nature genetics

Epistasis analysis with global transcriptional phenotypes

Nancy Van Driessche[1,2], Janez Demsar[3], Ezgi O Booth[1,4], Paul Hill[1], Peter Juvan[3], Blaz Zupan[3], Adam Kuspa[1,2,5] & Gad Shaulsky[1,2,4]

Classical epistasis analysis can determine the order of function of genes in pathways using morphological, biochemical and other phenotypes. It requires knowledge of the pathway's phenotypic output and a variety of experimental expertise and so is unsuitable for genome-scale analysis. Here we used microarray profiles of mutants as phenotypes for epistasis analysis. Considering genes that regulate activity of protein kinase A in *Dictyostelium*, we identified known and unknown epistatic relationships and reconstructed a genetic network with microarray phenotypes alone. This work shows that microarray data can provide a uniform, quantitative tool for large-scale genetic network analysis.

In classical epistasis analysis, two genes are mutated in the same strain and the phenotype of the double mutant is compared with those of the corresponding single mutants. The prevailing phenotype defines the epistatic mutation[1,2]. Genetic networks, outlining the details of a biological process, are then constructed by integrating the epistatic relationships between several pairs of genes. This task depends on the measured phenotype, and so the analysis of different systems requires a variety of experimental expertise. In addition, the rules of epistasis cannot be applied consistently if the experimental procedures are not identical for all pairs of genes in a certain pathway[2].

The genomic era introduced another challenge in applying epistasis analysis to network construction. Although methods exist for identifying and mutating genes as individuals or in pairs on a genomic scale[3–5], some cellular functions do not provide easily scored phenotypes, such as unusual appearance or sensitivity to certain culture conditions. For many other genes, we do not know what phenotype to measure. Therefore, we need systematic approaches to identifying the functions of new genes and their interactions to ensure progress from genome sequence to directed experimentation.

We tested whether whole-genome expression profiles could be used to determine relationships between genes in genetic networks, with the idea that transcriptional profiles of mutant strains could replace less objective phenotypes, such as morphology, in determining epistatic interactions between genes. Transcription profiles have already proven useful as phenotypes in the analysis of single-gene mutations[6–12]. The advantage of inferring epistatic relationships from the transcriptional phenotypes of mutants is that specific knowledge of the relationship between the gene function and the phenotype is not essential.

In our studies, we used the haploid soil amoeba *Dictyostelium discoideum*, which is amenable to microarray studies[13]. Upon removal of nutrients, *D. discoideum* executes a developmental program in which single cells aggregate and form multicellular organisms. Aggregation depends on chemotaxis towards extracellular cAMP, and the cells aggregate into centers ~8–10 h after starvation. Two cell types then differentiate and give rise to a fruiting body consisting of a spore mass on top of a stalk tube[14]. We focused on the interactions between genes in the protein kinase A (PKA) pathway, which is essential for development[15]. Mutations in this pathway have marked effects, including developmental arrest or attenuation, precocious development and aberrant sporulation and germination. The genes are dispensable for growth, and so null mutations can be studied. The PKA regulatory pathway has been elucidated by biochemical, cellular and genetic methods[2], but some relationships in the pathway have not been tested genetically.

To test the applicability of microarray profiling to epistasis analysis, we made ten combinations of single or double mutations in six genes (**Table 1**). We developed mutant cells and collected and analyzed RNA samples throughout the course of development. We then calculated the similarity between the expression profiles of the mutants and used the values to determine epistatic relationships between the mutated genes. We found that the epistatic relationships inferred from transcriptional profiles were in complete agreement with the relationships observed in classical studies. We also used the microarray profiles to infer two previously undetermined relationships between genes in the pathway. We propose that microarray profiles can be used to study systematically the relationships between genes in genetic pathways without existing knowledge of the pathway's phenotypic output.

RESULTS

Reconstruction of a known epistatic relationship

Classical genetic analysis showed that *pufA* is epistatic to *yakA*[16]. We tested whether epistasis analysis with expression profile phenotypes

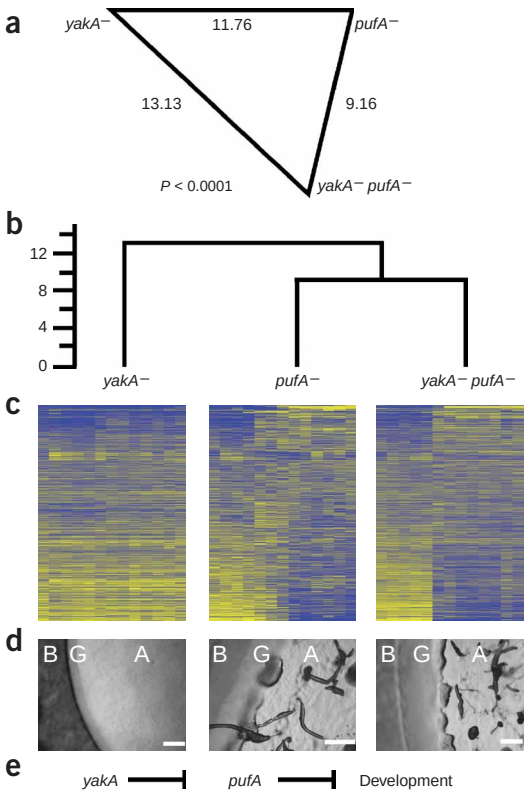
[1]Department of Molecular and Human Genetics and [2]Graduate Program in Developmental Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [3]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia. [4]Graduate Program in Structural Computational Biology and Molecular Biophysics and [5]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. Correspondence should be addressed to G.S. (gadi@bcm.tmc.edu).

