PERSPECTIVE

nature genetics

Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease

Venkateswara R Chintapalli, Jing Wang & Julian A T Dow

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

FlyAtlas, a new online resource, provides the most comprehensive view yet of expression in multiple tissues of *Drosophila melanogaster*. Meta-analysis of the data shows that a significant fraction of the genome is expressed with great tissue specificity in the adult, demonstrating the need for the functional genomic community to embrace a wide range of functional phenotypes. Well-known developmental genes are often reused in surprising tissues in the adult, suggesting new functions. The homologs of many human genetic disease loci show selective expression in the *Drosophila* tissues analogous to the affected human tissues, providing a useful filter for potential candidate genes. Additionally, the contributions of each tissue to the whole-fly array signal can be calculated, demonstrating the limitations of whole-organism approaches to functional genomics and allowing modeling of a simple tissue fractionation procedure that should improve detection of weak or tissue-specific signals.

Experimental reverse genetics is a powerful tool for understanding novel genes, a key goal of functional genomics[1,2], both for basic science and for the understanding of human disease[3,4]. Simple genetic models have vital roles in this endeavor because of the relative ease, power and cost of their reverse genetic techniques[2,5] compared with mouse. The announcement of the *Drosophila* genome was accompanied by prediction of its utility in understanding human genetic disease, and this optimism was accompanied by the formation—and subsequent closure or refocusing—of several companies. Now, a more prosaic approach pertains. The case for *Drosophila* as a model of human disease is based on a gene-by-gene argument, and the powerful genetics associated with this tiny fly is starting to produce useful insights. It has also proved possible to 'humanize' the fly by introducing human genes of interest and studying them in an organotypic context that can prove more informative than studies in cell culture.

However, recent data suggest that to study novel genes by reverse genetics (that is, studying gene function by generating mutants and studying the phenotypes), the *Drosophila* community may need to broaden its focus. It has long been realized that functional genomics demands functional phenotypes, whereas most model organisms were adopted for studies of development. The mismatch between the range of functions of an organism's genes and the range of phenotypes available in that organism has been termed the 'phenotype gap'[6–8]. The utility of *Drosophila* in studies of development is beyond dispute, but it is salutary to note that perhaps a third of a million researcher-years spent studying *Drosophila* (predominantly its development) had led to the identification of only 20% of its genes[8] before the release of the genome sequence[9]. So if most of the genes encoded by the *Drosophila* genome are not primarily developmental in function, where should one seek phenotypes for the rest?

Drosophila research in the UK has been helped by the Investigating Gene Function (IGF) initiative of the Biotechnology and Biological Sciences Research Council (BBSRC), which has invested £4 million ($6 million) in infrastructure for microarrays, proteomics and novel mutant collections. As part of this initiative, the UK *Drosophila* Affymetrix Microarray Facility has provided 1,000 arrays, as well as processing and analysis, free of charge to over 50 projects from the UK (http://www.mblab.gla.ac.uk/igf). As a service to the community, we have produced FlyAtlas (http://flyatlas.org), a microarray-based atlas of gene expression in multiple adult tissues. As well as providing some clear messages about the utility of whole-fly arrays, the database also helps delineate the phenotype gap by identifying those tissues in which specific genes of interest (and thus, the homologs of human disease genes) can be studied. The online data set thus provides an instant entrée into the field, not just for Drosophilists but also for scientists who can identify a likely homolog in the fly, regardless of the organism they use.

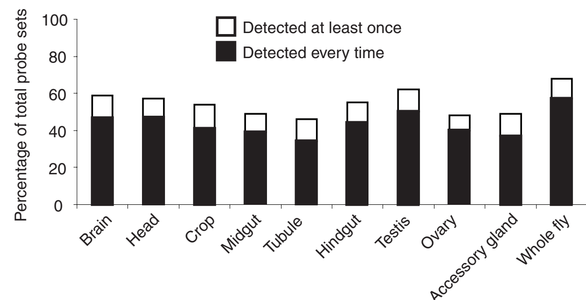

Figure 1 *Drosophila* tissues typically express around half the computed transcriptome. For each tissue, the number of probe sets giving at least one 'present' call is shown, together with those meeting the stricter criterion of giving four present calls out of four chips.

