Enterprise Biology Software: III. Research (2002)

ROBERT P. BOLENDER

Enterprise Biology Software Project, P. O. Box 303, Medina, WA 98039-0303, USA <u>http://enterprisebiology.com</u>

Summary

The Enterprise Biology Software (2002) updates the stereology literature database (research papers 1965 to 2001), refreshes the basic research tools (BIOLOGYtabs), and includes an upgrade of EBS (2001). This paper introduces the new software, includes further examples of using a literature database as a research tool, and continues to explore the relationship of stereology to discovery in the life sciences. Copies of the new software package are being sent to contributing authors and will be offered to first authors of stereology papers published in 2002.

The original EBS (Bolender 2001a, 2001b) described the development of a literature database for biological stereology. The first applications of the database included standardizing biological data, creating new data from old, and describing a mathematical platform for biology. In turn, the platform was used as a research tool for exploring two fundamental principles of biology: connectivity and stoichiometry. The project was launched by requesting reprints from first authors of stereology publications (1999-2001) and then returning copies of EBS (2001) to them.

Introduction

Background

The Enterprise Biology Software Project looks for ways of using mathematics and technology to accelerate learning and discovery in the life sciences. In the first release of the software (Bolender, 2001a, 2001b), a stereology literature database was introduced as a research tool for exploring complex problems in basic and clinical research. Three general observations came from that exercise.

- 1. Patterns of connectivity appeared routinely in research data when the effects of disruptive variables influencing stereological data were minimized. These troublesome variables included bias and time.
- 2. A stereology literature database can redefine the role of published data. Data from one research paper can influence and be influenced by all other data entries in the database. Moreover, data stored in a database can serve any purpose the user wishes thereby ensuring both the short and long-term benefits of a research publication.
- 3. The challenging task of unraveling gene function requires data analysis at the level of complex systems. The database technology included in the Enterprise Biology Software effectively upgrades stereological data to this level by adding a *Connection Model*, as shown below.

The current release of the software (EBS 2002) upgrades the stereology literature database through 2001 and

explores additional ways of using the database as a research tool. We begin our continuing story with a timely reality check.

Reality Check

In the current release, we address a controversial question. Can stereological data be trusted? The new designbased methods of stereology have been widely heralded as ushering in an era of unbiased stereology - a "modern stereology." These new methods have enjoyed great success and they are enthusiastically supported by the stereology community. However, a reality check forces us to ask what can only be regarded as an uncomfortable and unpleasant question. Do these new unbiased methods actually produce unbiased data? Consider cases A and B.



If we select Case A, then we can safely assume that all the conditions of the unbiased methods are met throughout the process of generating data – without exception. Here the stereological data estimates accurately reflect the structures in the living animal and there is no bias. If we select Case B, then we accept the view that the procedures used to capture data from living animals introduce bias. In this case, stereological data do not accurately reflect the structures in the living animal because the data carry a bias. Case A allows us to overlook the presence of the troublesome variables bias and time, Case B does not. Of the two, Case B seems closer to the truth.

Amazing Secret

Is the bias in biological data hiding an amazing secret? Probably, yes. A striking clue comes from recent discoveries in molecular biology. Genome sequences taken from a variety of animals reveal an astounding similarity. In several cases, the genetic blueprints are nearly identical (Waterston et al., 2002). Should this pattern of similarity persist, then structures in different animals might differ largely by size and phenotypic expression. In other words, all structures based on the same genetic code can be expected to be equivalent and scalable. A recent article would seem to support this view. When kidney precursor cells from a human were transplanted into a mouse a small human kidney developed and functioned normally (Dekel et al., 2003).

What is the secret? Perhaps, observed differences in biological data – within and across many animal species – represent little more than a disturbance created by methodological bias and phenotypic variation. If we remove the bias from the data and account for phenotype, then intense order would be expected to appear - everywhere we look. The secret so well hidden by our methods is that biology - like physics and chemistry - is also an $r^2=1$ player. In other words, its properties can likewise be defined mathematically.

Tantalizing Question

One of the most compelling advantages of a stereology literature database is that it allows us to explore the roots of complexity in biology. If, for example, we wish to learn about how biology is "wired" mathematically, then we can set a problem and develop a strategy for solving it. For example, how might we use stereology to unravel gene function? One approach – especially well suited to stereology – would be to predict gene function, as suggested below.



Such a scheme is of interest to us here because it poses tantalizing questions for stereologists. Are the secrets of gene function encrypted in the images we can see with light and electron microscopy? Can these images generate algorithms that mirror the functions of genes? If yes, then we have sufficient incentive to create new data resources that can be leveraged now and in the future.

Super-Complexity

Reading the stereology literature – from end to end - offers one a rare opportunity of seeing bits and pieces of the "big picture." For example, discovery in biology is a process often being influenced importantly by two general types of complexity – one associated with biology the other with the way we explore biology. These two complexities combine to form a super-complexity that defines – in our case - the stereological data we store in our journals and databases. The challenge for us here is to take these complex data apart and then reassemble them into new data capable of generating new applications.

In stereological data, we can identify one complexity for biology and two complexities for exploring:

- Phenotype (biology)
- Bias (exploring)
- Time (exploring)

Taken as one, the three complexities define a super-complexity. Let's begin with bias, because it has become such a lively topic among stereologists. Why are stereological data biased? Stop for a moment and think. When assessing the accuracy of biological data, only two points are actually important - the starting point and the end point. The starting point is the living animal and the end point is our research product – stereological data. A complexity appears as a bias when the same structures have different values at each location – the living animal and the journal article. The difference between the data at these two locations can be explained by our application of experimental methods that routinely add bias to the structures we are trying to estimate stereologically.

What – exactly – is this difference between these two points? Unfortunately, we don't know. Nonetheless, we can be confident that a difference does exist and that bias has a harmful effect on our data – especially in a complex setting (Weibel and Paumgartner, 1978; Weibel, 1979; Coggeshall and Lekan, 1996). This creates a dilemma in that we have to find a solution to this bias problem even when we suspect at the outset that no exact solution exists. The most practical course to follow is one of mitigation - prevent bias from forming in the first place and minimize it if it already exists.

Let's look at some examples. The design-based methods of stereology were developed specifically with the goal of avoiding bias (Baddeley et al., 1985; Gundersen, 1986). Although such methods help to minimize some of the overall bias, they cannot guarantee – by themselves - unbiased stereological data. For the most part, an unbiased method is applied to a structure already compromised by bias. This leads to a painful truth.

Sampling a biased structure with an unbiased sampling method produces an unbiased estimate of a biased structure.

Think about it. It's the only reasonable conclusion we can draw. Even the optical fractionator - long considered the gold standard of unbiased estimates – has been reported to carry a methodological bias of between 4 to 24% (Hatton and von Bartheld, 1999).

While avoiding bias in the first place is the best strategy, we are still left with the larger problem of dealing with all the unavoidable bias generated by our experimental methods. One solution to this seemingly intractable problem is to find a way of letting this methodological bias cancel out by itself – automatically. There are several advantages of this approach to reducing complexity. It can increase the accuracy of stereological data and - at the same time - eliminate the blurring effect of time (see EBS 2001). The net effect is that the unwanted complexity in stereological data – produced by bias and time - can be markedly reduced. In practical terms, the process of reducing complexity consists of upgrading the stereological literature from a change model to a connection model.

Next, let's consider the living animal as a source of complexity. Recall that a phenotype (organism, organ, ..., molecule) is a product of the genotype (genetic makeup) and that the same organism – including all its parts - can express one or more distinct phenotypes. Since the process of hunting for patterns with stereological data often involves grouping data from different animals with different phenotypes, we need to know when to group and when not to. In some cases, an inappropriate grouping may bias an outcome importantly, in other cases not. Knowing when to avoid the undesirable combinations can therefore have a beneficial effect.

The following short list of natural and artificial factors, which influence stereological data, serves to illustrate the many origins of super-complexity. Our task here is to simplify this complex picture by evaluating the natural factors and minimizing the artificial ones (see the appendix for examples).

1. Natural

- a. Animal Type i. Human
 - ii. Rat
 - iii. Mouse
- iv. Etc. b. Side

- i. Left
- ii. Right
- Sex
- i. Male ii. Female
- ii. d. Age

C.

- i. Embryo
 - ii. Fetus
 - iii. Neonate
 - iv. Juvenile
 - v. Adult
- vi. Elderly
- e. Environment
- f. Diet
- g. Etc.
- 2. Artificial (Artifacts)
 - a. Specimen Preparation
 - i. Fixation
 - 1. Fixative
 - 2. Emersion
 - 3. Perfusion
 - 4. Fixative Buffer
 - ii. Embedding Material
 - 1. Frozen
 - 2. Paraffin
 - 3. Methacrylate
 - 4. Epon; Araldite
 - iii. Section Characteristics
 - 1. Thickness
 - 2. Compression
 - 3. Lost Caps
 - iv. Stain
 - 1. Surface
 - 2. En Bloc
 - 3. Immunological
 - b. Collecting and Recording Data
 - i. Sampling Methods
 - 1. Unbiased
 - 2. Potentially Biased
 - ii. Collecting Data
 - 1. Identifying Boundaries of Structures
 - 2. Identifying Structures
 - 3. Etc.

Finally, this list can be used to put a number on the complexity generated by these natural and artificial factors. If we take 33 factors (subheadings) 12 (headings) at a time, we can estimate very roughly the number of possible combinations – each capable of defining a single structure with a unique complexity. Recall from EBS 2001 that $_{n}C_{r} = {}_{n}P_{r}/r!$. Substituting we get: ${}_{33}C_{12} = {}_{33}P_{12}/12! = 354,817,320$. Such a result suggests that the potential for complexity in stereological data – even when only a single structure is involved – can be very great indeed.

