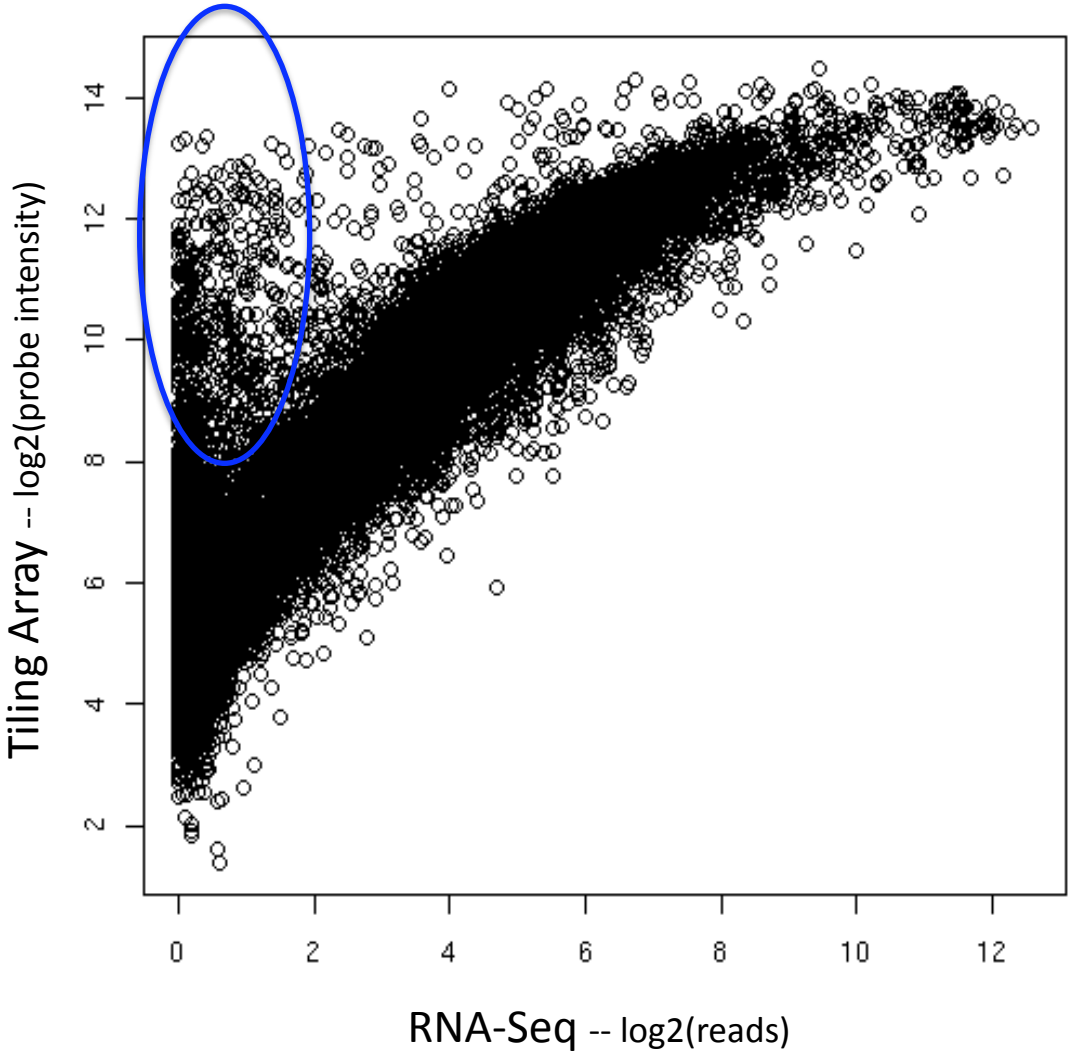
Transcriptome Comparison Between Tiling Arrays and Next Generation Sequencing on Matched Worm Samples, Towards Making Optimal Use of Arrays