Venkateswara R. Chintapalli, Jing Wang and Julian A.T. Dow are in the Division of Molecular Genetics, University of Glasgow, Glasgow G11 6NU, UK.
e-mail: j.a.t.dow@bio.gla.ac.uk

Published online 29 May 2007; doi:10.1038/ng2049

Table 1 Genes that show extreme specificity of expression and that serve to validate the quality and discrimination of the data set

		Mean signal level								
Gene	Expected region	Brain	Head	Crop	Midgut	Tubule	Hindgut	Ovary	Testis	Accessory gland
DD2R	Brain	**436**	39	5	1	3	1	0	1	3
trp	Head (eye)	74	**2,649**	2	5	2	1	0	3	5
CG8985	Crop	13	8	**364**	2	3	1	0	1	4
ser99Dc	Midgut	0	11	3	**6,335**	5	4	0	2	3
CG15408	Tubule	1	0	3	2	**4,337**	3	1	7	5
CG9993	Hindgut	3	4	2	11	13	**2,647**	2	2	9
Otu	Ovary	4	5	7	7	14	4	**1,353**	6	4
Mst84Dd	Testis	3	5	5	9	6	8	1	**2,525**	46
Acp76A	Accessory gland	3	5	4	3	5	3	4	40	**11,565**

Errors are omitted for clarity; these are in the range of 5%–10% and are given online at http://flyatlas.org/. Boldface indicates the maximum signal for each gene.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

FlyAtlas is composed of data covering nine distinct adult tissues (brain, head, crop, midgut, Malpighian tubules, hindgut, testis, ovary and male accessory glands), dissected from 7-day-old Canton S adults, and two larval tissues (tubule and fat body). Each tissue was processed with four biological replicates on Affymetrix *Drosophila* Genome 2 chips (with 18,880 probe sets for 18,500 transcripts) and compared with a matched whole-fly sample (**Supplementary Methods** online). The data set is thus composed of 900,960 individual expression values. FlyAtlas presents an excellent opportunity to study gene expression in multiple tissues and provides a complementary resource to published developmental data sets[10].

The transcriptional landscape of the adult

The data provide a good survey of gene expression in the adult. A similar fraction of the genome is expressed in each tissue (**Fig. 1**): in each case, around half of all probe sets demonstrate the presence of significant expression. Between them, the different tissues show expression of 85% of the computed *Drosophila* transcriptome, whereas only 67% of probe sets are detectable in the whole-fly sample (**Fig. 1**). Although the widespread practice of grinding up the whole organism for transcriptomic or proteomic studies is understandable, given the tiny size of the whole organism, the implication of this result, as we discuss later, is that a significant fraction of the genome will be missed or underrepresented in such samples.

Of course, the quality of the data set and the underlying tissue dissections is critical for such assertions. It is possible to perform an informal 'quality control' on the data set by seeking genes that are strongly enriched in particular tissues. For example, the photoreceptor channel *trp* should be enriched in the head sample but ideally should not show up strongly in the brain. The results (**Table 1**) confirm that there is excellent discrimination, even between physically adjacent tissues (such as brain and head; midgut, tubule and hindgut or testis and accessory gland).

Generalizing from these examples, although 25% of all probe sets are ubiquitously expressed ('housekeeping' genes), there is a significant fraction of the genome that is highly tissue-specifically expressed in the adult (**Table 2**). Although brain and testis are particularly distinctive[11], every tissue has a population of tens to thousands of genes that are detected nowhere else.

The clear implication of this result is that, for over a quarter of *Drosophila* genes, there is a single tissue in which study should be focused. Indeed, it may be counterproductive to embark on a now-standard reverse-genetic workup of a novel gene unless that tissue is studied, because an informative phenotype may be missed. Analogous to the Krogh principle of comparative physiology that "for every physiological problem, there is an animal uniquely suited by nature to study it"[12], we propose an analogous principle for functional genomics: for every novel gene, it is as sensible to study it where it is most abundantly expressed as to study it where it is first encountered. Thus, FlyAtlas is a first step toward connecting the researcher to the tissue.

These results also show the importance of taking an organismal (tissue) and ontogenetic (multiple–life stage) view of gene expression in order to prevent a danger we term 'shoehorning': the squeezing of a gene's declared functions into expected phenotypes (for example, assuming that a channel must be neural in function) or into the phenotypes familiar to a particular experimenter (behavioral or developmental, for example). Given that the vast majority of *Drosophila* research focuses on embryonic development, and most of the remainder on neurogenetics, there is a real danger that major functions might be overlooked because of the dearth of functional research outside these areas. Again, the availability of tissue-specific expression data should encourage a more global view of each gene's possible functions.

Meta-analysis

There are further useful meta-analyses possible for these data. For example, research in *Drosophila* has been revolutionized by the GAL4/UAS binary system for tissue-specific expression of transgenes[13]. Many GAL4 drivers are derived from enhancer-trap screens and may not be absolutely specific to a particular tissue or cell type. Genes that are absolutely specific to a single tissue of those studied so far can be extracted from the data set (**Supplementary Table 1** online); the control regions of these genes may prove ideal drivers for GAL4 expression in flies. Conversely, there is also a need to identify genes with abundant and relatively invariant expression. It has been shown that quantitative PCR normalization is better performed with a basket of invariant genes, from which the

Table 2 Genes expressed uniquely in specific tissues

Tissue	Unique expression	Predominant expression
Brain	292	779
Head	156	585
Crop	12	12
Midgut	117	316
Tubule	31	80
Hindgut	80	279
Testis	1,317	2,079
Ovary	81	241
Accessory glands	47	339
Total (out of 18,770)	2,121	4,698

For each tissue, 'unique expression' refers to the number of probe sets called 'present' at least twice but present nowhere else in the fly; 'predominant expression' refers to the number of probe sets called 'present' more often in a given tissue than in all other tissues combined.

