

Function-informed transcriptome analysis of *Drosophila* renal tubule

Jing Wang*, Laura Kean*, Jingli Yang*, Adrian K Allan*, Shireen A Davies*, Pawel Herzyk[†] and Julian AT Dow*

Addresses: *Division of Molecular Genetics, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G11 6NU, UK. [†]Sir Henry Wellcome Functional Genomics Facility, University of Glasgow, Glasgow G12 8QQ, UK.

Correspondence: Julian AT Dow. E-mail: j.a.t.dow@bio.gla.ac.uk

Published: 26 August 2004

Genome Biology 2004, 5:R69

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/9/R69>

Received: 14 May 2004

Revised: 25 June 2004

Accepted: 23 July 2004

© 2004 Wang et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Comprehensive, tissue-specific, microarray analysis is a potent tool for the identification of tightly defined expression patterns that might be missed in whole-organism scans. We applied such an analysis to *Drosophila melanogaster* Malpighian (renal) tubule, a defined differentiated tissue.

Results: The transcriptome of the *D. melanogaster* Malpighian tubule is highly reproducible and significantly different from that obtained from whole-organism arrays. More than 200 genes are more than 10-fold enriched and over 1,000 are significantly enriched. Of the top 200 genes, only 18 have previously been named, and only 45% have even estimates of function. In addition, 30 transcription factors, not previously implicated in tubule development, are shown to be enriched in adult tubule, and their expression patterns respect precisely the domains and cell types previously identified by enhancer trapping. Of *Drosophila* genes with close human disease homologs, 50 are enriched threefold or more, and eight enriched 10-fold or more, in tubule. Intriguingly, several of these diseases have human renal phenotypes, implying close conservation of renal function across 400 million years of divergent evolution.

Conclusions: From those genes that are identifiable, a radically new view of the function of the tubule, emphasizing solute transport rather than fluid secretion, can be obtained. The results illustrate the phenotype gap: historically, the effort expended on a model organism has tended to concentrate on a relatively small set of processes, rather than on the spread of genes in the genome.

Background

Microarrays allow the interrogation of the transcriptome, the set of genes transcribed in a particular cell type under a particular condition [1]. Arrays are particularly potent tools when their coverage is relatively comprehensive, based on a

completed and well annotated genome, such as that of *Drosophila* [2]. Commonly, they are used in time series, for example of development, of life events such as metamorphosis [3], of rhythmic behavior [4] or of responses to environment, such as aging or starvation [5,6]. In *Drosophila*, arrays

are frequently used for whole-organism studies, but in multicellular organisms the ease of experimentation must be balanced against two potential problems: sensitivity and opposing changes. In the first case, even large changes in gene expression in a small tissue will not significantly influence the overall levels in the whole organism; in the second, changes in opposite directions in roughly balanced populations of cells (for example, the sharpening of expression patterns of pair-rule genes) will cancel out at an organismal scale. It is thus vital to resolve gene expression not only over time but also over space. In practice, this means looking at gene expression in defined cell types and tissues as well as in the whole organism. Our assumption is that the expression of many putative genes will go undetected until such tissue-specific studies are performed [7] - with obvious consequences for post-genomics - and we illustrate this point in this paper.

We applied Affymetrix arrays in the context of a defined tissue with extensive physiological characterization, the Malpighian (renal) tubule of *Drosophila melanogaster*. The tubule is a valuable model for studies of both epithelial development and function. Developmentally, the tissue is derived from two distinct origins: an ectodermal outpushing of the hindgut and subsequent invasion (late in embryogenesis) by mesodermal cells [8]. Tubule morphology is very precisely and reproducibly specified; in the tiny tissue of 150 cells, there are altogether six cell types and six regions, specified to single-cell precision [9]. The transport processes that underlie fluid production in the tubule are known in extraordinary detail for so small an organism [10-12]. The dual origin of the cell types is reflected by dual roles for the ectodermal principal cells and mesodermal stellate cells in the mature tubule; the principal cell is specialized for active transport of cations, whereas the stellate cell appears to control passive shunt conductance [11,13,14]. Cell signaling pathways are also understood in considerable detail: several peptide hormones that act on tubule have been identified [15-17], and the second messengers cyclic AMP, cyclic GMP, calcium and nitric oxide have all been shown to have distinct roles in each tubule cell type [10,18-20].

This wealth of physiological knowledge provides a framework for the analysis of the results, and thus - unusually in genetic model organisms - a reality check on the usefulness of the experiment.

Results

The principle of the experiment was to compare the transcriptome of 7-day adult *Drosophila melanogaster* Malpighian (renal) tubules, for which defined state there is a wealth of physiological data, with matched whole flies. As described in Materials and methods, data were analyzed by Affymetrix MAS 5.0 software, or by dChip, or dChip and Significance Analysis of Microarrays (SAM) software. Both methods of identifying differentially expressed genes from dChip-

normalized data gave virtually the same results. Indeed, SAM analysis followed by further filtering produced 1,465 differentially expressed genes compared to 1,455 genes identified within filtering by dChip alone. Furthermore, the latter list is indeed a subset of the former one. For that reason we report only the list generated by dChip in comparison with MAS data.

Both MAS and dChip/SAM gave comparable views of the data, despite the radically different approaches to analysis. It has been shown that the average absolute log ratios between replicate arrays calculated with dChip are significantly lower than one calculated with Affymetrix software (Li and Wong [21]). This bias affecting fold-change calculations is the price of the increased precision that manifests itself in reduced variance, and consequently in the increased sensitivity of identification of differentially expressed genes. Nonetheless, the rank correlation is good (Spearman's $r = 0.6$, $p < 0.0001$). Taking genes called as significant by both systems, MAS5 'up' call or dChip t -test p -value of 0.01, and narrowing the list by setting an arbitrary cutoff of twofold enrichment and minimum mean difference of 100, MAS5 reported 683 genes and dChip reported 671. Furthermore, the dChip-reported genes overlap with 77% of MAS5-reported genes and this number increases to 91% if only the top 500 MAS5-reported genes are considered. Our confidence in the quality of the dataset is thus high. For simplicity, and because the two analyses produce concordant results, further analysis is restricted to the MAS5 results.

The full microarray data have been deposited in ArrayExpress [22]. The fly versus fly and tubule versus tubule samples were extremely consistent, despite the technical difficulty in obtaining the latter (30,000 tubules were dissected in total). In contrast, there was wide divergence between fly and tubule samples (Figure 1). Although a common set of housekeeping genes showed comparable abundance, there was a large set of genes enriched in the fly sample, and a smaller set of genes strongly enriched in the tubule sample. In detail, of 13,966 array entries, 6,613 genes were called 'present' in all five fly samples, compared with 3,873 in tubules. A total of 3,566 genes were present in both fly and tubule: 3,047 in fly only and 307 in tubule only. This illustrates the point that whole-organism views of gene expression are not necessarily helpful in reflecting gene-expression levels in individual tissues. The microarray data are summarized in Tables 1,2.

Validation of the microarray

Four genes were selected from each of three fly tubule expression classes: very highly enriched; uniformly expressed; and very highly depleted. The expression of each gene was verified by quantitative reverse transcription PCR (RT-PCR) and the data are presented in Table 3. The agreement between Affymetrix microarray and quantitative PCR determination is good, further increasing our confidence in the robustness of the dataset, and in the approximate correspondence between

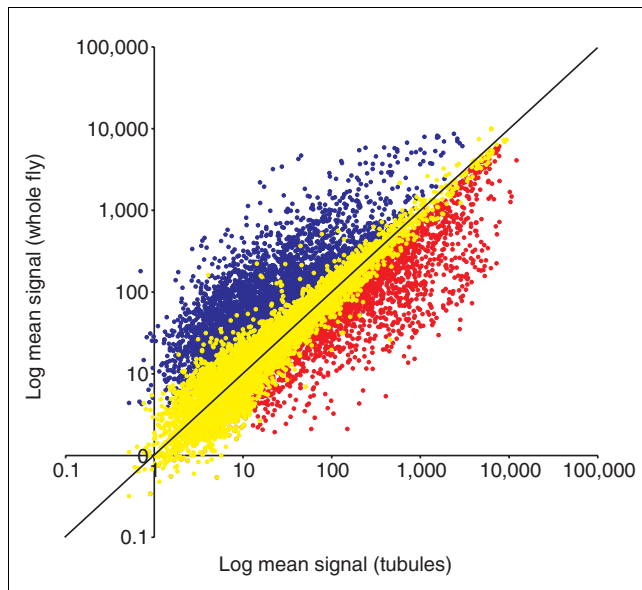


Figure 1
Scatterplot of mean whole fly vs tubule signal intensities. Genes called as significantly enriched in tubule compared with fly by MAS 5.0 are in red, those significantly depleted in blue, and those not significantly different in yellow.

signal and RNA abundance as a population average. It should be noted that the absolute sizes of the ratios are quite variable; this is a property of dividing a large number by a very small one. Nonetheless, genes scored as enriched or depleted on the arrays are invariably similarly scored by quantitative RT-PCR (QRT-PCR).

These data can also be used to validate the use of the normalized Affymetrix signal as a semi-quantitative measure of RNA abundance (Table 1). If the QRT-PCR dataset of Table 3 is normalized against corresponding signals for *rp49* (generally taken to be a ubiquitous gene with invariant expression levels in *Drosophila*), and compared with the globally normalized Affymetrix signal, the agreement is seen to be excellent (Figure 2), with a Spearman's r of 0.83 ($p < 0.0001$). With appropriate caution, the normalized Affymetrix signal can thus be taken as a reasonable estimate of expression levels between genes.

Table 1 shows the top 20 genes listed by mean Affymetrix signal intensity. Although this is only a semi-quantitative measure of transcript abundance, the identities of the known genes in the lists are illuminating, and persuade us that the approach has some informal value. Specifically, mRNAs for ribosomal proteins dominate the list, and transporters are conspicuous in the balance. For example, the V-ATPase that energizes transport by tubules is represented by one gene (other subunits are also abundant, but just below the cutoff for Table 1). The α -subunit of the Na^+ , K^+ ATPase is also highly abundant: this is more surprising, and is discussed

below. Two organic cation transporters are also very abundant. Alcohol dehydrogenase, long known to be expressed in tubules [23,24], is also a major transcript. There are also surprises: the most abundant signal is for metallothionein A. This is entirely consistent with our classical understanding of tubule function: it has long been known as a route for metal sequestration and excretion [25-30]. However, in the entire literature on Malpighian tubules, we are not aware of a physiological investigation of the role of metallothionein, other than documentation of expression [31,32]. The microarray results can thus potentially direct and inform future research.

Table 2 lists the 53 tubule-enriched genes that are enriched at least 25-fold, in comparison with the whole fly (the full list is provided as an additional data file). The conspicuous feature of these data is the extent to which tubule transcripts differ from any previously published profile. When comparing fly with tubule, there is a large set of genes that are downregulated and another large set of genes that are upregulated in tubule. The extent of the upregulation is also remarkable: the top gene is 99-fold enriched; the top 10 at least 50-fold enriched; and the top 100 at least 16-fold enriched in tubule compared to fly. The standard errors are also extremely low, meaning that we can be very confident (by two separate statistical measures) of the genes called significantly enriched in tubule.

The phenotype gap

Another prominent feature of the signal data in Table 1 is the relatively large fraction of novel genes (those for which there is not even a computer prediction of function) at the top of the list. Indeed, five of the top 10 genes by signal intensity are completely novel - that is, there are no known orthologs - and should provide tantalizing insights into tubule function. The 'phenotype gap' [33,34] is a key problem in functional genomics; that is, the genetic models preferred for genomics are historically not the organisms selected by physiologists. This can lead to a log-jam in reverse genetics, which depends critically on a wide range of phenotypes to identify effects of the mutation of target genes [12]. It has recently become possible to quantify the phenotype gap [35]. The present dataset elegantly exposes the phenotype gap in *Drosophila*, and shows that the tubule phenotype may go some way to closing it. Around 20% of *Drosophila* genes have been studied in sufficient detail to attract names (beyond the standard 'CG' notation for computer-annotated genes). Figure 3 shows that the fraction of anonymous genes in the tubule-enriched list is far higher than would be expected. That is, previous work has tended to overlook these genes. Conversely, because it is possible to perform detailed physiological analysis in tubules, it is possible to close the phenotype gap for these genes. There is a general implication from these data: that functional genomics, in *Drosophila* and other species, will rely increasingly on the study of specific tissues, as it is only in this context that expression of genes will be either measurable or explicable.

