

## Chapter #11

# VISUALISING THE SCIENCE OF GENOMICS

KATHY M. TAKAYAMA

*School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW Australia 2052*

**Abstract:** The term ‘genomics’ broadly refers to the study of the genome, or the complete genetic inheritance of an organism. The genome sequence of an organism provides the equivalent of a complete genetic map; yet, knowledge of the sequence itself does not reveal how this map manifests itself into the physical characteristics or phenotypes observed for an organism. Genomics research is dependent upon comparative analyses of extraordinary volumes of data. Whilst visualisations may facilitate the significance and understanding of such comparisons, the complexity and scope of the information provides a challenge for the classroom learner. This chapter examines the roles of representations in genomics ‘visual literacy’, and addresses the challenges associated with distilling a rapidly progressing research area into pedagogical frameworks that can accommodate the dynamic nature of the field. The chapter also presents an application of visualisations in genomics education within the context of a tertiary level international collaborative research project. The student-centred project, ‘Visualising the Science of Genomics’, presents a novel example of inquiry-based teaching in genomics in an online environment.

## 1. INTRODUCTION

The ‘biological sciences’ have traditionally been a visually-rich discipline, from the sub-microscopic to the biome level. Educators have depended upon the combination of theory with laboratory or field-based teaching to provide contextual frameworks for student learning. The era of genomics has introduced novel challenges and opportunities for the incorporation of visualisations in science education. The ‘exemplar phenomena’ as described by Gilbert in the first chapter of this book take on a different context in genomics. For, in genomics, we have not exemplar phenomena, but the field itself lays the foundation for the discovery of exemplars. It is an approach

whereby systematic discovery, integration and analysis of genetic information are transformed into multidisciplinary contexts. The models are cognitively stratified as the ‘interpretation of the genomic script’ continues to reveal new layers of complexity. Whilst scientists readily embrace rapid improvements in sophisticated visualisation and analytical technologies in genomics research, they are not practicable in the teaching laboratory. The research-teaching nexus risks disconnection if fidelity of context cannot be attained through an authentic learning experience. Visualisations can facilitate the understanding of genomics if they are strategically presented, learning with the visualisation is appropriately scaffolded, and educators consider the requirements for visual literacy of the learner in specific contexts.

## 2. GENOMICS: WHAT, WHY, HOW

The term ‘genomics’ broadly refers to the study of the genome, which encompasses the complete genetic inheritance (DNA) of an individual organism. Ever since the elucidation of the double helical structure of DNA by Watson and Crick in 1953 (Watson and Crick, 1953), scientists have been analysing individual genes from various species long before the term ‘genomics’ entered the public vernacular. Yet it was not until more than 40 years after Watson and Crick’s discovery that the first genome, from the bacterium *Haemophilus influenzae*, was deciphered, revealing the complete sequence of A, G, C and T bases comprising its DNA code (Fleischmann et al., 1995). The era of genomics was hence launched by this achievement, and has contextualised research and teaching in the life sciences into an entirely new framework. The availability of the complete DNA blueprint for an organism paves the way for global analysis of how the genetic, metabolic, and physiological processes in the organism are dictated, and how these networks interact. Subsequently, the much-anticipated announcement of the completion of the human genome sequence in 2001 (Lander et al., 2001) marked an historical milestone. As of September 2004 a total of 215 completed genome sequences had been published, and 963 genome sequencing projects were in progress (Bernal et al., 2001; Krypides, 1999, 2004) and the number of sequenced genomes continues to increase at a rate of one per week.

The genome sequence of an organism provides the equivalent of a complete genetic map. However, knowledge of the genome sequence itself does not reveal how this map manifests itself into the physical characteristics or behaviours (phenotypes) observed for an organism. Some would argue that we are now well ensconced in a *post-genomic* era, as we focus on the

analyses of the enormous datasets that comprise these genetic maps. Such analytic approaches include functional genomics (study of the expression patterns for all of the genes in a genome), structural genomics (study of physical characteristics of the genome), comparative genomics (comparisons of genomes from different organisms), proteomics (study of expression patterns for all proteins of a cell/organism) and pharmacogenomics (application of genomics toward the identification of drug targets). Taxonomic jargon notwithstanding, the enormity and scope of these approaches render them crucially dependent upon bringing together communities of diverse scientists contributing their distinctive talents. Consequently, the genomics/post-genomics eras have led to a major international shift in research and educational strategic initiatives, resulting in the development of interdisciplinary centres and degree programmes.