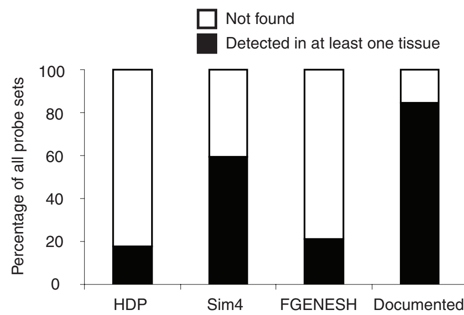

Figure 2 Evidence for novel transcription units in the *Drosophila* genome. Probe sets were classified according to their Affymetrix annotations; documented genes refer to probe sets that are annotated with a FBgn or CG reference number and are thus known to FlyBase. The remaining probe sets were then categorized by feature code (HDP, sim4 or FGENESH). To provide a strict test of expression, the number of tissues for which all four chips were called 'present' by Affymetrix software was scored for each probe set, and only those with at least one expressed tissue were scored as 'expressed'. Data are expressed as a percentage of all probe sets in that category.

best reference genes for a particular experiment can be deduced experimentally[14]. FlyAtlas offers good candidate reference genes for the adult (**Supplementary Table 2** online); interestingly, the *Drosophila* 'standard' reference gene r*p49* (*Rpl32*) ranks 3,172nd.

New transcription units

There are 18,880 *Drosophila* probe sets on the Affymetrix *Drosophila* 2 expression chip, covering around 13,500 genes. This is because many genes have additional, transcript-specific probe sets. However, several thousand probe sets were designed against features that were not considered sufficiently authoritative to justify annotation as a *Drosophila* gene, usually because of lack of compelling data that the putative gene was transcribed. It is clear from FlyAtlas that many of these features are genuinely transcribed (**Supplementary Table 3** online), frequently tissue-specifically; this may explain why they were not hit with multiple ESTs. For example, HDP feature 01001 has a prodigious signal of 9,570 in male accessory glands and is almost undetectable elsewhere, so it is not surprising that its expression was not previously validated. Furthermore, each class of feature has a characteristic hit rate (**Fig. 2**). Across FlyAtlas, 85% of probe sets against documented genes were found to be transcribed in at least one tissue; whereas 18%, 60% and 21% of HDP, sim4 and FGENESH features were found to be transcribed. Indeed, of 3,000 undocumented features on the Affymetrix array, 665 detect significant expression somewhere in the adult. Thus, microarrays of tissue-specific RNA samples may be useful tools in transcript discovery, even for a genome as well-documented as *Drosophila*.

Table 3 Are the genes identified in embryonic tissues by *in situ* analyses also expressed in adult tissues?

	Number of genes		
Region	Adult array	Embryonic *in situ*	Common
Brain	8,092	1,218	1,115
Midgut	6,770	1,452	1,255
Tubule	5,969	507	445
Hindgut	7,588	827	759

Adult regions selected were those for which there are recognizable late-embryonic precursors. A nonredundant list of CG numbers for genes called 'present' on four out of four chips was compared with CG numbers identified by embryonic *in situ* analyses to be either found in a given tissue, expressed ubiquitously or called 'faint ubiquitous' in stage 13–16 embryos. CG numbers common to both lists were counted.

An ontogenetic perspective

These data provide perhaps the clearest view yet of the transcriptional landscape of adult *Drosophila*; but does it bear any resemblance to other life stages? The Berkeley *Drosophila* expression database (http://www.fruitfly.org/cgi-bin/ex/insitu.pl) thus far contains information on the systematic determination of patterns of gene expression in *Drosophila* embryogenesis by RNA *in situ* for 6,138 genes. The Berkeley *Drosophila* Genome Project (BDGP) data set is searched by FlyAtlas so that users are notified if an embryonic expression pattern is available for a particular gene. Additionally, the BDGP *in situ* data have been annotated with Gene Ontology terms for distinct body parts, so we looked for persuasive overlap between embryonic and adult expression patterns. There is very little reason to assert *a priori* that the same genes expressed in a given embryonic tissue are still expressed in the adult, but if around half the genome were expressed in an embryonic tissue (as in adult), and if there were no connection between embryonic and adult expression patterns, one would predict that only 25% of *in situ* hits would be expressed in the corresponding adult tissue. In fact, the concordance is far higher for the several tissues sampled (**Table 3**), with typically 90% of the genes identified by embryonic *in situ* analyses still expressed in adult flies, as measured by microarray. This result suggests intriguingly that the mature transcriptional profile of many tissues is substantially established by the late embryo.

