# Structural and functional characterization of five bovine cell lines

D. Becker, G.C.M. Moreira, F. Hadlich, A. Sharma, V. Blanchet, E. Clark, C. Pfarrer, J. Vanselow, C. Charlier, D. Rocha, C. Kuehn

**BovReg Final Conference - Brussels**

*(14th February 2024)*

*This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 815668*

# Cell Lines in Animal Research

- Cell lines from farmed animal species are used as *in vitro* surrogates for *in vivo* experiments, but they lack functional annotation

- Cell lines with cell aneuploidy and heteroploidy often show functional consequences[1,2] → molecular characterization of cell lines is important to gain insight into possible alterations

## Objective

**Characterization of cell lines frequently used in bovine research at functional and structural level**

(1) de Vos et al. (2023) iScience 26(3):106252
(2) Williams et al. (2008) Science 322(5902): 703–709.

# The Five Bovine Cell Lines

(1) Rutter and Luther (1984) Vet Rec 114(16):393-396
(2) Hambruch et al. (2010) Placenta 31(1):67-74
(3) Madin and Darby (1958) Proc Soc Exp Biol Med 98(3):574-576
(4) Huynh et al. (1991) Exp Cell Res 197(2):191–199
(5) Sharma et al., unpublished

- **EBL**(1)

embryonic bovine lung cells: established from a *lung* of a 7-month old bovine fetus, spontaneously immortalized

- **F3**(2)

*Trophoblast* cells isolated from a bovine cotyledon of a male fetus (approx. 5 month of gestation), spontaneously immortalized

- **MDBK**(3)

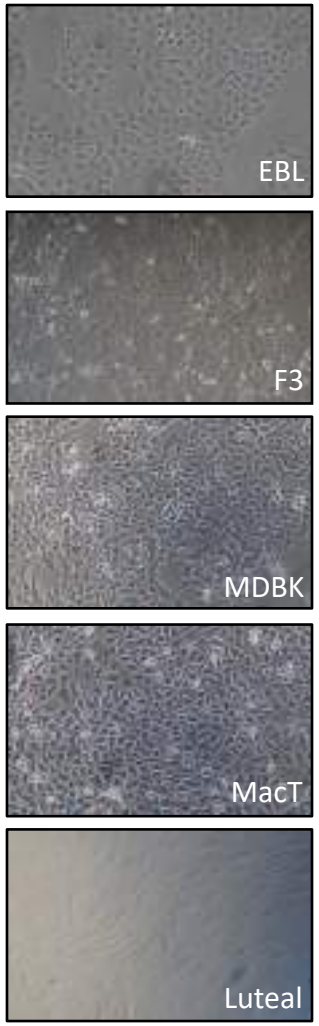
Madin-Darby Bovine Kidney cells: derived from a *kidney* of an adult steer, spontaneously immortalized

- **MacT**(4)

generated from *mammary alveolar cells* using SV40 large T antigen

- **Luteal cells**(5)

generated from ovary tissue from lactating Holstein cow



# Omics Data

- **Whole-genome data**
  - Genomic DNA of cell lines was sequenced on a NovaSeq6000
  - Data was analyzed using the nf-core Sarek workflow(1)
  - SNPs and small indels were called with HaplotypeCaller (GATK)(2)

- **Whole-transcriptome data**
  - polyA+ libraries from EBL, MacT, MDBK and Luteal cells were sequenced on a HiSeq2500 (Illumina), the F3 RNA-Seq library was sequenced on a NextSeq2000 (Illumina)
  - Data was analyzed using the nf-core rnaseq-3.3 and 3.4 pipelines

- **Epigenomic data**
  - Cell pellets distributed to partners
  - ATACseq libraries sequenced on NovaSeq, peak calling with BovReg nf-core ATACseq pipeline (ULIEGE)

(1) Garcia et al. (2020) F1000Res 9:63
(2) McKenna et al. (2010) Genome Res 20(9):1297–1303