Table 1**Most abundant genes in tubule, sorted by normalized Affymetrix signal strength**

Gene	Signal	Enrichment	Function
<i>MtnA</i>	12,114 ± 581	3.0 ± 0.0	Cu-binding
<i>CG7874</i>	10,672 ± 518	7.4 ± 0.4	
<i>CG14292</i>	10,392 ± 572	8.4 ± 0.5	
<i>CG3168</i>	10,199 ± 459	6.2 ± 0.3	Transporter
<i>RpS25</i>	9,368 ± 276	1.3 ± 0.0	Small-subunit cytosol ribosomal protein
<i>Adh</i>	8,895 ± 395	1.3 ± 0.0	Alcohol dehydrogenase; EC 1.1.1.1
<i>RpS20</i>	8,720 ± 226	1.2 ± 0.0	Small-subunit cytosol ribosomal protein
<i>CG13315</i>	7,818 ± 370	3.9 ± 0.6	
<i>CG14245</i>	7,767 ± 305	13.4 ± 2.3	
<i>RpL27A</i>	7,757 ± 198	1.3 ± 0.0	Large-subunit cytosol ribosomal protein
<i>CG18282</i>	7,711 ± 160	1.7 ± 0.0	
<i>RpL18A</i>	7,514 ± 200	1.4 ± 0.0	Large-subunit cytosol ribosomal protein
<i>RpL14</i>	7,483 ± 209	1.3 ± 0.0	Large-subunit cytosol ribosomal protein
<i>RpP2</i>	7,481 ± 283	1.3 ± 0.1	Cytosolic ribosomal protein
<i>CG6726</i>	7,307 ± 244	14.4 ± 0.5	Peptidase
<i>RpL23a</i>	7,284 ± 254	1.2 ± 0.1	Large-subunit cytosol ribosomal protein
<i>CG4046</i>	7,250 ± 165	1.1 ± 0.1	Structural protein of ribosome
<i>CG7084</i>	7,211 ± 329	36.8 ± 6.5	Transporter
<i>RpL3</i>	7,179 ± 105	1.4 ± 0.1	Large-subunit cytosol ribosomal protein
<i>CG9914</i>	7,088 ± 466	12.0 ± 1.4	Enzyme
<i>CG3203</i>	7,024 ± 219	1.3 ± 0.1	L17-like
<i>CG6846</i>	6,989 ± 177	1.3 ± 0.1	Structural protein of ribosome
<i>blw</i>	6,890 ± 142	1.7 ± 0.0	ATP synthase alpha subunit
<i>BcDNA:GH08860</i>	6,742 ± 278	5.0 ± 0.3	Enzyme
<i>RpS3</i>	6,709 ± 240	1.3 ± 0.1	DNA-(apurinic or apyrimidinic site) lyase
<i>CG5827</i>	6,603 ± 169	1.3 ± 0.1	Structural protein of ribosome
<i>CG15697</i>	6,543 ± 174	1.3 ± 0.1	Structural protein of ribosome
<i>RpS9</i>	6,502 ± 171	1.2 ± 0.0	Small-subunit cytosol ribosomal protein
<i>Rack1</i>	6,463 ± 105	1.3 ± 0.0	Protein kinase C binding protein
<i>vha26</i>	6,416 ± 190	3.1 ± 0.3	V-ATPase E subunit
<i>Ser99Da</i>	6,305 ± 2100	0.6 ± 0.2	Serine carboxypeptidase
<i>Ser99Db</i>	6,300 ± 2119	0.6 ± 0.2	Serine-type endopeptidase
<i>CG1883</i>	6,258 ± 172	1.2 ± 0.1	Structural protein of ribosome
<i>RpL32</i>	6,251 ± 217	1.3 ± 0.1	Large-subunit cytosol ribosomal protein
<i>Atpalpha</i>	6,240 ± 151	4.2 ± 0.1	Na, K-ATPase alpha subunit
<i>CG3270</i>	6,234 ± 167	32.3 ± 2.6	Sarcosine oxidase
<i>RpS26</i>	6,080 ± 151	1.3 ± 0.1	Small-subunit cytosol ribosomal protein
<i>sop</i>	6,070 ± 157	1.1 ± 0.0	Small-subunit cytosol ribosomal protein
<i>RpL7</i>	6,060 ± 113	1.2 ± 0.0	Large-subunit cytosol ribosomal protein
<i>CG3321</i>	6,034 ± 122	1.6 ± 0.0	Enzyme
<i>eIF-4a</i>	6,027 ± 270	1.9 ± 0.1	
<i>CG8857</i>	5,977 ± 309	1.4 ± 0.1	Structural protein of ribosome
<i>oho23B</i>	5,940 ± 176	1.3 ± 0.1	Ribosomal protein
<i>CG3762</i>	5,874 ± 79	4.2 ± 0.1	
<i>CG9091</i>	5,850 ± 281	1.2 ± 0.1	Structural protein of ribosome
<i>vha16</i>	5,845 ± 215	2.6 ± 0.1	V-ATPase c subunit
<i>CG18323</i>	5,820 ± 201	1.5 ± 0.1	

Table 2

Genes enriched more than 25-fold in tubules		
Gene	Product	MAS enrichment
CG13365		98.9
CG14957		95.9
CG13905		85.2
CG13836		80.6
<i>Irk3</i>	Potassium channel protein-like	80.3
CG14963		55.4
CG3014		54.0
CG13161		53.8
CG17043		49.9
CG18095		47.8
CG13656		45.5
CG13311		43.5
CG17817		40.9
CG9434		40.6
CG17522	Glutathione transferase	39.5
CG15359		38.7
CG7084	Organic cation transporter	36.8
CG8028	Monocarboxylate transporter-like	36.6
CG8951	Sodium-dependent multivitamin transporter-like	35.8
CG3690		34.8
CG15406	Sugar transporter	34.5
CG14293		33.5
CG17028		33.4
CG3285	Sugar transporter-like	33.0
CG3270		32.3
<i>scarlet</i>	ATP-binding cassette (ABC) transporter	32.3
CG6529	Sugar transporter-like	32.1
CG2680	4-nitrophenylphosphatase-like	31.2
CG8620		30.5
CG15279	Cation amino-acid symporter	30.1
CG9509		29.7
CG14539		29.3
CG3382	Organic anion transporter	29.3
CG6602		29.3
CG5361	Alkaline phosphatase-like	29.2
CG8957	Iodide symporter-like	29.1
CG10006		29.0
CG15155		28.9
CG10226	ATP-binding cassette transporter	28.3
CG2196	Sodium iodide symporter	27.7
CG16762		27.6
CG14195		27.4
CG8125	Aryldialkylphosphatase	27.4
CG7881	Sodium phosphate cotransporter	27.1
CG8934	Sodium iodide symporter-like	27.1
CG7402	N-acetylgalactosamine-4-sulfatase-like	26.9
<i>NaPi-T</i>	Na phosphate cotransporter	26.8

Table 2 (Continued)

Genes enriched more than 25-fold in tubules		
CG8791	Sodium phosphate cotransporter	26.8
CG8776	Cytochrome b561-like	26.6
CG3212		26.6
CG14857	Organic cation transporter-like	26.4
CG8932	Sodium-dependent multivitamin transporter-like	25.9
<i>Cyp6a18</i>	Cytochrome P450, CYP6A18	25.5

Reconciling array data with function

Many microarray experiments merely classify enriched genes to their Gene Ontology families. However, the uniquely detailed physiological data available on the Malpighian tubule allows a much more informative approach. The dataset can be validated by inspection, based on known molecular functions in the tissue and new functions can be inferred from abundant or enriched transcripts in the dataset. As the array is relatively comprehensive (corresponding to the 13,500 genes in release 1 of the *Gadfly* annotation), the results are also relatively authoritative.

Organic solutes

The housekeeping ribosomal transcripts vanish from the enrichment list (Table 2), which is now dominated by transporters. Intriguingly, these are not for the V-ATPase that is considered to dominate active transport by the tubule, but for organic and inorganic solutes. There is a range of broad-specificity transporters - for organic cations, anions, monocarboxylic acids, amino acids and multivitamins. There are also multiple inorganic anion co-transporters for phosphate and iodide. Most are not only very highly enriched, but also highly abundant. In more detail, the results are remarkable (Table 4). Nearly every class of transporter is represented, and almost all of these have at least one representative that is both abundant and enriched, implying a very specific renal role; indeed, this table contains the genes with the highest average enrichments of any class, frequently more than 30-fold. Some transporters have been documented implicitly as having a tubule role; many of the classical *Drosophila* eye-color mutants also have an effect on tubule color, and have since been shown to encode genes for transport of eye-pigment precursors [12,36]. These genes now turn out to be both abundant and enriched; among the ABC transporters are *scarlet* and *white*, and among the monocarboxylic acid transporters is *CG12286*, which we have recently argued to correspond to karmoisin, a probable kynurenine transporter [37]. Glucose and other sugar transporters are consistently abundant and enriched, implying that sugar transport is a major (and previously unsuspected) role of the tubule. Inorganic transporters are also included in the table; there are also copper and zinc transporters, which is consistent with electron-probe X-ray

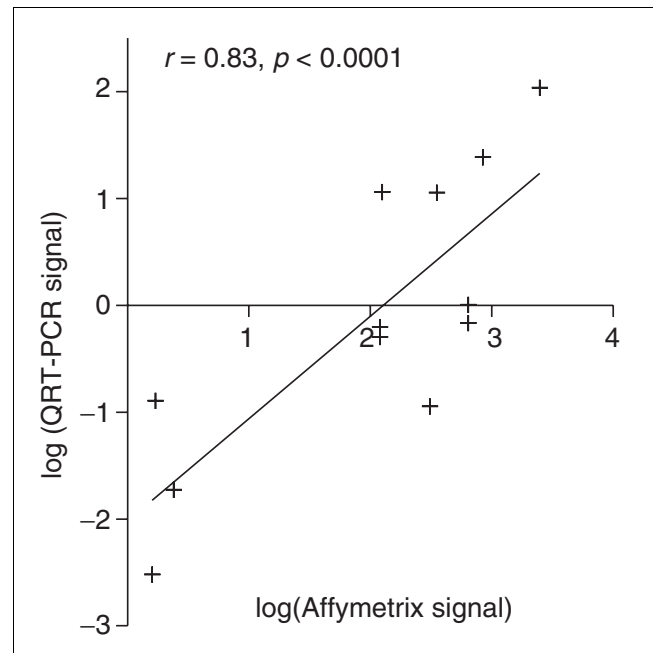
Table 3**Validation of array data by QRT-PCR**

Gene	MAS enrichment	SAM enrichment	QRT-PCR enrichment
Highly enriched			
<i>CG13665</i>	98.9	8.7	9.0
<i>CG14957</i>	95.9	21.9	23.8
<i>CG13905</i>	22.6	17.4	11.0
<i>CG13836</i>	80.6	30.1	11.7
Evenly expressed			
<i>CG17737</i>	1.0	0.9	0.74
<i>CG10731</i>	1.0	1.1	0.68
<i>CG8327</i>	1.0	0.8	1.2
<i>Arp66</i>	1.0	1.1	0.47
Highly depleted			
<i>CG13421</i>	0.00	0.067	0.19
<i>CG12408</i>	0.01	0.11	0.14
<i>Act88F</i>	0.01	0.14	0.03
<i>CG15575</i>	0.01	0.082	0.008

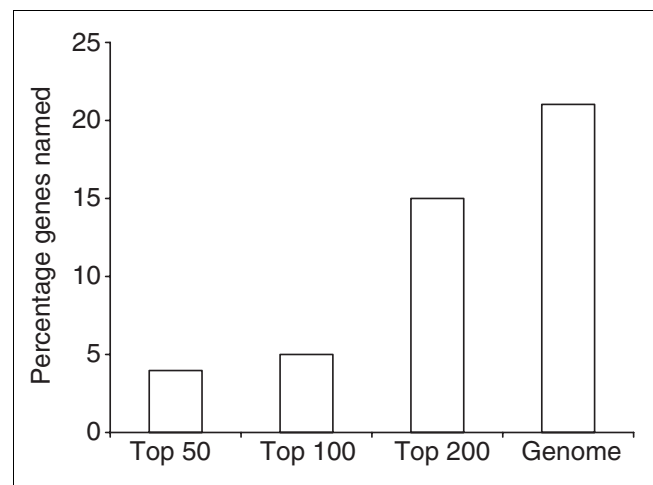
Enrichment in tubule mRNA compared to whole fly mRNA, computed from the microarray dataset with MAS 5.0 or SAM (see text), were compared with real values obtained by QRT-PCR. Four separate fly and tubule samples were run with primers for each gene, and for *rp49*, a ribosomal gene generally considered to be invariant. RNA quantities were calculated, and the gene:*rp49* ratio calculated for each sample pair. Tubule enrichment was calculated as the $(\text{gene:rp49})_{\text{tubule}} / (\text{gene:rp49})_{\text{fly}}$.

microanalysis data that heavy metals accumulate in tubule concretions [38,39], and with the extreme abundance of metallothionein A (Table 1).

As well as specific transporters, the tubule is enriched for several families of broad-specificity transporters (organic anion and cation transporters, multivitamin transporters, ABC multidrug transporters and an oligopeptide transporter). When combined these would be capable of excreting a huge majority of organic solutes. These results invite a substantial revision of our interpretation of the role of the tubule. Classically, it is considered to be the tissue that excretes waste material, both metabolites and xenobiotics, and provides the first stage of osmoregulation. However, nearly all work on insect tubules in the last half-century has focused on the ionic basis of fluid secretion and its control, as these are easily measured experimentally. Although there have been sporadic reports on the active transport of organic solutes such as dyes [40-42], the historical view was of a relatively leaky epithelium, with a paracellular default pathway for those solutes not recognized by specific transporters. While consistent with the more classical view of the tubule, our results also suggest that

**Figure 2**

Semi-quantitative inter-gene comparison is possible using Affymetrix signal. The 24 QRT-PCR results underlying Table 3 were normalized against *rp49*, and plotted against the Affymetrix signal globally normalized as in MAS 5.0. Spearman's *r* was calculated, and significance of the correlation assessed (one-tailed), using Graphpad Prism 3.0.

