ENCODE and Cancer Nov 17 Notes

> By Jason Bedford

Presentation by Peng Jiang from Shirley Liu Lab

Slide 3

We have a lot of data from ENCODE and TCGA we want to know: Which TFs drive gene specific expression

Slide 4

But there are some problems 1. Confounding factors of tumor gene expression. 2. Public datasets may not match the conditions of each other.

Slide 5

We know that tumor gene expression variation is driven by Copy Number Alteration and DNA Methylation (Li et al., 2013).

Slide 6

ENCODE ChIP-seq condition versus Tumor conditions Cell lines are not the same as tumors.

1. ENCODE ChIP-seq profiles are done in cell line conditions.
2. Each cancer type has unique physiological condition.

Slide 7

Rabbit: Regression analysis with background integration.

Slide 8

Identifying significant TFs using a multivariate linear model.

Gene CNA + Promoter methylation Promoter degree + CpG content + TF Regulatory score

Slide 9

R: TF Regulatory score B: Background factors

Slide 10

a review of alternative state of the art method LASSO

Slide 11

comparison of computational complexity and CV error.

Slide 13

applying Rabbit to some other data sets.

Slide 14

results of the transcriptional landscape

Slide 15

comparing to current cancer gene databases.

Slide 16

comparing to other data sets we find that

Slide 17

novel finding on breast cancer TF regulation

Slide 18

extending to rna binding motifs

Slide 20

the multivariate learner model for rna binding.

Slide 21

interesting finding.

Slide 22

overall summary.

Questions