Electronic Supplementary Material (ESI) for Molecular BioSystems

Symmetric Epistasis Estimation (SEE) and its application to dissecting interaction map of *Plasmodium falciparum*

Supplementary Online Materials

Yang Huang[1], Geoffrey Siwo[2], Stefan Wuchty[1], Michael T. Ferdig[2], Teresa M. Przytycka[1*]

[1]National Center for Biotechnology Information, NLM, NIH, 8600 Rockville Pike, Building 38A, Bethesda, MD 20894.

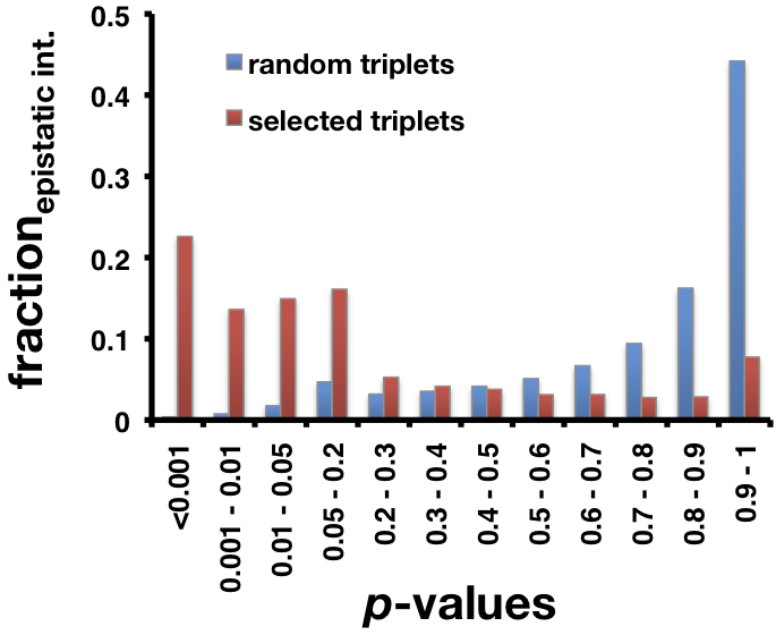
[2] Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Sciences, Notre Dame, IN 46556.

* corresponding author

Supplementary Results

Comparison with a naïve method

Applying a naïve method we would have to test all possible triplets for epistatic interactions, demanding huge computational resources and causing a serious multiple-testing issue. To verify that our selected candidate triplets were indeed enriched with epistatic interactions, we therefore randomly constructed 100,000 triplets from all possible triplets in Plasmodium dataset, demanding that loci were on different chromosomes. Fitting our models, we calculated and corrected *p*-values as described in the Methods. Repeating the process 10 times, we indeed observed higher fraction of significant, corrected *p*-values in the set of candidate triplets compared to random triplets, indicating enrichment of epistatic interactions in candidate triplets (Supplementary Fig. S1).



Supplementary Figure S1 Determining the enrichment of candidate triplets that showed epistatic interactions, we observed a decreasing trend of the fraction of triplets that had epistatic interactions with increasing corrected *p*-value. In contrast, we randomized triplets and found an inverse trend, confirming that candidate triplets predominately carried epistatic interactions.

Comparison with step-wise method

In order to compare the performance of our approach we also implemented a variation of Storey *et al.*'s step-wise search method [1]. For each gene expression trait, we selected a primary locus that provided the best association *p*-value. For each pair of an expression trait and its primary locus, we selected a secondary locus among all other loci, which allowed the best regression improvement using the epistatic interaction model (synergistic model) on the primary and secondary locus compared to the additive model. This way, we computed a secondary locus and a nominal *p*-value for each pair. We then generated a null distribution for the secondary locus as outlined in [1]. Finally, we compared their nominal *p*-values against the null distribution to estimate corrected *p*-values. In order to obtain the number of false positives, we calculated corresponding FDR [2]. By design, our method SEE is more likely to uncover an interaction for which both loci

contributed equally to the underlying synergistic effect. Since the step-wise methods detected eQTLs where one of the underlying two loci had a significant effect, we expected that results of both methods might overlap only to a small extent. Applying the step-wise method to our data, we obtained 7,665 candidate one-locus eQTLs and identified 74 locus pairs that had an interaction effect on a gene with $p \leq 0.01$ and FDR $\leq$ 0.24. In comparison to our set of 1,713 two-locus eQTLs, we indeed found a small overlap of four two-locus eQTLs. This relatively low overlap is hardly surprising since the step-wise approach calls for a primary locus with a maximal marginal effect among all possible loci, while our method looks for strong contributions from both loci — goals that are hard to achieve simultaneously given a relatively small progeny number.