ENCODE/modENCODE Meeting 2009

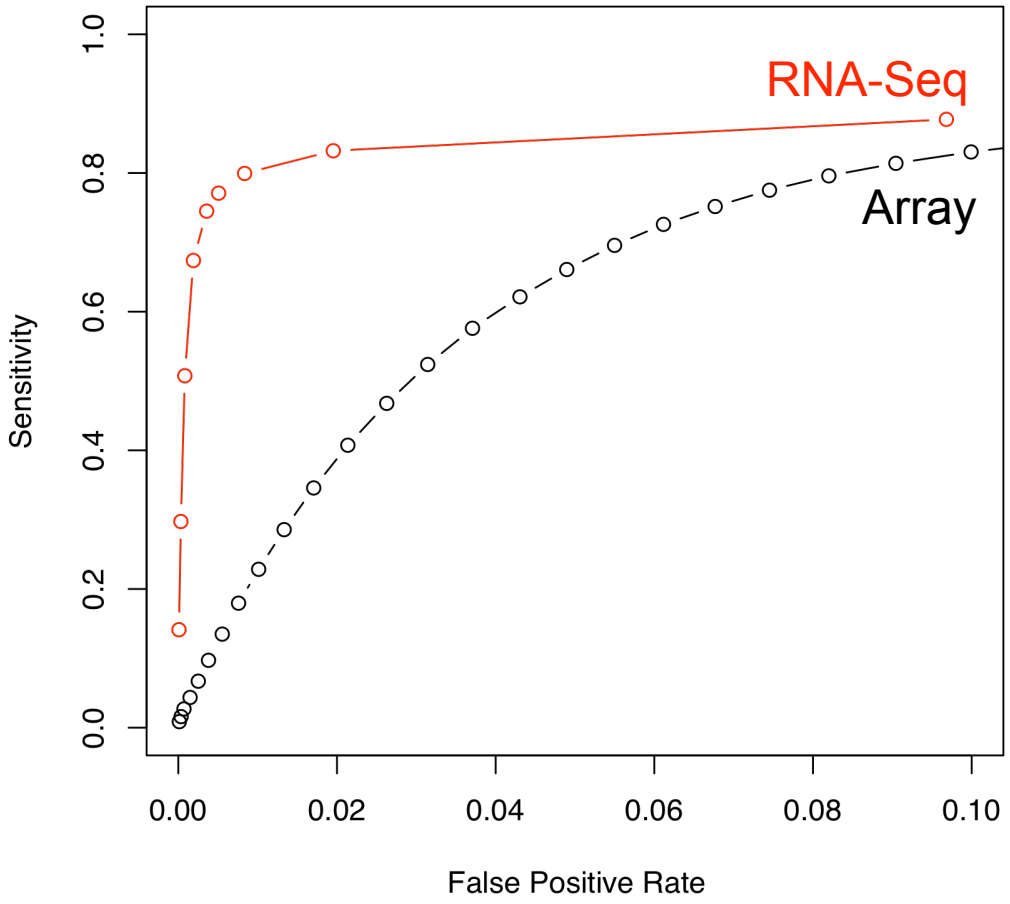
Ashish Agarwal, LaDeana W. Hillier, David Koppstein,
Joel Rozowsky, Andrea Sboner, Lukas Habegger, Jeanyoung Jo,
Michael Snyder, Philip Green, Valerie Reinke,
Robert H. Waterston, Mark Gerstein