**Figure 3**

The phenotype gap. Genes enriched in tubules are historically under-researched. The percentage of genes with explicit names (other than automatic CG annotations) is shown for the entire genome, and for the top 50, 100 and 200 genes (as judged by fold enrichment) from the tubule dataset.

the insect is emulating a leaky epithelium to produce the primary urine by incorporating a vast array of broad-specificity active transporters in the plasma membranes of what is electrically rather a tight epithelium. Indeed, this interpretation

Table 4

Transporters sorted by class		
Gene/class	Signal	Enrichment
ATP-binding cassette (ABC) transporter (6/46)		
<i>st</i>	1,521 ± 34	32 ± 2.8
<i>CG10226</i>	290 ± 25	28 ± 3.4
<i>CG9270</i>	422 ± 21	21 ± 2.7
<i>w</i>	798 ± 53	10 ± 1.4
<i>bw</i>	18 ± 2	4 ± 1.2
<i>CG17338</i>	72 ± 6	3 ± 0.2
Cationic amino-acid transporter (1/5)		
<i>CG7255</i>	308 ± 34	7 ± 0.8
Copper transporter (1/6)		
<i>CG7459</i>	374 ± 6	5 ± 0.6
Monocarboxylate transporter (4/14)		
<i>CG8028</i>	2,567 ± 82	37 ± 2.1
<i>CG8468</i>	1,377 ± 67	10 ± 0.7
<i>CG8389</i>	698 ± 38	4 ± 0.2
<i>CG12286 (kar)</i>	550 ± 15	3 ± 0.1
Multidrug efflux transporter (1/6)		
<i>CG8054 (now CG30344)</i>	1,366 ± 68	6 ± 0.4
Pyrimidine-sugar transporter of Golgi (1/1)		
<i>CG3874 (frc)</i>	877 ± 40	5 ± 0.3
Oligopeptide transporter (1/3)		
<i>CG9444</i>	517 ± 12	10 ± 1.2
Organic anion transporter (3/5)		
<i>CG3382</i>	1,076 ± 56	29 ± 3.3
<i>CG3380</i>	3,385 ± 126	24 ± 1.6
<i>CG6417</i>	678 ± 90	9 ± 2.4
Organic cation transporter (11/21)		
<i>CG7084</i>	7,211 ± 329	37 ± 6.5
<i>CG14857</i>	472 ± 13	26 ± 5.5
<i>CG17751</i>	1,331 ± 34	25 ± 4.2
<i>CG16727</i>	3,152 ± 200	23 ± 3.2
<i>CG17752</i>	4,847 ± 37	21 ± 2.1
<i>CG14856</i>	36 ± 5	7 ± 2.4
<i>CG3168</i>	10,199 ± 459	6 ± 0.3
<i>CG6231</i>	269 ± 30	5 ± 1.0
<i>CG7342</i>	20 ± 2	5 ± 1.5
<i>CG8654</i>	274 ± 29	4 ± 0.6

Table 4 (Continued)

Transporters sorted by class		
Reduced folate transporter (2/3)		
<i>CG14694</i>	584 ± 22	13 ± 1.6
<i>CG6574</i>	190 ± 8	4 ± 0.3
Sodium bicarbonate cotransporter (1/1)		
<i>CG4675 (Ndae1)</i>	531 ± 34	5 ± 0.5
Sodium-dependent inorganic phosphate cotransporter (1 / 20)		
<i>NaPi-T</i>	1,430 ± 428	27 ± 2.3
Sodium-dependent multivitamin transporter (4/5)		
<i>CG8951 (now CG31090)</i>	1,363 ± 30	36 ± 3.9
<i>CG8932</i>	2,106 ± 130	26 ± 1.4
<i>CG8451</i>	365 ± 10	4 ± 0.4
<i>CG10879 (now CG31668)</i>	6 ± 1	3 ± 0.7
Glucose transporter (3/17)		
<i>CG7882</i>	4,951 ± 171	16 ± 0.8
<i>CG8249</i>	302 ± 12	6 ± 1.0
<i>Glut1</i>	342 ± 24	3 ± 0.2
Sugar transporter (7/7)		
<i>CG15406</i>	5,322 ± 186	35 ± 2.8
<i>CG3285</i>	1,405 ± 55	33 ± 1.3
<i>CG6529 (now CG31272)</i>	3,774 ± 131	32 ± 4.8
<i>CG15407</i>	840 ± 44	25 ± 2.1
<i>CG14606</i>	1,210 ± 56	22 ± 2.0
<i>CG15408</i>	3,333 ± 194	21 ± 1.7
<i>CG8837</i>	1,277 ± 88	19 ± 2.9
Zinc transporter (4/6)		
<i>BG:DS07295.1 (now CG3994)</i>	3,608 ± 91	10 ± 1.0
<i>CG4334</i>	378 ± 19	5 ± 0.4
<i>CG17723</i>	919 ± 59	4 ± 0.3
<i>CG5130</i>	104 ± 10	4 ± 0.6

For brevity, only family members enriched by more than threefold are shown. For each grouping, the numbers in parentheses refer to the number of genes enriched in tubule, compared to the total number of such genes in the *Drosophila* genome, as classified by Gene Ontology. Where original gene names have been superseded by later annotations of the *Drosophila* genes, the new names are shown in parentheses.

is consistent with other independent data: the intercellular junctions in tubule are known to be of the pleated stellate variety, the invertebrate equivalent of tight junctions [43]; and, like salivary glands, tubule cells are known to be highly polytene [44-47] or even binucleate [48], adaptations that maximize the size of cells and thus maximize their area/circumference ratios.

Table 5**V-ATPase genes that are enriched in tubule**

Subunit	Copy number	Genes	Affymetrix reference	Signal	Enrichment
<i>V₁</i> sector					
A	3	<i>vha68-1</i> (CG12403)	I42380_at	9 ± 2	0.5 ± 0.1
		<i>vha68-2</i> (CG3762)	I46305_at	5,874 ± 79	4.2 ± 0.1
		<i>vha68-3</i> (CG5075)	I46306_at	2 ± 0	0.04 ± 0.02
B	1	<i>vha55</i> (CG17369)	I53041_at	2,304 ± 74	2.7 ± 0.1
SFD (H)	1	<i>vhaSFD</i> (CG17332)	I44191_at	2,671 ± 66	4.4 ± 0.2
C	1	<i>vha44</i> (CG8048)	I53422_at	1,400 ± 74	3.5 ± 0.1
D	3	<i>vha36-1</i> (CG8186)	I52480_at	2,846 ± 154	4.5 ± 0.4
		<i>vha36-2</i> (<u>CG13167</u>)	I47073_at	2 ± 0.4	0.1 ± 0.0
		<u>CG8310</u>	I44407_at	29 ± 4	0.6 ± 0.09
E	1	<i>vha26</i> (CG1088)	I51930_at	6,416 ± 190	3.1 ± 0.3
F	2	<i>vha14-1</i> (CG8210)	I43625_at	3,722 ± 105	3.2 ± 0.2
		<i>vha14-2</i> (CG1076)	I49368_at	5.6 ± 1.6	1.5 ± 1.1
G	1	<i>vha13</i> (<u>CG6213</u>)	I44156_at	2,952 ± 68	3.3 ± 0.1
<i>V₀</i> sector					
A	5	<i>vha100-1</i> (CG1709)	I53997_at	155 ± 8	0.8 ± 0.0
		<i>vha100-2</i> (CG7679, CG18617)	I42661_at	3,718 ± 157	5.4 ± 0.3
		<i>vha100-3</i> (CG30329)	not on array		
		<u>CG12602</u>	I46249_at	306 ± 26	1.3 ± 0.1
c	5	<i>vha100-4</i> (<u>CG7678</u>)	I41662_at	66 ± 3	0.24 ± 0.04
		<i>vha16</i> (CG3161)	I41528_at	5,845 ± 215	2.6 ± 0.1
		<i>vha16-2</i> (CG32089)/ <i>vha16-3</i> (CG32090)	I48578_at	32 ± 7	1.4 ± 0.22
		<i>vha16-4</i> (CG9013)	I47341_at	18 ± 4	1.4 ± 0.6
PPA1 (c")	2	<i>vhaPPA1-1</i> (CG7007)	I42158_at	1,895 ± 79	4.1 ± 0.2
		<i>vhaPPA1-2</i> (CG7026)	I49926_at	57 ± 7	0.9 ± 0.1
M9.7 (e, H)	3	<i>vhaM9.7-1</i> (CG11589)	I54011_at	101 ± 9	1.8 ± 0.0
		<u>CG1268</u>	I48161_at	14 ± 1	0.1 ± 0.0
AC39 (d)	2	<i>vhaM9.7-2</i> (CG7625)	I49187_at	3,101 ± 127	2.9 ± 0.1
		<i>vhaAC39-1</i> (CG2934)	I54279_at	2,082 ± 52	3.4 ± 0.1
		<i>vhaAC39-2</i> (<u>CG4624</u>)	I50428_at	13 ± 2	0.8 ± 0.12

All genes significantly similar to known human or yeast V-ATPase subunits were identified by BLAST search, extending our previously reported annotation of the V-ATPase family [53], by identifying the genes underlined above as V-ATPase subunits. For comparison, enrichment ratios significantly greater than 1 and signals over 1,000 are shown in bold. (*vha16-2* and *vha16-3* are in tandem repeat and share the same Affymetrix oligo set, and so cannot be distinguished here.)

V-ATPases

Physiological analysis of the tubule has concentrated on the secretion of primary urine, and the energizing transporter is a plasma membrane proton pump, the V-ATPase [13,49-51]. This is a large holoenzyme of at least 13 subunits, encoded by 31 *Drosophila* genes [52,53]. V-ATPases have two distinct roles, one carried out at low levels in endomembrane compartments of all eukaryotic cells and the other in the plasma membranes of specialized epithelial cells of both insects and vertebrates [54]. In such cells, the V-ATPases can pack the plasma membrane to such an extent that they resemble semi-

crystalline arrays when observed by electron microscopy [55]. It is clearly of interest to find out which genes contribute to the plasma-membrane role of the V-ATPase, though this would normally involve difficult and tedious generation of selective antibodies capable of distinguishing between very similar proteins. However, the mRNAs for those V-ATPase subunits enriched in epithelia should also be particularly abundant; one could thus predict that at least one gene encoding each V-ATPase subunit should show enrichment in tubule compared with the rest of the fly. This is indeed the case (Table 5): invariably, one gene for each subunit is both

significantly enriched, and far more abundant, than any other gene encoding that subunit. The reason that the enrichment is not higher is probably because the whole-fly samples contain other epithelia, each with enriched V-ATPase, as minor parts of the overall sample.

The array data thus allow a rapid and authoritative prediction to be made on the subunit composition of the plasma membrane V-ATPase. It will be interesting to extend these data to other epithelia in which V-ATPase is known to be functionally significant.

Na⁺, K⁺-ATPase

The role of the classical Na⁺, K⁺-ATPase in tubule is enigmatic. In nearly all animal epithelia, transport is energized by a basolateral Na⁺, K⁺-ATPase, which establishes a sodium gradient that drives secondary transport processes. By contrast, insect epithelia are energized by a proton gradient from the apical V-ATPase [56,57] and, consistent with this, many insect tissues are paradoxically refractory to ouabain, the specific Na⁺, K⁺-ATPase inhibitor [58]. Accordingly, models of insect epithelial function tend not to include the Na⁺, K⁺-ATPase. It is thus interesting to note that both *Atpalpha* and *Nervana 1* (encoding isoforms of the α and β subunits, respectively) are among the most abundant transcripts in tubule (Table 6). Both are about as enriched in tubule as the V-ATPase subunits, but are significantly more abundant (compare Table 5). By contrast, a novel alpha-like subunit (*CG3701*), and both *Nrv2* (the neuronal β -subunit) and other novel β -like subunits are at near-zero levels. As Na⁺, K⁺-ATPase has previously been documented as being particularly abundant in *Drosophila* tubule [59], it may thus be prudent to re-include the Na⁺, K⁺-ATPase as an important part of models of tubule function.

Table 6

Na ⁺ , K ⁺ -ATPase		
Gene	Signal	Enrichment
α -subunit		
<i>Atpalpha</i>	6,240 \pm 151	4.22 \pm 0.05
<i>CG3701</i>	6 \pm 1	0.85 \pm 0.17
β -subunit		
<i>Nrv1</i>	1,924 \pm 71	3.47 \pm 0.21
<i>Nrv2</i>	2 \pm 1	0.09 \pm 0.06
<i>CG11703</i>	7 \pm 2	0.46 \pm 0.18
<i>CG5250</i>	4 \pm 0	0.18 \pm 0.04
<i>CG8663</i>	20 \pm 1	0.1 \pm 0.01

Although the *Drosophila* Na⁺, K⁺-ATPase has classically been thought to be composed of a dimer of *Atpalpha* and either *Nrv1* or *Nrv2*, the other genes here are more similar by BLASTX to the corresponding alpha and beta subunits than any other gene (data not shown). They are thus included in the table as candidate alternative subunits.

Potassium channels

Potassium is actively pumped across the tubule, and the main basolateral entry step is via barium-sensitive potassium channels, both in tubule [50,60,61] and in other V-ATPase-driven insect epithelia [62,63]. Of the ion channels, the potassium channel family is by far the most diverse in all animals: in *Drosophila*, there are at least 28, and in human 255, K⁺-channel genes [64]. Inspection of the potassium channels on the array (Table 7) clearly identifies just four that are expressed at appreciable levels. *Irk3*, *Ir*, *Irk2* and *NCKQ* are all both very abundant and highly enriched in tubule. *Irk3* in particular is 80-fold enriched over the rest of the fly, implying a unique role in tubule. Three of these genes are members of the inward rectifier family of potassium channels: supporting the hypothesis that they are critical for potassium entry, these channels are known to be highly barium-sensitive [65]. An inward rectification of potassium current (meaning that potassium would pass much more easily into the cell than out) would be ideal for a basolateral entry step. Inward rectifier channels normally associate with the sulfonylurea receptor (SUR), an ABC transporter, in order to make functional channels [66,67]. In tubules, *SUR* mRNA is present at extremely low abundance (signal 6, enrichment 0.9 times). However, *CG9270*, a gene with very close similarity to *SUR* (1×10^{-28} by BLASTP) is very abundant in tubule (see Table 4), (signal 422, enrichment 21 times). A second very similar gene, *CG31793* (previously also known as *CG10441* and *CG17338*), is very much less abundant (signal 24, enrichment 0.5). We therefore predict that novel inward rectifiers, formed between *Irk3*, *Ir* or *Ir2* and *CG9270*, may provide the major basolateral K⁺ entry path in tubule. In contrast, the other classes of K⁺ channel, and the Na/K/Cl co-transporter that has been documented in tubule, are all relatively low in both abundance and enrichment.

