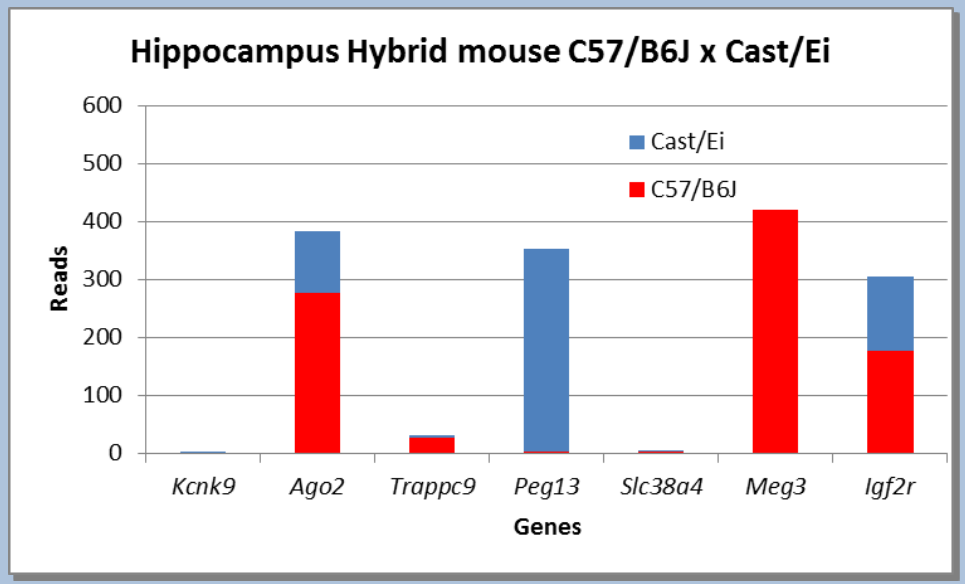
Meg3 – maternally expressed gene 3



Peg13 – paternally expressed gene 13


Cerebellum Hybrid mouse C57/B6J x Cast/Ei
600
500
400
300
200
100
0
Reads
Cast/Ei
C57/B6J
Kcnk9
Ago2
Trappc9
Peg13
Slc38a4
Meg3
Igf2r
Genes


Cortex Hybrid mouse C57/B6J x Cast/Ei
600
500
400
300
200
100
0
Reads
Cast/Ei
C57/B6J
Kcnk9
Ago2
Trappc9
Peg13
Slc38a4
Meg3
Igf2r
Genes


Hippocampus Hybrid mouse C57/B6J x Cast/Ei
600
500
400
300
200
100
0
Reads
Cast/Ei
C57/B6J
Kcnk9
Ago2
Trappc9
Peg13
Slc38a4
Meg3
Igf2r
Genes

Acknowledgements

Ulrich Zechner

Alexis Cooper

Stefan Diederich