Unraveling Gene Function with Stereology

In the simplest of terms, stereology serves as a mathematical interface between the biological data we collect and the same data as it actually exists in a living animal. Since we – along with our methods - are trained to look at biology through the lens of a change model, we are well equipped to answer the question: "Does it change - yes or no?" However, such a question may become less relevant as the focus of research shifts toward explaining gene function. When the shift occurs, we will need new methods and new training to ask the more challenging question: "Exactly, how does it change?"

Let's set the stage. Unraveling gene function is a very big question consisting of many smaller ones. It is a sufficiently difficult question because the number of small questions will remain largely asymptotic until we begin to understand the origins of emergent properties in biology – things like being alive, thinking, reasoning, creativity, etc. Emergent properties, you may recall, continue to be one of the least understood properties of complex systems.

One way of approaching a complex system like biology is to figure out a way of predicting its products with equations. The task of writing systems of equations (algorithms) for predicting gene function can be a relatively straightforward process when based on a structural hierarchy. However, the more difficult part of the job is to define – realistically - the boundaries of a given prediction model.

Our first attempts at assembling prediction models suggested that the underlying order of biology is more than sufficient to meet the demands of such a building task (EBS 2001). There appears to be at least two general approaches to the problem. We can try to collect new "standard reference data" for all organs of a single species, or

we can piece together a prediction model based on data currently available to us in the stereology literature database. The advantage of the first option is that all the regression curves can have coefficients of determination equal to one $(r^2=1)$, whereas the disadvantage of the second option is that the curves will have r^2 only approaching one $(r^2\approx1)$. Since our only option today is the literature database, we will pursue a strategy that asks "small" questions of likely importance to both approaches.

How do we use stereology to predict gene function – from molecules to organisms and organisms to molecules?

We already know from DNA sequence analysis that genomes across animals show remarkable similarities, which no doubt explains our finding of widespread order – repeatedly - across biology (EBS 2001; 2002).

Recall that a genotype is the genetic complement of an individual, whereas a phenotype is the genetic expression of physical characteristics that define an individual. In short, genotypes produce phenotypes. A single genotype can produce multiple phenotypes (e.g., development; exposures) and many similar phenotypes can come from different animals. In short, animals share similarities and differences in genotype and phenotype. This, presumably, explains the presence (or absence) of structural and functional analogies across biology.

All animals (e.g., mammals)



Once again, we encounter the familiar problem of having to unfold complexity before we can tackle our problem. Now as before, our strategy consisted of first peeling away layers of complexity to get at the underlying principles and then putting the biology back together with equations. In the original EBS (2001), this approach allowed us to identify biological principles of connectivity and stoichiometry. Here we apply the same discovery strategy as a way of evaluating complexity in stereological data.

Methods

Searching for Patterns with Stereology

As introduced in EBS (2001), the process being used here to unravel gene function with stereology consists of finding patterns in biology that can produce equations, which, in turn, can predict events upstream and downstream from the genome. The most promising approach uncovered thus far consisted of reformatting the published data to accommodate questions that involve the complex interaction of many parts. Success with this approach depended on minimizing the effect of bias on stereological data, which was accomplished by generating new data – the connection types 1 to 4. An important limitation of these data, however, is that we may never know how much bias remains or how much the residual bias differs from one paper to another.

The EBS Project generates new information from the stereology literature and then uses it to ask complex questions about biology. A summary of the overall process appears below.



A standard strategy of the EBS Project consists of identifying a problem and then building a platform specific to its solution. In general, a platform depends on generating connections between and among data in ways that can minimize the disruptive effects of phenotype, bias, and time. The basic building blocks of these "designer" platforms include the data of four connection types.

Connection type 1: Plots one structure against another at several time points – at one hierarchical level – for one paper.

Build: The points are fitted to a regression line and a close connection is identified by a coefficient of determination approaching 1.0 ($r^2 \approx 1.0$). The results are expressed as the equation of a regression line, which includes linear (y=mx+a) and power ($y=bx^a$).

Application: : The data type is used to look for the connection patterns of several structures within a single paper.

Limitation: The patterns detected apply only to the data of a single paper.

Disruptive Variables:

Bias: Since data come from the same sections, bias will be similar for the two structures except when the size and shape of the structures introduce variability (Weibel and Paumgartner, 1978; Weibel, 1979). Similar bias has no effect on the equation of the curve, whereas a variable bias does (see appendix; Connection Types).

Time: The effect of time cancels out.

Phenotype: The phenotype may or may not change over time.

Examples: EBS 2001, 2002; appendix: Minimizing Bias in Stereological Data

Connection type 2: Plots many structures against many structures at one or more time points – at one hierarchical level – for one paper.

Build: The points are fitted to a regression line and a close connection is identified by a coefficient of determination approaching 1.0 ($r^2 \approx 1.0$). The results are expressed as the equation of a regression line, which includes linear (y=mx+a) and power ($y=bx^a$).

Application: The data type is used to look for connections between several structures within a single paper. Such information can: (1) identify groups of structures that change as a group over time, (2) point to mechanisms of coordinated control, (3) suggest the sequence of gene expression, (4) identify patterns in settings where genes are being turned on and off, and (5) distinguish between qualitative and quantitative changes in gene expression - new genes products vs. more of the same gene products. *Limitation:* The patterns detected apply only to the data of a single paper.

Disruptive Variables:

Bias: Since data come from the same sections, bias will be similar for the structures being connected except when the size and shape of the structures introduce variability. Similar bias has no effect on the equation of the curve, whereas a variable bias does (see appendix; Connection Types).

Time: The effect of time cancels out.

Phenotype: The phenotype may or may not change over time.

Examples: EBS 2001, 2002; appendix: Minimizing Bias in Stereological Data

Connection type 3: Plots many structures against many structures at one or more time points – at many hierarchical levels – for one paper.

Build: The points are fitted to a regression line and a close connection is identified by a coefficient of determination approaching 1.0 ($r^2 \approx 1.0$). The results are expressed as the equation of a regression line, which includes linear (y=mx+a) and power ($y=bx^a$).

Application: The data type is used to look for connections between several structures within a single paper. Such information can: (1) identify groups of structures that change as a group over time, (2) point to mechanisms of coordinated control, (3) suggest the sequence of gene expression, (4) distinguish between qualitative (new genes products) and quantitative (more of the same gene products) changes in gene expression, (5) detect patterns in settings where genes are being turned on and off, (6) spot shifts in bias produced by collecting data from different sets of sections (light vs. electron microscopy), and (7) sort out multiple sets of order, as defined by separate regression curves.

Limitation: The patterns detected apply only to the data of a single paper.

Disruptive Variables:

Bias: When data come from the same sections, bias will be similar for the structures being connected except when the size and shape of the structures introduce variability. Data collected from different sets of sections will be expected to carry bias unique to each section set. Similar bias has no effect on the equation of the curve, whereas a variable bias does (see appendix; Connection Types).

Time: The effect of time cancels out.

Phenotype: The phenotype may or may not change over time.

Examples: EBS 2001, 2002; appendix: Minimizing Bias in Stereological Data

Connection type 4 (data pairs): Plots one structure against another – at one hierarchical level – for one or more papers.

Build: The points are fitted to a regression line and a close connection is identified by a coefficient of determination approaching 1.0 ($r^2 \approx 1.0$). The results are expressed as the equation of a regression line, which includes linear (y=mx+a) and power ($y=bx^a$).

Application: The data type is used to look for connections between pairs of structures within a single paper and across all papers - in the stereology database. Such information can: (1) identify common connections across biology, (2) be used to build prediction machines (data replicators), (3) serve as building blocks for assembling sets of interlocking equations (biological algorithms), and (4) be used to predict gene function

starting from a single value and moving both upstream and downstream – all the way from molecule to organism.

Limitation: Each data pair may carry a residual bias unique to each structure and to each study. *Disruptive Variables:*

Bias: Since data come from the same sections, bias will be similar for the two structures being connected except when the size and shape of the structures introduce variability. Similar bias has no effect on the equation of the curve, whereas a variable bias does (see appendix; Connection Types).

Time: The effect of time cancels out.

Phenotype: The phenotype may or may not change over time.

Examples: EBS 2001, 2002; appendix: Minimizing Bias in Stereological Data.

Predicting Gene Functions and Products with Stereology

Prediction is a by-product of unfolding complexity. Progress toward solving a complex problem can often be advanced by breaking a large intractable problem into several smaller and simpler ones. For example, to predict gene function we can begin by imagining an outcome, transforming our imagination into a software machine, figuring out how it works, and then trying it out on real-world data.

What can we learn from such an exercise? A software machine designed to predict gene products works best in a real-world setting when four conditions are met: (1) the process runs in both directions – upstream and down stream (toward and away from the gene), (2) the biological algorithms have coefficients of determination (r^2) equal to 1.0, (3) bias is eliminated or at least minimized, and (4) the time variable is avoided.

1. Predict structures upstream and downstream

Folder 6 of the BIOLOGYtabs program includes eleven tabs: animal, adrenal gland, brain, heart, kidney, liver, lung, ovary, pancreas, pituitary, and testis. Each tab includes at least two sets of biological algorithms (upstream; downstream) assembled from the data of a published paper. The biological algorithm represents a set of power equations (y=bx^a) connected across levels of the biological hierarchy by sharing a similar x or y value. Each level of the hierarchy may contain one or more power equations. Here, the coefficient of determination of all the regression equations equals 1.0 and the algorithms predict the published data exactly – paper by paper.