Signal Correlation of Transcribed Regions (TARs) on Matched Samples



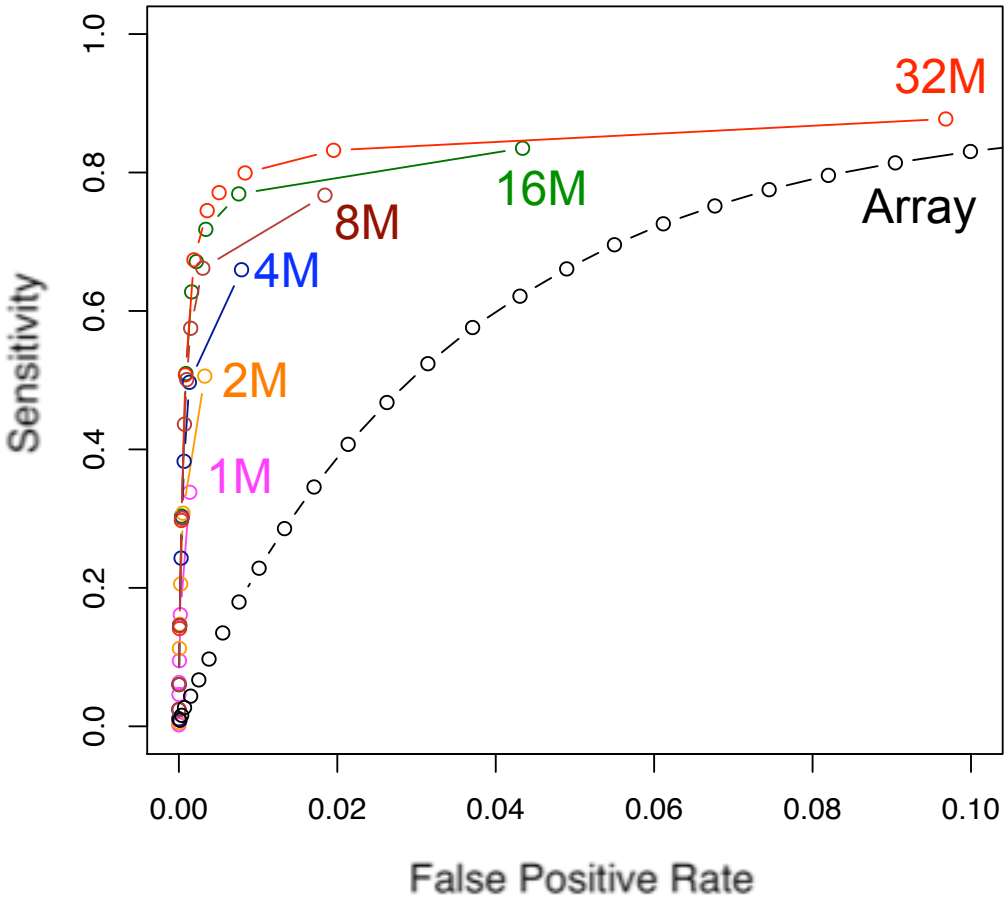
- *C. elegans* L2 larval stage
  - similar results for L3, L4, and young adult
- polyA enriched RNA
- raw signal correlation
  Spearman = 0.90
- Regions potentially affected by cross-hybridization evident