Published online 10 April 2005; doi:10.1038/ng1545

Table 1 Genes and mutations

Gene	Protein	Function	Knockout	Overexpression	References
yakA	YakA	Minibrain-like protein kinase	Aggregationless	Growth defect, precocious development	35
pufA	PufA	Pumilio-like RNA binding protein, inhibits pkaC mRNA translation	Suppressor of *yakA* knockout		16
acaA	AcaA	Aggregation-stage adenylyl cyclase	Aggregationless		23
regA	RegA	Response-regulator controlled cAMP-specific phosphodiesterase	Precocious sporulation		20,45
pkaR	PkaR	cAMP-binding regulatory subunit of PKA	Precocious sporulation	Aggregationless	33,46
pkaC	PkaC	Protein kinase catalytic subunit of PKA	Aggregationless	Precocious sporulation	21,47

could also identify that relationship. We developed *yakA*$^-$ cells, *pufA*$^-$ cells and double mutant *yakA*$^-$ *pufA*$^-$ cells for 24 h, collected RNA samples at 2-h intervals and interrogated the samples with a microarray representing 5,624 genes. We identified similarities between the transcriptional profiles of the three strains by computing the Euclidean distances from 5,624 array targets over 13 time-point measurements. The distance between the double mutant *yakA*$^-$ *pufA*$^-$ and the single mutant *pufA*$^-$ was smaller than that between *yakA*$^-$ *pufA*$^-$ and *yakA*$^-$ or that between the two single mutants (**Fig. 1a**). We also used hierarchical clustering to evaluate the data and represent the results with a dendrogram (**Fig. 1b**). In this representation, two strains are more similar to each other than either of them is to a third strain if the two are directly connected on the tree structure. The distance between *pufA*$^-$ and *yakA*$^-$ *pufA*$^-$ was the smallest, and so they clustered together leaving *yakA*$^-$ as the out-group (**Fig. 1b**). The statistical significance of this clustering by bootstrapping[17,18] analysis was >0.99. Assuming a linear signaling pathway, if separate mutations in two genes give different phenotypes and the phenotype of the double mutant strain is similar to that of one of the single mutants, then that single mutation is epistatic[1,2]. Therefore, this analysis supports the conclusion that *pufA* is epistatic to *yakA*. This conclusion is identical to that reached by classical genetic analysis[16] and provides a proof-of-principle for the use of microarray phenotypes in epistasis analysis.



We carried out a detailed comparison of the profiles of the *pufA*$^-$ mutant, the *yakA*$^-$ mutant and the double mutant (**Fig. 1c**). We first clustered the 5,624 gene trajectories from wild-type *D. discoideum* development[19] using the k-means algorithm and then ordered the clusters to emphasize developmental progression (data not shown). We imposed that order on the expression trajectories of the three mutant strains (**Fig. 1c**). In the wild type, ~40% of the genes are developmentally regulated, and so, when presented in a heat map (**Fig. 1c**), their trajectories showed a shift in color at ~6–12 h of development[19]. Such shifts were evident at the top 10% and the bottom 20% of the *yakA*$^-$ *pufA*$^-$ and the *pufA*$^-$ charts. On the other hand, *yakA*$^-$ cells, which cannot aggregate, showed little or no change in expression throughout the 24-h course of development (**Fig. 1c**). To illustrate the morphological defects, we cultured the three strains on bacterial lawns. As the cells grew and depleted the bacteria, they starved and developed to their capacity. *yakA*$^-$ cells did not develop and remained unaggregated after depletion of the food source (**Fig. 1d**). Inactivation of *pufA* in wild-type or in *yakA*$^-$ cells resulted in aggregation and development, albeit with attenuation at the finger stage (**Fig. 1d**). Therefore, *pufA*$^-$ is a suppressor of *yakA*$^-$ and *pufA* is epistatic to *yakA*[16] (**Fig. 1e**).

Derivation of unknown epistatic relationships

Next, we tested whether microarray data could detect unknown genetic relationships. Because biochemical analysis showed that PufA binds the 3′ untranslated region (UTR) of the *pkaC* mRNA and inhibits its translation[16], we expected *pkaC* to be epistatic to *pufA*.

Figure 1 Transcriptional profiling to test whether *pufA* is epistatic to *yakA*. RNA samples from developing *yakA*$^-$, *pufA*$^-$ and *yakA*$^-$ *pufA*$^-$ cells were analyzed. (**a**) Euclidean distances between the three mutants are presented as edges in a triangle (numbers indicate length). The null hypothesis of uniform distribution was tested by the χ^2 test and rejected (*P* value indicated). (**b**) The dendrogram represents the distances between the different mutants (arbitrary units on scale bar). Vertical distances represent the dissimilarity between leaves and joints. Confidence levels (from bootstrap analysis) were >0.99. (**c**) Heat maps representing expression changes between mutants. Genes are ordered according to their expression in the wild type[19] (not shown). The order from top to bottom is identical in all mutants. Columns represent time points (0–24 h in 2-h intervals); rows represent genes. Blue indicates lower-than-average levels of expression; yellow, higher-than-average. (**d**) Morphological phenotypes. Single cells deposited on bacterial lawns propagated and formed plaques. Cells in the center starved and developed but cells in the periphery continued to grow. The opaque bacterial lawn (B) is on the left; in the growing zone (G), amoebae are consuming bacteria; starving amoebae (A) are on the right. The *yakA*$^-$ cells did not aggregate. The *pufA*$^-$ and *yakA*$^-$ *pufA*$^-$ cells developed to the finger stage. Scale bar, 1 mm. (**e**) The genetic interaction between *yakA* and *pufA*.

