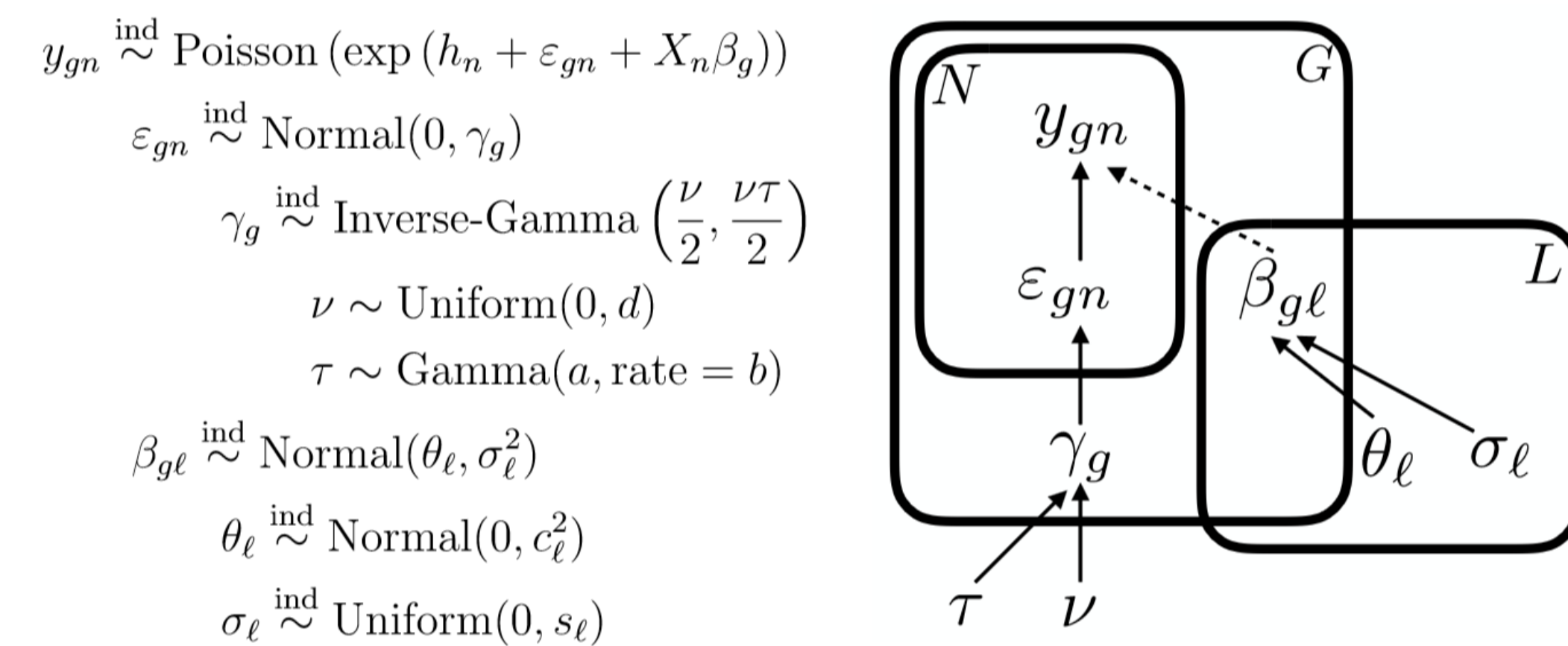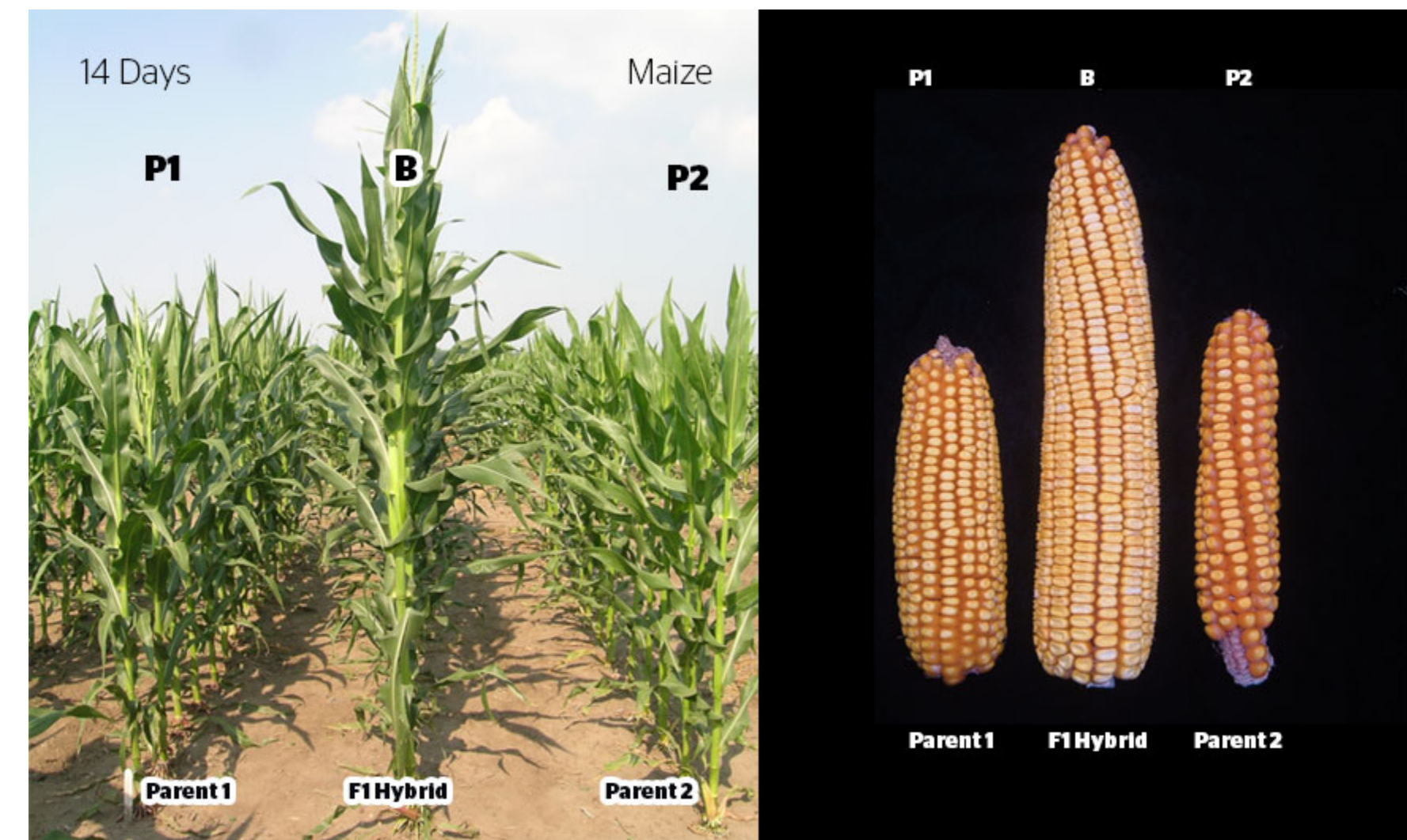Jarad Niemi, Will Landau, and Dan Nettleton