Chloride and water flux

In a fluid-secreting epithelium, a necessary correlate of the active transport of cations must be the provision of a shunt pathway for anions and a relatively high permeability to water. In *Drosophila* tubules, a hormonally regulated chloride conductance pathway has been shown to occur in the stellate cells, although the molecular correlate of the currents has not been determined. There are three ClC-type chloride channels in the *Drosophila* genome, and RT-PCR has shown that all three are expressed in tubule [12]. The array data present a prime candidate (Table 8). Although all three genes are expressed, only one (*CG6942*) is both very abundant and enriched in tubule (signal 251, enrichment 4). It is thus an obvious candidate partner to provide a shunt pathway for the epithelial V-ATPase.

Water flux through the tubule is also phenomenally fast: each cell can clear its own volume of fluid every 10 seconds [12]. Although traditionally it was thought that only a leaky epithelium could sustain such rates, the identification of aquaporins (AQP) (the predominant members of the major intrinsic pro-

Table 7

Potassium channels and symporters		
Gene	Signal	Enrichment
Potassium channels		
<i>Irk3</i> (CG10369)	2771 ± 145	80.31 ± 7.75
<i>Ir</i> (CG6747)	1302 ± 112	14.19 ± 1.58
<i>Irk2</i> (CG4370)	527 ± 33	5.69 ± 0.24
<i>KCNQ</i> (CG12215)	101 ± 0	6.44 ± 2.31
<i>KCNQ</i> (CG12915)	111 ± 10	2.84 ± 0.46
<i>CG10864</i>	29 ± 7	3.74 ± 1.12
<i>CG32770</i> (CG6952)	5 ± 2	2.6 ± 1.05
<i>elk</i>	5 ± 3	2.23 ± 1.19
<i>CG9361</i>	6 ± 2	2.31 ± 0.84
<i>CG12214</i>	101 ± 11	2.15 ± 0.51
<i>CG7640</i>	12 ± 5	1.48 ± 0.76
<i>eag</i>	8 ± 1	1.59 ± 0.39
<i>Shaker cognate b</i>	6 ± 1	1.38 ± 0.54
<i>CG4450</i>	4 ± 0	1.62 ± 0.28
<i>Shaw</i>	26 ± 4	1.21 ± 0.54
<i>CG1756</i>	15 ± 4	1.31 ± 0.35
<i>Shaker</i>	26 ± 4	1.42 ± 0.23
<i>CG9637</i>	3 ± 1	1.32 ± 0.22
<i>Shal</i>	29 ± 4	1.19 ± 0.29
<i>CG3367</i>	6 ± 1	1.09 ± 0.09
<i>CG8713</i>	41 ± 3	0.9 ± 0.1
<i>Sh</i>	7 ± 3	0.65 ± 0.25
<i>CG9194</i>	8 ± 1	0.77 ± 0.12
<i>CG15655</i>	13 ± 3	0.45 ± 0.13
<i>Ork1</i>	28 ± 2	0.32 ± 0.03
<i>sei</i>	10 ± 2	0.21 ± 0.04
<i>Shab</i>	6 ± 1	0.21 ± 0.04
<i>CG12904</i>	4 ± 1	0.14 ± 0.07
<i>CG17860</i>	5 ± 2	0.1 ± 0.04
<i>Hk</i>	4 ± 1	0.07 ± 0.01
Calcium-activated potassium channels		
<i>CG10706</i>	21 ± 5	1.93 ± 1.06
<i>slo</i>	2 ± 0	0.11 ± 0.02
<i>CG4179</i>	4 ± 1	1.55 ± 0.91
Potassium-dependent sodium-calcium exchangers		
<i>CG14744</i>	39 ± 1	1.32 ± 0.12
<i>CG1090</i>	35 ± 2	0.81 ± 0.13
<i>CG14743</i>	5 ± 1	0.48 ± 0.19
<i>Nckx30C</i>	31 ± 5	0.38 ± 0.05
<i>CG12376</i>	8 ± 2	0.24 ± 0.07
<i>Nckx30C</i>	31 ± 5	0.11 ± 0.05
Sodium/potassium/chloride symporter		
<i>EG:8D8.3</i>	132 ± 6	2.46 ± 0.36
<i>CG10413</i>	185 ± 25	1.75 ± 0.22

Table 7 (Continued)

Potassium channels and symporters		
<i>CG5594</i>	65 ± 5	0.8 ± 0.07
<i>CG2509</i>	60 ± 6	0.42 ± 0.06
<i>CG4357</i>	12 ± 4	0.12 ± 0.04

Table 8

Chloride channels		
Gene	Signal	Enrichment
<i>CG6942</i>	251 ± 9	4 ± 0.29
<i>CG8594</i>	57 ± 5	0.86 ± 0.09
<i>CG5284</i>	100 ± 5	2.2 ± 0.16

These are the three genes with clear similarity to the CIC gene family of vertebrates [12].

Table 9

Aquaporins and other major intrinsic proteins		
Gene	Signal	Enrichment
<i>CG4019</i>	1666 ± 167	2.7 ± 0.3
<i>CG17664</i>	705 ± 91	7.9 ± 0.9
<i>DRIP</i>	318 ± 16	3.6 ± 0.4
<i>CG7777</i>	243 ± 11	0.6 ± 0.06
<i>CG12251</i> (AQP)	22 ± 3	0.5 ± 0.04
<i>CG5398</i>	8 ± 1	0.2 ± 0.05
<i>bib</i>	2 ± 1	1.1 ± 0.3

tein (MIP) family) as major water channels in both animals and plants [68] provides an obvious counter-explanation. There is physiological and molecular data for the presence of aquaporins in *Drosophila* tubule [69], and AQP-like immunoreactivity has been demonstrated in stellate cells [12]. Table 9 shows that only four of the seven AQP/MIP genes are abundant, and only three enriched. One can thus tentatively assign an organism-wide role to *CG7777* (signal 243, enrichment 0.6), but tubule-specific roles to *CG4019*, *CG17664* and *DRIP*. In particular, *CG17664*, is both highly abundant and very highly enriched (signal 705, enrichment 7.9).

Control of the tubule

The hormonal control of fluid secretion is well understood. The major urine-producing region of the tubule is the main segment [70], and is composed of two major cell types, principal and stellate cells [9,13,71]. Active cation transport in the

Table 10**Receptors called as upregulated in tubule, with enrichments more than threefold**

Gene	Signal	Enrichment
CG3212	85 ± 11	27 ± 11
CG17415 (calcitonin-like)	633 ± 48	17 ± 2
CG17084	288 ± 27	14 ± 2
CG1147 (neuropeptide Y-like)	34 ± 2	13 ± 8
CG14575 (CapaR)	311 ± 24	11 ± 1
CG7431 (octopamine-like)	40 ± 4	8.5 ± 0.9
CG12414 (nAChR α @ 80B)	9 ± 3	8 ± 3.6
CG7589 (ligand-gated Cl channel)	564 ± 35	7 ± 0.9
CG12370 (diuretic hormone-like)	203 ± 17	6.7 ± 9
CG15556	221 ± 12	6.4 ± 0.5
CG11340 (glycine-gated channel-like)	143 ± 8	5.0 ± 0.9
CG14593 (bombesin)	59 ± 13	5 ± 2
CG6390 (insulin-like growth factor)	85 ± 8	4.3 ± 0.6
CG8222 (Pvr, vascular endothelial growth factor-like)	294 ± 26	4.2 ± 0.5
CG6536	42 ± 5	4 ± 1.7
nAcR α	24 ± 4	4 ± 1.5
CG7404 (steroid-like)	239 ± 21	3.5 ± 0.4
CG10626 (LkR)	142 ± 7	2.9 ± 0.4

principal cell is stimulated by the hormones calcitonin-like peptide and corticotrophin releasing factor (CRF)-like peptide, both of which act through cyclic AMP (cAMP). Another peptide family, the CAPA peptides, act through intracellular calcium to stimulate nitric oxide synthase and thus raise cyclic GMP (cGMP), an unusual autocrine role for nitric oxide [20,72]. In the stellate cell, the chloride shunt conductance is activated by leucokinin [17,73], and a role for tyramine as an extracellular signal has also been proposed [74]. So far, the CAPA and leucokinin receptors have been identified [75,76]; both are prominent among the receptors enriched in tubule (Table 10). The CAPA receptor appears much more highly enriched in tubule than the leucokinin receptor, which is consistent with our understanding of each: the tubule is the only known target of CAPA, whereas leucokinin receptors are widely distributed in the adult gut, gonad and nervous system [75].

There are many other receptors that are reasonably abundant and enriched in tubule. As well as candidate receptors for calcitonin-like and other neuropeptides, there are two glycine/GABA-like receptors that might be expected to form ligand-gated chloride channels, together with good matches to vascular endothelial growth factor-like, insulin-like and bombesin-like receptors. The localization of, ligands for, and functional roles of these receptors will be of great interest. It

should be noted in this context that all hormones characterized so far act on one of the two main cell types in the principal section of the tubule. There are, however, six genetically defined cell types and six regions in the adult tubule [9], and it is likely that there will at least be ligands acting on the initial segment to stimulate calcium excretion, and others acting to regulate reabsorption by the lower tubule. If any of these receptors maps to these regions, they would be prime candidates for such roles.

Overall, the main surprise from these data is the sheer range of candidate ligands that could be inferred; this more than doubles the size of the endocrine repertoire so far postulated for insect tubules.

On a more general level, it is possible to trace out the key genes in all three intracellular signaling pathways that have been studied in detail in *Drosophila* tubule (Table 11). The results for signaling genes tend not to be as clear-cut as for transporters, as many are rather widely distributed, and so do not show enrichment, and many do not require high standing levels of protein (and implicitly mRNA) to achieve their effects. Nonetheless, it is possible to identify genes that are at least present, and frequently enriched, in tubule. For the cAMP pathway, it is possible to identify adenylate cyclases, protein kinase A catalytic and regulatory subunits, and a phosphodiesterase (*dunce*). For cGMP, there are both soluble and membrane guanylate cyclases, implying that the tubules may produce cGMP directly in response to novel ligands, as has recently been suggested [77]. Both *Drosophila* genes encoding protein kinase G are expressed in tubule, and one is highly enriched. This is consistent with the renal phenotype observed both in *foraging* mutants [78], and in tubules in which protein kinase G is overexpressed [79]. There is also a PDE11-like phosphodiesterase. For calcium, two genes for phospholipase C, one for calmodulin, and one for protein kinase C and for calcium/calmodulin-dependent protein kinase are apparent. There are also a number of interesting modulatory or anchoring proteins, such as 14-3-3 zeta, A-kinase anchoring proteins, and receptors for activated C-kinase (*Rack1*).

How is the tubule specified?

The developmental origin of the tubule has been reviewed in detail [80-82]. Briefly, four unique 'tip cells', specified by a cascade of neurogenic genes, control cell division in four out-pushings (anlagen) of the hindgut, to form the Malpighian tubules. Late in embryogenesis the tubule is invaded by mesodermal cells, which intercalate between the future principal cells, and which then differentiate to form stellate cells [8]. In the adult, there are known to be at least six cell types and six tubule regions [9]. These regions are specified to great precision, and it is clear that each cell in the tubule has a precise positional identity. How does this identity persist throughout the lifetime of the animal? Presumably, combinations of transcription factors interact to provide both regional and cell-

Table 11**Major genes of the cAMP, cGMP and calcium signaling pathways**

Function	Gene name	Signal	Enrichment	Comments
cAMP				
Adenylate cyclase	<i>rutabaga</i>	121 ± 12	1.4 ± .2	
	<i>Ac78C</i>	44 ± 5	7.2 ± 1.6	
	<i>Ac13E</i>	106 ± 4	4.1 ± 0.5	
Protein kinase A	<i>Pka-C3</i>	88 ± 9	1.7 ± 0.2	Catalytic subunit
	<i>Pka-R1</i>	183 ± 13	1.2 ± 0.1	Regulatory subunit
PDE	<i>dunce</i>	147 ± 6	3.9 ± 0.6	cAMP-specific
Calcium				
CamKinase	<i>Caki</i>	112 ± 10	1.7 ± 0.2	
Phospholipase C	<i>Small wing</i>	46 ± 6	1.1 ± 0.2	
	<i>Plc21C</i>	58 ± 5	1.1 ± 0.1	
Calcium release channels	<i>Itpr83A</i>	11 ± 2	1.2 ± 0.2	InsP ₃ receptor
Calmodulin	<i>Calmodulin</i>	1,019 ± 57	0.9 ± 0.06	
Protein kinase C	<i>Pkc98E</i>	217 ± 15	1.7 ± 0.2	
cGMP				
Guanylate cyclase	<i>CG14885</i>	13 ± 3	6 ± 2.5	Probably soluble beta subunit
	<i>Gyc76C</i>	410 ± 23	2.9 ± 0.4	Membrane form
	<i>CG4224</i>	23 ± 4	0.8 ± 0.2	Membrane form
	<i>CG9873</i>	137 ± 5	2.0 ± 0.7	Membrane form
	<i>Gycbeta100B</i>	20 ± 4	0.8 ± 0.1	Cytoplasmic, beta subunit
	<i>CG5719</i>	9 ± 3	3.5 ± 1.4	Membrane form
PDE	<i>CG10231</i>	182 ± 4	3.7 ± 0.6	cGMP-specific, PDE1 I-like
Protein kinase G	<i>foraging</i>	91 ± 2	0.3 ± 0.01	
	<i>Pkg21D</i>	448 ± 20	15.7 ± 2.3	
Serine/threonine protein phosphatases				
	<i>Cgl7746</i>	258 ± 32	4.3 ± 0.6	PPA-2C like
	<i>puckered</i>	228 ± 11	2.5 ± 0.2	Multifunctional
	<i>twins</i>	209 ± 11	2.0 ± 0.1	PPA-2A like
	<i>Pp2A-29B</i>	738 ± 28	1.9 ± 0.2	PPA-2A like
	<i>Microtubule star</i>	997 ± 46	1.3 ± 0.1	PPA-2A like
	<i>Pp1-87B</i>	318 ± 17	1.1 ± 0	PPA-I like
	<i>Pp1alpha-96A</i>	332 ± 8	1.1 ± 0	PPA-I like
Accessory proteins, associated with anchoring, cellular localization or modulation of signaling				
	<i>Akap550</i>	136 ± 6	2 ± 0.3	A-kinase anchor protein
	<i>AKAP200</i>	414 ± 30	0.35 ± 0.02	
	<i>14-3-3-zeta</i>	1,789 ± 42	2.6 ± 0.2	Diacylglycerol-activated PKC inhibitor
	<i>CG32812</i>	42 ± 4	2.7 ± 0.4	Calcineurin
	<i>Rack1</i>	6,463 ± 105	1.3 ± 0	Receptor for activated C-kinase

type coordinates and, after early establishment, these combinations must persist into adulthood. The microarray data allow the identification of transcription factors that are either

highly abundant or highly enriched in tubule. Although this is by no means a complete list of transcription factors that are of importance to tubules, it is a good starting point. Further-