In principle, *in situ* and microarray analyses are complementary; the former provides great spatial resolution (down to the single-cell level), whereas the latter is much more sensitive and quantitative. This can be seen in the relatively low numbers of genes detected by *in situ* analyses, even taking into account the 6,138 genes in the data set. However, it is clear that a joint *in situ*/array approach is powerful.

Surprising expression

FlyAtlas provides further provocative reminders that genes may not be expressed merely where they are expected. **Table 4** shows some exciting insights into some well-known genes, selected to cover a range of gene

Table 4 Some genes that are predominantly expressed in unexpected places

		Mean signal levels							
Gene	Described in	Brain	Head	Midgut	Tubule	Hindgut	Ovary	Testis	Accessory gland
cry	Circadian behavior	279	575	267	**1,972**	868	7	25	205
fas2	Neuronal fasciculation	129	66	49	**1,676**	78	5	9	53
obp56d	Olfaction	71	4,045	1	1	**5,663**	1	106	5
kelch	Nurse cell	181	**185**	22	16	22	5	6	6
rpk	Sensory neurons	1	0	1	7	0	**904**	47	0
toe	Eye, thorax	10	68	7	8	14	8	13	**3,725**
vnd	Embryonic CNS	6	4	**289**	5	6	3	2	8
dsx	Sex determination	21	119	89	**140**	106	9	1	8

Boldface indicates the maximum signal for each gene.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Table 5 *Drosophila* genes expressed in tissues analogous to those involved in human disease

	Tissue signals									
Gene	Brain	Head	Midgut	Hindgut	Tubule	Ovary	Testis	Accessory gland	Human gene	OMIM entry
Brain										
kek2	**156**	19	4	5	8	2	3	5	*SLITRK1* (*KIAA1910*)	Tourette syndrome (137580)
CG5594	**2,267**	867	441	142	35	133	53	124	*SLC12A6*	Agenesis of the corpus callosum with peripheral neuropathy (218000)
CG1909	**1,265**	306	2	24	2	6	0	9	(*RAPSYN*)	Congenital myasthenic syndrome associated with AChR deficiency (608931)
Lcch3	**748**	173	2	3	3	2	8	5	Gaba-A receptor gamma-2	Myoclonic epilepsy, severe, of infancy (607208)
CG7971	**1,722**	437	117	62	70	31	23	171	*NIPBL* (delangin)	Cornelia de Lange syndrome (122470)
Midgut										
CG6295	3	21	**5,461**	4	2	1	1	2	*LIPI* (*PRED5*)	Hypertriglyceridemia, susceptibility to (145750)
CG31636	2	19	**159**	2	1	1	7	1	*c17orf79* (*TTP1*)	Ataxia with isolated vitamin E deficiency (277460)
ε*Try*	2	5	**5,967**	6	6	3	5	3	Proenterokinase	Enterokinase deficiency (226200)
Tubule										
CG3762	441	1,213	4,665	5,497	**6,242**	1,906	354	1,170	Vacuolar proton pump	Renal tubular acidosis with deafness (267300)
ry	80	142	97	49	**770**	2	18	7	Xanthine oxidase	Xanthinuria type I (278300)
Irk3	328	123	4	48	**4,932**	1	1	11	Renal outer-medullary potassium channel; ROMK	Bartter syndrome, antenatal, type 2 (241200)
CG5284	568	451	292	361	**1,334**	540	75	357	Chloride channel CLCN5	Dent disease 1, nephrolithiasis, X-linked (30008)
CG17752	1	1	2	1	**6,341**	1	2	3	*SLC22A12* (*URAT1*)	Hypouricemia, renal (220150)
Ovary										
Pi3K92E	283	171	182	188	108	**395**	92	185	PI 3-kinase, alpha	Ovarian cancer (604370)
Testis										
bol	104	19	1	9	6	1	**1,778**	82	*DAZL* (*SPYGLA*)	Spermatogenic failure, susceptibility to (601486)
Ubp64E	294	221	238	514	334	196	**752**	367	*Drosophila* fat facets-related	Azoospermia (415000)
CG17150	2	2	5	3	5	2	**391**	6	*ITLN1* (*HL1*)	Kartagener syndrome (244400)

The entire Homophila database was downloaded from http://superfly.ucsd.edu/homophila/ and merged with the array database. We then selected probe sets against genes with Homophila annotations that showed signal enrichment in tissues analogous to those implicated in the human disease. Boldface indicates the maximum signal for each gene.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

class and function.