The biology student studying genomics is presented with a contextually rich perspective on how living systems function at a global molecular and cellular level. That is, whilst the development of techniques in molecular biology over several decades facilitated our understanding of how specific genes or pathways functioned individually, genomics provides us with a holistic view of how all of these pathways function simultaneously. In so doing, genomics enables the elucidation of entire networks, and for the educator, genomics presents a forum for the exploration of a broad range of contexts. Hence, biology education in the 21<sup>st</sup> century has evolved towards what has come to be known as ‘systems biology’, embracing multi- and interdisciplinary frameworks.

### **3. SIGNIFICANCE OF VISUAL LITERACY IN GENOMICS EDUCATION**

In research practice, the scope of the field of genomics has enhanced interactions: i) between the sciences, including biology, mathematics, computer science, chemistry, and statistics; ii) between the basic and clinical sciences; and iii) between the life sciences, social sciences and humanities. In practicality, however, an authentic learning experience that reflects this multidisciplinaryity is not as readily created in a ‘traditional’ classroom or course structure. The challenge for the biology educator to teach genomics in a contextually relevant framework which simultaneously is reflective of research practice and societal value becomes dependent on alternate approaches incorporating visualisation and application.

Genomics research is dependent upon visual comparative analysis of extraordinary amounts of complex information. Most of this primary visual analysis, however, is performed by high-precision detection equipment and

computers. The complexity and scope of such data, whilst daunting enough for the researcher, is prohibitively impractical for classroom students. The 'visual literacy' of the learner is significantly distinct from that practiced by the genomics researcher. Such visual literacy skills are acquired through extensive experience and contextual practice. According to Christopherson, visual literacy includes the ability to interpret and comprehend the meaning and significance of visual representations, effective communication through the application of the basic principles and concepts of visual design, the production of visual messages through appropriate technologies, and the application of visual thinking toward problem-solving (Christopherson, 1997).

The effectiveness of visual representations in genomics education pivots on distillation of the essential and useful 'visual messages' from the breadth and depth of options available. Metavisual competence as described in Chapter 1 by Gilbert, may be described in the area of genomics as specifically encompassing the following practices:

- Recognition and comprehension of modes of visual representation of genomic data.
- Communication of genomic information through visual representation.
- Comparative application of visual representations between sets of genomic data.
- Transfer of genomic information from one visual mode of representation into another visual mode of representation.
- Development of models representing relationships based on quantitative and/or qualitative comparisons of visual genomic data analysis.
- Predictions of behaviour of new genomic data based on previous visual models.

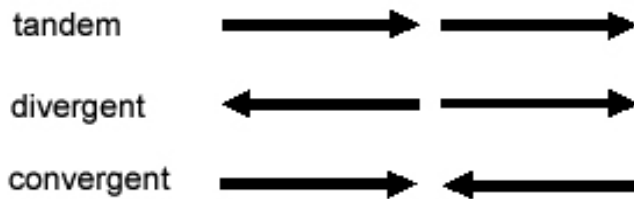
These skills are gradually acquired through active experience in genomics research, whereby the researcher's understanding evolves as knowledge is synthesised within appropriate contexts. However, the classroom student who does not have the opportunity to access this experiential learning must develop these metavisual skills through alternate frameworks that are often subject to limitations of time and resources.

The development of genomics visual literacy for the learner is facilitated through structured contextual exemplars. For example, given a new dataset of genome sequences from a novel bacterial species, the student may first utilise bioinformatics tools to analyse the data. S/he then learns to recognise specific visual motifs in the output that indicate functional, taxonomic, or structural information about the genomic sequence. This process might be conducted within the context of investigating evolutionary relationships

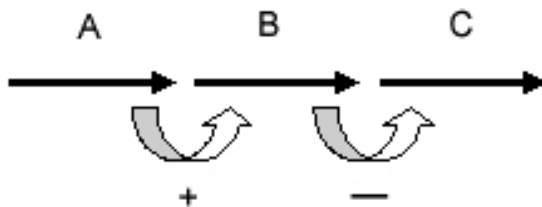
between genes or organisms, or within the context of following the epidemiology of a disease-causing pathogen. Such frameworks provide relevance and application, key criteria for authentic learning experiences (Chinn and Malhotra, 2002; Herrington and Herrington, 1998; Kolb, 1984; Meyer, 1992). Oftentimes the visual exemplars employed under the umbrella of genomics education do not represent first order concepts that stand alone, but rely inherently on the integration of conceptual understanding, identification of relationships between genomic concepts, and applications of several concepts toward the discovery of new information or the formulation of new hypotheses. The cognitive skills associated with this visual literacy are indicative of a higher order level. The student engaged in these visual representations would not only have to comprehensively integrate several genomics concepts associated with the visual representation, but would also proceed to apply this relational interpretation toward the synthesis of a new visual schematic. The complex nature of visual literacy in science and its dependence on the synthesis of knowledge from different contexts has indeed been highlighted by others (Ferk et al., 2003; Vrtacnik et al., 2000a, 2000b; Wu and Shah, 2004).