Fully Bayesian analysis of hierarchical count regression models applied to RNA-seq

Hierarchical Count Regression Model

$$y_{gn} \overset{\text{ind}}{\sim} \text{Poisson}\left(\exp\left(h_n + \varepsilon_{gn} + X_n\beta_g\right)\right)$$

$$\varepsilon_{gn} \overset{\text{ind}}{\sim} \text{Normal}(0, \gamma_g)$$

$$\gamma_g \overset{\text{ind}}{\sim} \text{Inverse-Gamma}\left(\frac{\nu}{2}, \frac{\nu\tau}{2}\right)$$

$$\nu \sim \text{Uniform}(0, d)$$

$$\tau \sim \text{Gamma}(a, \text{rate} = b)$$

$$\beta_{g\ell} \overset{\text{ind}}{\sim} \text{Normal}(\theta_\ell, \sigma_\ell^2)$$

$$\theta_\ell \overset{\text{ind}}{\sim} \text{Normal}(0, c_\ell^2)$$

$$\sigma_\ell \overset{\text{ind}}{\sim} \text{Uniform}(0, s_\ell)$$

Heterosis (hybrid vigor)

RNA-seq expression counts

Gene ID	B73				Mo17				B73 x Mo17				Mo17 x B73			
GRMZM2G107839	26	17	32	35	30	32	41	43	63	44	116	101	30	31	69	47
GRMZM5G899787	62	57	38	33	91	78	66	69	58	84	42	43	74	70	53	51
GRMZM5G899800	150	238	12	6	198	392	11	15	187	433	8	10	414	291	11	13
GRMZM2G301485	24	12	29	32	20	14	32	46	5	3	6	6	2	3	3	7
GRMZM5G899836	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...																

GPU-accelerated Markov chain Monte Carlo algorithm

Algorithm 1 MCMC for hierarchical RNA-seq model

1. **In parallel,** sample the ε_{gn}'s.

2. **In parallel,** sample the γ_g's.