# Genomic Features – SNPs and small Indels

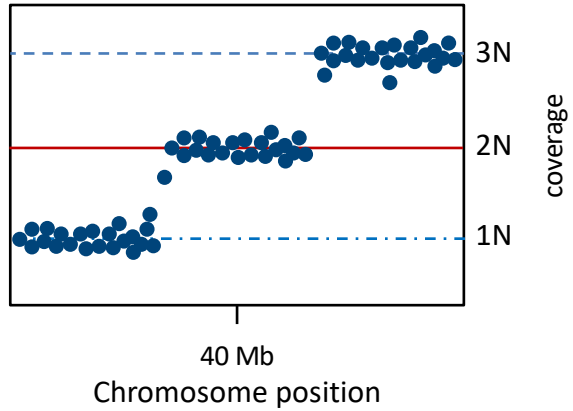
| | EBL | F3 | MacT | MDBK | Luteal |
|---|---|---|---|---|---|
| SNPs | 7,361,455 | 8,870,406 | 7,327,940 | 7,804,410 | 8,295,578 |
| Indels | 1,399,610 | 1,402,483 | 1,513,700 | 1,507,973 | 1,529,942 |
| **Total** | 8,761,065 | 10,272,889 | 8,841,640 | 9,312,386 | 9,825,520 |

- **Ø 7.8 million single nucleotide polymorphisms (SNPs) and 1.45 million small indels**
- **No clear difference between spontaneously immortalized and transfected cell lines**

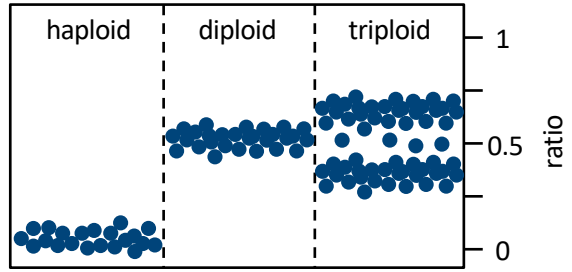
# Genomic Features – large structural aberrations

## Criteria

- **Read depth:** coverage of each chromosome should be similar to the average coverage




- **Allele support for heterozygous SNPs:** ratio of reads that support the alternative allele for heterozygous SNPs should be around 0.5 for diploid chromosomes