Seq vs Array: Annotation as Gold Standard

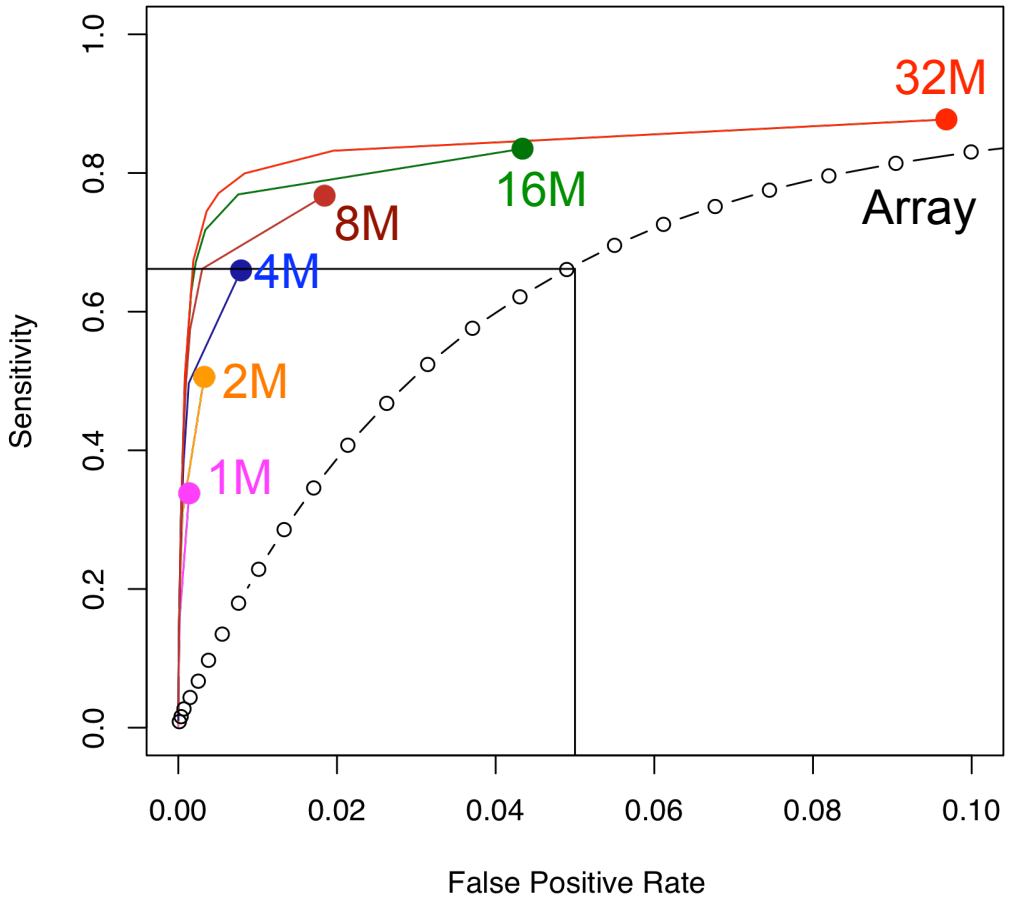


- Sensitivity and FPR substantially improved in RNA-Seq

Seq vs Array: Effect of Sequencing Depth