Table 12**Transcription factors and DNA-binding proteins that are abundant or enriched in tubule**

Gene	Signal	Enrichment
<i>CG10278</i>	175 ± 7	24.1 ± 11.5
<i>CG5093</i>	50 ± 4	19.3 ± 6.2
<i>pnt</i>	63 ± 5	17.5 ± 4.8
<i>CG2779</i>	5771 ± 317	16.8 ± 0.7
<i>Ptx1</i>	183 ± 8	12.7 ± 2.2
<i>Ets21C</i>	51 ± 17	9.8 ± 3.1
<i>CG4548</i>	91 ± 4	8.8 ± 4.3
<i>HLH4C</i>	6 ± 1	7.7 ± 6.9
<i>fkf</i>	266 ± 26	7.2 ± 1.1
<i>hth</i>	162 ± 13	7.2 ± 0.7
<i>CG4566</i>	17 ± 2	7.1 ± 4.2
<i>bowl</i>	71 ± 5	7.1 ± 0.7
<i>CG4037</i>	5 ± 1	6.7 ± 2.4
<i>tap</i>	5 ± 1	6.0 ± 3.0
<i>CG6913</i>	5 ± 2	6.0 ± 5.5
<i>CG3950</i>	287 ± 21	5.4 ± 0.9
<i>Awh</i>	21 ± 4	4.8 ± 1.4
<i>CG1162</i>	8 ± 1	4.7 ± 2.1
<i>ct</i>	145 ± 12	4.6 ± 0.8
<i>CG14202</i>	10 ± 1	4.6 ± 1.5
<i>tsh (ae)</i>	65 ± 5	4.6 ± 0.8
<i>CG9952</i>	45 ± 11	4.5 ± 0.6
<i>sv</i>	16 ± 2	4.3 ± 1.8
<i>fd59A</i>	11 ± 3	4.3 ± 1.7
<i>CG11914</i>	31 ± 4	4.2 ± 1.7
<i>slp2</i>	4 ± 2	4.1 ± 3.1
<i>Lim3</i>	13 ± 3	4.0 ± 1.1
<i>CG6419</i>	18 ± 3	4.0 ± 0.4
<i>Tis11</i>	337 ± 17	3.9 ± 0.6
<i>ny</i>	27 ± 4	3.9 ± 1.1

more, there are enhancer trap or reporter gene constructs available for many transcription factors. Accordingly, the top transcription factors and DNA-binding proteins were identified from the array dataset (Table 12).

Some of these transcription factors are already known to be present in tubule, and their presence is confirmed: *cut*, which is known to be required for development of, and expressed in adult Malpighian tubules [83]; and *forkhead* and *homothorax*, both implicated by expression or mutational analysis to be involved in tubule development [84,85]. *Teashirt*, which has recently been shown to be stellate-cell specific in the late embryo [8], is also present in the adult, with fairly high enrichment (4.6 times).

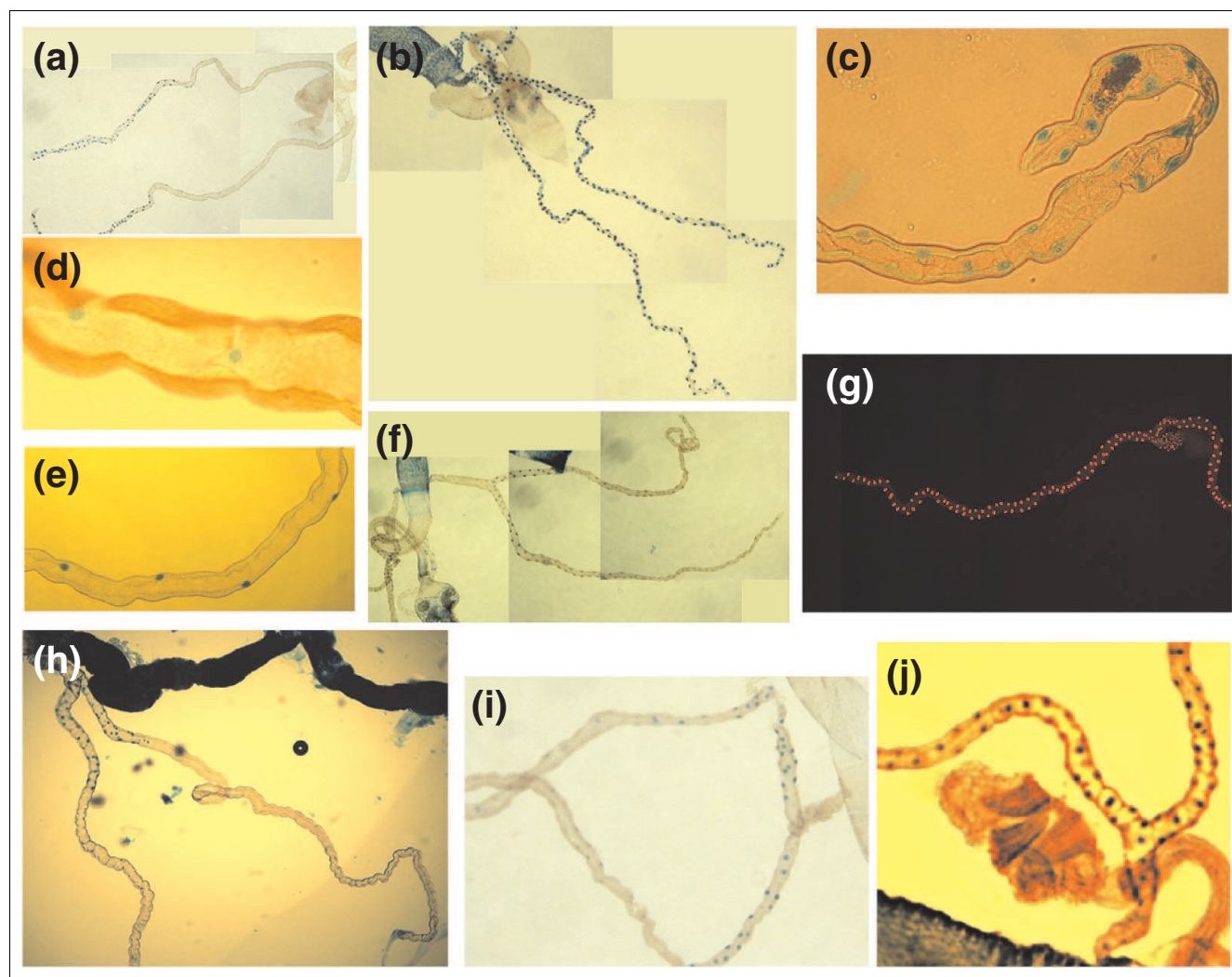
The array results also implicate a further set of transcription factor genes (*ETS21C*, *CG4548*, *bowl*, *sequoia*, *tap*, *CG1162*, *pnt*, *shaven*, *forkhead domain 59A*, *sloppy paired 2*, *lim3*) as important in adult. Significantly, these mainly encode transcription factors implicated in development of the nervous system (another ectodermal tissue), so their reuse in the adult tubule is not too surprising. Once the binding sites for these factors are known, it will be interesting to model gene expression in different tubule regions.

As transcription factors have been studied experimentally in some detail, they are relatively well represented by enhancer trap and other *in vivo* construct lines. Although individual lines do not necessarily represent the complete expression pattern of their cognate genes, a collection of such lines can provide a rapid first validation of a gene list (Table 12). Accordingly, representative reporter gene lines were ordered from the Bloomington Stock Center [86], and their adult staining patterns in tubule and gut are shown in Figure 4. The results are exciting: most lines showed patterned staining in tubule that is consistent with our original genetically derived map of the tubule [9]. For example, *homothorax* marks out the initial, main and transitional segments of the tubule, whereas *CG7417* marks the complementary lower tubule domain. The latter line is widely used as a highly specific mushroom body *GAL4* driver line in brain, and it is interesting that the two known lower tubule *GAL4* driver lines (*c507* and *c232*) are both insertions in *alkaline phosphatase 4*, a gene which is only expressed in lower tubule and the ellipsoid bodies of brain (next to the mushroom bodies) [87]. There is also a cell-type-specific transcription factor: *corto* is found only in stellate cells. Several other transcription factors show ubiquitous, rather than patterned, expression in the tubule, but this is nonetheless consistent with their identification in the microarray dataset.

Another interesting aspect of the data in Table 12 is the number of anonymous CG genes implicated in tubule function. These genes have been annotated as transcription factors because of DNA-binding domains, for example, but have not been characterized functionally. The epithelial phenotype gap is thus evident even in this most intensely studied group of genes.

Exceptions to the rule

The whole premise of microarray work is that an abundant or enriched signal indicates the importance of a gene product in a particular context. This hypothesis is normally both untested and unchallenged. The unusual depth of functional understanding of the tubule allows a more rigorous appraisal. In fact, the majority of the genes implicated in tubule function are found well up the list. There are, however, several conspicuous exceptions (Table 13). The calcium channels *trp* and *trpl* are normally considered to be eye-specific, and have an essential role in phototransduction [88-90]. It is thus not surprising to find both genes almost at the bottom of the gene

**Figure 4**

Expression patterns in tubules of some of the transcription factor genes indicated by the microarray data as being expressed in tubules. **(a)** *homothorax* (*hth*⁰⁵⁷⁴⁵), principal and stellate cells of initial and transitional segments only; **(b)** *polyhomeotic proximal* (*ph-p*), all cells of tubule, and midgut; **(c)** *pointed* (*pnt*¹²⁷⁷), principal and stellate cells of initial and transitional segments only; **(d)** *corto* (*corto*^{07128b}), stellate cells only; **(e)** *teashirt* (*tsh*⁰⁴³¹⁹, a kind gift of H. Skaer), stellate cells only; **(f)** *bunched* (*bnc*⁰⁰²⁵⁵), principal cells, whole tubule; **(g)** *cut* (immunocytochemistry, antibody a kind gift of Jan lab), whole tubule, principal cells only; **(h)** *CG7417* (*CG7417*^{201Y}), lower tubule (and midgut - strong); **(i)** *arc* (*arc*^{k11011b}), lower tubule, not ureter; **(j)** *Stat92E*⁰⁶³⁴⁶, all tubule cells and midgut.

list. We have shown, however, that fluid secretion is severely compromised by mutations in either gene. Similarly, nitric oxide synthase (NOS) is a major signal transducer in tubule [20,72]. Nonetheless, all three genes are within the 'bottom' 20 of the whole array, with signals that are barely detectable and significant depletion compared with the whole fly. This is a cautionary example: while abundant or enriched signals can be taken as reliable indicators of functional significance, the converse is not necessarily true.

The tubule and human disease

Consequent to the demonstration of the phenotype gap, there are some intriguing, abundant and enriched genes which by virtue of their non-uniform expression, are likely to be

important in (and best studied in) tubule. A systematic approach was taken by combining the tubule-enriched gene list with the homophila database of *Drosophila* genes with known human disease homologs. The results (Table 14) show the 50 human diseases with *Drosophila* homologs that are upregulated at least threefold in tubules. Intriguingly, several of these genes have human kidney phenotypes. Some are extremely well studied: for example, *rosy* (one of the first *Drosophila* mutations recorded) encodes xanthine oxidase, and mutation in either human or fly produces severe nephrolithiasis with concomitant distortion of tubules (reviewed in [12]). The distension of tubules is remarkable (Figure 5). In both species, lethal effects can be ameliorated by a high-water, low-purine diet. Other diseases, although less well doc-

Table 13**Genes with known significance to tubule function, but very low abundance/enrichment scores**

Gene name	Signal	Enrichment
<i>NOS</i>	1 ± 0	0.2 ± 0.04
<i>trpl</i>	9 ± 2	0.03 ± 0.01
<i>trp</i>	3 ± 1	0.02 ± 0.01

umented, have plausible renal phenotypes: for example, antenatal Bartter syndrome, a severe salt-wasting renal disease, associated with mutations in the ROMK channel (homolog *ir*); Dent disease, caused by mutation in *CLC5* (homolog *CG5284*); proximal renal tubular acidosis, caused by mutation in the NDAE co-transport (homolog *ndae1*); nephropathic cystinosis, caused by mutation in a lysosomal cystine transporter (homolog *CG17119*); mucopolysaccharidosis type IV, caused by mutation in galactosamine-6-sulphatase, an enzyme enriched in both human and fly kidney (homolog *CG7402*). Overall, there is a clear message that human and fly renal function may be relatively similar over quite a wide range of properties.