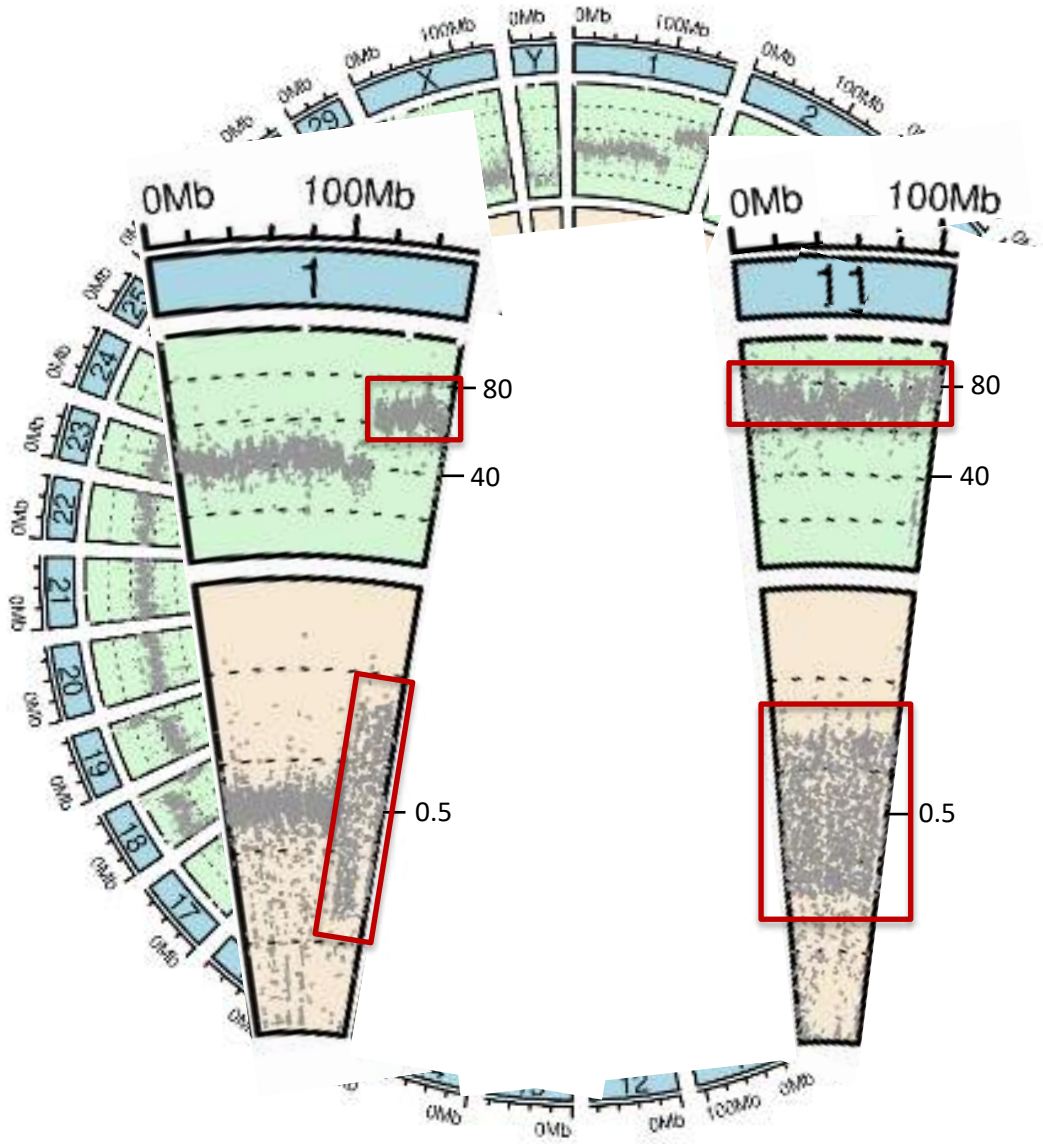


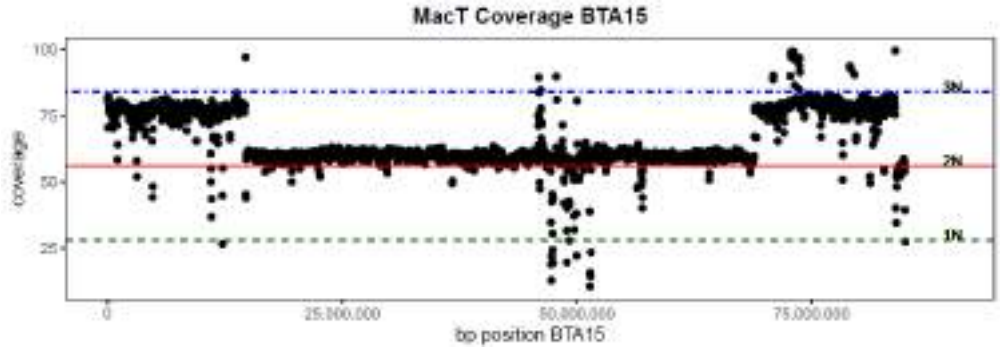
Read depth

Allele support ratio

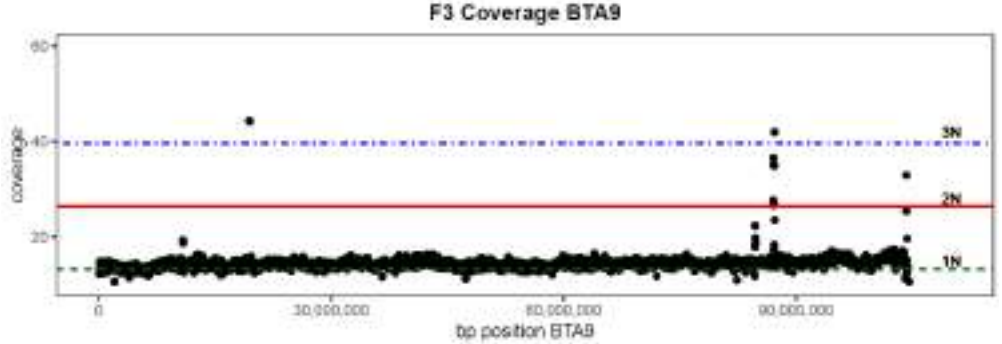
(in 50kb bins)