Cryptochrome, important in circadian function[15,16], is functionally significant in peripheral tissues as well as in brain[17]. However, the signals in the Malpighian tubule and hindgut are much higher than in brain, supporting reports that in this simple organism, tissues run autonomous clocks[18]. *Fasciclin2*, implicated in neural functions from axonal pathfinding to short-term memory, is predominantly expressed in the tubule. Given the importance of homoeostasis to small organisms, the tubule enrichment could help to explain the low viability of some *fas2* alleles, and it serves as a reminder of the non-neuronal significance of cell junctions[19].

The large family of odorant-binding proteins provides specificity to olfactory sensing in insects and, accordingly, has been described in the context of olfaction or gustation in the head[20,21]. However, one of the odorant-binding protein genes (*Obp56d*) shows expression not just in the head but also at extremely high levels in the hindgut, providing an unexpected opening for olfactory research. Another gene (*Obp56e*) is expressed in just the head and accessory gland, whereas *Obp56f* and *Obp22a* are expressed exclusively in the accessory gland but not in the head (data not shown).

Several developmental genes are reused in very unexpected places in the adult: for example, *twin of eyegone*, important in visual and thorax development[22,23], is virtually specific to the male accessory gland in the adult. Similarly, *ventral nerve system defective* (*vnd*)[24] is reused very specifically in the adult midgut. The sex determination gene *doublesex* is primarily an epithelial gene in the adult, suggesting that adult homeostasis may be sexually dimorphic.

Kelch, a component of the ring canal[25], is expressed at much higher levels in brain and head than ovary. This is appropriate for a homolog of gigaxonin, a human gene associated with neuropathy[26]. *Kelch* may thus provide a better model for human disease than mere sequence similarity might predict.

Organotypic disease models

Can this analogy be taken further? Are there other human genetic disease loci with *Drosophila* homologs that are expressed preferentially in tissues functionally analogous to the human target, increasing confidence in the validity of the *Drosophila* gene as a model? Combining the FlyAtlas data set with the Homophila database[27] provides a unique opportunity to test the idea. In fact, cursory examination uncovers multiple loci with this persuasive combination of evidence (**Table 5**).

Although these genes are plausible candidates for investigation, *Drosophila* is neither a mammal nor a vertebrate. Does a physiological approach to *Drosophila* yield insights into human gene function that are not obtainable in humans or mice? In fact, our understanding of many human genes (and not just those involved in development) has emerged from *Drosophila*. Behavioral screens in *Drosophila* identified the role of cyclic AMP in memory[28], identified the first clock genes[29] and found the prototype for the *Shaker* family of potassium channels[30]. Of course, neurogenetics is an accepted area of functional similarity between human and fly, but the approach can extend to other tissues that are much less explored. The roles of the Malpighian tubule, for