The tubule phenotype may also prove highly informative for other genes implicated in disease. Recently, a small 10 kDa protein, bc10, was shown to be downregulated in the transition from early-stage to invasive bladder carcinoma [91]. The normal function of this protein is not yet established, but its homolog (*bc10*) is highly abundant (893 ± 50) and moderately enriched (1.9 ± 0.09) in tubule, and a P-element insertion within the gene P{GT1}BG02443, is available from stock centers.

This comparative approach can be extended to non-human species. For example, *CG4928* represents an abundant and enriched transcript (3,778, 13 times enriched), that is highly similar (1.9×10^{-75}) to the *C. elegans* gene *unc-93* [92]. This is associated with a 'rubber-band' phenotype, in which motor co-ordination is sluggish; it is thus taken to be a myogenic or neuromuscular gene. The discovery that a close homolog is highly enriched in renal tissue opens new lines of investigation for this gene.

Discussion

These data have value at two distinct levels: specific and general. Specifically, we have found out more about the operation of the Malpighian tubule than in any single published piece of work since the very first pioneering days: a summary is given in Figure 6. This tissue is of great interest, both for developmental studies and for integrative physiological study of epithelial function. Despite 990 papers on Malpighian tubules since the start of the twentieth century, and a really rather

good understanding of ion and water transport, the microarray data provide strong indications that these are only minor properties of the tubule. Whole families of transporters are represented by abundant mRNAs and transport solutes that have yet to be studied in the context of tubule. Some datasets implicate particular genes in processes that have been studied in great physiological detail, and the presence of known genes with the novel can only increase our confidence in the result. In this context, the demonstrated abundance of transporters for almost every class of organic and inorganic solute dramatically diminishes the number of solutes for which a nonspecific paracellular pathway need be invoked. The data thus allow the conceptual view of the epithelium to alter from leaky to tight in a physiological-transport sense: this is consistent with electrophysiological data [93].

There are two areas where microarray data deserve comment. Firstly, more than 300 genes are expressed in tubule but called as absent in whole-fly samples. Although there is an obvious convenience and consistency in employing whole-organism samples for array studies, it is important to recognize that the approach is very likely to suppress the detection of those interesting genes that are not widely expressed. Secondly, the premise that abundance on an array (or more generally, abundance of an RNA species) necessarily correlates with functional significance can be spectacularly refuted by three examples, the *trp* and *trpl* channels and NOS. It is, however, probably significant that these are cell-signaling molecules, where a relatively small number of molecules can have a disproportionate influence on cell behavior. By contrast, the transport genes for which the tubule is so enriched are much more likely to exert effects proportional to their abundance.

Conclusions

Reverse genetics is a vital tool in functional genomics, but the 'phenotype gap' has hampered widespread implementation of this approach [35]. As the tubule presents a range of easily assayed phenotypes [12], this work specifically identifies those genes that are likely to be best studied in tubule by virtue of their very high enrichment. In addition to the obvious transport genes, it is interesting that many transcription factors and human disease gene homologs fall into this category.

Table 14***Drosophila tubule* as a model for human genetic disease**

Gene	Affymetrix signal	Enrichment	Blast probability	OMIM reference	Human disease	Available fly stocks
<i>CG10226</i>	290 ± 25	28.3	1.00E-184	171050	Colchicine resistance	
<i>CG7402</i>	99 ± 4	26.9	2.00E-40	253000	Mucopolysaccharidosis IVA	
<i>lr</i>	1,302 ± 112	14.2	1.00E-76	600359	Bartter syndrome, antenatal, 601678	
<i>ry</i>	655 ± 44	13.0	1.00E-184	607633	Xanthinuria, type I, 278300	
<i>Ptx1</i>	183 ± 8	12.7	6.00E-38	602669	Anterior segment mesenchymal dysgenesis and cataract, 107250	
<i>Fmo-1</i>	131 ± 11	12.0	9.00E-27	136132	[Fish-odor syndrome], 602079	
<i>CG4484</i>	504 ± 50	12.0	1.00E-49	606202	Oculocutaneous albinism, type IV, 606574	
<i>DS00004.14</i>	759 ± 54	10.6	1.00E-123	603470	Citrullinemia, 215700	
<i>CG9455</i>	355 ± 40	9.0	1.00E-42	107400	Emphysema; emphysema-cirrhosis, hemorrhagic diathesis due to	
<i>CG5582</i>	825 ± 49	8.5	1.00E-69	607042	Ceroid-lipofuscinosis, neuronal-3, juvenile, 204200	
<i>Cyp4d2</i>	1,008 ± 70	8.3	1.00E-27	107910	Gynecomastia, familial, due to increased aromatase activity	
<i>CG7433</i>	1,364 ± 50	7.4	1.00E-153	137150	GABA-transaminase deficiency	
<i>CG1140</i>	894 ± 26	7.3	1.00E-176	245050	Ketoacidosis due to SCOT deficiency	
<i>CG9547</i>	860 ± 34	7.0	1.00E-164	231670	Glutaricaciduria, type I	
<i>PhKgamma</i>	2,665 ± 152	6.9	1.00E-111	172471	Glycogenosis, hepatic, autosomal	
<i>CG4623</i>	382 ± 37	6.8	4.00E-28	606598	Charcot-Marie-Tooth disease, mixed axonal and demyelinating	I(3)j7B3
<i>CG12370</i>	203 ± 17	6.7	5.00E-40	138033		
<i>CG15556</i>	221 ± 12	6.4	6.00E-12	602851	Convulsions, familial febrile, 4, 604352	
<i>KCNQ</i>	101 ± 0	6.4	1.00E-108	602235	Epilepsy, benign, neonatal, type I, 121200; myokymia with neonatal	
<i>CG17119</i>	852 ± 28	5.7	6.00E-74	606272	Cystinosis, atypical nephropathic; cystinosis, late-onset juvenile	
<i>CG7408</i>	168 ± 6	5.6	3.00E-27	300180	Chondrodysplasia punctata, X-linked recessive, 302950	
<i>Spat</i>	724 ± 39	5.1	2.00E-88	604285	Hyperoxaluria, primary, type I, 259900	EP(x)1365
<i>CG8743</i>	1,001 ± 44	4.9	1.00E-100	605248	Mucopolipidosis IV, 252650	
<i>CG14593</i>	59 ± 13	4.9	2.00E-33	131244	ABCD syndrome, 600501; Hirschsprung disease-2, 600155	
<i>CG1673</i>	911 ± 142	4.8	1.00E-100	113530	Hypervalinemia or hyperleucine-isoleucinemia (?)	
<i>Ndae1</i>	531 ± 34	4.7	1.00E-184	603345	Renal tubular acidosis, proximal, with ocular abnormalities, 604278	
<i>CG7834</i>	3441 ± 106	4.3	8.00E-80	130410	Glutaricaciduria, type IIB, 231680	EP(2)2553, I(2)k00405
<i>Pvr</i>	294 ± 26	4.2	6.00E-69	164770	Myeloid malignancy, predisposition to	
<i>CG12030</i>	887 ± 51	4.1	1.00E-124	606953	Galactose epimerase deficiency, 230350	
<i>Mdr49</i>	239 ± 25	4.0	1.00E-184	171060	Cholestasis, familial intrahepatic, of pregnancy, 147480	I(2)k05224
<i>CG4685</i>	563 ± 19	4.0	1.00E-129	271980	Succinic semialdehyde dehydrogenase deficiency	EP(2)2545, I(2)k08713
<i>CG12338</i>	774 ± 16	3.9	4.00E-40	124050		
<i>CG12582</i>	183 ± 14	3.8	1.00E-142	248510	Mannosidosis, beta-	I(2)k10108
<i>Reg-3</i>	463 ± 24	3.8	1.00E-184	274270	Thymine-uraciluria	
<i>Cyp12c1</i>	73 ± 5	3.8	2.00E-34	124080	Aldosterone to renin ratio raised; hypoaldosteronism, congenital,	
<i>Fur1</i>	724 ± 29	3.7	1.00E-163	162150	Obesity with impaired prohormone processing, 600955	
<i>Cyp9c1</i>	258 ± 14	3.7	7.00E-53	274180	Thromboxane synthase deficiency	I(3)05545
<i>Drip</i>	318 ± 16	3.6	1.00E-37	154050	Cataract, polymorphic and lamellar, 604219	
<i>CG8654</i>	274 ± 29	3.6	2.00E-62	607096	Hypouricemia, renal, 220150	

Table 14 (Continued)***Drosophila* tubule as a model for human genetic disease**

<i>Cyp9f2</i>	1,700 ± 60	3.6	1.00E-69	124010	CYP3A4 promoter polymorphism; CYP3A4-V	
<i>ERR</i>	239 ± 21	3.5	5.00E-29	313700	Androgen insensitivity, 300068; breast cancer, male	EP(3)3340
<i>CG3603</i>	94 ± 7	3.4	5.00E-20	222745	DECR deficiency (2) (?)	
<i>CG9232</i>	877 ± 20	3.4	1.00E-118	606999	Galactosemia, 230400	
<i>CG8417</i>	502 ± 31	3.2	3.00E-71	154550	Carbohydrate-deficient glycoprotein syndrome, type Ib, 602579	EP(2)0844, EP(2)2192, EP(2)2358, I(2)05428, I(2)k06503
<i>CG4663</i>	439 ± 14	3.2	2.00E-29	601789	Adrenoleukodystrophy, neonatal, 202370; Zellweger syndrome, 214100	
<i>Cat</i>	4,316 ± 88	3.2	1.00E-184	115500	Acatalsemia	
Prominin-like	308 ± 24	3.0	1.00E-20	604365	Retinal degeneration, autosomal recessive, prominin-related	EP(2)0740

Genes that are abundant (Affymetrix signal > 50) and enriched (> 3 times) in tubule, and which are also closely similar (Blast probability < 10⁻²⁰) to genes mutated in human genetic diseases, as described in the Homophila database [99]. OMIM reference refers to entries in the Online Mendelian Inheritance in Man database [100].

This work thus stresses the importance of systematic, fine-grained, tissue-specific microarray analysis in closing the phenotype gap for multicellular model organisms.

Materials and methods

Flies

Drosophila melanogaster were kept on standard diet at 25°C and 55% relative humidity on a 12:12 h photoperiod. Malpighian tubules were dissected from 7-day-old adults, for compatibility with the extensive physiological literature on the tubule [10,11,13,15,17,19,20,39,70,75,94-96]. At this stage, the tubules are in a relatively stable state after adult emergence, and their secretion parameters do not change detectably between 3 and 14 days post-emergence.

Microarrays

Tubules were dissected in batches of 1,000 by a group of eight experimenters. Tubules were aggregated into Trizol every 15 min to minimize the distortion of the transcriptome by the trauma of dissection and *in vitro* incubation. Care was taken to sever the tubules from the gut at the lower ureter so that no other tissue was included in the sample. For each experimental point, whole flies from the same culture were homogenized in Trizol in batches of 100, to permit a matched pair comparison. Six repeats were performed. RNA was extracted according to standard protocols, and quality was assessed with an Agilent RNA Bioanalyzer. Samples of 20 µg total RNA were reverse-transcribed, then *in vitro* transcribed, according to Affymetrix standard protocols. The quality of the complementary RNA (cRNA) was also checked on an Agilent RNA Bioanalyzer, with a sample in which the broad cRNA peak exceeded the height of the low molecular weight degradation peak taken to be satisfactory. Samples were then run on the Affymetrix *Drosophila* genome array under standard conditions. Quality control was at several levels: the Affymetrix

MAS 5.0 software provided evidence of successful sample preparation, with test genes providing a 3':5' signal ratio of less than 3. dChip [97] provided an alternative view, with a direct oligo-by-oligo view on the success of hybridization across the array surface; slides with both single-probe and probe-set outlier rates of less than 5% were taken as satisfactory. Only arrays in which both results were in range were accepted. In this case, 11 of 12 arrays were satisfactory; the first tubule array failed both MAS and dChip criteria, and so the first experimental pair was discarded to leave a five-sample paired design. As will be seen from the results, this design

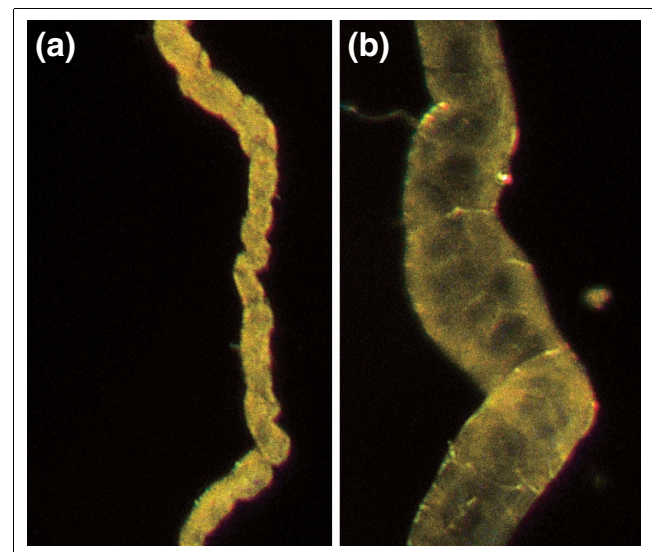


Figure 5
Recapitulation of human xanthinuria type 1 by *rosy* mutants. (a) Wild-type tubule; (b) tubule from adult *ry2* homozygous fly. Both micrographs are at the same magnification, and the diameter of the wild-type tubule can be taken as 35 µm.