EBL Ø coverage = 49.8X

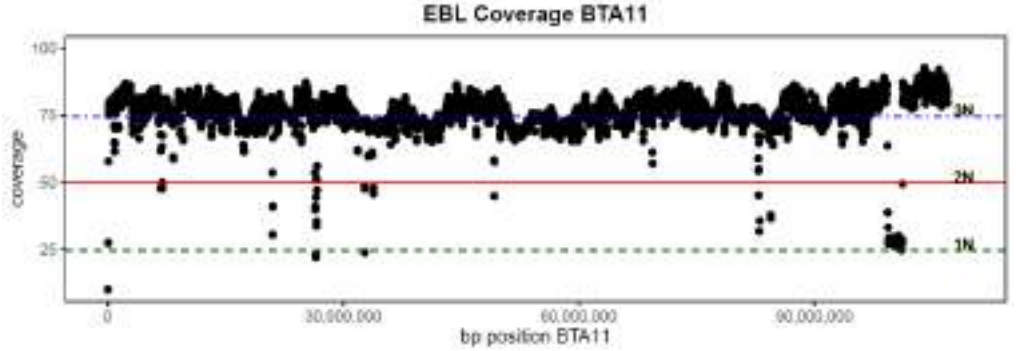





MacT BTA15 partly triploid

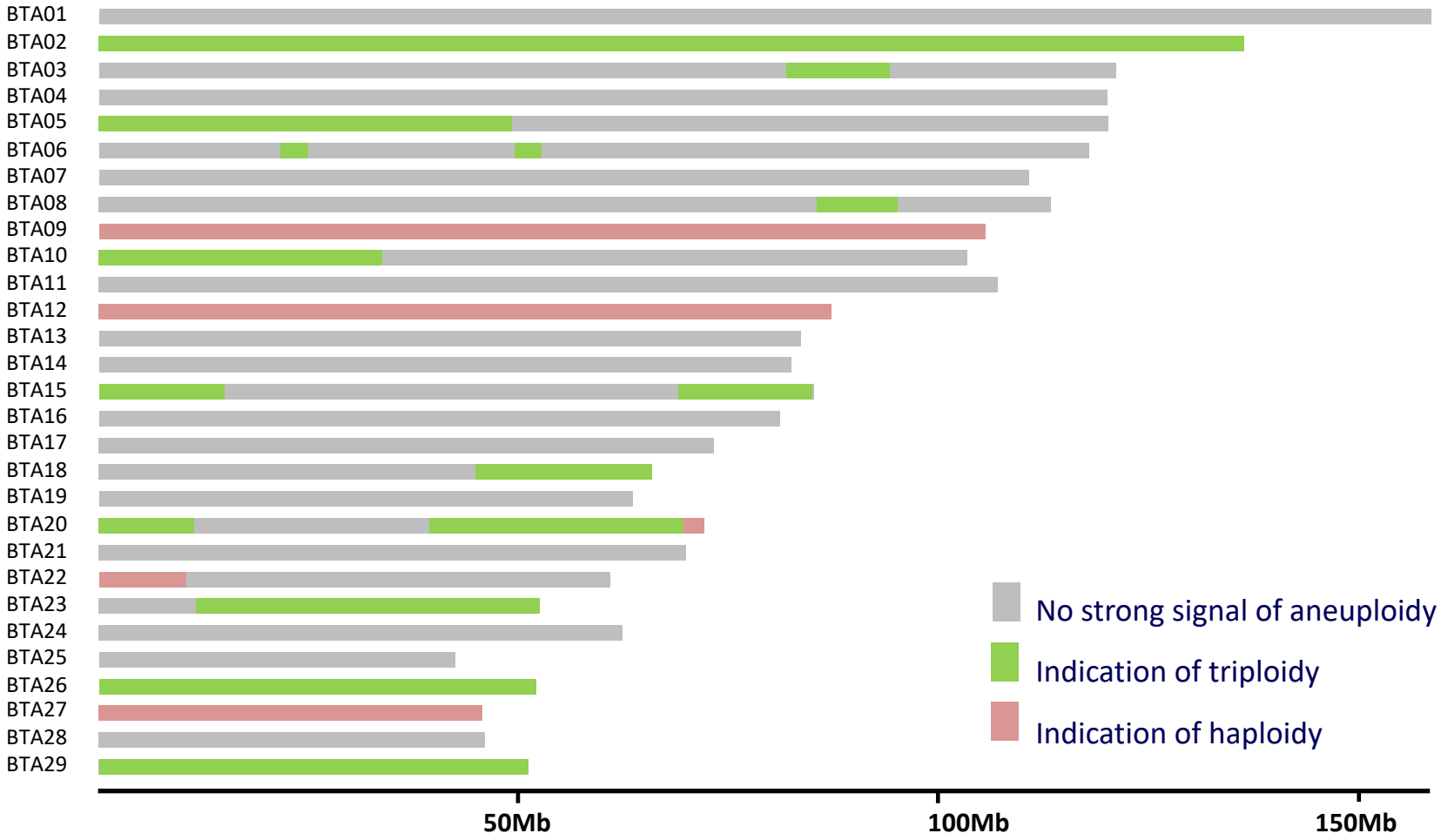



F3 BTA09 haploid




EBL BTA11 triploid

# MacT - large structural aberrations



# Genomic Features – large structural aberrations

| Cell Line | Aneuploidy (whole chromosome) | Structural variation (part of the chromosome) |
|---|---|---|
| EBL | 1 | 11 |
| F3 | 6 | 2 |
| MacT | 6 | 11 |
| MDBK | 2 | 7 |
| Luteal | 0 | 7 |

- **Aneuploidy and structural variations are common in all characterized cell lines**