The eleven tabs in folder 6 include about 500 power equations all of which can be connected by rule. Enter a single value into any one of the 500 or so data entry fields (x) and we can predict the structure of an animal, which, in the example, includes eleven organs (adrenal gland, brain, heart, kidney, liver, lung, ovary, pancreas, pituitary, and testis). Folder 6 shows how we could generate an entire animal or a specific gene product from a single seed value – located anywhere within the animal. Recall the key point. To establish a standardized bias across all structures, all data would have to be collected from the same set of animals.

2. Coefficients of determination equal to 1.0

Recall from EBS (2001) that the biological algorithms were generated from data pairs developed with control data taken from one or more papers. Although the coefficients of determination (r^2) often approached 1.0, they were seldom equal to 1.0. Herein lies the limitation of a literature database approach to prediction algorithms. An almost best solution can be expected to work reasonably well over several connections, but certainly not when the number of connections extends to hundreds or thousands. In such a setting, even a slight deviation from 1.0 will promptly diminish the reliability of a prediction.

The best prediction model includes *only* biological algorithms with coefficients of determination equal to one $(r^2=1.0)$. This is a helpful piece of information because it defines the problem to be solved. In EBS (2001), we discovered that control data, which typically clumped around a mean, could be fitted to a straight line by plotting data using a collection of data types - density, structure and mean structure. Recall, however, that for a given level of the hierarchy these three types of data differ only by a constant and that one data type can be readily calculated from another. This means that the ratio of the data pairs remains constant across the different data types. For example,

density/density = structure/structure = mean structure/mean structure

$$(V_i/V_{ref})/(V_j/V_{ref}) = V_i/V_j = meanV_i/meanV_j.$$

An analogous result occurs when we start with two data values and divide them by 10 and by 100 before taking the ratios:

$$(V_i/V_j) = (V_i/10)/(V_j/10) = (V_i/100)/(V_j/100).$$
 (2)

The key point to note here is that both expressions (1, 2) can give the same power curves, because the curves differ only by constants that factor out. Therefore, we can express each data point within a hierarchy as a power curve, having an r^2 equal to 1.0. This allows us to write a set of biological algorithms for a given paper that will predict all the data of that paper from any single value – exactly as published. All the algorithms in folder 6 were written accordingly. *Bear in mind, however, that all the algorithms in folder 6 came from different animals, each carrying a unique complexity.*

(1)

3. Minimize bias

A key pattern emerging from the stereological database is that structures consistently demonstrate a strict adherence to rules of organization when bias is minimized. The evidence for this observation comes directly from regression curves. Data combined from similar and different animals routinely gave curves with r^2 approaching 1.0. One conclusion to take from this observation is that the design plan for organs - including the structures within - is highly conserved across many different animals.

4. Avoid time

By plotting only one structure against another, the time variable can be excluded from the result.

Results

Enterprise Biology Software (2002)

EBS (2002) includes a stand-alone set of programs called **BIOLOGYtabs**, as well as an upgrade to the original package - EBS (2001). In summary, the new software includes (1) screens for accessing published stereological data quickly and easily, (2) new views of the stereology literature (four connection types), (3) tools for finding patterns in biology, and (4) a working model for predicting large data sets with biological algorithms – consistent with gene function.

BIOLOGYtabs: The software includes a collection of folders summarizing the literature of biological stereology from 1965 to 2001. **BIOLOGYtabs** can be installed two ways: (1) as a stand alone program without a client/server configuration (the data are bundled with the program code) and (2) as an addition to the appendix of EBS (2001), as part of the EBS Upgrade (2002).

BIOLOGYtabs includes seven folders that can be opened by clicking on one of the yellow command buttons located on the table of contents (see below). The **install reader** button runs a program that installs a runtime viewer for PDF files. The **help** button (1) displays directions for running the programs and (2) includes practice exercises for advanced users interested in running complex database queries.



Each folder has several tabs, e.g., 1.1 and 1.2. Clicking on a tab displays a new screen.

1.0 Papers



1.1 by citation: The citation screen includes tools for finding papers using any one of the items stored in two database tables: citation and author. To select papers, type items into data entry fields (yellow), pick them from drop down lists (gray buttons), or use the sort and filter buttons (user-defined). The sort screen uses drag and drop; one or more items can be moved from the "Source Data" window to the "Columns" window and either an ascending (checked) or descending (not checked) sort can be selected. Click on the OK button to run the sort. If you have Internet access and wish to read an abstract for a paper, click on the Abstract button.

The **filter screen** uses scripts written by the user. A script consists of a column name, relational operator (=, >, <, <>, etc) and values against which column values are compared. Boolean expressions can be connected with logical operators AND and OR. Text values are surrounded by quotation marks, whereas numerical values have none. Examples of simple scripts appear below.

Query: Get citation number 3171 **Filter script:** cit_citation_1_cit_nu = 3171

Query: Get all papers with citation numbers greater than 3000 **Filter script:** cit_citation_1_cit_nu > 3000

Query: Get all papers published by Australian authors in 1999 **Filter script:** cit_citation_1_cit_year = 1999 AND aut_author_aut_country = "AUSTRALIA" **Query:** Get all papers published by Swiss authors with data entered into the literature database. **Filter script:** aut_author_aut_country = "SWITZERLAND" and cit_citation_1_cit_data_transfer = "yes"

1.2 **by method:** The methods screen links three database tables: methods, citation, and authors. Use the buttons to select items from lists or filter the data files with scripts. For example:

Query: Get all the papers on the brain using frozen sections, the optical fractionator, and published since 1995.

Filter script: co_organ_co_organ_name = "brain" AND met_method_embedding = "frozen" AND met_method_counting_method = "optical fractionator" AND cit_citation_1_cit_year >= 1995

Query: Get all the papers using glutaraldehyde fixative in 1999. Hint: Select glutaraldehyde from the fixative list, click on the Filter button, and add <AND cit_citation_1_cit_year = 1999> Filter script: match(met_method_fixative, '[g][1][u][t][a][r][a][1][d][e][h][y][d][e]') AND cit_citation_1_cit_year = 1999

With a little practice, you can learn to write scripts that locate exactly the papers you want to read - very quickly and efficiently.

2.0 Hierarchy

The hierarchy folder contains three picture buttons that run two updated programs (2.1, 2.2) and a new program (2.3). Together, these programs have defined the hierarchical framework of the stereology literature database.



- 2.1 **by organ system:** The hierarchy browser, which was first introduced in the **Human Biology Course** (EBS 2001), displays a structural hierarchy similar to those found in textbooks. It along with the brain/cord browser served as a preliminary guide for setting up the hierarchies used for entering data into the literature database.
- 2.2 **by brain and spinal cord:** The brain/cord browser includes six hierarchical levels extending from organ to organ subcompartment five (osc5). When using it to identify a hierarchy for data entry, the easiest way of finding an item is to run a global search. This consists of typing in the name of a brain structure (or the first few letters thereof; e.g., hippo for hippocampus) into the global search field and pressing Enter. In response, several white data entry fields turn yellow. Click on these yellow fields until the structure of interest appears (e.g., hippocampus), along with its associated hierarchy.
- 2.3 by literature database: The hierarchy tables of tabs 2.1 and 2.2 above provided a theoretical framework for standardizing data entry. In practice, however, the hierarchies eventually used in the database represented a compromise between these theoretical schemes and the preferences of authors. This browser lists all the hierarchies written thus far for control data entry, which amounts to 7,463. When entering data from a new paper, the format used earlier for a paper with similar data can be quickly found by running a global search and then reused. A word of caution. When using this screen, expect to find inconsistencies in that the same structure may appear at different hierarchical levels. Such a problem is unavoidable when standardization is a dynamic process, as it is here. Moreover, it may be advantageous to use more than one hierarchy for the same organ to accommodate the type of data being entered (e.g., kidney anatomy vs. kidney physiology). In short, deciding on the "best" hierarchy for a given paper is a trade off, one that has both advantages.

3.0 Control Data

The folder contains lists of control data ordered hierarchically and standardized.



3.1 by data and data point: To use these screens efficiently, you need to know exactly in what hierarchy to look for the structure you want. This is accomplished by consulting a hierarchy screen, which is called by clicking on one of the three buttons located on the right side of the screen. Once you know where to look in the structural hierarchy, browse through the file using the drop down list (located in upper left hand corner of the screen), select specific data points with the "Search_<level>" field (enter a word; press Enter), or advanced users can write scripts (click on the Filter Button). Examples of the scripts include:

Query: Get all the volume-weighted mean volumes (MeanVv) for the nucleus in the cell compartment (CC).

Filter script: co_data_1_co_mean_v_weighted >0

Query: Get all data for hepatocytes (use the cell (C) screen).

Type hepatocyte into the "Search_<level>" field and press Enter. Alternatively, click on the drop down list box and scroll to a specific type of hepatocyte.

3.2 **by paper:** Click on the picture button to view standardized papers for the control data. When the screen displays, type a citation number in the data entry field (cit_nu) in the upper left hand corner of the screen. If you need to look up a citation number, click on the **List Citations** button. Caution! This program runs only when EBS 2001 is installed on the same computer or when BIOLOGYtabs is run from the Appendix of EBS 2002. The **by paper** program requires the database engine.

4.0 Experimental Data

The folder contains lists of experimental data ordered hierarchically and standardized. The text given for the control data of tab 3.0 applies here as well.



- 4.1 by data and data point: See 3.1 above.
- 4.2 **by paper:** See 3.2 above.

4.3 by percentage change

- 4.4 **by control and experimental data:** Use this screen to view individual values of controls matched to experimental values.
- 4.5 by control and experimental data: Use this screen to view percentage changed color coded: increase (red), decrease (blue), and no change (green).