3. **Reduction** to calculate $\sum_{g=1}^{G}\left[\log\gamma_g + \frac{\nu}{\gamma_g}\right]$. Then sample ν from its full conditional density, which is proportional to

$$\exp\left(-G\log\Gamma\left(\frac{\nu}{2}\right) + \frac{G\nu}{2}\log\left(\frac{\nu\tau}{2}\right) - \frac{\nu}{2}\sum_{g=1}^{G}\left[\log\gamma_g + \frac{\nu}{\gamma_g}\right]\right).$$

4. **Reduction** to calculate $\sum_{g=1}^{G}\frac{1}{\gamma_g}$. Then sample $\tau \sim \text{Gamma}\left(a + \frac{G\nu}{2}, \text{rate} = b + \frac{\nu}{2}\sum_{g=1}^{G}\frac{1}{\gamma_g}\right)$.

5. For $\ell = 1, \ldots, L$, **in parallel,** sample $\beta_{1\ell}, \ldots, \beta_{G\ell}$.

6. **Reduction** to calculate means and variances of the relevant $\beta_{g\ell}$'s. Then sample $\theta_1, \ldots, \theta_L$.

7. **Reduction** to calculate the shape and scale parameters of the inverse-gamma distributions. Then sample $\sigma_1, \ldots, \sigma_L$.

RNA-seq experiments

Select RNA fraction of interest (poly(A), ribo-minus and others)

Fragment and reverse transcribe

Sequence, map onto genome

Quantitate (relative, absolute, nonmolar and others)

3× 2× 1×
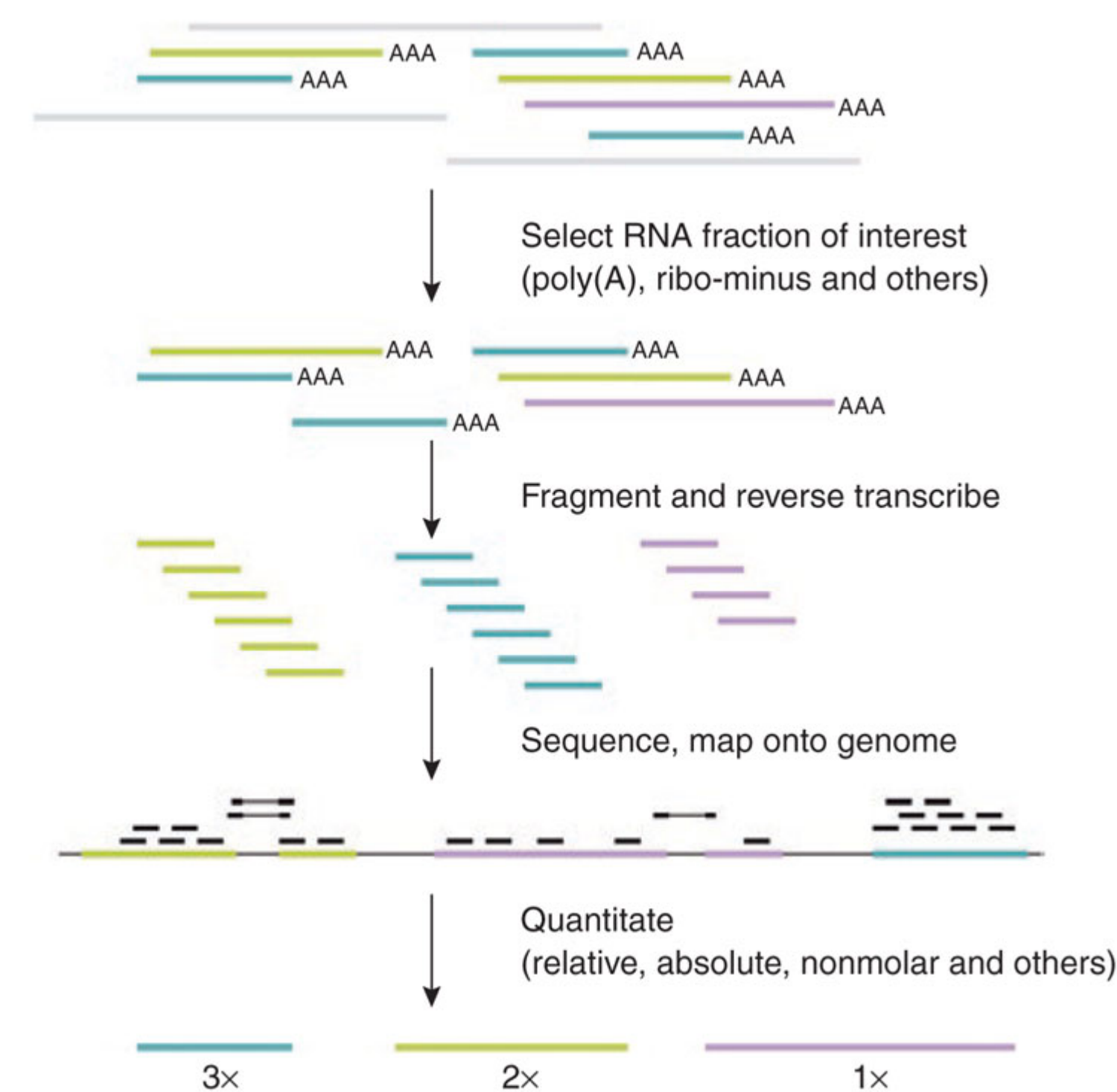
(Pepke, Wold, and Mortazavi (2009)
http://www.nature.com/nmeth/journal/v6/n11s/fig_tab/nmeth.1371_F5.html)