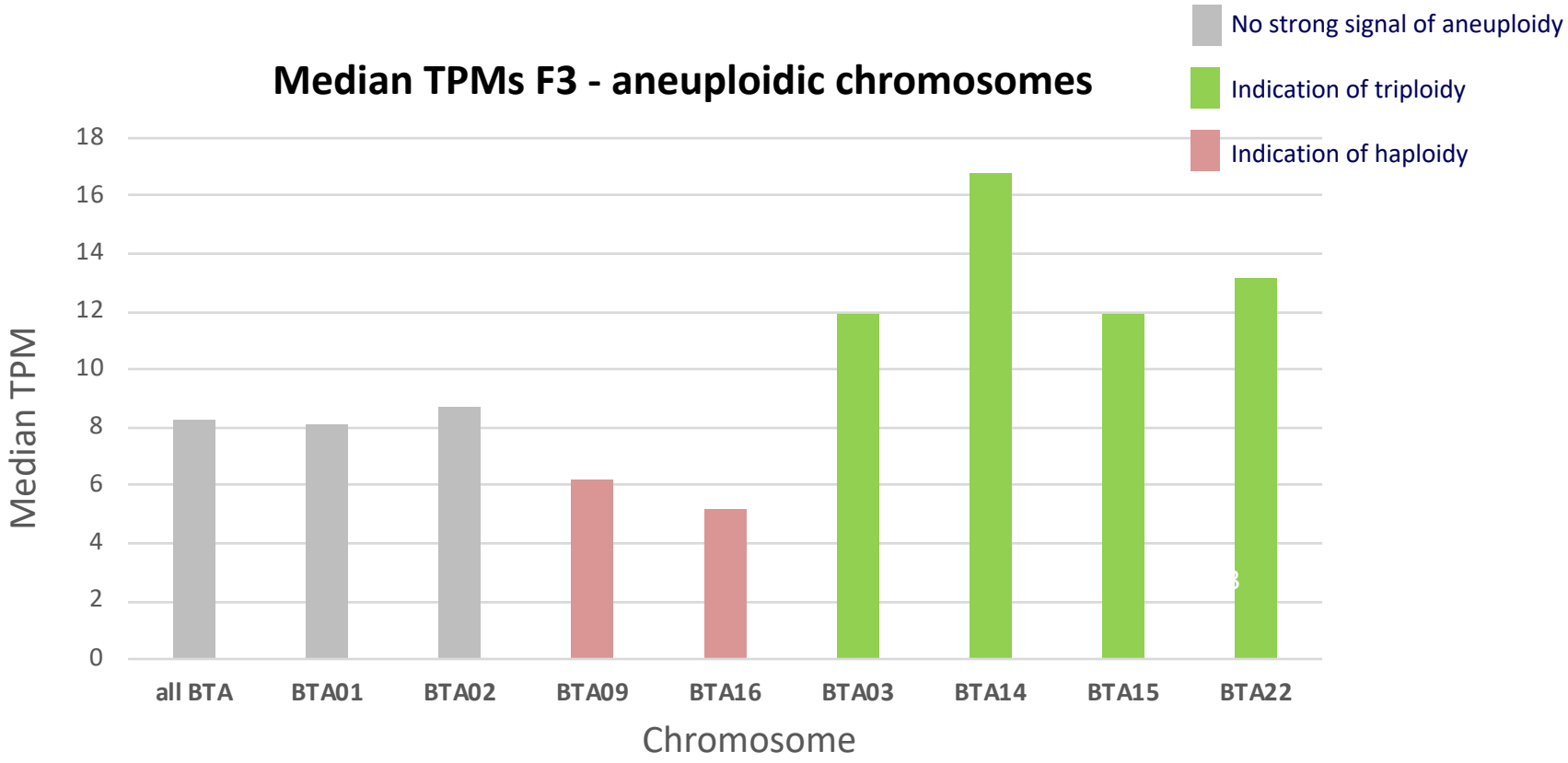
# Expression Activity

| | TPM > 1[a] | TPM > 0.1[b] | Median TPM[c] |
|---|---|---|---|
| **EBL** | 12,205 | 14,423 | 12.26 |
| **F3** | 11,348 | 13,280 | 8.28 |
| **MacT** | 11,478 | 13,709 | 9.90 |
| **MDBK** | 12,273 | 14,568 | 11.88 |