5.0 Order in Biology

Use this folder to browse stereological data expressed as connections (local; global)



- 5.1 **by connection (local; connection types 1-3):** Use this screen to look for quantitative relationships within a given paper expressed as linear regressions. For each paper, all possible combinations of data were compared row-by-row and column-by-column by calculating regression curves. Each row of the database table includes an equation that defines the relationship between the x and y variable. Entering a value for x value will generate a value for y.
- 5.2 by data pair (connection type 4): A data pair includes two pieces of data expressed as an x and y value. When data pairs having similar x and y names are selected and used to calculate a regression curve, the closeness of the relationship between these variables can be determined within and across animals. The position of the plotted point identified by the x and y values is determined by the relative amount of each value plus the bias accompanying each value. If the bias is largely the same for both values then it has little or no effect on the proportion of two values. In other words, bias can move the point up or down the regression curve, but the point remains steadfastly on the curve (see the worked examples in the appendix). Such an analysis allows us to look for quantitative relationships between two structures when bias is minimized. The updated literature database now includes about 12,000 data pairs.
- 5.3 by connection (global): Use this screen to view examples of linear regressions formed from type 4 connections (data pairs). The data pairs for a given curve may have come from one or several animals of similar or different species. As in 5.1, each row of the table includes an equation that defines the relationship between the x and y variable. Entering a value for x value will generate a value for y.

6.0 Predicting Gene Function

This folder includes a system of equations that together can define the structure of an organism. By introducing a single data value at any point within the animal (here the option includes 500 points at 11 locations) the remaining

499 points can be predicted. Predictions can be made upstream (toward the genes) and downstream (away from the genes). The values already entered came from the original papers.

The process of building an "organism" from a single seed value consists of predicting data downstream and upstream from the seed and using the organ volume to predict the volumes of other organs. Note: Additional connections between organs can be found in the literature database.



- 6.1 animal
- 6.2 adrenal gland
- 6.3 brain
- 6.4 heart
- 6.5 kidney
- 6.6 liver
- 6.7 lung
- 6.8 ovary6.9 pancreas
- 6.10 pituitary
- 6.11 testis

7.0 Enterprise Biology Software Project



- 7.1 Progress Report (2002) 7.1.1 Research – EBS 2002
- 7.2 Progress Reports (2001)
 7.2.1 Research from EBS 2001
 7.2.2 Education– from EBS 2001
- 7.3 Installation: includes instructions and trouble shooting
- 7.4 Licenses
 - 7.4.1 Single license
 - 7.4.2 Site license

EBS Upgrade (2002): The upgrade to EBS 2001 is distributed on the EBS (2002) CD in two folders. A list of the files contained therein can be found in the installation document (readme_install_bt.pdf), along with instructions for the upgrade. The main features of EBS (2002) include:

- Upgrade of stereology literature database (1965 to 2001)
- BIOLOGYtabs
- Annotations for structural hierarchy (control data)
- Improved error correction (control data)
- New tools for generating connection data
- Improved interface screens
- New PDF files

Discussion

Enterprise Biology Software (2002)

The first release of the stereology literature database (EBS; 2001) hinted at a biology overflowing with mathematical order. In this release, the pattern persisted while gaining greater focus and resolution. This should not come as a surprise. In a population of animals with largely similar genes, finding order that could be translated into general equations was more or less expected. Given the recent sequencing of the mouse genome, we now know that for nearly every gene in man a counterpart can be identified in the mouse. Both species have about 30,000 protein coding genes with only 300 unique to either – a difference of merely 1% (Waterston et al., 2002).

If the genome and the mathematical expression of that genome are widely similar across animals, then working out the equations for an animal seems a natural extension of the DNA sequencing projects. Such reasoning creates an opportunity for us in that stereology becomes the method of choice for showing that animals run largely on the same – or similar - sets of equations. For example, building a prediction machine for a mouse similar to the prototype of folder 6 could become a powerful platform for stereology – one that can interface directly with modern genetics. Indeed, such a connection might allow us to explore events at a level of detail far beyond the current reach of either stereology or molecular biology.

Unfolding complexity: The discussion in the appendix of this paper describes a biological stereology bristling with bias. Moreover, it appears that much of this bias reflects an *indivisible complexity* because the bias cannot be separated into its component parts. For a given stereological estimate, there are simply too many corrupting variables - each displaying a variability unique to its immediate setting.

This raises an interesting question. Given the remarkable similarity of genomes across different animals, how do we explain a persistent "biological" variation of roughly 10 to 20% for the same animals? Is it possible that roughly 1% of this variation comes from differences in phenotype, with the rest coming from the bias of our experimental methods? Soon, such a question might be answerable experimentally. For example, it should be possible to separate bias from phenotype by collecting stereological data - at the same time - from both cloned and wild type mice. All other things being equal (including bias), the differences between the two animal estimates may reveal the true biological variation.

The EBS Project looks - persistently - for ways of unfolding complexity. The first release standardized the stereology literature by organizing the data within the framework of a relational database. This operation changed the literature fundamentally in that it allowed stereological data to assume a more active role in unfolding complexity. Here we used the database to focus on a *divisible complexity* by unfolding three disruptive variables in stereological data: time, bias, and phenotypic variation. This was accomplished by selecting papers based on design-based sampling methods, calculating all possible outcomes of an experiment, and generating new data types based on connections.

Refolding complexity: Starting with the unfolded data - minimized for the effects of bias, time, and phenotype - new data products were assembled and tested with specific goals in mind. Examples of these new assemblies – defined as data connections - included data replicators and biological algorithms.

Furthermore, these new connection data allow stereology to serve as a wide-angle lens for viewing gene function. For example, each structural change that occurs in parallel (e.g., time_i vs. time_j for data types 2 and 3) predicts an underlying parallel activity of hundreds (thousands?) of genes programmed explicitly to produce growth. A key point to take from this hierarchical assembly process is that stereological data – as downstream information – can summarize enormous amounts of gene function with a single equation.

Since we already know that biology routinely uses a variety of backup systems, these equations and the connections they identify may become quite useful for detecting overlapping control mechanisms. Moreover, connections may prove invaluable for identifying coordinated regulation between and among assemblies of genes during expression.

Shifting perspectives: An intriguing property of the EBS Project is that it continually shifts our perspective of biology from one reference platform to another. When we move data from the pages of a journal to the tables of a relational database, for example, static data become active and interactive. In turn, stereological data designed for a change model can be reconfigured to accommodate a connection model. A connection model becomes a platform for finding patterns that lead to a new platform of biological algorithms consisting of interlocking equations. Each time we move to a new platform, we find a new opportunity for discovery.

One lesson to take from this excursion is that discovery in biology is largely an exercise in understanding the roots of complexity – both natural and artificial. For example, we can be reasonably certain that stereological data carry multiple layers of complexity – such as bias, time, and phenotype - that together obscure underlying patterns. Separate these complexities one from another and patterns appear practically everywhere we look. In effect, shifting from one platform to another is analogous to looking at new things in a new place through a new set of eyes.

Predicting Gene Function

One of the promising new applications of stereology will be to diagnosis and predict gene function. Thus far, the Enterprise Biology Software (Bolender, 2001a, 2001b, 2002) has presented numerous strategies, models, and examples of using stereological data well beyond the limits currently imposed by the change model. However, a major challenge remains. It now seems clear that large scale prediction will depend – by necessity – on our ability to assemble systems of equations that all have $r^2=1$.

Folder 6 of BIOLOGYtabs demonstrated that we could generate biological algorithms for a research paper, which, in turn, could be used to predict data with a high degree of reliability ($r^{2}=1$). We also know from EBS (2001, 2002) that these algorithms can generalize across different animal types, as shown by r^{2} values equal to or approaching 1.

If we set $r^{2}=1$ as the starting condition for the prediction algorithms, then several courses are open to us. The direct approach would be to collect large amounts of stereological data from the structures of a single animal. To this end, the mouse might be a good choice because the genome has been sequenced and clones will soon become readily available. Such a data set would provide $r^{2}=1$ predictions for an animal having a standardized bias. Moreover, clones may offer a promising new strategy for tiding up the problem of bias in biological stereology. Alternatively, we could use data already in the stereological database, but figure out a way to identify general equations with r^{2} s closer to one. Although both options present a formidable challenge, prospects for success would now seem to be clearly in our favor.

Concluding Comments

Checking on the bottom line: One of the most noticeable trends in biological stereology is that the number of papers reporting "no change" as the principal finding is increasing – especially for the nervous system. This observation encourages us to estimate just how effective the model really is. If we take the number of data points in the database showing a change and divide that by the total number of data points, then an estimate of effectiveness can be calculated as: change data/total data = 3019/18351 = 0.1645. The result tells us that the change model as a discovery platform is successful 16% of the time – or that the odds of getting a positive result (increase, decrease) for a stereological estimate are one out of six. However, 16% may be a generous estimate because of the nature of statistics. While a statistical test can demonstrate a significant difference between two data points, it cannot determine whether a difference was produced by a change or a bias – or both.

One way of improving our chances for a successful experimental outcome is to create new uses for stereological data by employing additional analysis models. This strategy appears to work. Using a connection model, for example, 15,000 new pieces of data were produced that defined new applications, with new outcomes, at a cost of practically zero. Indeed, it now seems likely that the amount of new data generated by the database will soon exceed that of the original data published in the stereology literature.

While all this talk about discovery platforms may be interesting, a far more entertaining approach would be for you to assemble a new platform – on your own. If we assume that the bottom line in science is discovery, then the change model was successful 3019 times out of a possible 18,351 (see above). In contrast, type 4 data of the connection model may offer a more compelling strategy. If we agree that it takes three points to identify a regression curve with an $r^2 \approx 1$, then we can estimate the number of possible combinations that can be taken from the 12,000 data pairs as: ${}_{n}C_{r} = {}_{n}P_{r}/r! = {}_{12,000}P_{3}/3! = 287,928,004,000$. In other words, a data set designed to look for connections is potentially 15,690,004 times larger than one designed to look for changes. There's more. Generating combinations from the data pairs gives a connection library that can be screened for specific outcomes or merely serendipitous ones. One might discover, for example, entirely unexpected relationships between structures located all cross the biological hierarchy. Moreover, such a library may prove to be a critical resource when it becomes important to verify or predict the structural and functional consequences of modifying genomes.