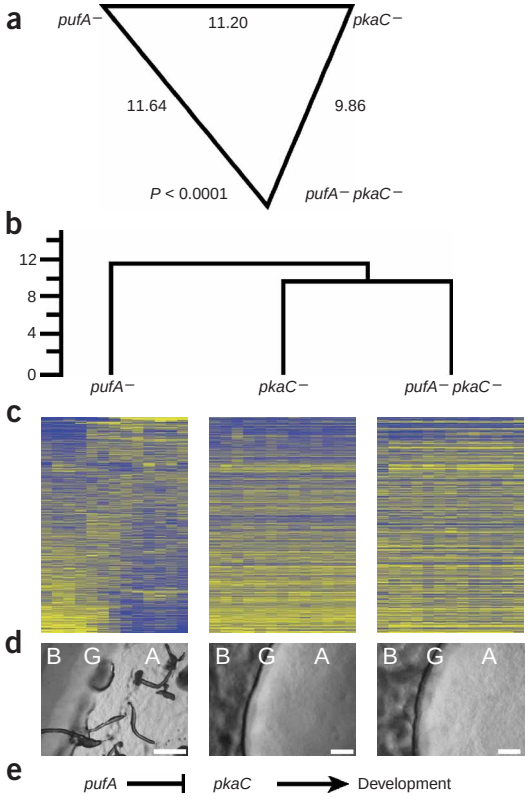


Figure 2 Transcriptional profiling to test whether *pkaC* is epistatic to *pufA*. RNA samples from developing *pkaC*$^-$, *pufA*$^-$ and *pufA*$^-$ *pkaC*$^-$ cells were analyzed. Data for *pufA*$^-$ cells are the same as in **Figure 1**. (**a**–**c**) Euclidean distances (**a**), dendrogram (**b**) and heat maps (**c**) as in **Figure 1 a**–**c**. (**d**) Morphological phenotypes as in **Figure 1d** for *pufA*$^-$, *pkaC*$^-$ and *pufA*$^-$ *pkaC*$^-$ cells. *pufA*$^-$ cells developed to the finger stage. *pkaC*$^-$ and *pufA*$^-$ *pkaC*$^-$ cells failed to aggregate. Scale bar, 1 mm. (**e**) The genetic interaction between *pkaC* and *pufA*.

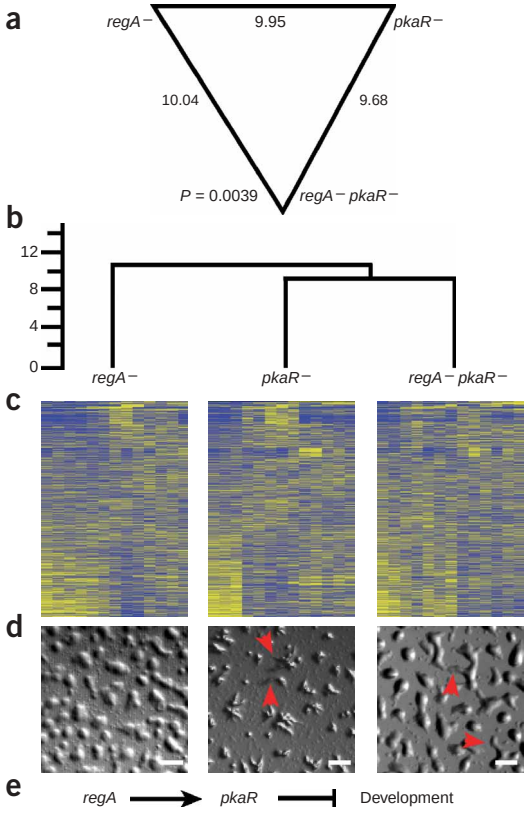


Figure 3 Transcriptional profiling to test whether *pkaR* is epistatic to *regA*. RNA samples from developing *regA*$^-$, *pkaR*$^-$ and *regA*$^-$ *pkaR*$^-$ cells were analyzed. (**a**–**c**) Euclidean distances (**a**), dendrogram (**b**) and heat maps (**c**) as in **Figure 1 a**–**c**. (**d**) Developmental phenotypes of the three mutant strains developed on agar. *regA*$^-$ cells aggregated without streaming. *pkaR*$^-$ and *regA*$^-$ *pkaR*$^-$ cells aggregated with few short streams (arrows). Scale bar, 5 mm. (**e**) The genetic interaction between *pkaR* and *regA*.

To test that possibility, we generated the double mutant *pufA*$^-$ *pkaC*$^-$ and compared its microarray profile with those of *pufA*$^-$ and *pkaC*$^-$. The profile of *pufA*$^-$ *pkaC*$^-$ was most similar to that of *pkaC*$^-$, and both were different from that of *pufA*$^-$ (**Fig. 2a**,**b**). This finding is consistent with the prediction that *pkaC* is epistatic to *pufA*.

Visualizing the gene expression data further supports that notion (**Fig. 2c**). Genes expressed during growth were not downregulated, and developmentally induced genes were not upregulated, in *pkaC*$^-$ or in *pufA*$^-$ *pkaC*$^-$ cells, suggesting that these mutant strains did not develop. Morphological examination of *pkaC*$^-$ and *pufA*$^-$ *pkaC*$^-$ cells confirmed that both mutants failed to aggregate upon starvation, whereas *pufA*$^-$ cells developed well (**Fig. 2d**). These results verify the microarray-based finding that *pkaC* is epistatic to *pufA* (**Fig. 2e**).

Epistasis between mutations with similar phenotypes

Classical epistasis analysis is most effective when the phenotypes of the individual mutations contrast with each other[1,2]. We tested whether microarray phenotypes could provide sufficient resolution for epistasis analysis even with mutations that cause similar morphological phenotypes by analyzing the precocious development mutants *regA*$^-$ and *pkaR*$^-$ (refs. 20–22). Owing to the similarity between the phenotypes of the two gene knockouts, the original epistatic relationships between them were determined using a prespore-specific dominant negative allele[20]. The genetic interaction between *regA* and *pkaR* during aggregation has not been determined, but their biochemical activities suggest that *pkaR* should be epistatic to *regA*.

To test the relationship between *pkaR* and *regA*, we constructed and examined double mutant *regA*$^-$ *pkaR*$^-$ and corresponding single mutant strains. The expression profiles of *pkaR*$^-$ and *regA*$^-$ *pkaR*$^-$ were more similar to each other than to that of *regA*$^-$, indicating that *pkaR* is epistatic to *regA* (**Fig. 3a**,**b**). The distances between these profiles were smaller than those observed in experiments with *yakA*$^-$ *pufA*$^-$ and *pufA*$^-$ *pkaC*$^-$ (**Figs. 1** and **2**) but were significantly different both by the χ^2 test (**Fig. 3a**) and by the bootstrap analysis (**Fig. 3b**). As expected, the gene trajectories of the three mutants were similar (**Fig. 3c**). *pkaR*$^-$ and *regA*$^-$ *pkaR*$^-$ showed a precocious shift in gene expression at 4–6 h instead of 8 h of development, and all three strains showed a precocious loss of coherent gene expression pattern at 14–18 h of development. Such a loss is characteristic of the postsporulation phase of development[19].