| **Luteal** | 11,780 | 14,186 | 8.98 |

- **Between 13,280 (F3) to 14,568 (MDBK) genes were expressed per cell line**

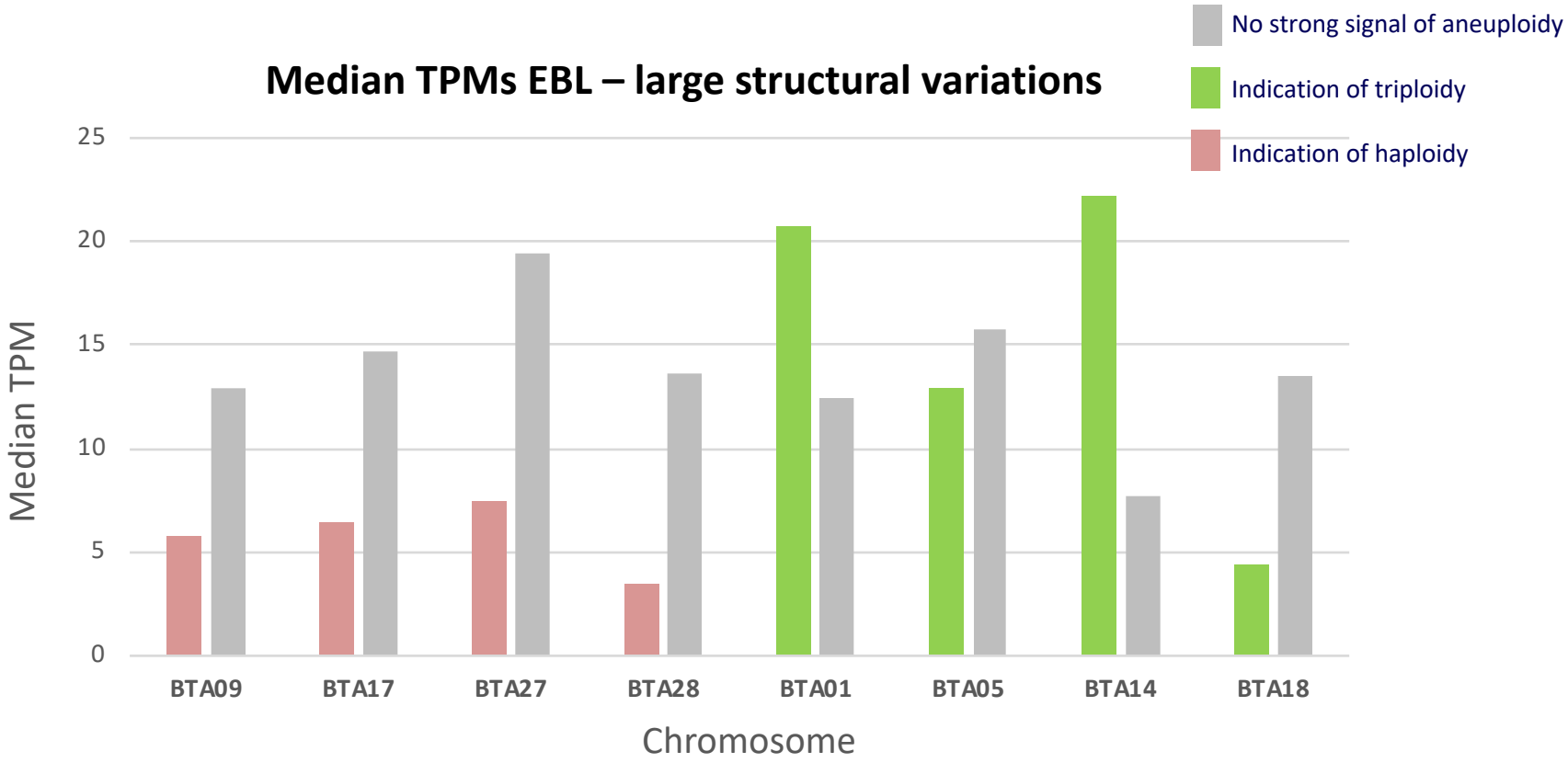
a: number of genes with TPM > 1
b: number of genes with TPM > 0.01
c: median TPM (only genes with TPM>0)