Epistatic interaction in yeast defined by different statistical models

We investigated the impact of the choice of different statistical models on the number of detected epistatic interactions in yeast. In particular, we implemented two variations of the step-wise method: (i) In analogy to the original step-wise approach, we first selected a primary locus that provided the best association *p*-value for each gene expression trait. For each pair of an expression trait and its primary locus, we selected a secondary locus, which allowed the best regression improvement using the synergistic model (eq. 1) compared to the one-locus model on the primary locus. We obtained 617 interactions in yeast when requiring corr. P $\leq$ 0.05. (ii) In a second approach we started with the same initial selection of a primary locus. Then, for each pair of an expression trait and its primary locus, we selected a secondary locus, using a more stringent statistical model that compares the synergistic (eq. 1) to the additive model (eq. 2), only detecting 277 interactions with P $\leq$ 0.05 (FDR < 0.9). Although both approaches utilized a step-wise filtering step, our statistical method returned fewer epistatic interactions, suggesting that the majority of epistatic interactions in yeast may be dominated by a significant one-locus main effect.

Comparing epistatic interactions with one-locus eQTLs

Considering each epistatic interaction *l'*, *l''*, *g* as two one-locus eQTLs, *l'*, *g* and *l''*, *g*, we compared our set of 1,713 epistatic interactions to previously detected one-locus eQTLs.

Using one-locus eQTLs from GeD [3], we found 110 one-locus eQTLs that appeared in epistatic interations. Compared to the smaller set of one-locus eQTLs detected by Gonzales *et al.* [4] we found 17 epistatic interactions where the corresponding loci showed significant main effects on the gene expression trait. By definition, single loci in epistatic interaction locus pairs detected by us were not required to exert a main effect on the expression of the underlying gene, a condition that is more consistent with the ways actual molecular interactions can influence traits non-additively.

Tuning SEE parameters to detect more epistatic interaction

Lowering the threshold of the minimal number of strains in an association clique, allows us to increase the number of nodes in the eQTL association graph as well as candidate triplets. As a consequence, however, the percentage of significant interactions over all candidate triplets decreases. In yeast, such a percentage dropped from 7% to 6% when we demanded 12 instead of 13 strains in each association clique while $sp_{D}^{'''} \geq 2$ and $sp_{IC}^{'''}$ ³ 35. Utilizing all association cliques with at least 6 strains we indeed found 8 epistatic interactions that were detected by Zhang *et al.*[5]. As a consequence, the percentage of significant interactions drastically drops while an enormous eQTL association graph required massive computational resources.

Supplementary Tables

See the Excel file for Supplementary Table S1, which contains 1,713 epistatic interactions in *P. falciparum* detected by SEE

Supplementary Table S2 14 epistatic interaction hotspots that were clustered into four hotspot regions based on their genomic proximity.

hotspot region	locus1	locus2	target genes
1	3_0	7_5.8	MAL13P1.189, MAL13P1.247, MAL6P1.126, MAL7P1.151, PF07_0038, PF08_0066, PFA0200w, PFD0340c, PFD0415c, PFI0260c, PFI1415w, PFL1675c, PFL1765c
1	3_2.9	7_5.8	MAL13P1.189, MAL13P1.247, MAL6P1.126, MAL7P1.151, PF07_0038, PF08_0066, PFA0200w, PFC0980c, PFD0415c, PFI0260c, PFI1415w, PFL1675c, PFL1765c
1	3_2.9	7_2.9	MAL13P1.88, MAL7P1.147, MAL7P1.151, PF07_0038, PF14_0616, PFB0290c, PFD0415c, PFE0410w, PFF1325c, PFI0260c, PFI1415w
1	3_8.6	7_2.9	MAL13P1.247, MAL13P1.88, MAL6P1.126, MAL7P1.147, MAL7P1.151, PF07_0038, PF11_0195, PF13_0282, PF14_0156, PF14_0258, PFB0290c, PFD0415c, PFE0410w, PFI0370c
1	3_8.6	7_5.8	MAL13P1.189, MAL13P1.247, MAL6P1.126, MAL7P1.151, PF07_0038, PF08_0066, PF14_0258, PF14_0710, PFB0290c, PFD0415c, PFI0260c, PFL1675c, PFL1765c
2	5_20	9_8.7	MAL13P1.142, PF08_0005, PF10_0135, PF10_0316, PF11_0048, PF14_0233, PF14_0526, PF14_0664, PFC0805w, PFE0415w, PFL0510c
2	5_20	9_11.6	MAL13P1.142, PF08_0005, PF10_0135, PF10_0316 PF11_0256, PF14_0233, PF14_0526, PF14_0664, PFE0415w, PFL0510c
2	5_25.8	9_8.7	MAL13P1.142, MAL13P1.46, PF08_0121, PF11_0438, PF13_0117, PF13_0257, PF14_0233, PF14_0526, PFC0805w, PFE0415w, PFI0980w, PFL0510c
2	5_25.8	9_11.6	MAL13P1.142, MAL13P1.46, PF08_0121, PF11_0438 PF13_0117, PF13_0257, PF14_0233, PF14_0526, PFE0415w, PFL0510c