These results suggest that *pkaR* is epistatic to *regA* during aggregation. Closer examination of the three strains showed that *regA*$^-$ cells aggregated without streams whereas *pkaR*$^-$ and *regA*$^-$ *pkaR*$^-$ cells aggregated with streams (**Fig. 3d**), although the streams were less elaborate than those of wild-type strains (data not shown) and somewhat different from each other. This morphological phenotype supports the microarray-based conclusion that *pkaR* is epistatic to *regA* during aggregation (**Fig. 3e**).

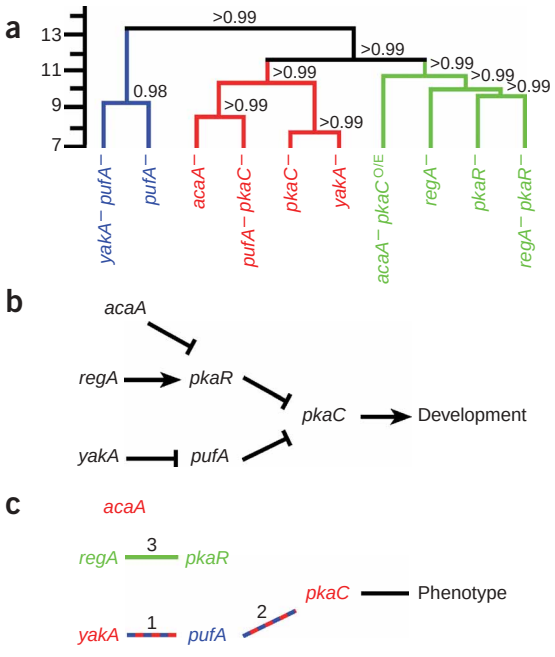


Figure 4 Pathway construction. RNA samples from developing *acaA*$^{-}$ and *acaA*$^{-}$ *pkaC*$^{O/E}$ cells were analyzed. All other data are the same as in **Figures 1–3**. (**a**) Distances were calculated between the ten mutants followed by hierarchical clustering. The dendrogram represents the distances (arbitrary units on scale bar) between the mutants. In this dendrogram, two strains are more similar to each other than to any other strain if they are connected directly. Vertical distances represent the dissimilarity between leaves and joints. Confidence levels (from bootstrap analysis) are shown next to each joint. The blue cluster contains *pufA*$^{-}$ and the double mutant *yakA*$^{-}$ *pufA*$^{-}$, both of which attenuate at the finger stage. The red cluster contains mutants that fail to aggregate; the green cluster, mutants that develop precociously. (**b**) The published PKA regulatory network[2]. Arrows, activation; barred lines, inhibition. (**c**) The genetic network inferred from the microarray data. Colors indicate relationship to the data in **a**.

Reconstruction of the PKA pathway

The above examples demonstrate the application of microarray phenotypes to epistasis analysis with three strains at a time. Next we tested the robustness of the relationships with ten mutants. We first collected microarray data from two additional strains, *acaA*$^{-}$ and *acaA*$^{-}$ *pkaC*$^{O/E}$, and considered them together with the eight mutants described above. *acaA*$^{-}$ cells do not develop due to lack of cAMP[23], but that defect is suppressed by overexpression of PkaC (*acaA*$^{-}$ *pkaC*$^{O/E}$); therefore, *pkaC* is epistatic to *acaA*[24].

Clustering the microarray profiles of the ten strains yielded a dendrogram (**Fig. 4a**) that is consistent with the relationships determined from the two-gene analyses, indicating the robustness of the relationships. *pufA*$^{-}$ and *yakA*$^{-}$ *pufA*$^{-}$ clustered together, separate from *yakA*$^{-}$; *pkaC*$^{-}$ and *pufA*$^{-}$ *pkaC*$^{-}$ clustered together, separate from *pufA*$^{-}$; and *pkaR*$^{-}$ and *regA*$^{-}$ *pkaR*$^{-}$ clustered together, separate from *regA*$^{-}$ (**Fig. 4a**), supporting results from the corresponding two-gene analyses (**Figs. 1–3**).

Finally, we tested whether the published pathway (**Fig. 4b**) could be reconstructed from microarray data alone (**Fig. 4c**). We inferred that *pufA* is epistatic to *yakA* (**Fig. 4c**) from the similarity between the *yakA*$^{-}$ *pufA*$^{-}$ double mutant and the *pufA*$^{-}$ mutant (**Figs. 1** and **4a**). We inferred that *pkaC* is epistatic to *pufA* (**Fig. 4c**) from the similarity between the *pufA*$^{-}$ *pkaC*$^{-}$ double mutant and the *pkaC*$^{-}$ mutant (**Figs. 2** and **4a**). We inferred the linear path from *yakA* through *pufA* to *pkaC* (**Fig. 4c**) from these relationships assuming that a linear combination is the most simple and from the lack of conflicting relationships, as described[2]. We inferred that *pkaR* is epistatic to *regA* (**Fig. 4c**) from the similarity between the double mutant *regA*$^{-}$ *pkaR*$^{-}$ and the *pkaR*$^{-}$ mutant (**Figs. 3** and **4a**). The data could also indicate that *pkaC* is epistatic to *acaA*[24] because the double mutant strain clustered with the *regA*$^{-}$ and *pkaR*$^{-}$ mutants, both of which have activated PkaC (**Fig. 4a**). That relationship is not formally complete, however, because we did not analyze the *pkaC*$^{O/E}$ strain.