Starting with the 12,000 data pairs (see tab 5.2 of BIOLOGYtabs), you can build a discovery platform of your own by writing a small program that selects all those combinations that produce curves with $r^2 \approx 1$. In turn, you can use these results to hunt for patterns by inspection or perhaps use them as a training set for a neural network.

Bias as the defining variable: Once we define bias as the difference between a value for a structure in a living animal and a journal publication, then we can argue successfully that all stereological estimates for biology are likely to carry an unknown bias. This means that certain data collected with modern stereological methods may be vastly superior to data collected with classical methods – in theory, but only slightly better in practice. Such an outcome would be expected when the principal source of bias is produced by the histological preparations – not by the stereological sampling methods.

Making a case for using classical data sets may become a practical necessity. A clear and current trend in biological stereology is to count cells with light microscopy and little more. Publishing a detailed ultrastructural analysis of a control organ has become unfashionable and may no longer be fundable. In fact, we may have to make do with what we already have.

On being a data driven discipline: Biological stereology now qualifies as a scientific discipline with standardized research data that can be viewed and interpreted with a variety of software applications. The implication of this new resource is that a discipline traditionally driven by technology (methods) can now be driven by data as well. In other words, we can explore biology by doing experiments with animals *and* by running experiments with data in the database. Eventually, as our database grows in size and maturity, predicting outcomes of proposed wet lab experiments will become routine. The move toward a data driven discipline suggests that we may have reached the

point where we now believe that the data of a single experiment can no longer capture – by itself - the complexity of the questions we are trying to answer

Looking to the future: If we can make the case - convincingly - that experimental animals running on the same genes also run on the same equations, then we may find ourselves in the fortunate position of being able to solve wonderfully difficult puzzles.

References

Baddeley, A. J., H. J. G. Gundersen, and L. M. Cruz-Orive. 1986 Estimation of surface area from vertical sections. J. Microsc. 142:259-276.

Bertram, J. F., P. D. Sampson and R. P. Bolender. 1986 Influence of tissue composition on the final volume of rat liver blocks prepared for electron microscopy. J. Electron Microsc. Tech. 4: 303-314.

Bolender, R. P. 2001a Enterprise Biology Software I. Research (2001) In: Enterprise Biology Software , Version 1.0 © 2001 Robert P. Bolender

Bolender, R. P. 2001b Enterprise Biology Software II. Education (2001) In: Enterprise Biology Software , Version 1.0 © 2001 Robert P. Bolender

Coggeshall R.E. and H.A. Lekan. 1996 Methods for determining numbers of cells and synapses: a case for more uniform standards of review. J Comp Neurol: 364(1):6-15.

Dekel B., T. Burakova, F.D. Arditti, S. Reich-Zeliger, O. Milstein, S. Aviel-Ronen, G. Rechavi, N. Friedman, N. Kaminski, J.H. Passwell, Y. Reisner. 2003 Human and porcine early kidney precursors as a new source for transplantation. Nat Med 9(1):53-60

Gundersen, H. J. G. Stereology of arbitrary particles. 1986 A review of unbiased number and size estimators and the presentation of some new ones in memory of William R. Thompson. J. Microsc. 143: 3-45.

Hatton W.J. and C.S. von Bartheld. 1999 Analysis of cell death in the trochlear nucleus of the chick embryo: calibration of the optical disector counting method reveals systematic bias. J Comp Neurol 409(2):169-186.

Waterston R. H., et. Al. 2002 Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520-562

Weibel, E.R. 1979 Stereological Methods, Vol. 1. Practical Methods for Biological Morphometry. Academic Press, London.

Weibel E. R. and D. Paumgartner. 1978 Integrated stereological and biochemical studies on hepatocytic membranes. II. Correction of section thickness effect on volume and surface density estimates. J Cell Biol 77(2):584-597.

Appendix

Unbiased Stereology - Putting It to the Test

Search the Internet for "unbiased stereology" and one is treated to a fine summary of the many and excellent benefits of design-based methods. After reading this material, however, one is left with the uncomfortable impression that the application of these methods produced only unbiased results. Indeed, little or no distinction is made between the methods (preparative; sampling) and the products (data). As an alternative, search on "fixation bias" or "fixation artifacts" and a more realistic picture will appear.

Our discussion of bias depends on how we decide to package it. If we agree to a common starting point, then we can avoid much of the usual controversy. Here, for example, we have defined two reference points as being essential to our discussion of bias – the living animal (start) and the published data (finish).

If, however, we decide to evaluate bias from a starting point other than the living animal, then we may be accepting as true – unwittingly - one or more of the following assumptions. Although some or all of the examples in the following list may appear ludicrous, all of us may have accepted – tacitly - one or more of them in our papers.

Question: Can unbiased methods produce unbiased data?

Answer: Yes, but only when the following statements are true or do not apply to the data being produced.

- Estimates made by different investigators for the same structure(s) are all unbiased.
- Estimates taken from an organ embedded with different embedding materials are all unbiased.
- Estimates taken from sections with section thickness are unbiased.
- Estimates taken from sections with section compression are unbiased.
- Estimates taken from organs with shrinkage or swelling are unbiased.
- Estimates taken from organs fixed with different fixatives are all unbiased.
- Estimates taken from sections stained with different stains are all unbiased.
- Estimates taken from control and experimental samples are unbiased even when the shapes and sizes of organelles have changed.

Given our definition of the start and finish points, the relationship between unbiased methods and unbiased results can be summarized as follows. *Unbiased methods applied to unbiased structures produce unbiased data, but unbiased methods applied to biased structures produce biased data.* Once again, biology resembles physics in that it may be impossible to observe anything without interacting with it and thus changing it. In other words, biology appears to have its own version of the Heisenberg *"uncertainty principle.*"

Sources of Bias in Stereological Data

Bias: An unbiased estimate exists when the mean of the estimates converges on the true mean. Bias can be defined as the deviation of results from the truth. In other words, with bias the mean does not converge on the true mean. The critical element of these definitions is the meaning of the word *true*. *True* may be likened to that which occurs in the living organism or to some stage of tissue preparation. Here, *true* refers to the living organism. The point of the following discussion is not to offer a comprehensive review of bias, but rather to list examples of familiar methods producing bias. Such a discussion seems relevant because of the success of the connection model. In a complex setting, one way of managing bias is to make it as uniform as possible – across the sampling space.

- 1. **Estimating volume:** The volume of a structure before (living) and after fixation (dead + fixed) are not necessarily the same.
 - a. **Displacement:** When applying the Archimedes's Principle, large structures are not usually influenced by the evaporation of water from containing vessel, whereas small structures are (Bertram et al., 1986). Movement of water into or out of the tissue initiated by the osmotic properties of the displacement fluid or fixative can influence volumes and subsequent estimates.
 - b. *Cavalieri:* Cavalieri estimates of volume can be influenced by the state of the tissue (fresh, fixed, frozen, embedded) and by the ability of the investigator to identify accurately the edges of the structure. This ability to define boundaries may depend on the methods of fixation, staining, embedding, and sectioning.
- 2. **Fixing tissue:** Typically, the overall effect of fixation is shrinkage, but both shrinkage and swelling can occur simultaneously in different compartments (Bertram et al., 1986). For example, the volume of an organ before and after fixation can be identical when the shrinkage of some compartments (e.g., cells) is balanced by the swelling of other compartments (e.g., connective tissue).
 - a. **Perfusion; installation; emersion:** The volumes of structures, tissue blocks, and tissue compartments (cells; interstitium) can be influenced by the procedure used to expose the living tissue to the fixative. For example, perfused tissue can be expected to exhibit uniform fixation throughout a tissue block, whereas tissue fixed by emersion typically displays a variable fixation. Indeed, the quality of the fixation determines the reliability of the subsequent stereological estimates.
 - b. *Ionic concentration:* The tonicity (isotonic, hypertonic, and hypotonic) of the fixative can lead to important volume changes in the tissue. Moreover, the different compartments within a tissue can respond differently to the fixative some swelling and others shrinking.