To ultimately facilitate visual literacy at this level, it is imperative to ensure that the process is appropriately scaffolded by assessing competence in 'prerequisite' visual literacy in the genomic context. That is, in order to perform cognitive operations in a spatial domain, the learner must be competent in the visuospatial skills that are required for each of the conceptual steps that comprise genomics visual literacy. If one were to deconstruct the metavisual processes described above into specific elements of 'basic' visual cognition, they could be described as follows:

- Gene structure: the visual conceptualisation of the structural and regulatory components of a gene, the basic unit of a genome. A corollary to this is the visual cognition of the structural components of a protein, which is a product encoded by a gene.
- Gene orientation and organisation: the ability to recognise or identify the orientation and organisation of genes relative to each other. There are three possibilities for genetic arrangement (orientation), as depicted below, whereby the arrows indicate the direction in which a gene is 'read':



- Gene relationships: integration of spatial and temporal conceptualisations of the regulation of one or more genes by another gene(s). For example, in the diagram below, Gene A can positively influence the expression of Gene B. Gene B in turn negatively regulates Gene C. However, Gene A cannot be said to directly negatively regulate Gene C, because if a mutation results in the loss of Gene B, Gene A has no direct control over Gene C:



- Sound understanding of which visual representation/application/model (or a combination thereof) is most appropriate for conveying specific genomic concepts, observations, or relationships.
- Genome relationships: visual recognition of patterns of similarity or differences between: i) genome sequences; ii) patterns of the expression of genome sequences; or iii) functional consequences of alterations in genome sequences.

The development of visual literacy in genomics education builds upon these foundations. Layers of complexity are added to contribute contextual depth and breadth to one's metavisual cognition through the integration of

experience and practice. The significance of this learning process becomes apparent if we reflect on the scientific method itself.

The practice of scientific inquiry is traditionally represented as a progression of distinct steps which are connected in a mainly sequential order:

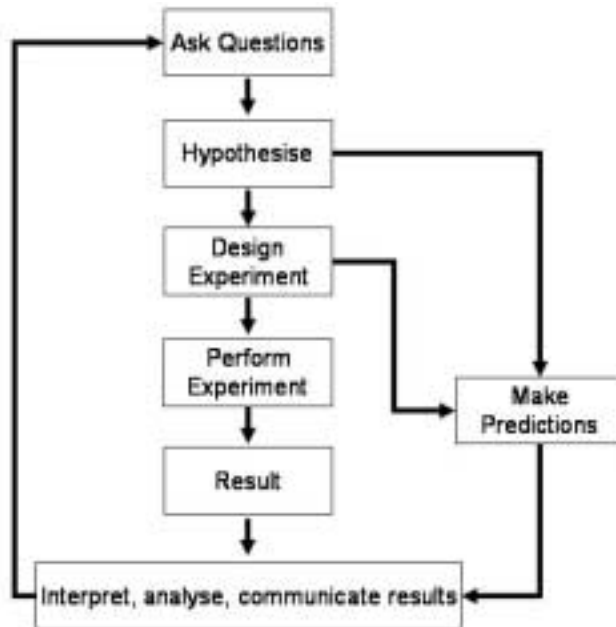
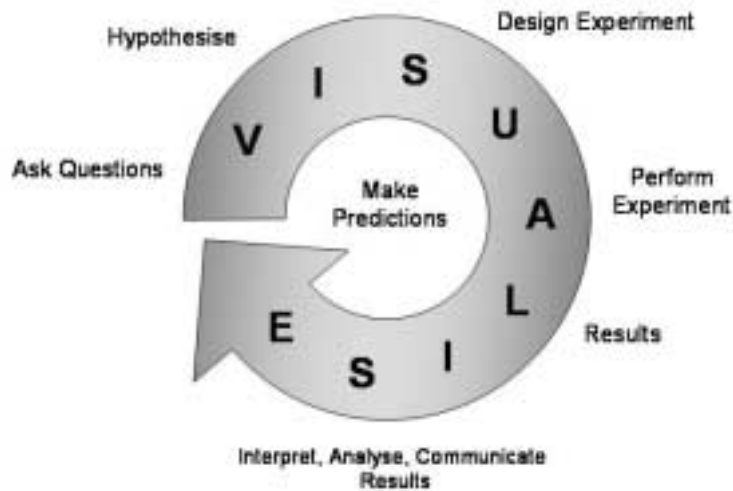


Figure #11-1. [Traditional representation of the process of scientific inquiry.]

The same process, when considered from a metavisual perspective, becomes highly interactive through the inherently dynamic role that visual cognition plays throughout every phase of inquiry (Figure 2). The visualisations created and employed in the process are themselves fluid in that they both inform, and are informed by each stage of the cycle. Furthermore, the formulation of predictions becomes more central in a metavisual context, because visual predictions can be conceptual (mental models; as discussed by Gilbert in Chapter 1, and reviewed in Gilbert, 2002) and/or factual (visual representation of data).