Our microarray-based pathway (**Fig. 4c**) is incomplete relative to the published pathway[2] (**Fig. 4b**) because we did not test the relationships between *acaA* and *pkaR* or between *pkaR* and *pkaC*. Nevertheless, we were able to construct the pathway (**Fig. 4c**) from microarray phenotypes alone (**Figs. 1–3** and **4a**), using the basic rules of epistasis analysis[1,2], showing that, in principle, a correct pathway can be constructed from microarray phenotypes of mutant strains alone.

The ten-strain dendrogram (**Fig. 4a**) also shows that profiles of mutants with similar physiologies cluster together, further supporting the validity of transcriptional profiles as phenotyping tools. Specifically, one cluster contains two mutants with attenuated development at the finger stage (**Fig. 4a**). The aggregationless mutants (*acaA*$^{-}$, *pkaC*$^{-}$, *yakA*$^{-}$ and *pufA*$^{-}$ *pkaC*$^{-}$) clustered together, and the precocious developers (*pkaR*$^{-}$, *regA*$^{-}$, *regA*$^{-}$ *pkaR*$^{-}$ and *acaA*$^{-}$ *pkaC*$^{O/E}$) were in a separate cluster (**Fig. 4a**).

Complexity of the transcriptional phenotype

pkaC is epistatic to *pufA*[16] (**Fig. 2**), but although *pkaC*$^{-}$ and *pufA*$^{-}$ *pkaC*$^{-}$ have indistinguishable morphologies (**Fig. 2d**), the difference

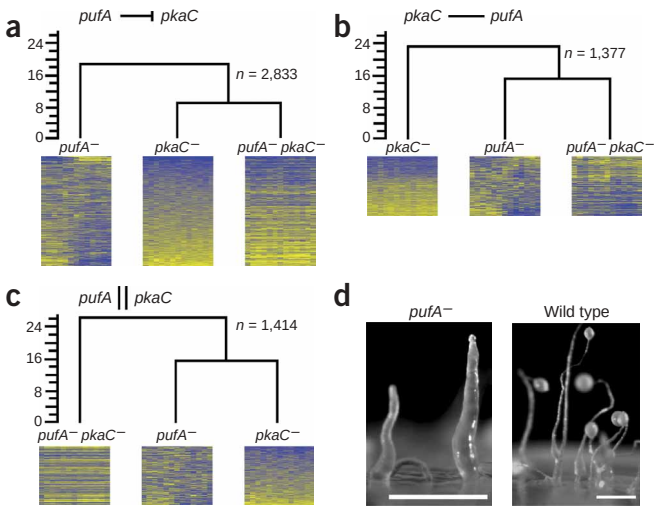


Figure 5 Additional genetic relationships between *pkaC* and *pufA*. Data from the single mutant strains *pkaC*$^{-}$ and *pufA*$^{-}$ and from the double mutant strain *pkaC*$^{-}$ *pufA*$^{-}$ (the same as in **Fig. 2**) were searched for genes that support the ideas that *pkaC* is epistatic to *pufA* (**a**), *pufA* is epistatic to *pkaC* (**b**) and *pkaC* is parallel to *pufA* (**c**). The dendrograms represent the clustering of the three mutants based on these genes; the number of genes supporting each relationship is indicated. The heat maps represent the gene trajectories in the three mutants, ranked by their order in the *pkaC*$^{-}$ strain. Developmental morphologies of the *pufA*$^{-}$ mutant (left) and wild type 24 h after starvation on nitrocellulose filters are shown (**d**) to illustrate the finger-stage attenuation of the mutant. Scale bar, 1 mm.

between their microarray profiles was not zero (**Figs. 2b** and **4a**). Experimental noise may partly account for that difference, but we suspected that the relationship between *pufA* and *pkaC* might be more complicated than expected. Formally, there are two other possible genetic relationships: *pkaC* and *pufA* could function independently (*pkaC* is parallel to *pufA*) or *pufA* could be epistatic to *pkaC*. We searched the microarray data for gene trajectories that would support either of these two relationships.

We calculated the Euclidean distances between the three time courses for each gene trajectory and divided the genes into three groups according to the relationship they best support (**Fig. 5**). The largest group, 2,833 genes, showed the smallest distance between *pkaC*$^-$ and *pufA*$^-$ *pkaC*$^-$, supporting the idea that *pkaC* is epistatic to *pufA* (**Fig. 5a**). We found 1,377 genes whose trajectories supported the idea that *pufA* is epistatic to *pkaC* (**Fig. 5b**) and 1,414 genes that supported the idea that *pkaC* is parallel to *pufA* (**Fig. 5c**). The 1,377 genes that support the idea that *pufA* is epistatic to *pkaC* had a coherent pattern in the *pkaC*$^-$ mutant and almost no pattern in the *pufA*$^-$ mutant or in the double mutant. In that case, the clustering was influenced by the distance of *pkaC*$^-$ from *pufA*$^-$ and the double mutant more than by any similarity between *pufA*$^-$ and the double mutant (**Fig. 5b**). In the parallel relationship, there is no expectation that the two single mutant profiles would be similar, only that the double mutant would be different from both (**Fig. 5c**). We therefore propose that *pufA* and *pkaC* may have parallel functions in addition to their expected epistatic relationship.

That idea is also supported by the developmental attenuation of the *pufA*$^-$ strain[16]. After 24 h of development, *pufA*$^-$ cells were at the finger stage of development, whereas wild-type cells had developed mature fruiting bodies (**Fig. 5d**). Inactivation of *pufA* leads to hyperactivation of PkaC[16]. Therefore, if things were simple, *pufA*$^-$ cells should have developed precociously, like *pkaR*$^-$, *pkaC*$^{O/E}$ and *regA*$^-$ cells. The finger-stage attenuation is suggestive of a *pufA*-dependent developmental process that is independent of *pkaC*, consistent with the idea of a parallel pathway. For example, PufA may suppress YakA translation late in development[16]. Our data (**Fig. 5**) show that relationships between genes are dependent on the measured phenotype, that microarray data provide a comprehensive phenotype in one experiment and that microarray-based epistasis provides insight into the biological function of genes.