- c. *Fixative buffer:* The buffer used for the fixative influences the appearance of structures, which, in turn, influences our ability to recognize structures and to collect data from images of sections. For example, the same tissue fixed with a collidine, phosphate, or cacodylate buffer often appears differently.
- d. *Fixatives:* Different fixatives fix the same tissues differently. The choice of fixative and fixative buffer can influence stereological estimates. For example, a fixation protocol designed to maximize the identification of the ER (en bloc staining) can be expected to give higher estimates than one that does not. In contrast, Karnovsky fixative by retaining cytoplasmic matrix protein can obscure small membranes such as the smooth-surfaced ER. This would produce an underestimate.
- 3. **Embedding tissue:** Embedding tissue blocks typically results in shrinkage the loss of volume. Paraffin embedding produces the greatest shrinkage, with smaller amounts for methacrylate, Epon, and Aryldite. Since each embedding material can affect a tissue volume differently, combining volumes from the same tissues embedded differently or using volumes coming from embedded and non-embedded tissues can influence stereological estimates importantly. In fact, combining incompatible reference volumes is a common practice in biological stereology (Weibel, 1979).
- 4. Sectioning tissue: Unbiased stereological methods using two-dimensional probes assume that data are collected from representative sections that have no thickness. However, most of the sections used for collecting stereological data are cut on a microtome and have a real thickness. When viewed in a light or electron microscope, all or part of the contents of the section can be seen in the image depending on the magnification and the microscope. This means that stereological data collected from microtome sections typically overestimate the true value. Sectioning compression and staining also introduce problems. When sections are compressed, more information is concentrated into a smaller area. This leads to overestimates for surface and length densities, but usually not for volume and numerical densities. Recently, a new sectioning artifact has been reported for the optical disector method (Hatton and von Bartheld, 1999). Finally, staining determines what we can see in a section and how much. Surface staining penetrates some distance into the section, whereas en block staining tends to stain a tissue block throughout. This explains at least in part why different staining procedures can give such different estimates.
- 5. Collecting data from sections and micrographs: Two people using the same set of micrographs or slides will not necessarily get the same set of stereological estimates. This holds true for people working in the same or different laboratories. Recognizing structures in sections is an acquired skill and one that is subject to great variation. This variation or counting bias from paper to paper may be the major source of bias for many structures.
- 6. **Publishing data:** Making mistakes when reporting data in publications occurs with some unknown frequency, but we as readers generally accept published data as being correct.
- 7. Entering data into the database: Reading a paper, interpreting it correctly, translating graphs into numbers, making data tables, recalculating results, and typing values into data entry forms are all subject to error. Since I am continually finding and correcting data entry errors, my skill at making them presumably remains untarnished.

Minimizing Bias in Stereological Data

Since there are so many potential sources of bias and so few design-based methods for correcting them, we are left with basically two choices: accept defeat and ignore bias or minimize the effects of bias.

The familiar stereological equation used to estimate structural data is more complicated than it first appears. For example, the volume of mitochondria in a liver is calculated as:

 $V_{(mitochondria, liver)} = V_{(liver)} \times V_{V(parenchyma/liver)} \times V_{V(hepatocyte/parenchyma)} \times V_{V(mitochondria/hepatocyte)}.$

In reality, however, an unknown and complex bias accompanies each variable:

 $V_{(\text{mitochondria,liver})} = \{V_{(\text{liver})} \times \Sigma \text{ unknown bias } \#1\} \times \{V_{V(\text{parenchyma/liver})} \times \Sigma \text{ Unknown bias } \#2\} \times \{V_{V(\text{hepatocyte/parenchyma})} \times \Sigma \text{ unknown bias } \#3\} \times \{V_{V(\text{mitochondria/hepatocyte})} \times \Sigma \text{ unknown bias } \#4\}.$

An unknown bias includes a collection of individual biases – each capable of producing an overestimate or an underestimate. It gets worse. A significant difference between control and experimental estimates for mitochondrial volume increases the uncertainty of an unknown bias from four to eight sources (4 control + 4 experimental). If we accept a significant difference as true, then we accept – by default – that the effect of bias on the outcome is unimportant.

Experimental data: When running an experiment that looks for a change, we have learned to ignore the influence of bias on the assumption that both the control and experimental data are similarly biased.

True control value \times bias_{control} = control value (published data)

True experimental value \times bias_{experimental} = experimental value (published data)

Change = experimental value (published)/control value (published data)

where we assumed that: $bias_{control} \approx bias_{experimental}$

This convention is widely practiced in biological stereology and based on the assumption that the control and experimental bias will be similar and therefore cancel – at least some of the time. The uncertainty attached to such

an argument is that we never know when it is true and when not. In short, we knowingly put our results in harms way by attaching the validity of our experimental results to the mercy of an unknown bias. The constant threat of an unknown bias can produce very annoying questions. For example, is it possible that a significant difference between a control and experimental time point could be explained entirely by a difference in the control and experimental bias? The answer, of course, is yes it can.

Control and experimental data: In biological stereology, bias accumulates throughout the process of specimen preparation. Preparation bias is variable from paper to paper and may be largely responsible for many of the inconsistencies seen throughout the stereological literature. Although we can be certain that most stereological data will carry a bias, the amount of the bias remains entirely unknown – at least at present.

Correction factors for bias may not be very effective and even may contribute additional bias; recall the earlier discussion of tissue shrinkage and swelling. However, we have two options. We can use methods that were designed to minimize bias, such as the fractionator, or we can encourage some of the bias to cancel out.

When two data values share a similar bias (i.e., they share the same specimen preparation), dividing one value by the other removes bias. In effect, bias is treated as a constant that cancels. Consider structures X and Y having a similar bias:

 $(Y \times bias_Y) / (X \times bias_X) = Y / X,$

assuming that $bias_X \approx bias_Y$

The remaining bias will be related to differences between the two structures (e.g., size and shape) and to differences in specimen preparation (control vs. experimental data). In the literature database, examples of stereological data with minimized bias include the percentage change data and the four connection types. The following worked examples show how the process of minimizing bias works.

Connection Type 2 (many structures vs. many structures at one level); Connection Type 3 (many structures vs. many structures at many levels): Connection types 2 and 3 pairs avoid bias by plotting two sets of values (t_0 and t_1) with similar bias and then treating the bias as a constant. A plot of true and biased values illustrates the effect of bias on the connection types. Consider true values at control (X_1, X_2, X_3) and experimental (Y_1, Y_2, Y_3 ,) time points (t_0, t_1) where:

 t_0 Values X₁ (true) = 10 X₂ (true) = 12 X₃ (true) = 14

 t_1 Values Y_1 (true) = 12 Y_2 (true) = 14

 Y_3 (true) = 16

The control values t_0 are plotted against the control values (as a reference) and against the experimental values (t_1). The plot:

- Shows the experimental curve $(t_0 \text{ vs. } t_1)$ parallel to the reference curve $(t_0 \text{ vs. } t_0)$ and
- Indicates an increase



Next, a 10% bias (×1.1) is added to the control and experimental data and plotted.

Control Data X_1 (biased) = 10 × 1.1 = 11 X_2 (biased) = 12 × 1.1 = 13.2 X_3 (biased) = 14 × 1.1 = 15.4 Experimental Data Y_1 (biased) = $12 \times 1.1 = 13.2$ Y_2 (biased) = $14 \times 1.1 = 15.4$ Y_3 (biased) = $16 \times 1.1 = 17.6$

Notice that the addition of a uniform bias changes the relative position of the curves, but that the curves remain parallel and display a roughly similar change: y=x+2 vs. y=x+2.2. Here the useful information is that the slope (x) does not change. This tells us that there is simply more of the same things (e.g., uniform growth).



Data pairs (Connection Type 4): Data pairs avoid bias by plotting two values (X and Y) with presumably similar bias and then treating the bias as a constant. A plot of true and biased values will help to illustrate the effect of bias on the data pairs. Consider four true values X_1 , Y_1 , X_2 , Y_2 where:

 $X_1 (true) = 10$ $Y_1 (true) = 20$ $X_2 (true) = 40$ $Y_2 (true) = 80$

A plot of these true values gives the following curve.



If we add a shrinkage bias of 10%, then the observed values underestimate the true ones.

 X_1 (biased) = $10 \times 0.9 = 9$ Y_1 (biased) = $20 \times 0.9 = 10$ X_2 (biased) = $40 \times 0.9 = 36$

 Y_2 (biased) = 80 × 0.9 = 72



When the two curves are combined, we can see the effect of the bias on the true curve. In fact, the equation of the line does not change – it is still y=4x. The biased curve is superimposed on the true curve, but displaced down and to the left because of the bias.



However, if the amount of bias in the X and Y values differ, then the data pairs may over or under estimate the true values. For example, a:

- 5% difference between the biased data underestimates the true value by 6%.
- 10% difference between the biased data underestimates the true value by 11%.
- 15% difference between the biased data underestimates the true value by 17%.

Bias differs by 5%

- X_1 (biased; 10%) = 10 × 0.9 = 9
- Y_1 (biased; 15%) = 20 × 0.85 = 17
- X_2 (biased; 10%) = 40 × 0.9 = 36
- Y_2 (biased; 15%) = 80 × 0.85 = 68

Bias differs by 10%

 $\begin{array}{l} X_1 \text{ (biased; 10\%)} = 10 \times 0.9 = 9 \\ Y_1 \text{ (biased; 20\%)} = 20 \times 0.8 = 16 \\ X_2 \text{ (biased; 10\%)} = 40 \times 0.9 = 36 \\ Y_2 \text{ (biased; 20\%)} = 80 \times 0.8 = 64 \end{array}$

Bias differs by 15% X₁ (biased; 10%) = $10 \times 0.9 = 9$ Y₁ (biased; 25%) = $20 \times 0.75 = 15$ X₂ (biased; 10%) = $40 \times 0.9 = 36$

 Y_2 (biased; 25%) = 80 × 0.75 = 60



In summary, this example shows that the addition of a constant bias simply moves the points up or down on the true line. This means that the ratio of the two structures can remain the same with or without bias. If the ratio of two structures were the same from one type of animal to another, then we would expect that data collected from different animals containing a different but constant bias will all fit on the same true line. This is exactly what we observe using real-world data pairs. When we plotted the same data pairs from different animals and calculated regressions, the r² approached 1.0. However, when the bias of one estimate is not similar to that of the other, the difference between the two biases remains.

Consider the future possibilities. Given what we now know about bias, experiments can be designed for separating two components of complexity in stereological data – phenotype and bias.

Comparing Phenotypes

Summary: Phenotypes can be compared and contrasted by plotting one phenotype against another. Grouping data from different animals (phenotypes) can produce regression curves wherein some points fall closer to the line than others. This variation might be explained by differences between phenotypes, by the presence of bias, or both. We can look for differences between phenotypes by plotting X and Y values and compare the results to an equivalency line (45°) wherein X=Y. The results of such an analysis will be of help to us in deciding whether or not a point should be included in a regression group.