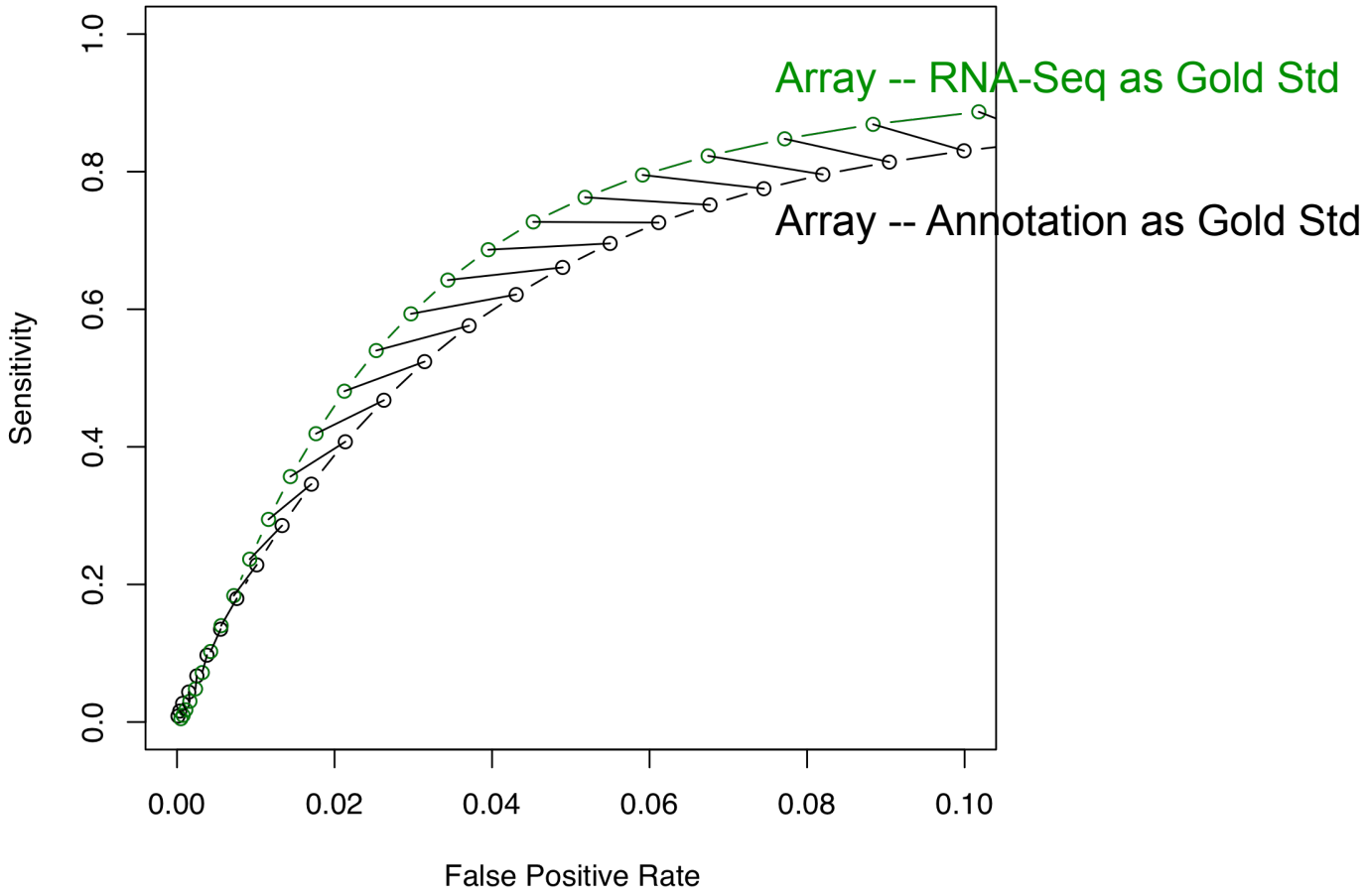


Seq vs Array: Sequencing Depth Required to Match Array



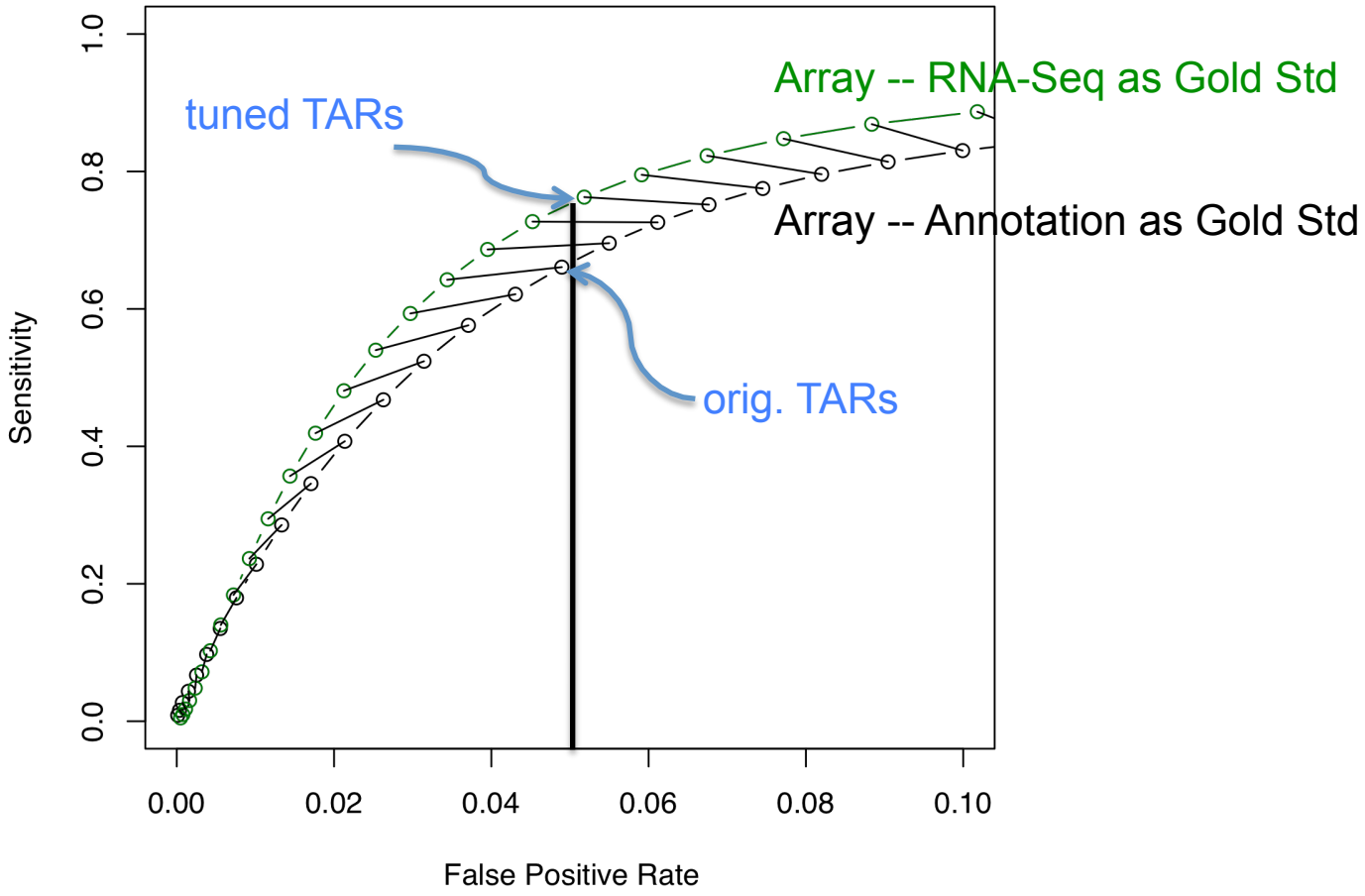
- At desired FPR of 0.05, RNA-Seq matches tiling array with 4M reads, about 1/2 lane