DISCUSSION

Our data show that microarray profiles can be used as phenotypes in epistasis analysis. This approach has several advantages over classical methods. First, existing knowledge of the phenotype is not needed, whereas in classical genetics, investigators must decide which phenotypes to measure. In our case, one would measure aggregation, culmination, sporulation, etc., and data from the different assays might not be compatible or consistent. Second, our approach allows derivation of epistatic relationships without knowledge of the phenotype. This characteristic provides an enormous advantage for mutations in certain genes that do not result in a readily measurable phenotype. The third advantage is quantification: classical genetics does not define how similar two mutants must be for epistasis to be determined, because it is based on many different assays[1,2]. We used Euclidian distance in this report, but other distance metrics could be used as well (we obtained qualitatively identical results using Pearson's correlation instead of Euclidian distance calculations; data not shown). Therefore, the microarray profile provides a quantitative and uniform measurement that could be used to compare all possible mutants in a given genome. Microarray profiling of mutant strains could be used in the analysis of large-scale mutant collections[3–5] and in conjunction with methods for automated experimental design and analysis[2,25].

We also found that transcriptional profiles may be more sensitive than morphology or biochemistry. We found that the distance between the phenotypes of the epistatic mutation and the respective double mutation was never zero. Part of this distance must be due to experimental noise, but the case of *pufA* and *pkaC* shows that it may also be biologically relevant. The idea that *pufA* and *pkaC* may have independent roles in addition to their epistatic relationship during aggregation was first suggested by the morphological attenuation of the *pufA*$^-$ mutants at the finger stage[16]. Most of the microarray data supports the early developmental relationship between *pufA* and *pkaC*, but hundreds of gene trajectories suggest that *pufA* may have an independent role later in development. These findings may explain why *pufA*$^-$ cells, which have high PkaC activity, do not sporulate precociously like *regA*$^-$ and *pkaR*$^-$ cells, which also have high PkaC activity[16,20–22]. *D. discoideum* PufA belongs to the PUF family of highly conserved RNA binding proteins. PUFs are characterized by an RNA-binding domain that recognizes and binds specific 3′ UTRs of mRNA sequences[26]. The fly homolog of PufA, Pumilio, acts together with Nanos to bind the 3′ UTR of hunchback mRNA, specifying the abdomen in fly embryos[27–30]. The interaction between the two proteins also controls cell division in the fly germ line by binding the 3′ UTR of cyclin B mRNA[31]. Our finding that hundreds of genes are not in agreement with the idea that *pkaC* is epistatic to *pufA* suggests that PufA might bind mRNA targets other than the *pkaC* transcript. This idea is supported by findings in *Saccharomyces cerevisiae* that each of the five yeast Puf proteins binds dozens of functionally related mRNAs[32]. Hence, an advantage of determining epistatic relationships between genes based on the transcriptional profiles of mutants is that one assay can show the complexity of the relationships between genes, which is impossible with less-detailed assays.

Another demonstration of the high resolution provided by the microarray phenotype is the relationship between *regA* and *pkaR*. Knockout mutations in either gene result in precocious development[20,33]. Determining the epistatic relationship between such mutants is nearly impossible with classical genetics because two mutations must have different, preferably opposite, phenotypes[1,2]. The microarray analysis allowed us to determine that *pkaR* is epistatic to *regA* using two knockout mutations that result in similar phenotypes. The finding was statistically significant and robust because the relationship between every two single-mutation strains and the corresponding double-mutation strain remained the same even in the context of seven other strains.

Although microarray phenotypes provide many advantages, they also have a distinct disadvantage. In classical genetics, the influence of one gene on another or of a gene on a phenotype can be described as excitatory or inhibitory, depending on the effects of the mutation[1,2]. For example, if knocking out *yakA* leads to lack of aggregation, *yakA* is said to excite aggregation. The microarray phenotype can also be interpreted in that way, but the interpretation requires knowledge of the relationship between the microarray profile and the biology of the organism. This type of interpretation is not possible for every mutation.

Epistasis analysis with microarray data also suffers from the conceptual problems that plague classical epistasis analysis. For example, we assume that the genes function in a signaling pathway, where the epistatic mutation is the downstream one, assuming a linear pathway[2]. In biochemical or developmental pathways, the situation is reversed, and there is no simple way to decide which rules to apply to a given problem. The assumption of linearity may also be incorrect, resulting in situations where the epistatic mutation is upstream, even in a

signaling pathway[1]. Because the microarray data do not offer solutions to these problems, detailed analysis of individual pathways is still required after general pathways are described by microarray profiling.

We previously showed that epistasis analysis with classical, qualitative genetic data could be automated[2,34]. We are currently working on automating the task of epistasis analysis with quantitative microarray data as well. Automated genetic analysis of uniform, quantitative microarray phenotypes should be useful for constructing genetic networks on a genome scale.

METHODS

Strains and transformation. The following *D. discoideum* strains were previously described: *yakA*$^-$ (ref. 35), *pufA*$^-$ (ref. 16), *yakA*$^-$ *pufA*$^-$ (ref. 16), *pkaC*$^-$ (ref. 36), *regA*$^-$ (ref. 37), t108d2 [AX4 *pkaR*$^-$][38] and *acaA*$^-$ and *acaA*$^-$ *pkaC*$^{O/E}$ (*acaA*$^-$ *act15/pkaC*), in which *pkaC* is overexpressed under regulation of the actin 15 promoter in *acaA*$^-$ cells[24]. Gene functions are described in **Table 1**.

We used gene disruption by homologous recombination and selection for Blasticidin S resistance to generate the double mutants *regA*$^-$ *pkaR*$^-$ and *pufA*$^-$ *pkaC*$^-$. We used the knockout vector for *regA* described previously[37] to transform strain AK0636 (AX4 *pkaR*$^-$)[38]. We transformed *pufA*$^-$ cells with the knockout vector for *pkaC*, p292 (provided by H. K. MacWilliams; Ludwig-Maximilians-Universität München). We confirmed gene disruptions by Southern-blot analysis with gene-specific probes (data not shown).