*Figure #11-2.* [The process of scientific inquiry represented from a metavisual perspective. The model is a modification of that proposed by Keefe et al. (2004), in which visualisation is one of the sequential steps of the scientific process.]

Kirby and co-authors emphasise the importance of the definition of explicit visualisation goals in order to create effective visualizations (Keefe et al., 2004). In the model above, the visual goals would evolve with each cycle of the scientific process. Indeed, the definition of visualisation goals has been described as an iterative process dictated by the underlying scientific applications (Brooks Jr., 1996), and as progress is made in the elucidation of the scientific problem, the goals for the visualisations that address or communicate the problem are modified accordingly. The field of genomics presents a prime example of how such a formative metavisual approach continues to inform both research practice and pedagogy.

#### **4. THE ROLE OF VISUALISATION IN TEACHING GENOMICS**

The rapidly increasing number of sequenced genomes continues to add to the enormity and diversity of genomic information in global databases. The researcher or classroom learner is challenged with interpreting and communicating a study focused on a gene(s) or genome of interest with

peers who may approach the analysis of a gene/genome from a different contextual perspective. The Gene Ontology Consortium has developed a taxonomic system to facilitate a universal dialogue based on ontologies describing the products of genes in terms of their: i) molecular functions; ii) associated biological processes; and iii) cellular components (Consortium, 2000). According to the Consortium, the *molecular function* describes *what* the gene product does; that is, its molecular activity. The *biological process* refers to the *significance (how)* of this activity within “one or more ordered assemblies of molecular functions” (Consortium, 2000). Whilst a biological process is not equivalent to a pathway, it is dependent upon several or large networks of pathways to achieve a broader aim. Examples range from specific processes such as pyrimidine metabolism to broader processes such as cell growth and maintenance. The *cellular component* simply describes *where* the activity of the gene product takes place in the cell (Consortium, 2000).

These groupings are comprehensively inclusive for any gene regardless of species. I would propose that all genomic visualisations can also be categorised under the same contextual groupings. In identifying the category(ies) to which a given visualisation belongs, the educator or learner develops a metacognitive awareness of the goals and interpretive practices invoked by the visualisation. This in turn facilitates the processes associated with conceptual understanding of the content embedded in the visualisation.

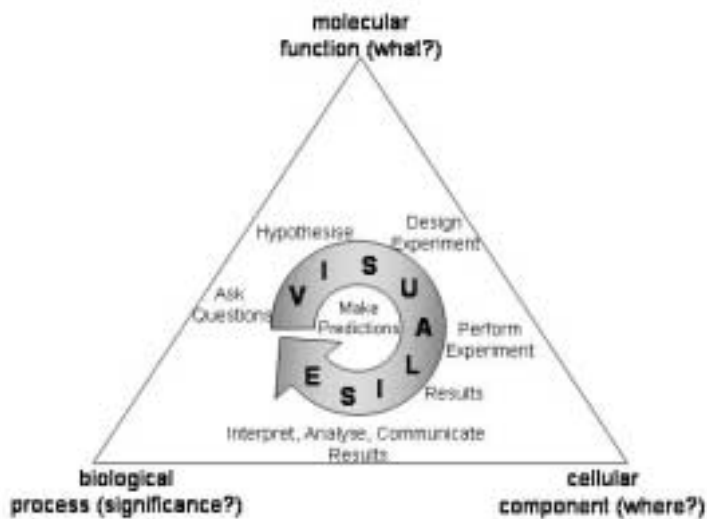


Figure #11-3. [Integration of the metavisual process within the context of the pedagogical goals for genomics visualisations.]

In the following sections, commonly used examples of visualisations in genomics research are discussed from a pedagogical perspective. The examples selected are by no means exhaustive; especially, as the field of genomics continues to rapidly advance, modelling environments will also continue to evolve to accommodate the vast datasets that are generated.

## 4.1 Visual representations of genomic information: contextual, visual and pedagogical goals

### 4.1.1 Sequence analysis: comparative genomics

A genomic sequence can be compared to other genomic sequences for various investigative purposes. Comparisons can also be performed for protein (gene product) sequences. Examples of such visualisations are shown in Figure 4A and B (a colour version of this figure is available in the accompanying CDROM).