# Expression Activity – Aneuploidic chromosomes

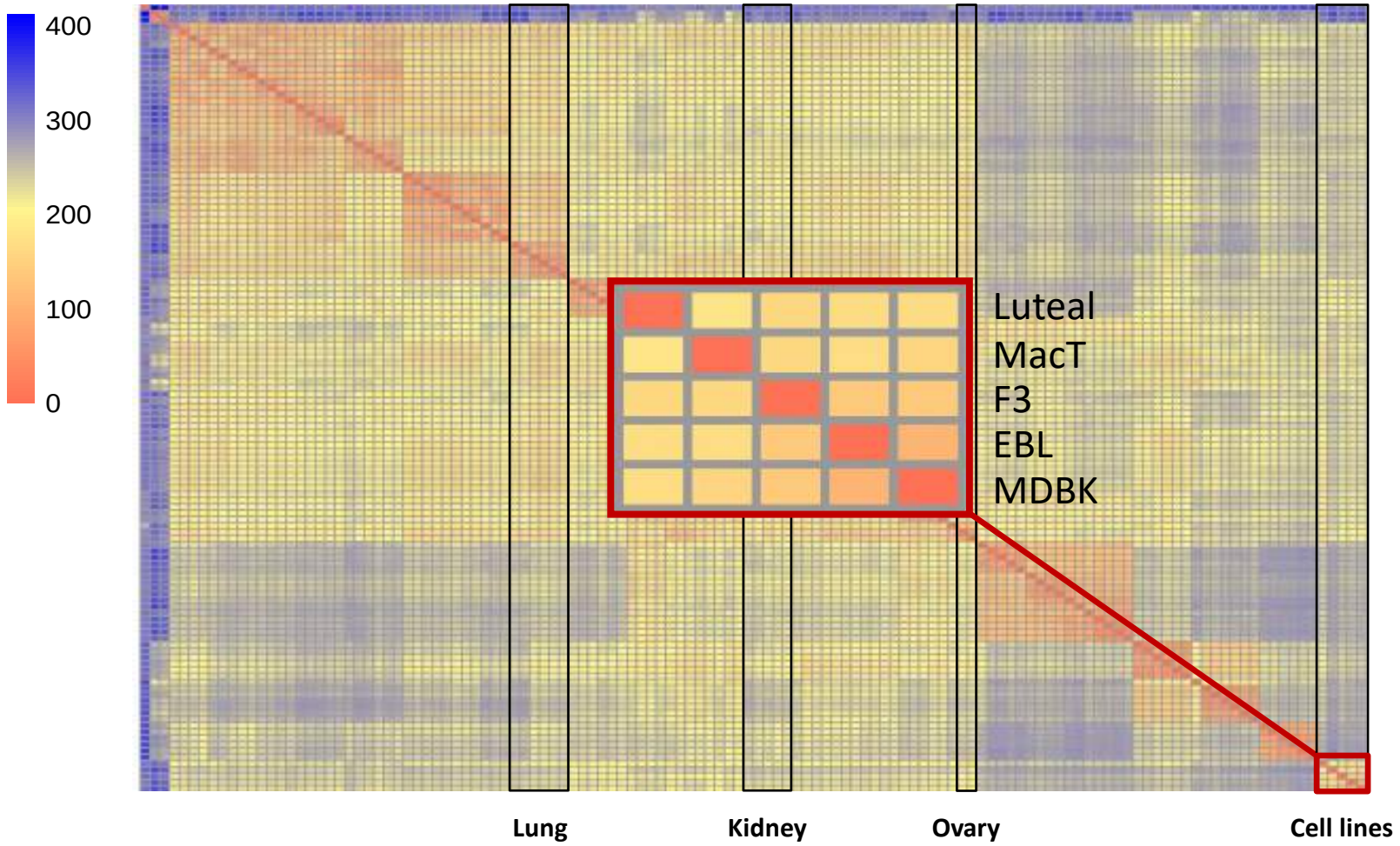



- **Aneuploidy is reflected in the transcriptome**

# Expression Activity – large structural variations



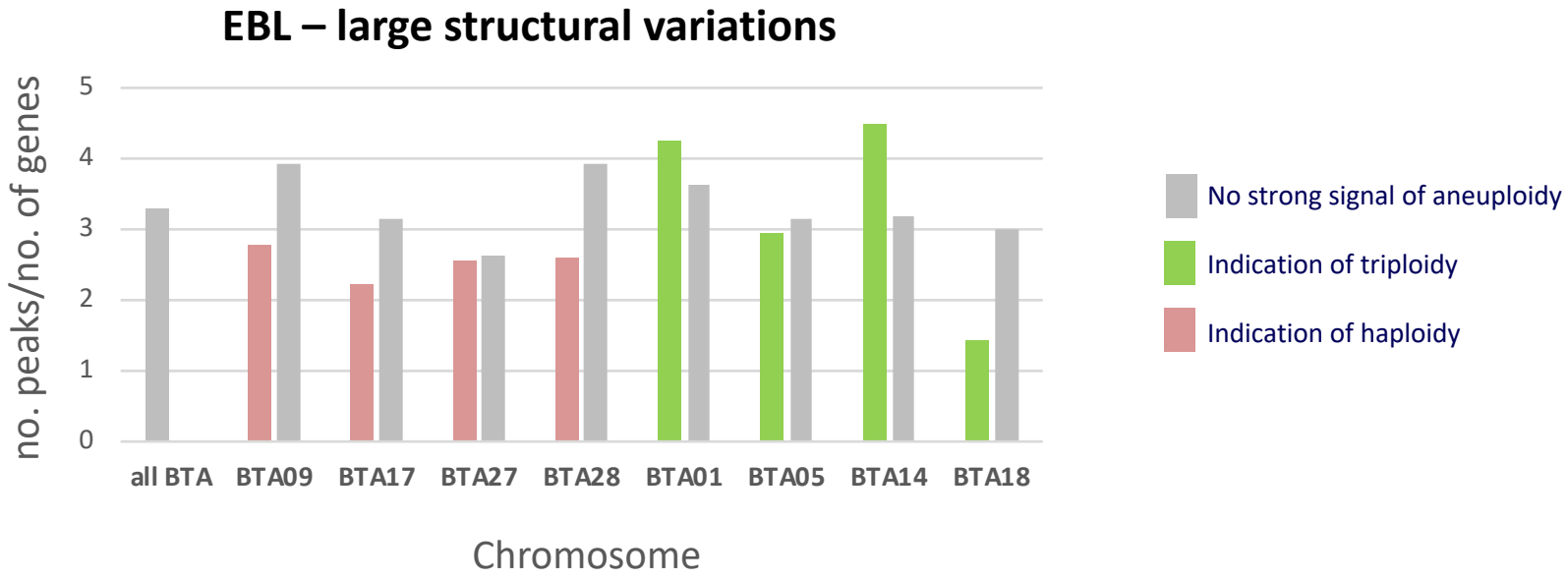

- **Structural variations are reflected in the transcriptome, but duplications do not necessarily lead to a higher expression, whereas deletions seem to decrease the expression activity**




- **Cell lines cluster together and only show low correlation with tissue of origin**

# Epigenomic data - ATACseq

| | EBL | F3 | MacT | MDBK | Luteal |
|---|---|---|---|---|---|
| Total Number of ATACseq peaks | 86,123 | 112,554 | 86,207 | 96,037 | 127,284 |




- **Chromosomal structural variation is partly reflected in chromatin accessibility**

# Summary

- **Aneuploidy and structural variations are common in all characterized cell lines**
  - Identification of various chromosomes that were either (partly) haploid or triploid

- **Different omics data provide information about the genome structure, transcriptome and chromatin accessibility of five bovine cell lines**
  - Reference of the genome architecture of these cell lines for future functional studies
  - Further omics data will deliver even more details (WGMS, CHiPseq, CAGE, HiC)

- **Chromosomal structural variation is reflected in transcriptomic profile and partly in the epigenomic profile**

*„All models are wrong, but some are useful.“*
G. Box, Statistician

# PARTNERS

*Thank you for your attention*

**www.bovregproject.eu**

*This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 815668*