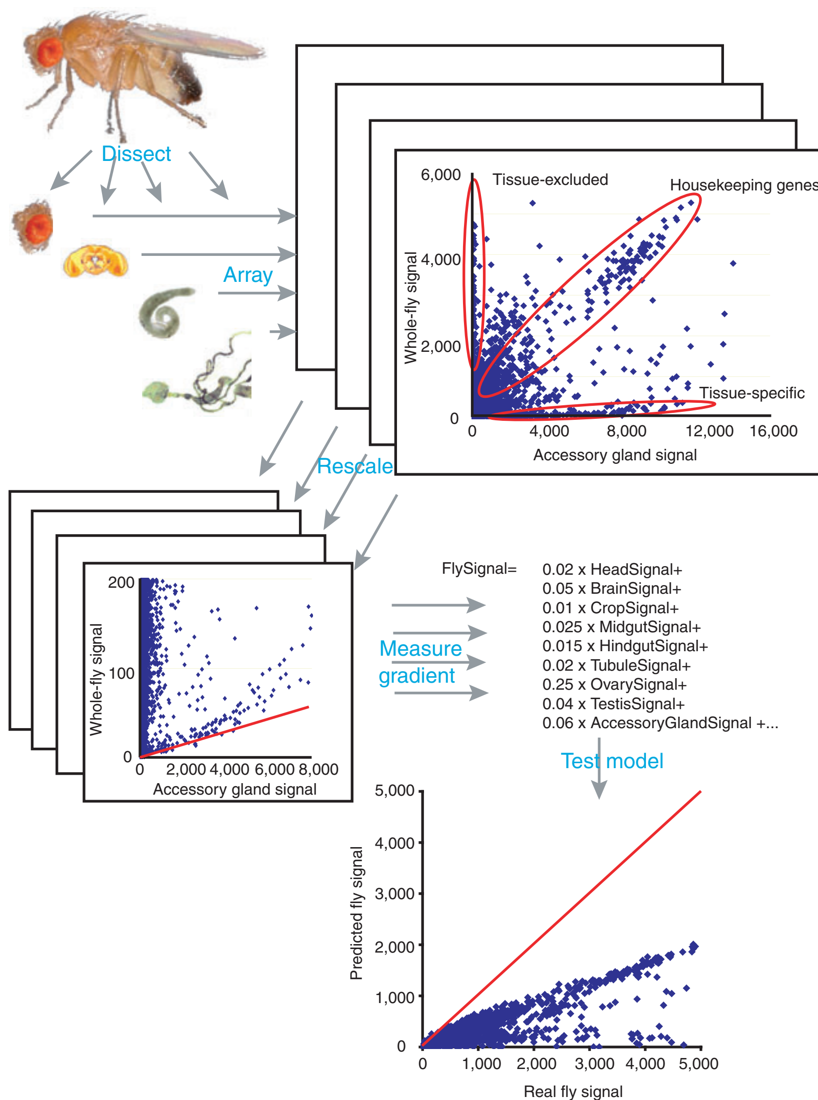

Figure 3 Calculating the equation of the fly. Tissues are dissected from the adult fly, and their signals are plotted against the whole fly. Although all degrees of specificity are represented, most genes tend to be generally expressed, tissue excluded or tissue specific. By rescaling and plotting the lowest gradient that runs through experimental data, coefficients for the whole-fly array simulation can be estimated. These can then be tested iteratively against real-world results as new tissues are added.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

example, are analogous to those of the human kidney, liver and innate immune system[31]. As well as generating a primary urine, it is loaded with cytochrome P450s and glutathione transferases that imply a major role in detoxification[32], and it is capable of mounting a robust immune response independent of the cardinal insect immune tissue, the fat body[33,34]. The tubule data set contains genes with homologs well-known in the mammalian renal literature. The V-ATPase proton pump, for example, has been shown to be essential[35], and mutations in genes encoding all 13 V-ATPase subunits implicated in plasma membrane transport have resulted in a renal phenotype in *Drosophila*[36], presaging the discovery of a renal phenotype in humans with a similar mutation[37]. All three inward-rectifier K^+ channel genes (analogous to the ROMK channel associated with a form of Bartter syndrome) are strongly enriched in the tubule and hindgut, implying an epithelial transport role rather than the neural role that is usually sought for channels in *Drosophila*[38]. Perhaps most famously, *rosy*, the second mutant ever identified in *Drosophila*[39], exactly recapitulates the symptoms of xanthinuria type I, the human disease associated with xanthine oxidase mutations[40,41]. Thus, there are ample scope and informative phenotypes to explore renal function in this simple organism[42].

The other tissues listed in **Table 4** also show persuasive similarities, but the physiological study of other tissues is at an earlier stage. However, the data clearly illustrate the importance of a tissue-centered view in functional genomics, both *a priori* in the design of transcriptomic or proteomic experiments and *post hoc* in the central goal of functional genomics: the elucidation of the major functions of all the genes encoded by a genome. To this end, FlyAtlas provides a valuable tool to focus efforts directed at both goals.

We suggest an important and salutary meta-analysis of this data set. Many *Drosophila* microarray studies have used whole-organism RNA samples (from embryos, larvae or adults). Clearly, given the tissue specificity of expression we describe above, this is less than ideal. However, it is possible to demonstrate exactly how bad it is by modeling the whole-organism transcriptome from the transcriptomes of its constituent tissues—that is, drawing up an 'equation of the fly'. In principle, the whole-fly transcriptome is made up of the sum of the transcriptomes of each tissue, multiplied by a coefficient that is effectively the fraction of organismal mRNA contributed by that tissue:

$$\text{Signal}_{\text{Fly}} = \sum_{\text{Tissues}} Signal_n \times c_n$$

A set of hybridizations with tissue-specific mRNA preparations allows the coefficients to be calculated. In principle, for each gene with truly tissue-specific expression, the whole-fly signal represents the contribution of that tissue to the whole fly. However, this approach depends on the accurate identification of truly tissue-specific genes. A simpler approach (**Fig. 3**) is to plot tissue signal against whole-fly signal for each tissue and take the lowest gradient through the experimental points; this represents the best estimate of the coefficient for that tissue. The resulting equation can be tested against the real whole-fly transcriptome. As each tissue is added, the average abundance of widely expressed genes approaches the theoretical 1:1 line, and highly tissue-specific genes disappear from the lower region of the plot as they reach the trend line. As can be seen, with just these nine tiny tissues, just under half of the real signal can be modeled. The major missing contribution is probably from the thoracic and abdominal carcass (cuticle, wings, legs and muscle).

The implication of the equation for whole-organism transcriptomics is severe; it shows that most tissues each contribute less than 5% to the organismal signal. So, considering a typical array signal of 100, and taking a 50% change as a threshold for detection, a gene expressed in a single tissue would have to change its signal by at least 1,000 for the change to be detected in a whole-fly hybridization. Thus, a whole-fly array is capable of detecting only orchestrated changes in widely expressed genes

or truly exceptional changes in genes with more restricted expression patterns. Whole-fly arrays are thus a trade-off between convenience and sensitivity, reporting a severely attenuated subset of the true changes occurring between experimental groups. As the body plan and relative tissue sizes of the adult are largely established by the late embryo (with the conspicuous exception of those adult tissues generated by proliferation of the imaginal discs), the argument can be extended to whole-organism arrays or proteomic studies of other life stages.