2	5_31.5	9_11.6	MAL13P1.46, PF08_0121, PF10_0135, PF11_0124, PF11_0438, PF13_0117, PF13_0257, PF14_0233, PF14_0526, PFE0415w, PFL0510c
3	7_14.4	8_48.9	PF10_0208, PF11_0002, PF11_0256, PF11_0486, PF14_0648, PFE1625c, PFL0305c, PFL1545c, PFL1740w, PFL2325c,
3	7_20.2	8_48.9	PF10_0208, PF11_0124, PF11_0256, PF11_0360, PF11_0486, PF14_0648, PFE1625c, PFL0305c, PFL1740w, PFL2575c
3	7_23.1	8_48.9	MAL13P1.273, PF10_0208, PF11_0124, PF11_0256, PF11_0486, PFB0820c, PFE1625c, PFL0305c, PFL1740w, PFL2575c
4	12_28.6	14_0	MAL13P1.142, MAL13P1.45, MAL6P1.52, PF10_0407, PF11_0048, PF11_0329, PF11_0331, PF13_0254, PFB0290c, PFD1070w, PFE0410w, PFF0940c, PFI1415w

Supplementary Table S3 110 epistatic interactions in yeast detected by SEE (corr. P ≤ 0.05).

locus1	locus2	target gene	locus1	locus2	target gene
7722_at_x06	9908_at_x05	YPR184W	10539_at_x08	7351_at_x00	YJL078C
4294_at_x14	4708_at_x08	YPR184W	5626_at_x00	7006_at_x07	YJL078C
6829_at_x01	7207_at_x04	YGR041W	10403_at_x15	7313_at_x09	YJL078C
6829_at_x01	7176_at_x09	YGR041W	10396_at_x15	7313_at_x09	YJL078C
2821_at_x00	9924_at_x09	YPR184W	10539_at_x08	7313_at_x09	YJL078C
8125_at_x07	9924_at_x09	YPR184W	10396_at_x15	6979_at_x06	YJL078C
6121_at_x01	8666_at_x04	YOL130W	10539_at_x08	6979_at_x06	YJL078C
6973_at_x00	8666_at_x04	YOL130W	10396_at_x15	4990_at_x00	YJL078C
7135_at_x15	8666_at_x04	YOL130W	4023_s_at_x01	5241_at_x12	YBR132C
7135_at_x15	8675_at_x06	YOL130W	4023_s_at_x01	7132_at_x02	YBR132C
7135_at_x15	8650_at_x11	YOL130W	4023_s_at_x01	6969_at_x06	YBR132C
4530_at_x00	9055_at_x14	YFL026W	6900_at_x13	7222_at_x09	YKL178C
10424_at_x12	4272_at_x07	YKL178C	7722_at_x06	8499_at_x05	YKL178C
4556_at_x05	4708_at_x08	YKL178C	4472_s_at_x03	9567_at_x12	YKL178C
4272_at_x07	4680_at_x10	YKL178C	4472_s_at_x03	9592_at_x03	YKL178C
4272_at_x07	5259_at_x15	YKL178C	10238_at_x12	7346_at_x10	YKL178C
10298_at_x05	4294_at_x14	YKL178C	4431_at_x09	9962_at_x02	YKL178C
10340_at_x07	4949_at_x12	YKL178C	2128_s_at_x04	9962_at_x02	YKL178C
10424_at_x12	4556_at_x05	YKL178C	4651_at_x15	9779_at_x13	YOL126C
10668_at_x04	4556_at_x05	YKL178C	2826_at_x02	7524_at_x01	YEL021W
10439_at_x07	4556_at_x05	YKL178C	4550_at_x02	8004_at_x01	YFL026W
7078_at_x01	8958_at_x14	YKL178C	4545_at_x13	6969_at_x06	YKL178C
10322_at_x15	10831_s_at_x15	YKL178C	4556_at_x05	7114_at_x13	YKL178C
10831_s_at_x15	9974_at_x15	YKL178C	11254_at_x00	5282_at_x01	YKL178C
3326_at_x07	9893_at_x10	YGL009C	11215_at_x14	5282_at_x01	YKL178C
3326_at_x07	9882_at_x08	YGL009C	5241_at_x12	7794_at_x09	YCL018W
2821_at_x00	9962_at_x02	YPR184W	5241_at_x12	7808_at_x12	YCL018W
3129_at_x00	7777_at_x07	YNL327W	5241_at_x12	7808_at_x02	YCL018W
4255_at_x03	6973_at_x00	YOR390W	6929_at_x07	9055_at_x14	YCL018W
10726_at_x09	6048_at_x15	YER150W	4708_at_x08	9192_at_x06	YDR033W
11289_at_x04	5384_at_x08	YPL029W	4020_at_x12	8125_at_x07	YEL021W
10746_at_x11	2334_s_at_x03	YER150W	4949_at_x12	5984_at_x08	YKL178C
5626_at_x00	9240_at_x07	YNL327W	11254_at_x00	5435_at_x15	YKL178C
5626_at_x00	9518_at_x11	YNL327W	11215_at_x14	5435_at_x15	YKL178C
7078_at_x01	7808_at_x12	YNL327W	2186_at_x01	6973_at_x00	YKL178C
7078_at_x01	7777_at_x07	YNL327W	4550_at_x02	5054_at_x01	YKL216W
10769_at_x04	7722_at_x06	YPL029W	10331_at_x04	6310_at_x14	YKL216W
4294_at_x14	5094_at_x08	YKL178C	5384_at_x08	5569_at_x12	YKL216W
6893_at_x00	7006_at_x07	YKL178C	5411_at_x14	5569_at_x12	YKL216W
7006_at_x07	9842_at_x07	YKL178C	4495_at_x03	5544_at_x08	YKL216W