Growth, development and RNA preparation. We grew mutant *D. discoideum* strains in HL5 liquid broth[39]. We washed exponentially growing cells free of nutrients, deposited them on nitrocellulose filters at 3×10^6 cells cm^{-2} and developed them at 22 °C (ref. 40). We collected samples (1×10^8 cells each) at 2-h intervals for 24 h, resuspended cells in 1 ml of TRIZOL reagent (Life Technologies) and extracted total RNA in accordance with the manufacturer's recommended protocol. In addition, for each time point collected, we determined the proportion of spores as described[41]. We repeated filter development experiments of each mutant strain at least twice (biological replications).

For documentation of morphology, we plated single cells on SM-nutrient agar plates in association with bacteria (*Klebsiella aerogenes*)[39] and photographed the emerging plaques. We documented aggregation morphology by depositing HL5-grown cells on non-nutrient agar and following their development by time-lapse video microscopy with transmitted light as described[42].

Microarray experiments, normalization and multi-array scaling. We carried out microarray experiments as described[19]. We analyzed each sample by a two-color assay where the common reference was a total RNA sample made from several time points of developing wild-type (AX4) cells and the hybridization targets consisted mainly of cDNA. We applied a single-chip normalization procedure to the quantified data to correct for spatial effects, to reject irreproducible data based on the variability of replicate log-ratios and to bring the data to a common measurement scale to allow for multiarray comparisons as described[19,43].

For each mutant, we carried out the experiment with three kinds of replication. First, we printed each hybridization target twice on the array, allowing for single-chip normalization (**Supplementary Note** online). In addition, we carried out at least two hybridization experiments from each RNA extraction to allow correction for hybridization variation (technical replication). Finally, we carried out most of the assays with two biological replications. We averaged both the technical and biological replicates to form normalized time course data.

When comparing time courses of different mutants, we brought the data to a common measurement scale to allow for multiarray comparisons. We scaled the averaged time course data sets for the different mutants to multiple arrays by dividing the log(ratio) of each gene by the median log(ratio) of all the genes at a given time point. We also analyzed all data with the Bioconductor software[44] (**Supplementary Figs. 1–5** online) and found that it was qualitatively indistinguishable from the data shown here (**Figs. 1–5**), indicating that our conclusions are independent of the normalization procedure. Data analysis with the Bioconductor software is described in **Supplementary Note** online.

Detection of epistatic relationships. To detect epistatic relationships, we centered the mutant time course data to the wild-type time course[19]: for each gene, we created an expression vector consisting of the $\log_2$ ratios of that gene at the 13 time points. We scaled each 13-element gene vector by subtracting the median $\log_2$ ratio of the gene in the wild-type time course. This operation is equivalent to comparing all the gene expression values in the mutants to their respective values in the wild-type.

To calculate dissimilarities between mutant time-course experiments, we calculated Euclidean distances both between the wild-type–centered time courses and between the smooth fit coefficients. To calculate the Euclidean distances, we converted the 5,624-row × 13-column matrix of each mutant time course to a $5{,}624 \times 13$ vector. We used the dist() function in the statistical software package R to calculate pairwise Euclidean distances. We imported the resulting distance matrix into the hclust() function in R using complete-linkage clustering to connect the nodes of the trees. We visualized the results as hierarchical trees (dendrograms). We calculated confidence levels by 10,000 rounds of bootstrapping[17,18] with randomly selected (with replacement) 5,624 genes.

Because dendrograms obscure some distances between mutants, we also calculated and presented the results as triangles where the nodes represent strains and the edges represent the Euclidean distances between the microarray profiles of the respective strains. We also used the χ^2 statistic to test the distribution of microarray targets supporting each possible genetic relationship against the null hypothesis of uniform distribution.

Detection of genes that violate epistatic relationships. To identify genes that were not consistent with the idea that *pkaC* is epistatic to *pufA*, we calculated Euclidean distances between the 13-element gene expression vectors of the *pufA*$^-$, *pkaC*$^-$ and *pufA*$^-$ *pkaC*$^-$ mutants and divided the 5,624 genes into three groups. In one group, the distance between the gene expression vectors of *pufA*$^-$ and *pkaC*$^-$ was the smallest, supporting the idea that *pufA* is parallel to *pkaC*. In the second group, the distance between the expression vectors of *pufA*$^-$ and *pufA*$^-$ *pkaC*$^-$ was the smallest, supporting the idea that *pufA* is epistatic to *pkaC*. The last group of genes supported the idea that *pkaC* is epistatic to *pufA*. We carried out hierarchical clustering of the three groups of genes as described above and plotted the results as dendrograms.

ArrayExpress accession numbers. Array, A-MEXP-161; experiment, E-TABM-6.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank V. Lundblad for discussions and for critical reading of the manuscript, R. Guerra for help with statistical analysis and for discussions and E. Holloway for assistance with data deposition. This work was supported by a grant from the National Institute of Child Health and Human Development. N.V.D. and E.O.B. are supported in part by training fellowships from the W.M. Keck Foundation of the Gulf Coast Consortia through the Keck Center for Computational and Structural Biology. J.D., P.J. and B.Z. are supported in part by the Slovene Ministry of Education, Science and Sports.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 15 December 2004; accepted 10 February 2005
Published online at http://www.nature.com/naturegenetics/

1. Avery, L. & Wasserman, S. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet.* **8**, 312–316 (1992).
2. Zupan, B. *et al.* GenePath: a system for automated construction of genetic networks from mutant data. *Bioinformatics* **19**, 383–389 (2003).
3. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
4. Tong, A.H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
5. Winzeler, E.A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
6. Perou, C.M. *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* **96**, 9212–9217 (1999).