Array Tuning: Using RNA-Seq as Gold Std.



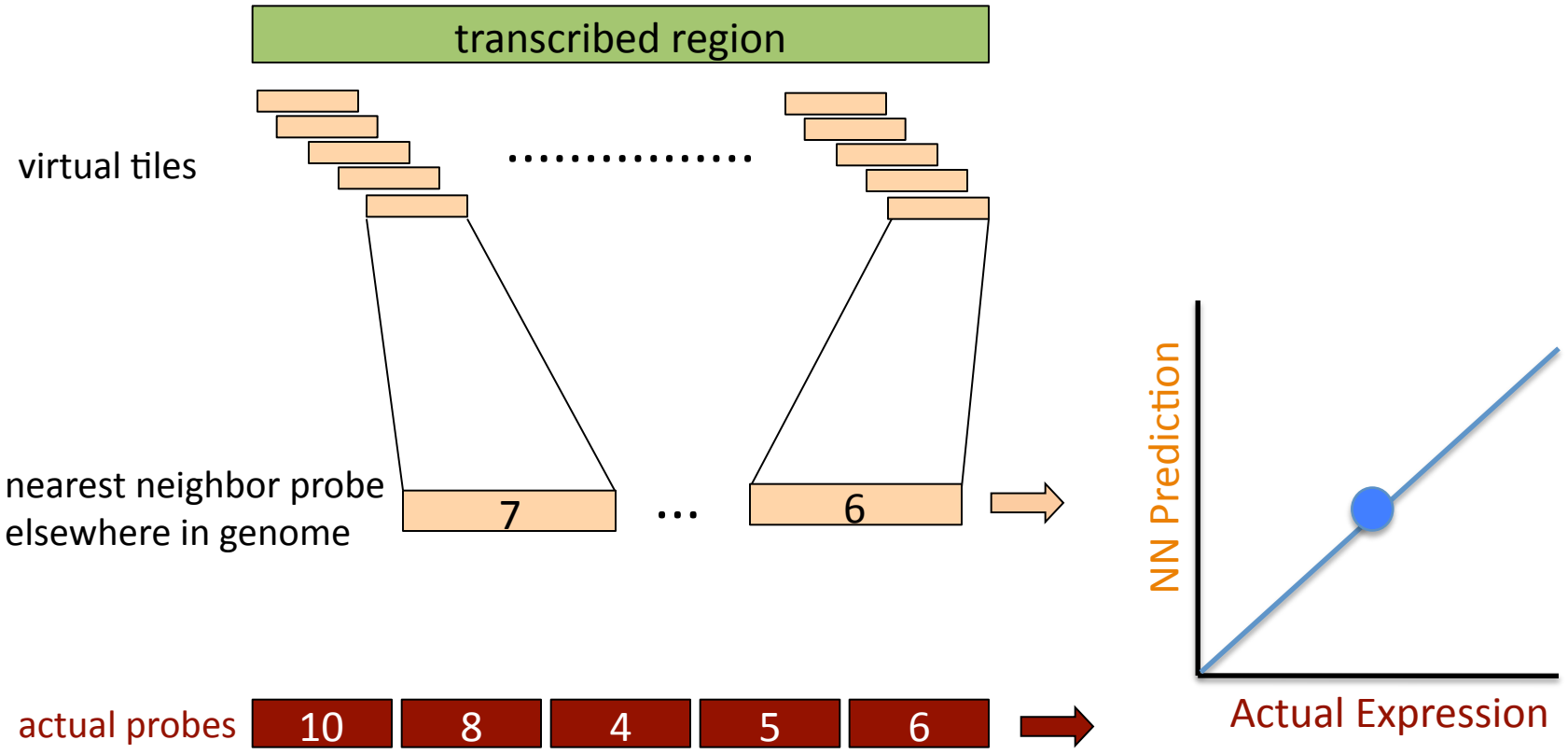
- Matched RNA-Seq data serves as a better gold standard