4272_at_x07	5297_at_x09	YKL178C	4508_at_x14	5544_at_x08	YKL216W
4556_at_x05	5297_at_x09	YKL178C	10331_at_x04	2315_at_x06	YKL216W
10322_at_x15	2821_at_x00	YKL178C	5584_at_x06	9354_at_x14	YKL216W
4651_at_x15	6900_at_x13	YKL209C	4294_at_x14	5020_at_x08	YKL216W
4653_at_x00	6900_at_x13	YKL209C	4294_at_x14	4985_at_x15	YKL216W
4395_at_x11	6632_at_x05	YKL209C	4294_at_x14	4990_at_x00	YKL216W
10298_at_x05	6632_at_x05	YKL209C	10539_at_x08	4990_at_x00	YKL216W
4395_at_x11	6623_at_x08	YKL209C	10059_at_x00	10397_at_x00	YKL216W
4395_at_x11	6626_at_x11	YKL209C	4251_at_x08	9868_at_x08	YKL216W
5808_i_at_x00	9709_at_x03	YFL026W	10340_at_x07	4251_at_x08	YKL216W
10539_at_x08	7290_at_x01	YJL078C	10340_at_x07	4253_i_at_x00	YKL216W
10955_at_x05	7290_at_x01	YJL078C	10340_at_x07	4255_at_x03	YKL216W
10539_at_x08	7292_at_x05	YJL078C	4550_at_x02	5027_at_x06	YKL216W
10403_at_x15	7351_at_x00	YJL078C	4272_at_x07	5027_at_x06	YKL216W
10396_at_x15	7351_at_x00	YJL078C	6242_at_x13	7794_at_x09	YLL013C

References

1 Storey, J. D., Akey, J. M. & Kruglyak, L. Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol* **3**, e267 (2005).

2 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-9445 (2003).

3 Huang, Y., Wuchty, S., Ferdig, M. T. & Przytycka, T. M. Graph theoretical approach to study eQTL: a case study of Plasmodium falciparum. *Bioinformatics* **25**, i15-20 (2009).

4 Gonzales, J. M. *et al.* Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Biol* **6**, e238 (2008).

5 Zhang W, Zhu J, Schadt EE, Liu JS. A Bayesian partition method for detecting pleiotropic and epistatic eQTL modules. PLoS Comput Biol 6: e1000642 (2010).