Is there a strategy that would maximize the chances of detecting a spatially restricted change, at least in the adult, for finite extra effort? From the data available at present, perhaps a five-way split would provide a working compromise: head, alimentary canal, male and female genitalia (separately) and carcass. 'Whole-head' would report on head, brain and fat body, with perhaps a threefold drop in sensitivity for the individual tissues; similarly, 'alimentary canal' would report on midgut, tubule and hindgut, with similar performance. Gonads would combine the distinctive transcriptomes of (for example) testes and accessory glands, without swamping either, and the carcass would report on cuticle, muscle and associated tissues. The increased cost could then be balanced against hugely increased authority and utility of the data.

Thus, FlyAtlas hints at some exciting new directions for *Drosophila* functional genomics. Of course, this approach needs continual refinement; as new tissues and developmental stages are added to FlyAtlas, the authority of the data set will increase. Ideally, an atlas should include every domain that can be identified by a GAL4 enhancer trap, perhaps using poly-A binding protein technology to allow selective purification of mRNAs[43]. Similarly, as the computational transcriptome of the fly is successively refined and the knowledge rolled out in new generations of array (such as tiling or exon arrays[44]), it will be useful to update the data iteratively. Overall, though, the emphasis on tissues and organ systems will help to redirect and focus efforts toward organism-level systems biology in this useful model organism.

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

ACKNOWLEDGMENTS
This work was funded by the UK's Biotechnology and Biological Sciences Research Council (BBSRC). We are most grateful to S. Terhzaz, P. Cabrero and L. Aitchison for their guidance in dissections and S.-A. Davies and S. Goodwin for their critical reading of the manuscript.

COMPETING INTERESTS STATEMENT
The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