To illustrate the methods, regressions were used to analyze closely related species (Lewis rat vs. Fischer 344 rat), sidedness (left vs. right), sex (male vs. female), age (young vs. old; time vs. time; stage vs. stage), and molecular constituent (neuron i (+) vs. neuron j (+)).

We begin with a brief summary of the results followed by specific examples taken from the stereology literature. Additional examples of these analysis methods can be found in the previous version of the software (Case Studies).

Animal Types: Data pairs, which included a collection of similar data collected from different animals, were plotted as power curves with r^2s .

- **Observation:** A surprisingly large number of curves were found with $r^2 \approx 1$.
- *Interpretation:* Organisms appear to build structures according to a similar set of genetic instructions. In other words, different animals can display remarkably similar mathematical phenotypes. This suggests that a structure defined by one or more equations can apply equally well to several different animals. Thus far, the data suggest that the major difference among animals is largely quantitative not qualitative. In other words, larger animals have more of the same components in the same proportions than smaller ones.

Sides (left vs. right): Components of the left side of the brain (x axis) are plotted against similar components on the right side (y axis).

- **Observation:** Although paired organs can have different sizes (e.g., lung and kidney), data are frequently published for only one side or the other. In grouping data, it therefore becomes important to be aware of phenotypic variations when combining data from different animals and different sides.
- *Interpretation:* Evaluating the consequence of grouping data from different sides will probably have to be determined on a case by case basis.

Sex (male vs. female): Male values (x axis) are plotted against similar female values (y axis) - as a power curve.

- *Observation:* Although males and females display different amounts of structures, quantitative relationships between the same structures remained more or less the same.
- Interpretation: A curve with an r²≈1 suggests that structures in males and females reflect similar phenotypes. A quantitative difference between the sexes can be seen as more (a curve above a 45° line; x=y) or less (a curve below a 45° line; x=y).

Age (young vs. old; aging over time; growth; embryonic development): Patterns of development and aging are treated as a data set consisting of rows (structure) and columns (time). All possible combinations can be tested for patterns.

Age	t ₁	t ₂	t ₃	t _n
Structure i	i ₁	i ₂	i ₃	i _n
Structure j	j ₁	j ₂	j ₃	j _n
Structure k	k ₁	k ₂	k ₃	k _n

- **Observation:** (a) In general, the r²s were consistently better for comparisons of adjacent time points (all structures at t_i vs. all structures at t_{i+1}) than for comparisons of two structures over time (t_i ... t_n). This suggests that genetic regulation coordinates large sets of related structures as a group step by step. (b) The typical hallmark of change consists of two parallel curves, t_i vs. t_i.
- *Interpretation:* Parallel curves with an r²≈1 suggest that the relative amounts of different structures remained more or less constant during development, whereas the absolute amounts changed. Once the development program was up and running, the structure grew larger by increasing the y intercept (scaled up) or grew smaller by decreasing it (scaled down). Once the relationship of the structures was established, growth perhaps could be controlled very efficiently by the action of a single "y-intercept growth gene."

Molecules (patterns across cells):

- **Observation:** Molecular data presented earlier in EBS 2001 showed regressions with r²≈1. Here, however, cells labeled for specific molecules present a somewhat clouded picture. One can find strong relationships between cells expressing different molecules, but only after hunting for them in an otherwise diffuse data set.
- *Interpretation:* Caution should be exercised in interpreting such data because the "extracted" curves could be merely fortuitous. Treating such curves as clues not conclusions would seem a sensible course to follow.

Plots of Phenotypes Taken from the Stereology Literature

Animal Types: When volume of the brain is compared to the volume of area 10 in six different primates, a pattern of similarity would seem to appear in a log-log plot.

Citation 3109: Semendeferi K, Armstrong E, Schleicher A, Zilles K, Van Hoesen GW. 2001 Prefrontal cortex in humans and apes: a comparative study of area 10. Am J Phys Anthropol 114(3):224-241.



However, a linear plot of the same data shows that the regression is being influenced importantly by the human data point (circled).



Remove the human data point and the pattern of similarity disappears. The point to be made here is that both linearlinear and log-log plots are both equally important when interpreting data with regressions.



Remove the gibbon and the chimpanzee and a new curve now suggests a considerable similarity for the brains of the bonobo, gorilla, and orangutan.



V brain	V area 10	animal
1158300	14217.7	human
393000	2239.2	chimpanzee
378400	2804.9	bonobo
362900	1942.5	gorilla
356200	1611.1	orangutan
88800	203.5	gibbon

The point of the graphic immediately above is to show that regressions, which at first do not show an r^2 very close to one, can be induced to yield multiple interpretations – some of which may display a better r^2 .

Closely Related Species: Here we can show that counts of some neurons from Lewis and Fischer 344 rats - for all practical purposes – are identical. The r^2 value is 0.9999 and all the points fall on the reference line (x = y; Fischer 344 Rat = Lewis Rat)

Citation 2969: Schmitz C, Dafotakis M, Heinsen H, Mugrauer K, Niesel A, Popken GJ, Stephan M, Van de Berg WD, von Horsten S, Korr H. 2000 Use of cryostat sections from snap-frozen nervous tissue for combining stereological estimates with histological, cellular, or molecular analyses on adjacent sections. J Chem Neuroanat 20(1):21-29.



Fischer 344 rat	Lewis rat	Cell Number
604457	603439	pyramidal cell (CA1-3) - hippocampus
48738689	51848939	granule cell - cerebellum
144647	132742	Purkinje cell - cerebellum

Left Side vs. Right Side: The stereology literature database includes many papers wherein data are given for only one organ of an organ pair or from the left or right side of an unpaired organ. Here we compare volumes of compartments taken from the left and right sides of the brain. The results suggest that the left and right sides are remarkably similar – at least for the compartments being compared. Moreover, the similar pattern applies to both male and female animals. Notice, however, that the regression curve sits just below the equality line (x = y; Left Side = Right Side), suggesting that the left side may be slightly larger than the left.

Citation 2734: Highley JR, McDonald B, Walker MA, Esiri MM, Crow TJ. 1999 Schizophrenia and temporal lobe asymmetry. A post-mortem stereological study of tissue volume. Br J Psychiatry 175:127-134.



Left	Right	Structure Volumes (female)
11.35	10.64	temporal lobe
4.72	4.27	inferior temporal gyrus
6.2	6.46	middle temporal gyrus
8.14	7.65	superior temporal gyrus
42.56	40.39	total gray matter
15.48	13.96	white matter



Left	Right	Structure Volumes (male)
12.93	12.74	temporal lobe
5.1	4.98	inferior temporal gyrus
7.33	7.43	middle temporal gyrus
9.44	8.58	superior temporal gyrus
49.03	47.63	total gray matter
19.56	17.34	white matter
1	1	Standard
10	10	
100	100	

Sex (Dimorphism): Sexual dimorphism refers to somatic differences between males and females of the same species, typically arising from differences in sexual maturation. Here we compare males to females in diestrus and proestrus (parts of the estrus cycle). The results suggest that little difference exists between the males and females – for the structures being compared.

Citation 3131: Madeira MD, Ferreira-Silva L, Paula-Barbosa MM. 2001 Influence of sex and estrus cycle on the sexual dimorphisms of the hypothalamic ventromedial nucleus: stereological evaluation and Golgi study. J Comp Neurol 432(3):329-345.



Male	Female (diestrus)	Female (proestrus)	Neuron Number
55800	55700	56800	neuron (ventromed n)
2850	2560	2810	neuron (ventromed n;anterior)
18800	18800	19200	neuron (ventromed n;dorsomed)
14000	14300	13300	neuron (ventromed n;central)
17200	17000	21200	neuron (ventromed n;ventrolat)
1	1	Standard	
1000	1000		

100000 100000

Sex (Male vs. Female): Here we compare the volume of the neuropil of the medial preoptic nucleus in male and female animals at different ages. The results suggest that the male has more neuropil than the female (the curves are below the equivalency line: male=female), but that the three regression curves are parallel. However, a comparison of neurons yields a different result for the 30 month; medial preoptic nucleus comparison.





Male	Female	Time	Medial Preoptic Nucleus - Volumes
0.112	0.076	6m	lateral
0.136	0.079	6m	medial
0.015	0.007	6m	central
0.127	0.096	24m	lateral
0.174	0.113	24m	medial
0.017	0.009	24m	central
0.128	0.104	30m	lateral
0.155	0.114	30m	medial
0.016	0.009	30m	central
0.001	0.001	Standard	
0.1	0.1		
1	1		



Male	Female	Time	Medial Preoptic Nucleus – Number of Neurons
13600	11500	6m	lateral
21200	18100	6m	medial
4940	2430	6m	central
13400	10600	24m	lateral

22200	19300	24m	medial
5010	2390	24m	central
14300	11800	30m	lateral
23700	17800	30m	medial
4540	2980	30m	central
0	0	Standard	
15000	15000		
25000	24000		
25000	24000		

Age (Young vs. Old): Of all things being compared here, age appears to be the most variable parameter across both animals and phenotypes. In this example - taken from young and old rats – the number of neurons in the hippocampus appears to remain constant, whereas the volumes of the compartments containing these cells appear to change.

Citation 3333: Sousa N, Almeida OF, Holsboer F, Paula-Barbosa MM, Madeira MD. 1998 Maintenance of hippocampal cell numbers in young and aged rats submitted to chronic unpredictable stress. Comparison with the effects of corticosterone treatment. Stress 2(4):237-249.