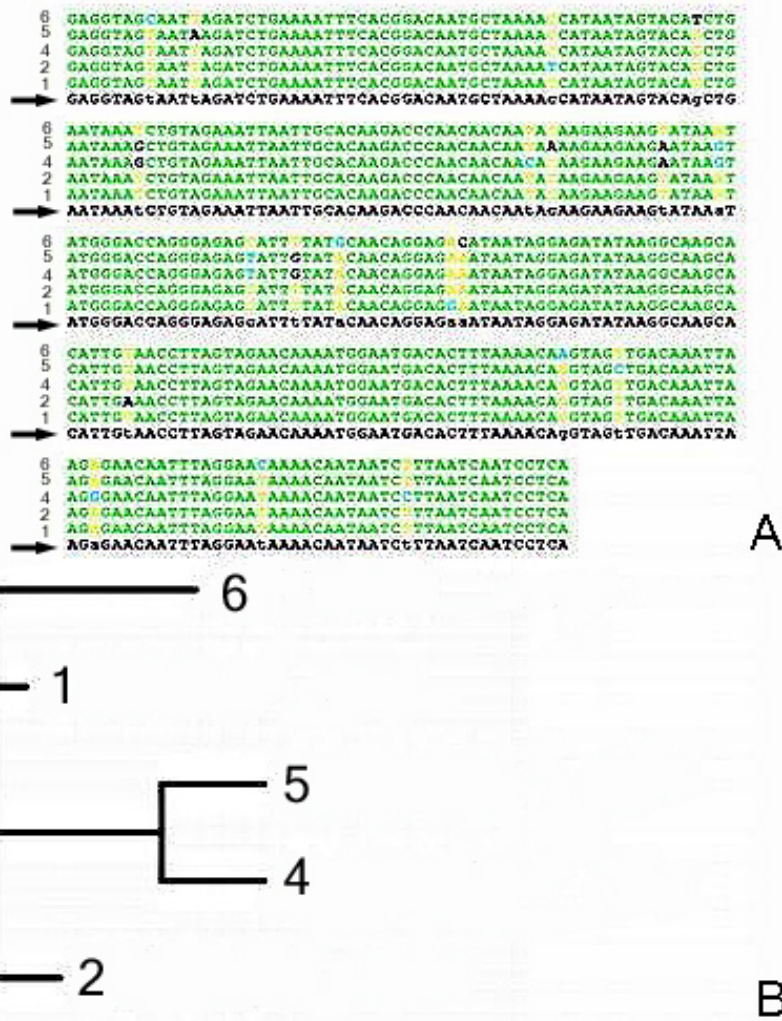


Figure #11-4. [A: Alignment of nucleic acid sequences from a hypervariable (highly prone to mutation) region of a specific Human Immunodeficiency Virus (HIV-1) clone (subspecies) found in a patient. The alignment continues from the top set of grouped rows through to the bottom set of grouped rows, and is read from left to right. Each row represents the nucleic acid sequence of the virus present in a blood sample at a specific visit. The visits began at seroconversion<sup>1</sup> and continued at 6-month intervals; the numbers refer to visits in sequential order. The arrow indicates a consensus sequence, which represents the most common nucleotides found at each position. B: Phylogenetic tree diagram representing the evolutionary relationships amongst the HIV-1 clones from Figure 4A. Branch lengths and

<sup>1</sup> Seroconversion: After an infection with HIV-1, antibodies to the virus can be detected in the blood. This is called seroconversion (converting from HIV-negative to HIV-positive).

proximity are indicative of how related specific clones are to one another. Refer to the text for further explanation.]

The contextual goal of the visualisation representing the comparison in Figure 4A would be to demonstrate i) similarities/differences in the gene/protein sequences between species or within a single species; or ii) identification of conserved sequence motifs, which can in turn be extrapolated and analysed to assign function. In the specific case presented here, which represents alignments of protein sequences from the Human Immunodeficiency Virus (HIV-1), the first goal is realised. The visual goals are to illustrate patterns of similarity or differences amongst a set of sequences. For the student to navigate the figure the metavisual knowledge required includes:

- knowing the convention for reading sequences: left to right;
- knowing that each line of the sequence comparisons, or alignments, represents an individual dataset (whether it represents a gene(s) or protein sequence) in the comparison;
- knowing that each letter represents a 'basic unit' of the sequence (whether it is a nucleotide (for DNA) or an amino acid (for proteins))

The pedagogical goal of the figure would be a merger of these two goals, and would be summarised: to identify patterns of similarity/differences in a sequence alignment that would represent changes/conservation of genome sequences.