Heterosis detection
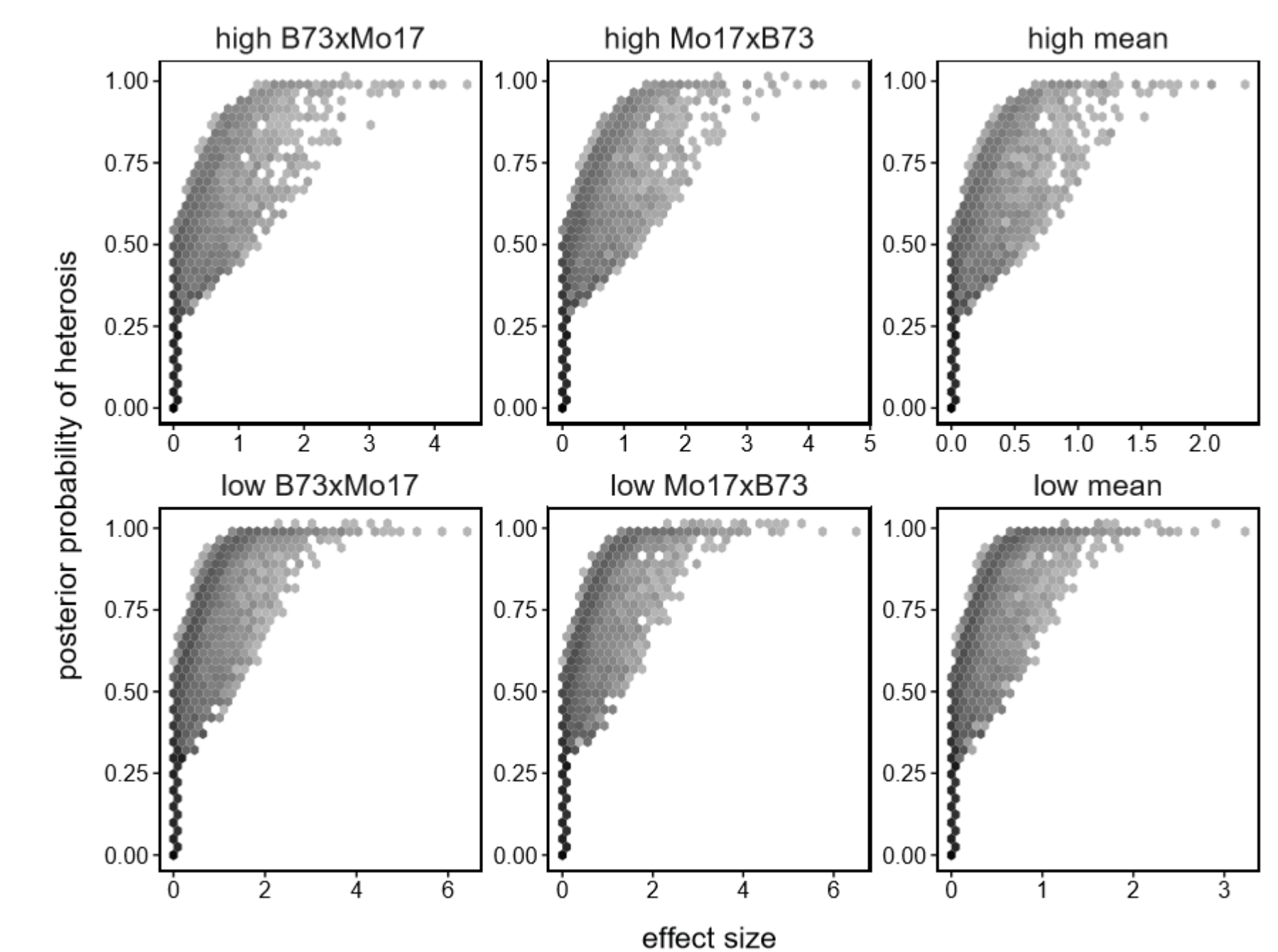
Will Landau, Jarad Niemi, and Dan Nettleton (2018) Fully Bayesian analysis of RNA-seq counts for the detection of gene expression heterosis. to appear *Journal of the American Statistical Association*