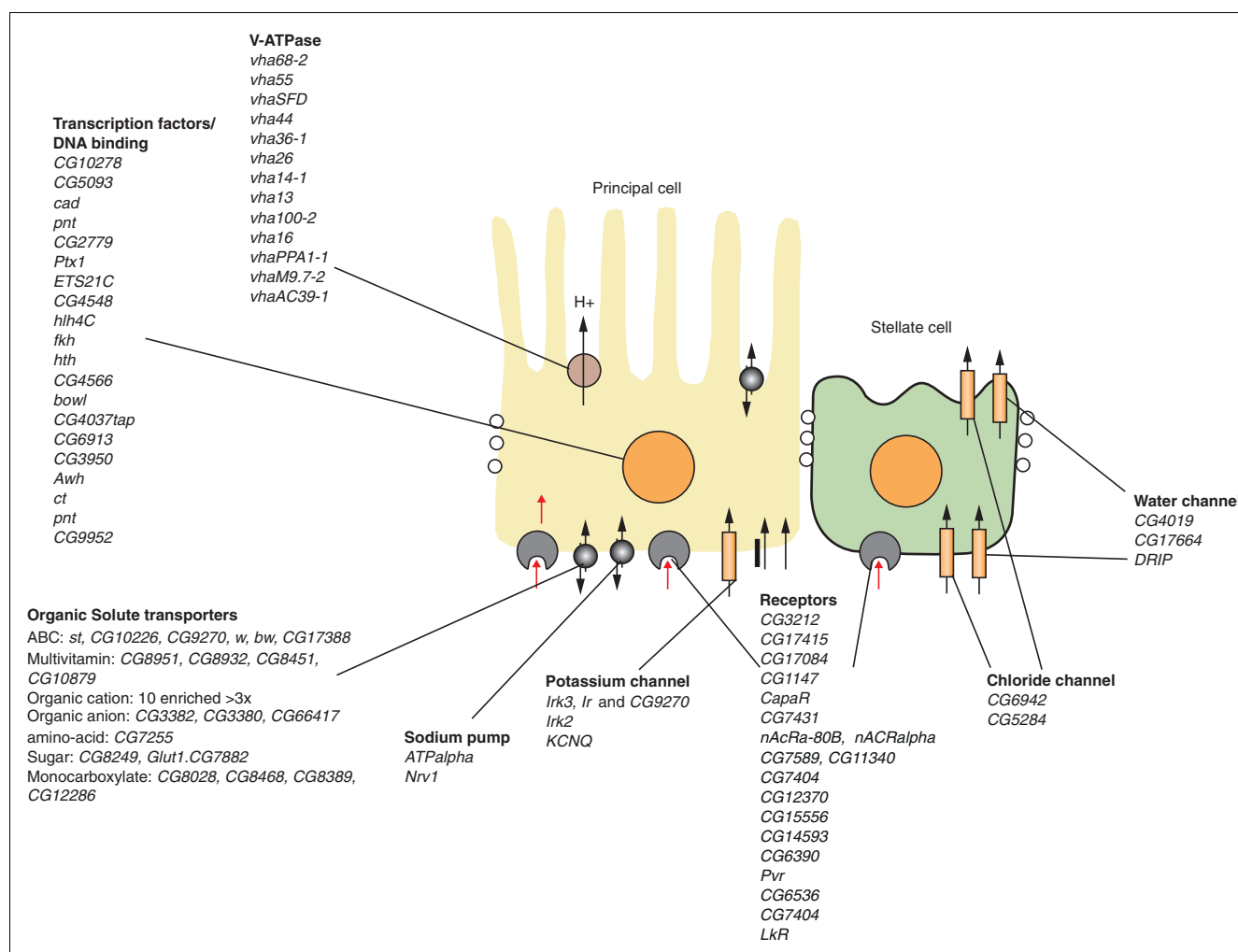


Figure 6
 Summary of major genes enriched in tubule. Genes shown are upregulated at least threefold.

was sufficient to identify tubule-enriched genes with a high level of confidence. As sample collection extended over the whole day, array results from morning versus afternoon samples were compared (data not shown), but no difference was found between the two groups at this very broad time resolution.

Bioinformatics

Microarray samples were analyzed by two independent routes. The first was low-level analysis with the Affymetrix MAS 5.0 suite and identification of differentially expressed genes using the Affymetrix Data Mining Tool. The second was low-level analysis using dChip software [97] followed by assessment of significance using SAM software [98] followed by post-analysis by dChip. The MAS5 low-level analysis consisted of background subtraction followed by robust conversion of probe-level perfect match-mismatch (PM-MM) expression values into probe-set-level signals followed by linear multi-chip normalization (scaling). Tubule enrichment

was based on an Affymetrix 'up' call, and a critical level of $p < 0.05$. In this analysis method, tubule and fly samples were taken as matched pairs, reflecting their biological origin. The dChip-based low-level analysis consisted of background correction followed by the multi-chip, 'invariant-set' nonlinear normalization at probe level followed by the calculation of model-based expression indices using PM expression values only. Differentially expressed genes between two groups of five replicates were identified within dChip by filtering data using the following criteria: lower 90% confidence bound of fold-change [21] > 2 ; difference between group means on antilog scale > 100 and p -value for t -test of equal group means < 0.01 . Alternatively, the differentially expressed genes were identified using SAM software with 1,000 sample permutations and false-discovery rate cutoff of 1%. These were then post-filtered using two first criteria from the dChip analysis mentioned above. Fold change was calculated as a ratio of group means. Outputs were saved as Excel files, and parsed by hand-coded Perl scripts.

Additional data file

A list of genes (Additional data file 1) called as upregulated in tubule by Affymetrix SAM 5 software, and with more than two-fold enrichment is available with the online version of this article.

Acknowledgements

We thank the staff of the Sir Henry Wellcome Functional Genomics facility in Glasgow, for their help and training in Affymetrix technology. We thank the following members of the Dow/Davies lab for their assistance in the 'ten thousand tubule days': Laura Kean, Valerie Pollock, Shirley Graham, Kate Broderick, Matthew Macpherson, Kostas Stergiopoulos and Pablo Cabrero. This work was funded by BBBSRC GAIN grants to J.A.T.D and S.A.D.

References

- Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: **Yeast microarrays for genome wide parallel genetic and gene expression analysis.** *Proc Natl Acad Sci USA* 1997, **94**:13057-13062.
- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al: **The genome sequence of *Drosophila melanogaster*.** *Science* 2000, **287**:2185-2195.
- White KP, Rifkin SA, Hurban P, Hogness DS: **Microarray analysis of *Drosophila* development during metamorphosis.** *Science* 1999, **286**:2179-2184.
- McDonald MJ, Rosbash M: **Microarray analysis and organization of circadian gene expression in *Drosophila*.** *Cell* 2001, **107**:567-578.
- Zou S, Meadows S, Sharp L, Jan LY, Jan YN: **Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2000, **97**:13726-13731.
- McCarroll SA, Murphy CT, Zou S, Pletcher SD, Chin CS, Jan YN, Kenyon C, Bargmann CI, Li H: **Comparing genomic expression patterns across species identifies shared transcriptional profile in aging.** *Nat Genet* 2004, **36**:197-204.
- Andrews J, Bouffard GG, Cheadle C, Lu J, Becker KG, Oliver B: **Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis.** *Genome Res* 2000, **10**:2030-2043.
- Denholm B, Sudarsan V, Pasalodos-Sanchez S, Artero R, Lawrence P, Maddrell S, Baylies M, Skaer H: **Dual origin of the renal tubules in *Drosophila*: mesodermal cells integrate and polarize to establish secretory function.** *Curr Biol* 2003, **13**:1052-1057.
- Sözen MA, Armstrong JD, Yang MY, Kaiser K, Dow JAT: **Functional domains are specified to single-cell resolution in a *Drosophila* epithelium.** *Proc Natl Acad Sci USA* 1997, **94**:5207-5212.
- Dow JAT, Maddrell SHP, Görtz A, Skaer NV, Brogan S, Kaiser K: **The Malpighian tubules of *Drosophila melanogaster*: a novel phenotype for studies of fluid secretion and its control.** *J Exp Biol* 1994, **197**:421-428.
- Dow JAT, Davies SA: **The *Drosophila* Malpighian tubule: an epithelial model for integrative physiology.** *Comp Biochem Physiol* 1999, **124A**:S49-S49.
- Dow JAT, Davies SA: **Integrative physiology and functional genomics of epithelial function in a genetic model organism.** *Physiol Rev* 2003, **83**:687-729.
- O'Donnell MJ, Dow JAT, Huesmann GR, Tublitz NJ, Maddrell SHP: **Separate control of anion and cation transport in Malpighian tubules of *Drosophila melanogaster*.** *J Exp Biol* 1996, **199**:1163-1175.
- The *Drosophila melanogaster* Malpighian tubule** [http://fly.to/tubules/]
- Coast GM, Webster SG, Schegg KM, Tobe SS, Schooley DA: **The *Drosophila melanogaster* homologue of an insect calcitonin-like diuretic peptide stimulates V-ATPase activity in fruit fly Malpighian tubules.** *J Exp Biol* 2001, **204**:1795-1804.
- Kean L, Pollock VP, Broderick KE, Davies SA, Veenstra J, Dow JAT: **Two new members of the CAP_{2b} family of diuretic peptides are encoded by the gene *capability* in *Drosophila melanogaster*.** *Am J Physiol* 2002, **282**:R1297-R1307.
- Terhaz S, Oconnell FC, Pollock VP, Kean L, Davies SA, Veenstra JA, Dow JAT: **Isolation and characterization of a leucokinin-like peptide of *Drosophila melanogaster*.** *J Exp Biol* 1999, **202**:3667-3676.
- Dow JAT, Maddrell SHP, Davies S-A, Skaer NJV, Kaiser K: **A novel role for the nitric oxide/cyclic GMP signalling pathway: the control of fluid secretion in *Drosophila*.** *Am J Physiol* 1994, **266**:R1716-R1719.
- Davies SA, Huesmann GR, Maddrell SHP, O'Donnell MJ, Skaer NJV, Dow JAT, Tublitz NJ: **CAP_{2b}, a cardioacceleratory peptide, is present in *Drosophila* and stimulates tubule fluid secretion via cGMP.** *Am J Physiol* 1995, **269**:R1321-R1326.
- Rosay P, Davies SA, Yu Y, Sozen MA, Kaiser K, Dow JAT: **Cell-type specific calcium signalling in a *Drosophila* epithelium.** *J Cell Sci* 1997, **110**:1683-1692.
- Li C, Wong WH: **Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application.** *Genome Biol* 2001, **2**:research0032.1-0032.11.
- EBI databases: ArrayExpress home** [http://www.ebi.ac.uk/arrayexpress]
- Fibla J, Enjuanes L, Gonzalez-Duarte R: **Inter-specific analysis of *Drosophila* alcohol dehydrogenase by an immunoenzymatic assay using monoclonal antibodies.** *Biochem Biophys Res Commun* 1989, **160**:638-646.
- Anderson S, Brown M, McDonald J: **Tissue specific expression of the *Drosophila Adh* gene: a comparison of *in situ* hybridization and immunocytochemistry.** *Genetica* 1991, **84**:95-100.
- Aoki Y, Suzuki KT, Kubota K: **Accumulation of cadmium and induction of its binding protein in the digestive tract of fleshfly (*Sarcophaga peregrina*) larvae.** *Comp Biochem Physiol C* 1984, **77**:279-282.
- Marchal-Segault D, Briancon C, Halpern S, Fragu P, Lauge G: **Secondary ion mass spectrometry analysis of the copper distribution in *Drosophila melanogaster* chronically intoxicated with Bordeaux mixture.** *Biol Cell* 1990, **70**:129-132.
- Wessing A, Zierold K: **Metal-salt feeding causes alterations in concretions in *Drosophila* larval Malpighian tubules as revealed by X-ray microanalysis.** *J Insect Physiol* 1992, **38**:623-632.
- Rabitsch WB: **Tissue-specific accumulation patterns of Pb, Cd, Cu, Zn, Fe, and Mn in workers of three ant species (Formicidae, Hymenoptera) from a metal-polluted site.** *Arch Environ Contam Toxicol* 1997, **32**:172-177.
- Schofield RMS, Postlethwait JH, Lefevre HW: **MeV-ion microprobe analyses of whole *Drosophila* suggest that zinc and copper accumulation is regulated storage not deposit excretion.** *J Exp Biol* 1997, **200**:3235-3243.
- Ballan-Dufrancais C: **Localization of metals in cells of pterygote insects.** *Microsc Res Tech* 2002, **56**:403-420.
- Bonneton F, Wegnez M: **Developmental variability of metallothionein *mtn* gene-expression in the species of the *Drosophila melanogaster* subgroup.** *Dev Genet* 1995, **16**:253-263.
- Durliat M, Bonneton F, Boissonneau E, Andre M, Wegnez M: **Expression of metallothionein genes during the post-embryonic development of *Drosophila melanogaster*.** *Biometals* 1995, **8**:339-351.
- Brown SD, Peters J: **Combining mutagenesis and genomics in the mouse - closing the phenotype gap.** *Trends Genet* 1996, **12**:433-435.
- Bullard DC: **Mind the phenotype gap.** *Trends Mol Med* 2001, **7**:537-538.
- Dow JAT: **The *Drosophila* phenotype gap - and how to close it.** *Briefings Funct Genomics Proteomics* 2003, **2**:121-127.
- Dow JAT, Davies SA: **The *Drosophila melanogaster* Malpighian tubule.** *Adv Insect Physiol* 2001, **28**:1-83.
- FlyBase error report for CG12286 and karmoison on Thu Dec 6 07:17:59. 2001.**
- Zierold K, Wessing A: **Mass dense vacuoles in *Drosophila* Malpighian tubules contain zinc, not sodium. A reinvestigation by X-ray microanalysis of cryosections.** *Eur J Cell Biol* 1990, **53**:222-226.
- Wessing A, Zierold K: **The formation of type I concretions in *Drosophila* Malpighian tubules studied by electron microscopy and X-ray microanalysis.** *J Insect Physiol* 1999, **45**:39-44.
- Meulemans W, De Loof A: **Transport of the cationic fluorochrome rhodamine 123 in an insect's Malpighian tubule: Indications of a reabsorptive function of the secondary cell type.**