7. Golub, T.R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
8. Gray, N.S. *et al.* Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **281**, 533–538 (1998).
9. Holstege, F.C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
10. Alizadeh, A.A. & Staudt, L.M. Genomic-scale gene expression profiling of normal and malignant immune cells. *Curr. Opin. Immunol.* **12**, 219–225 (2000).
11. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
12. Roberts, C.J. *et al.* Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–880 (2000).
13. Shaulsky, G. & Loomis, W.F. Gene expression patterns in Dictyostelium using microarrays. *Protist* **153**, 93–98 (2002).
14. Loomis, W.F. *Dictyostelium discoideum: A Developmental System* (Academic, New York, 1975).
15. Loomis, W.F. Role of PKA in the timing of developmental events in *Dictyostelium* cells. *Microbiol. Mol. Biol. Rev.* **62**, 684 (1998).
16. Souza, G.M., da Silva, A.M. & Kuspa, A. Starvation promotes *Dictyostelium* development by relieving PufA inhibition of PKA translation through the YakA kinase pathway. *Development* **126**, 3263–3274 (1999).
17. Efron, B., Halloran, E. & Holmes, S. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93**, 13429–13434 (1996).
18. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
19. Van Driessche, N. *et al.* A transcriptional profile of multicellular development in *Dictyostelium discoideum*. *Development* **129**, 1543–1552 (2002).
20. Shaulsky, G., Fuller, D. & Loomis, W.F. A cAMP-phosphodiesterase controls PKA-dependent differentiation. *Development* **125**, 691–699 (1998).
21. Mann, S.K.O., Yonemoto, W.M., Taylor, S.S. & Firtel, R.A. DdPK3, which plays essential roles during *Dictyostelium* development, encodes the catalytic subunit of cAMP-dependent protein kinase. *Proc. Natl. Acad. Sci. USA* **89**, 10701–10705 (1992).
22. Simon, M.N., Pelegrini, O., Veron, M. & Kay, R.R. Mutation of protein kinase A causes heterochronic development of *Dictyostelium*. *Nature* **356**, 171–172 (1992).
23. Pitt, G.S. *et al.* Structurally distinct and stage-specific adenylyl cyclase genes play different roles in *Dictyostelium* development. *Cell* **69**, 305–315 (1992).
24. Wang, B. & Kuspa, A. *Dictyostelium* development in the absence of cAMP. *Science* **277**, 251–254 (1997).
25. King, R.D. *et al.* Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252 (2004).
26. Spassov, D.S. & Jurecic, R. The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? *IUBMB Life* **55**, 359–366 (2003).
27. Wharton, R.P. & Struhl, G. RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. *Cell* **67**, 955–967 (1991).
28. Murata, Y. & Wharton, R.P. Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in Drosophila embryos. *Cell* **80**, 747–756 (1995).
29. Wharton, R.P., Sonoda, J., Lee, T., Patterson, M. & Murata, Y. The Pumilio RNA-binding domain is also a translational regulator. *Mol. Cell* **1**, 863–872 (1998).
30. Sonoda, J. & Wharton, R.P. Recruitment of Nanos to hunchback mRNA by Pumilio. *Genes Dev.* **13**, 2704–2712 (1999).
31. Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K. & Kobayashi, S. Maternal Pumilio acts together with Nanos in germline development in Drosophila embryos. *Nat. Cell Biol.* **1**, 431–437 (1999).
32. Gerber, A.P., Herschlag, D. & Brown, P.O. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* **2**, E79 (2004).
33. Abe, K. & Yanagisawa, K. A new class of rapid developing mutants in *Dictyostelium discoideum*: Implications for cyclic AMP metabolism and cell differentiation. *Dev. Biol.* **95**, 200–210 (1983).
34. Demsar, J. *et al.* GenePath: a computer program for genetic pathway discovery from mutant data. *Medinfo* **10**, 956–959 (2001).
35. Souza, G.M., Lu, S. & Kuspa, A. YakA, a protein kinase required for the transition from growth to development in *Dictyostelium*. *Development* **125**, 2291–2302 (1998).
36. Wetterauer, B.W. *et al.* A protein kinase from *Dictyostelium discoideum* with an unusual acidic repeat domain. *Biochim. Biophys. Acta* **1265**, 97–101 (1995).
37. Shaulsky, G., Escalante, R. & Loomis, W.F. Developmental signal transduction pathways uncovered by genetic suppressors. *Proc. Natl. Acad. Sci. USA* **93**, 15260–15265 (1996).
38. Wang, B. & Kuspa, A. CulB, a putative ubiquitin ligase subunit, regulates prestalk cell differentiation and morphogenesis in *Dictyostelium* spp. *Euk. Cell* **1**, 126–136 (2002).
39. Sussman, M. Cultivation and synchronous morphogenesis of *Dictyostelium* under controlled experimental conditions. in *Methods in Cell Biology* (ed. Spudich, J.A.) 9–29 (Academic, Orlando, Florida, 1987).
40. Shaulsky, G. & Loomis, W.F. Cell type regulation in response to expression of ricin-A in *Dictyostelium*. *Dev. Biol.* **160**, 85–98 (1993).
41. Kibler, K. *et al.* A novel developmental mechanism in *Dictyostelium* revealed in a screen for communication mutants. *Dev. Biol.* **259**, 193–208 (2003).
42. Kibler, K., Svetz, J., Nguyen, T.L., Shaw, C. & Shaulsky, G. A cell-adhesion pathway regulates intercellular communication during *Dictyostelium* development. *Dev. Biol.* **264**, 506–521 (2003).
43. Katoh, M. *et al.* An orderly retreat: dedifferentiation is a regulated process. *Proc. Natl. Acad. Sci. USA* **101**, 7005–7010 (2004).
44. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
45. Thomason, P.A. *et al.* An intersection of the cAMP/PKA and two-component signal transduction systems in *Dictyostelium*. *EMBO J.* **17**, 2838–2845 (1998).
46. de Gunzburg, J., Franke, J., Kessin, R.H. & Veron, M. Detection and developmental regulation of the mRNA for the regulatory subunit of the cAMP-dependent protein kinase of *Dictyostelium discoideum* by cell-free translation. *EMBO J.* **5**, 363–367 (1986).
47. Anjard, C., Pinaud, S., Kay, R.R. & Reymond, C.D. Overexpression of DdPK2 protein kinase causes rapid development and affects the intracellular cAMP pathway of *Dictyostelium discoideum*. *Development* **115**, 785–790 (1992).