1. Kaiser, K. From gene to phenotype in *Drosophila* and other organisms. *Bioessays* **12**, 297–301 (1990).
2. Adams, M.D. & Sekelsky, J.J. From sequence to phenotype: reverse genetics in *Drosophila melanogaster. Nat. Rev. Genet.* **3**, 189–198 (2002).
3. Orkin, S.H. Reverse genetics and human disease. *Cell* **47**, 845–850 (1986).
4. Ruddle, F.H. Reverse genetics as a means of understanding and treating genetic disease. *Adv. Neurol.* **35**, 239–242 (1982).
5. Bargmann, C.I. High-throughput reverse genetics: RNAi screens in *Caenorhabditis elegans. Genome Biol.* **2**, REVIEWS1005 (2001).
6. Brown, S.D.M. & Peters, J. Combining mutagenesis and genomics in the mouse–closing the phenotype gap. *Trends Genet.* **12**, 433–435 (1996).
7. Bullard, D.C. Mind the phenotype gap. *Trends Mol. Med.* **7**, 537–538 (2001).
8. Dow, J.A.T. The *Drosophila* phenotype gap - and how to close it. *Brief. Funct. Genomic. Proteomic.* **2**, 121–127 (2003).
9. Adams, M.D. *et al.* The genome sequence of *Drosophila melanogaster. Science* **287**, 2185–2195 (2000).
10. Arbeitman, M.N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster. Science* **297**, 2270–2275 (2002).
11. Andrews, J. *et al.* Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res.* **10**, 2030–2043 (2000).
12. Krogh, A. The progress of physiology. *Am. J. Physiol.* **90**, 243–251 (1929).
13. Brand, A.H. & Perrimon, N. Targetted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
14. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).
15. Stanewsky, R. *et al.* The cryb mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila. Cell* **95**, 681–692 (1998).
16. Emery, P., So, W.V., Kaneko, M., Hall, J.C. & Rosbash, M. CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* **95**, 669–679 (1998).
17. Ivanchenko, M., Stanewsky, R. & Giebultowicz, J.M. Circadian photoreception in *Drosophila*: functions of cryptochrome in peripheral and central clocks. *J. Biol. Rhythms* **16**, 205–215 (2001).
18. Giebultowicz, J.M., Stanewsky, R., Hall, J.C. & Hege, D.M. Transplanted *Drosophila* excretory tubules maintain circadian clock cycling out of phase with the host. *Curr. Biol.* **10**, 107–110 (2000).
19. Carthew, R.W. Adhesion proteins and the control of cell shape. *Curr. Opin. Genet. Dev.* **15**, 358–363 (2005).
20. Graham, L.A. & Davies, P.L. The odorant-binding proteins of *Drosophila melanogaster*: annotation and characterization of a divergent gene family. *Gene* **292**, 43–55 (2002).
21. Hekmat-Scafe, D.S., Scafe, C.R., McKinney, A.J. & Tanouye, M.A. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster. Genome Res.* **12**, 1357–1369 (2002).
22. Dominguez, M., Ferres-Marco, D., Gutierrez-Avino, F.J., Speicher, S.A. & Beneyto, M. Growth and specification of the eye are controlled independently by Eyegone and Eyeless in *Drosophila melanogaster. Nat. Genet.* **36**, 31–39 (2004).
23. Aldaz, S., Morata, G. & Azpiazu, N. The Pax-homeobox gene eyegone is involved in the subdivision of the thorax of *Drosophila. Development* **130**, 4473–4482 (2003).
24. Jimenez, F. *et al.* vnd, a gene required for early neurogenesis of *Drosophila*, encodes a homeodomain protein. *EMBO J.* **14**, 3487–3495 (1995).
25. Robinson, D.N. & Cooley, L. *Drosophila* kelch is an oligomeric ring canal actin organizer. *J. Cell Biol.* **138**, 799–810 (1997).
26. Bomont, P. *et al.* The gene encoding gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is mutated in giant axonal neuropathy. *Nat. Genet.* **26**, 370–374 (2000).
27. Chien, S., Reiter, L.T., Bier, E. & Gribskov, M. Homophila: human disease gene cognates in *Drosophila. Nucleic Acids Res.* **30**, 149–151 (2002).
28. Byers, D., Davis, R.L. & Kiger, J.A. Defect in cyclic AMP phosphodiesterase due to the *dunce* mutation of learning in *Drosophila melanogaster. Nature* **289**, 79–81 (1981).
29. Konopka, R.J. & Benzer, S. Clock mutants of *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA* **68**, 2112–2116 (1971).
30. Salkoff, L. & Wyman, R. Genetic modification of potassium channels in *Drosophila Shaker* mutants. *Nature* **293**, 228–230 (1981).
31. Dow, J.A.T. & Davies, S.A. The Malpighian tubule: rapid insights from post-genomic biology. *J. Insect Physiol.* **52**, 365–378 (2006).
32. Yang, J. *et al.* A *Drosophila* systems approach to xenobiotic metabolism. *Physiol. Genomics* published online 8 May 2007 (doi:10.1152/physiolgenomics.00018.2007).
33. McGettigan, J. *et al.* Insect renal tubules constitute a cell-autonomous immune system that protects the organism against bacterial infection. *Insect Biochem. Mol. Biol.* **35**, 741–754 (2005).
34. Kaneko, T. *et al.* PGRP-LC and PGRP-LE have essential yet distinct functions in the *Drosophila* immune response to monomeric DAP-type peptidoglycan. *Nat. Immunol.* **7**, 715–723 (2006).
35. Davies, S.A. *et al.* Analysis and inactivation of *vha55*, the gene encoding the V-ATPase B-subunit in *Drosophila melanogaster,* reveals a larval lethal phenotype. *J. Biol. Chem.* **271**, 30677–30684 (1996).
36. Allan, A.K., Du, J., Davies, S.A. & Dow, J.A.T. Genome-wide survey of V-ATPase genes in *Drosophila* reveals a conserved renal phenotype for lethal alleles. *Physiol. Genomics* **22**, 128–138 (2005).
37. Karet, F.E. *et al.* Mutations in the gene encoding B_1 subunit of H^+-ATPase cause renal tubular acidosis with sensorineural deafness. *Nat. Genet.* **21**, 84–90 (1999).
38. Evans, J.M., Allan, A.K., Davies, S.A. & Dow, J.A.T. Sulphonylurea sensitivity and enriched expression implicate inward rectifier K^+ channels in *Drosophila melanogaster* renal function. *J. Exp. Biol.* **208**, 3771–3783 (2005).
39. Glassman, E. & Mitchell, H.K. Mutants of *Drosophila melanogaster* deficient in xanthine dehydrogenase. *Genetics* **44**, 153–162 (1959).
40. Dent, C.E. & Philpot, G.R. Xanthinuria: an inborn error of metabolism. *Lancet* **263**, 182–185 (1954).
41. Wang, J. *et al.* Function-informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol.* **5**, R69 (2004).
42. Dow, J.A.T. & Davies, S.A. Integrative physiology and functional genomics of epithelial function in a genetic model organism. *Physiol. Rev.* **83**, 687–729 (2003).
43. Yang, Z., Edenberg, H.J. & Davis, R.L. Isolation of mRNA from specific tissues of *Drosophila* by mRNA tagging. *Nucleic Acids Res.* **33**, e148 (2005).
44. Manak, J.R. *et al.* Biological function of unannotated transcription during the early development of *Drosophila melanogaster. Nat. Genet.* **38**, 1151–1158 (2006).