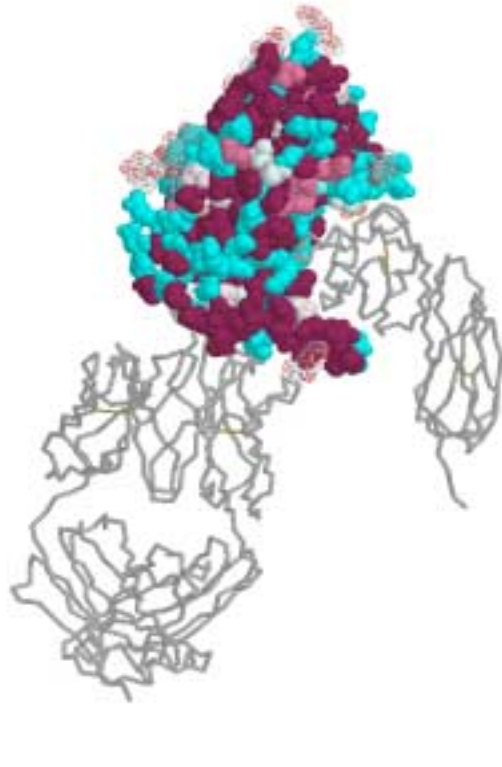
Interestingly, Figure 4B is a computer-generated visualisation that is derived from the sequence alignment data of Figure 4A. The contextual goal of such a figure may be: i) to reveal relative sequence similarities amongst genes or proteins; or ii) to reveal relative gene/protein sequence similarities, and in so doing, demonstrate the evolutionary (phylogenetic) relationships amongst species. The same visual representation is assigned distinct terminologies depending upon the contextual goal of the diagram. If the biological context is limited to the first goal, the diagram is called a 'dendrogram'; if the contextual goal extends to the latter, the visualisation is a 'phylogenetic tree'. The specific example presented here is a phylogenetic tree showing evolutionary relationships amongst mutational variants of HIV-1. Because a phylogenetic tree inherently results from comparisons of sequence similarities, it is a specific type of dendrogram. Dendrograms are an exemplar of the integration of the metavisual process within the scientific and pedagogical contexts (see Figure 3). They organise, using a visual process, the relationships between species/genes/proteins attained through

sequence comparisons. The specific visual literacy skills necessary for a student to interpret this organisational map are:

- to know that each node represents sequences from an individual species/gene/protein;
- to know that species/genes/proteins with the most similar sequences are closest together on the branched diagram;
- to know that the branch distances are a measure of how similar or divergent two sequences are.

The pedagogical aim of a visualisation such as one depicted in Figure 4B is to map and identify the relationships amongst a group of species/genes/proteins based on comparisons of sequence similarities. The visual organisation of these comparisons facilitates the learner's achievement of this goal. The dendrogram distills the computational process involved in determining relational data amongst a set of sequences, and presents the learner with the opportunity to interact with this complex analysis. The student can readily engage in comprehending, analysing, applying, and synthesising new models from data presented in this visual format.

The third visualisation in this series is presented in Figure 5 (a colour version of Figure 5, and a 3D movie file of a similar coloured model is available in the accompanying CDROM).



*Figure #11-5.* [Screen capture of 3D Protein Explorer representation of an HIV protein sequence alignment using the ConSurf program. For a coloured version of Figure 5, please refer to the CDROM.]

The image is a screen capture of a three-dimensional (3D) computer-derived structure representing the HIV-1 protein molecule encoded by the sequences from Figure 4A. The coloured 3D image can be viewed on the accompanying CD-ROM. Each amino acid residue of the protein structure is colour-coded to indicate whether that specific residue is identical amongst all HIV-1 sequences analysed, or whether it is highly variable. The colour gradations that fall between the two extremes are indicative of varying degrees of sequence variability. Collectively referred to as sequence homology, this type of qualitative and quantitative data can be obtained from the first visualisation in Figure 4A. The 3D figure, however, maps this information onto a biological structure. From a metacognitive perspective, this representation integrates the scientific process of analysing abstract data (in which letters represent protein sequences) into a more relevant visual context. For the learner, the outcomes of sequence variability or similarity

are modeled onto a tangible representation of the HIV-1 protein. The 3D visualisation draws on an additional component of visual literacy, that of spatial cognition and application. This chapter will not engage in a discussion of spatial literacy, as it has been comprehensively covered in other sections of this book and by numerous other studies (chapters x, y, z, etc- to be cross-referenced later; (Ben-Zvi et al., 1988; Keig and Rubba, 1993; Kozma and Russell, 1997; Seddon et al., 1985; Tuckey et al., 1991; Wu and Shah, 2004), except to succinctly recap in general terms, metacognitive spatial literacy requires the learner to:

- be able to translate fluently between modes of representation;
- be able to mentally and conceptually visualise changes in perspective for, as well as perform mental operations on, a 3D representation.