- J Cell Sci* 1992, **101**:349-361.
41. Maddrell SHP, Gardiner BOC, Pilcher DEM, Reynolds SE: **Active transport by insect Malpighian tubules of acidic dyes and of acylamides.** *J Exp Biol* 1974, **61**:357-377.
 42. Gaertner LS, Morris CE: **Accumulation of daunomycin and fluorescent dyes by drug-transporting Malpighian tubule cells of the tobacco hornworm, *Manduca sexta*.** *Tissue Cell* 1999, **31**:185-194.
 43. Skaer HLB, Maddrell SHP: **How are invertebrate epithelia made tight?** *J Cell Sci* 1987, **88**:139-141.
 44. Thomson JA, Gunson MM: **Developmental changes in the major inclusion bodies of polytene nuclei from larval tissues of the blowfly, *Calliphora stygia*.** *Chromosoma* 1970, **30**:193-201.
 45. Bedo DG: **Polytene chromosomes in pupal and adult blackflies (Diptera: Simuliidae).** *Chromosoma* 1976, **57**:387-396.
 46. Campos J, Andrade CF, Recco-Pimentel SM: **Malpighian tubule polytene chromosomes of *Culex quinquefasciatus* (Diptera, Culicinae).** *Mem Inst Oswaldo Cruz* 2003, **98**:383-386.
 47. Campos J, Andrade CF, Recco-Pimentel SM: **A technique for preparing polytene chromosomes from *Aedes aegypti* (Diptera, Culicinae).** *Mem Inst Oswaldo Cruz* 2003, **98**:387-390.
 48. Maddrell SHP, Lane NJ, Harrison JB, Gardiner BOC: **DNA replication in binucleate cells of the Malpighian tubules of Hemipteran insects.** *Chromosoma* 1985, **91**:201-209.
 49. Maddrell SHP, O'Donnell MJ: **Insect Malpighian tubules: V-ATPase action in ion and fluid transport.** *J Exp Biol* 1992, **172**:417-429.
 50. Weltens R, Leysens A, Zhang AL, Lohrmann E, Steels P, van Kerkhove E: **Unmasking of the apical electrogenic H pump in isolated Malpighian tubules (*Formica polyctena*) by the use of barium.** *Cell Physiol Biochem* 1992, **2**:101-116.
 51. Bertram G, Wessing A: **Intracellular pH regulation by the plasma-membrane V-ATPase in Malpighian tubules of *Drosophila* larvae.** *J Comp Physiol B* 1994, **164**:238-246.
 52. Dow JAT, Davies SA, Guo Y, Graham S, Finbow ME, Kaiser K: **Molecular genetic analysis of V-ATPase function in *Drosophila melanogaster*.** *J Exp Biol* 1997, **200**:237-245.
 53. Dow JAT: **The multifunctional *Drosophila melanogaster* V-ATPase is encoded by a multigene family.** *J Bioenerget Biomemb* 1999, **31**:75-83.
 54. Harvey WR, Maddrell SHP, Telfer WH, Wieczorek H: **H⁺ V-ATPases energize animal plasma membranes for secretion and absorption of ions and fluids.** *Am Zool* 1998, **38**:426-441.
 55. Wieczorek H, Brown D, Grinstein S, Ehrenfeld J, Harvey WR: **Animal plasma membrane energization by proton motive V-ATPases.** *BioEssays* 1999, **21**:637-648.
 56. Wieczorek H: **The insect V-ATPase, a plasma-membrane proton pump energizing secondary active transport - molecular analysis of electrogenic potassium transport in the tobacco hornworm midgut.** *J Exp Biol* 1992, **172**:335-343.
 57. Wieczorek H, Harvey WR: **Energization of animal plasma membranes by the proton-motive force.** *Physiol Zool* 1995, **68**:15-23.
 58. Anstee JH, Bowler K: **Ouabain sensitivity of insect epithelial tissues.** *Comp Biochem Physiol* 1979, **62A**:763-769.
 59. Lebovitz RM, Takeyasu K, Fambrough DM: **Molecular characterization and expression of the (Na⁺ + K⁺)-ATPase α -subunit in *Drosophila melanogaster*.** *EMBO J* 1989, **8**:193-202.
 60. Masia R, Aneshansley D, Nagel W, Nachman RJ, Beyenbach KW: **Voltage clamping single cells in intact malpighian tubules of mosquitoes.** *Am J Physiol Renal Physiol* 2000, **279**:F747-F754.
 61. Wiehart UJ, Klein G, Steels P, Nicolson SV, Van Kerkhove E: **K(+) transport in Malpighian tubules of *Tenebrio molitor* L.: is a K(ATP) channel involved?** *J Exp Biol* 2003, **206**:959-965.
 62. Zeiske W, Van Driessche W, Ziegler R: **Current-noise analysis of the basolateral route for K⁺ ions across a K⁺-secreting insect midgut epithelium (*Manduca sexta*).** *Pflugers Arch* 1986, **407**:657-663.
 63. Hanrahan JW, Wills NK, Phillips JE, Lewis SA: **Basolateral K channels in an insect epithelium. Channel density, conductance, and block by barium.** *J Gen Physiol* 1986, **87**:443-466.
 64. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: **Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
 65. Tanaka A, Tokimasa T: **Theoretical background for inward rectification.** *Tokai J Exp Clin Med* 1999, **24**:147-153.
 66. Inagaki N, Gonoi T, Clement JP 4th, Namba N, Inazawa J, Gonzalez G, Aguilar-Bryan L, Seino S, Bryan J: **Reconstitution of IKATP: an inward rectifier subunit plus the sulfonylurea receptor.** *Science* 1995, **270**:1166-1170.
 67. Inagaki N, Gonoi T, Clement JP, Wang CZ, Aguilar-Bryan L, Bryan J, Seino S: **A family of sulfonylurea receptors determines the pharmacological properties of ATP-sensitive K⁺ channels.** *Neuron* 1996, **16**:1011-1017.
 68. Engel A, Walz T, Agre P: **The aquaporin family of membrane water channels.** *Curr Opin Struct Biol* 1994, **4**:545-553.
 69. Dow JAT, Kelly DC, Davies SA, Maddrell SHP, Brown D: **A novel member of the major intrinsic protein family in *Drosophila* - are aquaporins involved in insect malpighian (renal) tubule fluid secretion?** *J Physiol* 1995, **489**:P110-P111.
 70. O'Donnell MJ, Maddrell SHP: **Fluid reabsorption and ion transport by the lower Malpighian tubules of adult female *Drosophila*.** *J Exp Biol* 1995, **198**:1647-1653.
 71. Wessing A, Eichelberg D: **Malpighian tubules, rectal papillae and excretion.** In: *The Genetics and Biology of Drosophila Volume 2c*. Edited by: Ashburner A, Wright TRF. London: Academic Press; 1978:1-42.
 72. Davies SA, Stewart EJ, Huesmann GR, Skaer NJV, Maddrell SHP, Tublitz NJ, Dow JAT: **Neuropeptide stimulation of the nitric oxide signaling pathway in *Drosophila melanogaster* Malpighian tubules.** *Am J Physiol* 1997, **42**:R823-R827.
 73. O'Donnell MJ, Rheault MR, Davies SA, Rosay P, Harvey BJ, Maddrell SHP, Kaiser K, Dow JAT: **Hormonally-controlled chloride movement across *Drosophila* tubules is via ion channels in stellate cells.** *Am J Physiol* 1998, **274**:R1039-R1049.
 74. Blumenthal EM: **Regulation of chloride permeability by endogenously produced tyramine in the *Drosophila* Malpighian tubule.** *Am J Physiol Cell Physiol* 2003, **284**:C718-C728.
 75. Radford JC, Davies SA, Dow JA: **Systematic G-protein-coupled receptor analysis in *Drosophila melanogaster* identifies a leukocinin receptor with novel roles.** *J Biol Chem* 2002, **277**:38810-38817.
 76. Iversen A, Cazzamali G, Williamson M, Hauser F, Grimmekhuijzen CJ: **Molecular cloning and functional expression of a *Drosophila* receptor for the neuropeptides capa-1 and -2.** *Biochem Biophys Res Commun* 2002, **299**:628-633.
 77. Kerr M, Davies SA, Dow JAT: **Cell-specific manipulation of second messengers: a toolbox for integrative physiology in *Drosophila*.** *Curr Biol* in press.
 78. MacPherson MR, Broderick KE, Graham S, Day JP, Houslay MD, Dow JAT, Davies SA: **The *dg2 for* gene confers a renal phenotype in *Drosophila* via cGMP-specific phosphodiesterase.** *J Exp Biol* 2004, **207**:2769-2776.
 79. MacPherson MR, Lohmann SM, Davies SA: **Analysis of *Drosophila* cGMP-dependent protein kinases and assessment of their *in vivo* roles by targeted expression in a renal transporting epithelium.** *J Biol Chem* 2004. Doi:10.1074/jbc.M405619200
 80. Lengyel JA, Liu XJ: **Posterior gut development in *Drosophila*: a model system for identifying genes controlling epithelial morphogenesis.** *Cell Res* 1998, **8**:273-284.
 81. Skaer H: **The alimentary canal.** In: *The Development of Drosophila melanogaster Volume 2*. Edited by: Bate M, Martinez Arias A. Cold Spring Harbor: Cold Spring Harbor Press; 1993:941-1012.
 82. Wan S, Cato AM, Skaer H: **Multiple signalling pathways establish cell fate and cell number in *Drosophila* malpighian tubules.** *Dev Biol* 2000, **217**:153-165.
 83. Blochlinger K, Jan LY, Jan YN: **Postembryonic patterns of expression of *cut*, a locus regulating sensory organ identity in *Drosophila*.** *Development* 1993, **117**:441-450.
 84. Tearle RG, Nusslein-Volhard C: **Tubingen mutants and stock list.** *Dros Inf Serv* 1987, **66**:209-269.
 85. Kurant E, Pai CY, Sharf R, Halachmi N, Sun YH, Salzberg A: **Dorsotals/homothorax, the *Drosophila* homologue of *meis1*, interacts with *extradenticle* in patterning of the embryonic PNS.** *Development* 1998, **125**:1037-1048.
 86. **Bloomington Stock Center homepage** [<http://fly.bio.indiana.edu>]
 87. Yang MY, Wang Z, MacPherson M, Dow JAT, Kaiser K: **A novel *Drosophila* alkaline phosphatase specific to the ellipsoid body of the adult brain and the lower Malpighian (renal) tubule.** *Genetics* 2000, **154**:285-297.
 88. Montell C: **New light on TRP and TRPL.** *Mol Pharmacol* 1997, **52**:755-763.
 89. Hardie RC: **Phototransduction in *Drosophila melanogaster*.** *J Exp Biol* 2001, **204**:3403-3409.
 90. Minkin B: **The TRP channel and phospholipase C-mediated**

- signaling. *Cell Mol Neurobiol* 2001, **21**:629-643.
91. Gromova I, Gromov P, Celis JE: **bc10: A novel human bladder cancer-associated protein with a conserved genomic structure downregulated in invasive cancer.** *Int J Cancer* 2002, **98**:539-546.
 92. Levin JZ, Horvitz HR: **The *Caenorhabditis elegans* unc-93 gene encodes a putative transmembrane protein that regulates muscle contraction.** *J Cell Biol* 1992, **117**:143-155.
 93. Beyenbach KW: **Regulation of tight junction permeability with switch-like speed.** *Curr Opin Nephrol Hypertens* 2003, **12**:543-550.
 94. Riegel JA, Farndale RW, Maddrell SHP: **Fluid secretion by isolated Malpighian tubules of *Drosophila melanogaster* Meig.: effects of organic anions, quinacrine and a diuretic factor found in the secreted fluid.** *J Exp Biol* 1999, **202**:2339-2348.
 95. Dube K, McDonald DG, O'Donnell MJ: **Calcium transport by isolated anterior and posterior Malpighian tubules of *Drosophila melanogaster* : roles of sequestration and secretion.** *J Insect Physiol* 2000, **46**:1449-1460.
 96. Rheault MR, O'Donnell MJ: **Analysis of epithelial K(+) transport in Malpighian tubules of *Drosophila melanogaster*: evidence for spatial and temporal heterogeneity.** *J Exp Biol* 2001, **204**:2289-2299.
 97. Li C, Wong WH: **Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.** *Proc Natl Acad Sci USA* 2001, **98**:31-36.
 98. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
 99. Chien S, Reiter LT, Bier E, Gribskov M: **Homophila: human disease gene cognates in *Drosophila*.** *Nucleic Acids Res* 2002, **30**:149-151.
 100. **Online Mendelian Inheritance in Man** [<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>]