Array Tuning: Developing an Optimized Array Scoring System

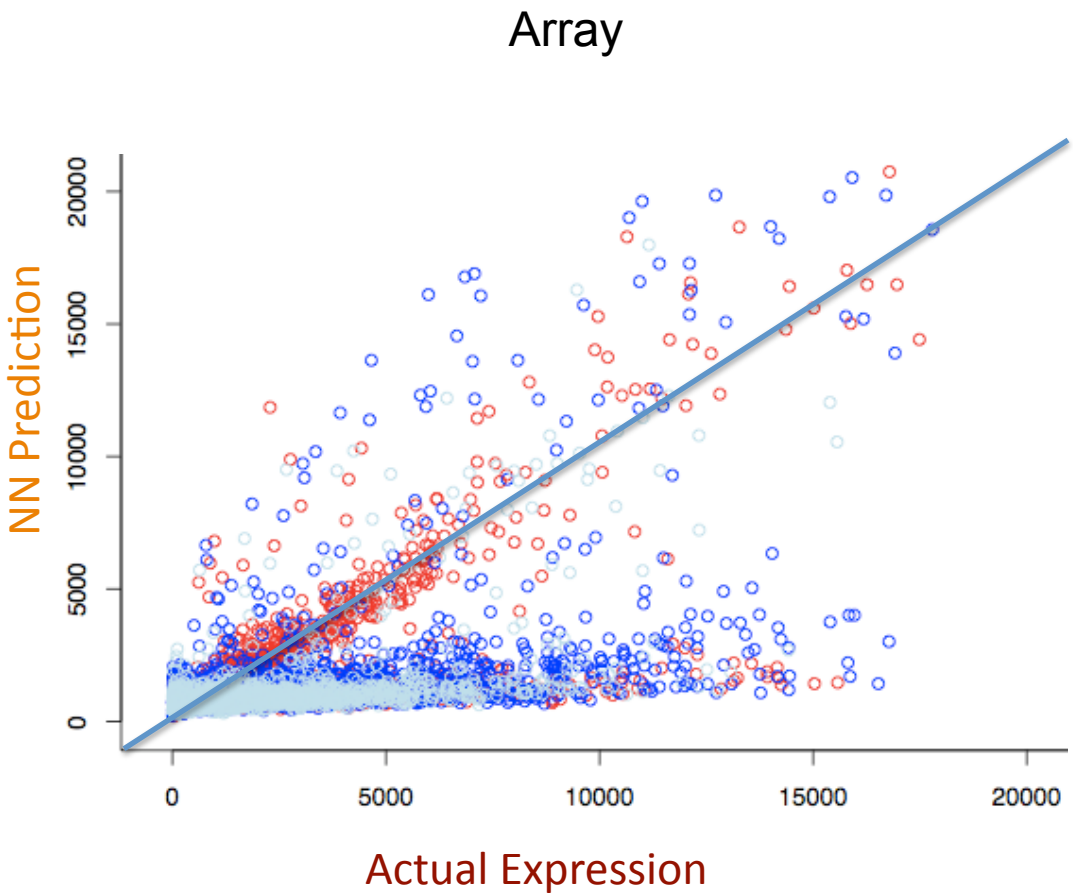


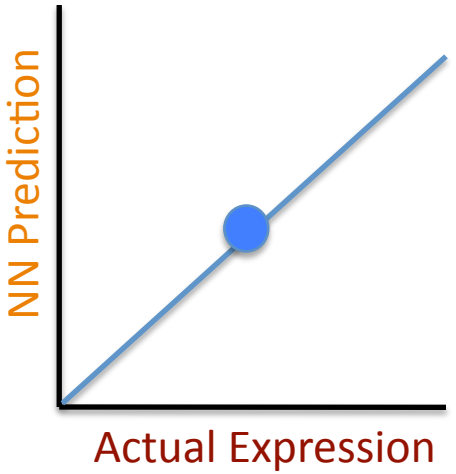
- Parameters can be tuned with knowledge of better gold standard
- This can be used to calibrate an optimal threshold and p-value for array analysis