Engagement with the 3D visualisation of the HIV-1 protein involves cognitive function at several levels. The student must be able to conceptually translate between the sequence comparison (Figure 4A) and the 3D model. The student must visually and conceptually recognise that each molecule in the 3D structure represents a 'variable position' whereby variability is indicated by colour. Furthermore, the student should be able to appreciate the conceptual implications when studying the 3D visualisation from several perspectives: i) focusing on variability of specific amino acid residues; ii) focusing collectively on variability vs conservation for the entire 3D structure; iii) focusing on relationship between variability of amino acid residues and topology of the molecule; iv) mentally considering biological function and interaction with other molecules in the context of amino acid variability.

Edelson and Gordin (1998) have highlighted the significance of scientific visualisations as a means for students to participate in the practice of science using an inquiry-based approach. The representation of genomic analysis in the above examples can be integrated into an inquiry-based learning process through the creation of appropriate contextual frameworks. The project discussed below, 'Visualising the Science of Genomics', revolves around this approach.

#### **4.1.2 Global expression analysis using microarrays: functional genomics**

From a biological perspective, the examples above may be considered visualisations of 'static' genomics analysis. That is, the visualisations capture what can be inferred from sequence data without taking into consideration the actual expression (or absence thereof) of these sequences.

The student has gained an appreciation for the inherent characteristics of a genome(s) but has yet to 'see' it in action. The molecular functions (what?), cellular components (where?), and/or biological processes (significance; how?) associated with the genomic information are predicted but not assessed experimentally or visually. Expression analysis introduces a new perspective on genomics for the student. In this case, the *what*, *where* and *how* become primarily relevant as the focus is directed to the products of gene sequences; i.e., whether the genes are expressed, when they are expressed, and to what extent they are expressed. These parameters comprise the goals of functional genomics. The student is now challenged with visualising the application of a genome sequence with added quantitative and qualitative cognitive layers.

DNA microarrays (also referred to as DNA chips) are utilised by researchers to simultaneously detect the level of expression (in the form of messenger RNA, or mRNA) of every gene in a genome. Each spot on a microarray represents a specific gene of the genome. The reader unfamiliar with microarray technology who wishes to gain further insight into the scientific methodology is referred to several excellent microarray teaching resources (Campbell, 2004; Campbell and Heyer, 2003; Shi, 1998). In brief, microarrays are hybridised with a target mRNA population (or a cDNA 'copy' of the population) that has been labelled with a fluorescent dye. If a gene is 'expressed', its corresponding mRNA would be present in the population, and this would be indicated on the microarray by hybridisation of the fluorescent mRNA molecule to a specific spot on the array corresponding to the gene sequence. Two different target mRNA populations would represent two different conditions to be tested, and the populations would be labelled with two distinct fluorescent dyes in order to distinguish between the two.

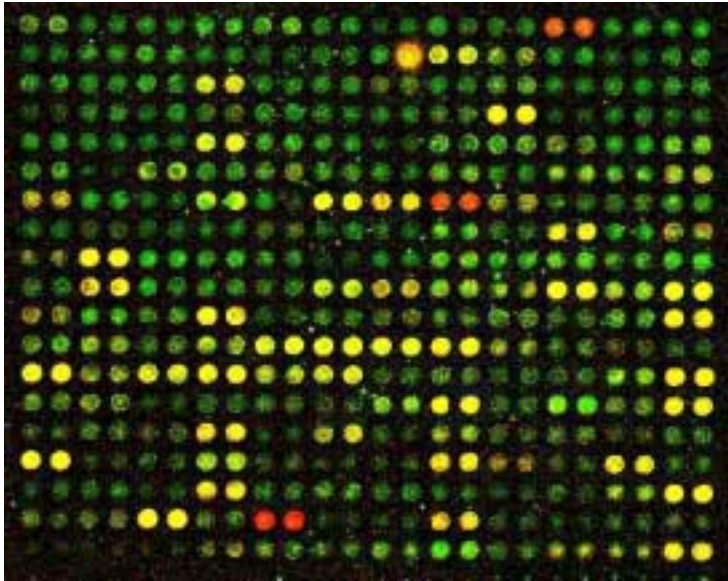


Figure #11-6. [Subsection of a microarray. Image courtesy of The Ramaciotti Centre for Gene Function Analysis. A coloured version of this image is available on the CDROM.]

Figure 6 is a subsection of a laser-scanned microarray (A coloured version of this figure is available on the CDROM). The experimental goal for this visualisation is to determine differences in the levels of expression of every single gene of a cell or organism when it is subjected to an experimental condition and a control (or reference) condition. For example, the microarray in Figure 6 was utilised to measure changes in global gene expression profiles for rat cells grown under a specific experimental condition A vs. condition B. Each spot represents an individual gene and the genes have been spotted in duplicate on the microarray slide to demonstrate reproducibility within the experiment. Redder spots indicate genes that are expressed at higher levels under condition A whilst greener spots represent genes expressed at higher levels under condition B. When expression levels between the two conditions are equivalent, the red and green are 'merged' and the spot is yellow.