Young	Old	Cell (Neuron Number)
1193000	1159000	granule cell
47165	45038	hilar cell
241875	243750	CA3 cell
427080	396500	CA1 cell
1 10000000	1 10000000	Standard



Young Old Structure Volume

6.73	6.3	dentate gyrus;molecular layer
1.84	2.03	dentate gyrus;granular layer
2.31	2.64	dentate gyrus;hilus
2.98	3.13	hippocampus;CA3;stratum oriens
2.48	2.51	hippocampus;CA3;stratum pyramidale
6.02	5.53	hippocampus;CA3;stratum radiatum+lacun
2.75	3.32	hippocampus;CA1;stratum oriens
1.64	1.78	hippocampus;CA1;stratum pyramudale
7.2	6.77	hippocampus;CA1;stratum radiatum+lacun
1	1	Standard
10	10	

Next, we can plot individually the data for each of the three brain areas: dentate gyrus, CA3 and CA1. Now we find three parallel curves with some separation. Perhaps there is a slight gain of volume in the dentate gyrus and CA1.



Young	Old	Structure Volume
6.73	6.3	dentate gyrus;molecular layer
1.84	2.03	dentate gyrus;granular layer
2.31	2.64	dentate gyrus;hilus
2.98	3.13	hippocampus;CA3;stratum oriens
2.48	2.51	hippocampus;CA3;stratum pyramidale
6.02	5.53	hippocampus;CA3;stratum radiatum+lacun
2.75	3.32	hippocampus;CA1;stratum oriens
1.64	1.78	hippocampus;CA1;stratum pyramudale
7.2	6.77	hippocampus;CA1;stratum radiatum+lacun
1	1	Standard
10	10	

Next, we can plot the data according to the best r^2 values.



Young	Old	Series 1
1.84	2.03	dentate gyrus;granular layer
2.31	2.64	dentate gyrus;hilus
2.75	3.32	hippocampus;CA1;stratum oriens
1.64	1.78	hippocampus;CA1;stratum pyramudale

In this analysis, the structures in Series 1 suggest a coordinated gain in volume. Series 2 may include two parts.

Young	Old	Series 2
2.98	3.13	hippocampus;CA3;stratum oriens
2.48	2.51	hippocampus;CA3;stratum pyramidale
6.02	5.53	hippocampus;CA3;stratum radiatum+lacun
6.73	6.3	dentate gyrus;molecular layer
7.2	6.77	hippocampus;CA1;stratum radiatum+lacun

Finally, we can simplify series 2 into series 3 and 4.



Young	Old	Series 3
2.98	3.13	hippocampus;CA3;stratum oriens
2.48	2.51	hippocampus;CA3;stratum pyramidale

Series three structures appear equal in young and old animals.

Young	Old	Series 4
6.02	5.53	hippocampus;CA3;stratum radiatum+lacun
6.73	6.3	dentate gyrus;molecular layer
7.2	6.77	hippocampus;CA1;stratum radiatum+lacun

However, the structures in series 4 suggest a slight loss of volume.

Age (Aging Over Time): One way of analyzing a continuous data set – extending over several years – is to look for patterns where none appear to exist. Using this approach, a complex data set will often reveal an underlying order.

Consider the curve below. The relationship between the two prefrontal layers looks murky at best, having an r^2 of just 0.65. With a little effort, we can readily extract three curves that carry quite respectable r^2 values (see series 2, 3, and 4). Now it appears that the relationship of one cortical layer to another changes importantly over time. Finally, series 5 and 6 add back two noise producing variables – time and bias – by plotting age against level volume. The mischief caused by the time and bias variables can be very great indeed! The r^2 of series 5 is a dismal 0.0311 and that of series 6 even worse at 0.0001.

Citation 2732: O'Donnell KA, Rapp PR, Hof PR. 1999 Preservation of prefrontal cortical volume in behaviorally characterized aged macaque monkeys. Exp Neurol 160(1):300-310.



Age	Layer I	Layers II-VI	Prefrontal Cortex Volume
7.3	20.4	121	Area 46
10.2	13.6	102.6	Area 46
10.8	13.4	101.7	Area 46
11	18.5	122.9	Area 46
11.4	10.8	93.9	Area 46
12.3	12.8	87.4	Area 46
20.4	22.5	106.3	Area 46
22.8	21.8	161.4	Area 46
23	17.4	101.7	Area 46
24	13.1	82.3	Area 46
24	18.1	102.8	Area 46
24	18.2	110.5	Area 46
24.8	18.4	118.9	Area 46
25.1	20.3	130.5	Area 46
25.8	21.9	139.5	Area 46
26	12.8	72.6	Area 46
26.9	14.1	86.8	Area 46
29.8	18.7	123.7	Area 46
29.9	18.2	116.3	Area 46
31.3	13.6	77.7	Area 46
31.9	16.1	105.2	Area 46





10.2	13.6	102.6	Area 46
10.8	13.4	101.7	Area 46
11	18.5	122.9	Area 46
11.4	10.8	93.9	Area 46
Age	Layer I	Layers II-VI	Prefrontal Cortex (Series3)
24.8	18.4	118.9	Area 46
25.1	20.3	130.5	Area 46
25.8	21.9	139.5	Area 46
Age	Layer I	Layers II-VI	Prefrontal Cortex (Series4)
26	12.8	72.6	Area 46
26.9	14.1	86.8	Area 46
29.8	18.7	123.7	Area 46
29.9	18.2	116.3	Area 46
31.3	13.6	77.7	Area 46
31.9	16.1	105.2	Area 46

Notice what happens to the r^2 values in the two figures below when we return the effects of the disruptive variables – time and bias.





Age (Growth): How does the growth of an organelle population occur over time? Here a set of five organelles at the first time point is plotted against the same organelles at the next time point. This gives us seven regression lines. The first curve (2m vs. 6m) is above the reference line (here 45°), indicating growth – an increase in volume. The second curve (6m vs. 9m) shows growth relative to the reference line and to the first curve. The third curve, however, shows a loss of volume relative to the reference line and to the second curve. The y intercept values indicate that the growth process is interrupted two times by a contraction in volume (9m vs. 12m and 21m vs. 24). Notice that the curves remain essentially parallel throughout the developmental period and that the r^2 are usually 0.99 or better. Such a high degree of order can be seen frequently – all across the stereology literature.

Citation 3136: Henrique RM, Rocha E, Reis A, Marcos R, Oliveira MH, Silva MW, Monteiro RA. 2001 Agerelated changes in rat cerebellar basket cells: a quantitative study using unbiased stereological methods. J Anat 198(Pt 6):727-736.



The behavior of these parallel curves suggests that a single gene may be controlling this set of organelles including all their molecular constituents – as a group. In effect, this illustrates one way in which stereology might be used to detect the downstream effects of a genetic control mechanism.



If, instead, we plot one organelle against another for all time points, the r^2 values are not very good (0.42 to 0.77). This can be explained – at least in part - by the up and down migration of the parallel curves. It also suggests that the ratio of one organelle to another may be changing during development.



9m	5.6	1.8	117	14
12m	5.1	1.5	93	11
15m	4.2	1.7	86	10
18m	4.8	2.4	106	11
21m	6.5	3.9	141	13
24m	6.5	34	128	11

Note: A very similar pattern is observed for citation 2345 (not shown), coming from the same lab.

Age (Embryonic Development): As done earlier for citations 2732 and citation 3136, we plot a data set at one time against a similar data set at a later time point. Here, however, the time between adjacent time points is shorter, being in days. In such a setting, one often observes sets of parallel curves alternating with sets of unparallel curves. The reference curve is calculated by plotting the youngest time point against itself. A curve above the reference curve signals an increase in volume, whereas a curve below a decrease.

When both curves are parallel, all the structures being compared change proportionately. When this is not the case and the curves are not parallel, it may tell us something about genetic control mechanisms. For example, genes may be activated sequentially, genes may produce structures at different rates, gene products may turnover at different rates, etc. In the d17 vs. d18 plot one can imagine that the activation order of genes might be glomerulus, ureteric epithelium, and mesenchyme – in that order. An alternative explanation might be that superimposed on top of a routine production of structures is the migration (i.e., differentiation) of one compartment into another. Notice that the r^2 values are very close to 1.0, indicating a consider degree of order in the kidney during development.

Citation 3317: Bertram JF, Young RJ, Spencer K, Gordon I. 2000 Quantitative analysis of the developing rat kidney: absolute and relative volumes and growth curves. Anat Rec 258(2):128-135.



comma vesicle

1

0.174

0.254



Molecules (patterns across cells): Do cells positive for one molecule form patterns with other cells positive for a different molecule? In the plot below, parvalbumin positive neurons are plotted against somatostatin-like positive neurons and the correlation between the two cell types is weak ($r^2=0.174$). However, it is possible to select points that give $r^2=0.99$. Admittedly, this is a risky procedure and the arrangement of the points might be entirely fortuitous. However, it offers one way of extracting patterns from data sets that might otherwise be without interpretation.

Citation 3135: Jinno S, Kosaka T. 2000 Colocalization of parvalbumin and somatostatin-like immunoreactivity in the mouse hippocampus: quantitative analysis with optical dissector. J Comp Neurol 428(3):377-388.





Dorsal	Dorsal	Subdivision - Neuron Number
neuron (parvalbumin+)	neuron (somatostatin-like+)	
152	352	CA1-SO
347	28	CA1-SP
33	10	CA1-SR
<u>5</u>	2	CA1-SLM
<u>61</u>	<u>48</u>	<u>CA3-SO</u>
<u>216</u>	<u>140</u>	CA3-SP
<u>32</u>	<u>18</u>	CA3-SR
15	99	Hilus
154	3	GL



Dorsal	Dorsal	Subdivision - Neuron Number
neuron (parvalbumin+)	neuron (somatostatin-like+)	
152	352	CA1-SO
347	28	CA1-SP
216	140	CA3-SP