Cross-Hyb: Developing Nearest-Neighbor Predictor

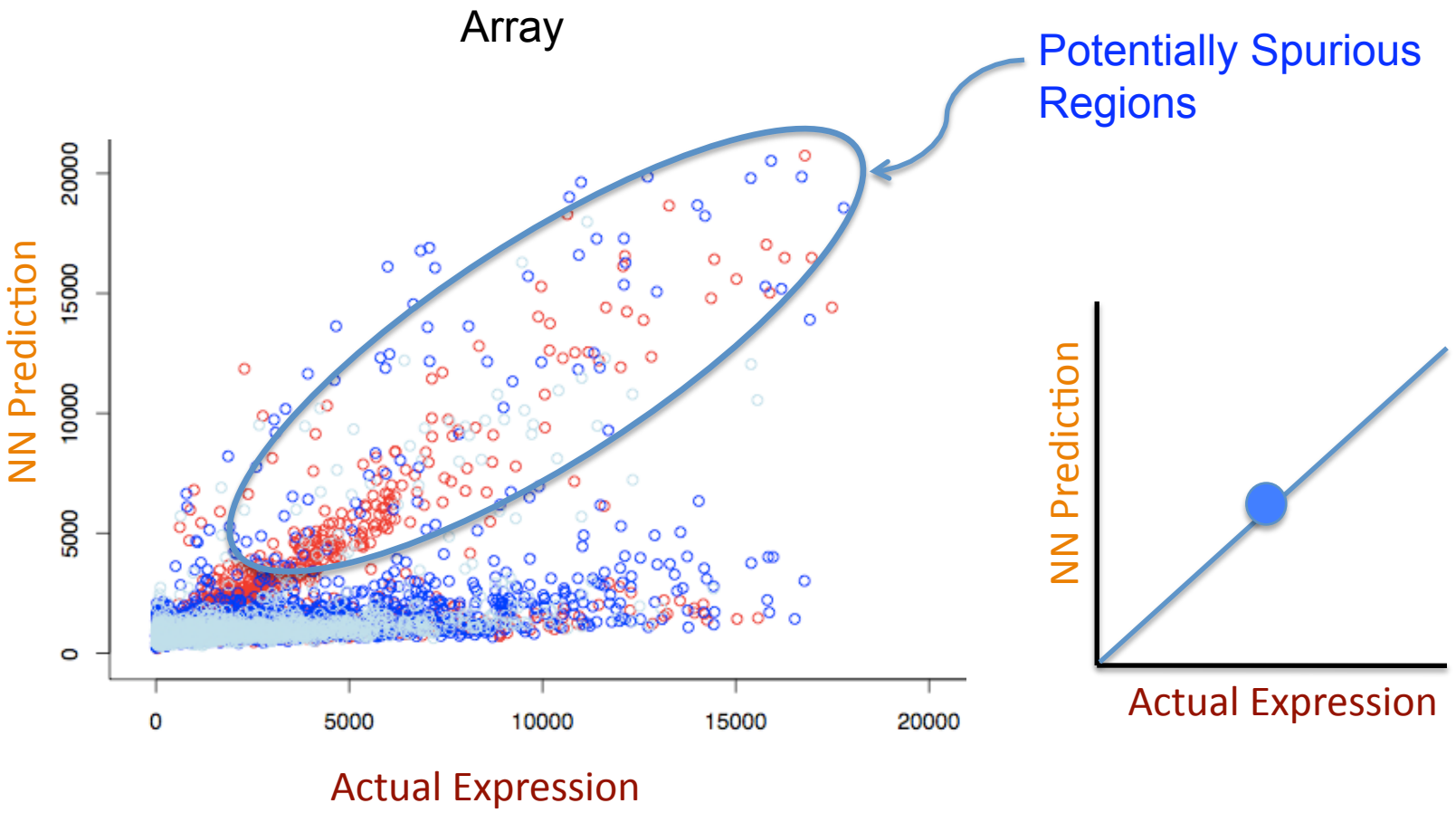


Cross-Hyb: Flagging Regions of Spurious Signal





Cross-Hyb: Flagging Regions of Spurious Signal



Cross-Hyb: No Spurious Signal in RNA-Seq