The visual goal of this representation is not inherently apparent, because it is, in effect, 'raw data'. The experimental and visual goals are only realised when the specific red vs. green intensities of each spot are converted into numerical ratios that are subsequently analysed by log transformations. Biological significance is portended by experts through practiced comparisons of significant trends in patterns of expression amongst networks of genetic pathways. Furthermore, the expert also may have prior knowledge of the significance of specific relationships between these

networks. Yet, educators will often present similar microarray visualisations in the classroom as a demonstration of functional genomics without providing the rest of the pedagogical framework. This framework requires the following components:

- conceptual and visual link between a genomic sequence and a microarray image.
- visual and temporal understanding of the procedures and steps involved in a functional genomics analysis study.
- conceptual understanding of the scientific inquiry process in microarray analysis.
- visual literacy with regard to identification of visual goals for microarray analysis representations and the ability to select appropriate visualisations to convey contextual goals.

Such an example of the categorical delineation of contextual teaching goals associated with functional genomics can also provide a framework to guide the development of mental practice tasks that can further enhance learning (Cooper et al., 2001). The metacognitive process becomes crucial for the student to benefit from the multitude and variety of visualisations that are directly taken from research environment to the classroom setting. The cognitive apprenticeship model described by Collins and colleagues (1991) is particularly apt in describing how this can be achieved at both cognitive and metavisual levels:

- identify the biological question and make it visible to the student
- provide an authentic context for an abstract task (like functional analysis of a genome) so students understand the relevance
- vary the diversity of situations, as well as the perspectives and contexts of visual presentations, and articulate common aspects so that students can transfer what they learn (Collins et al., 1991).

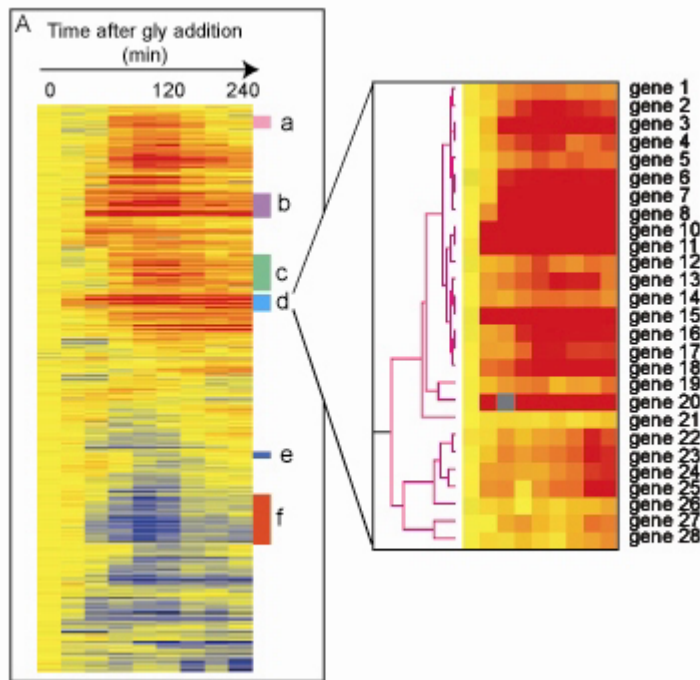


Figure #11-7. [Representation of the changes in the expression patterns for all genes of a yeast microarray over time. Please refer to the text for further details; a coloured version of this figure is available on the CDROM. Figure adapted from image courtesy of I. Dawes, The University of New South Wales and The Ramaciotti Centre for Gene Function Analysis]

Microarray raw data as exemplified in Figure 6 is quantified through computer analysis, and collectively presented as in Figure 7. A coloured version of Figure 7 is available on the CDROM. This image represents the changes in the expression patterns for all genes on a microarray over time. Visual cognitive load is significantly increased as there are several conceptual representations that need to be linked and integrated into the student's thought process. Each column in the left panel represents the expression status for all genes at a given timepoint. Reading from left to right, these columns represent microarray data analysed for samples from 0 minutes to 240 minutes after the induction of the experimental condition (addition of glycine). As in the scanned microarray, the expression status is depicted by coded colour. The panel on the right is a detailed 'zoom-in' of a specific section of the left panel. Here, the individual genes are indicated and their expression patterns over time (reading left to right) become more obvious. An additional level of visual representation is included toward the left of the righthand panel; this dendrogram indicates functional relationships

amongst these genes. Hence, the student must visually interpret and integrate into a framework, the categorical concepts of time, qualitative and quantitative expression, and functional relationships.

## 4.2 2D vs 3D visualisations in genomics and contextual authenticity

Functional relevance of genomics analysis can also be represented through the creation of 3D visualisations of molecular structures. In order for a 3D image to be pedagogically successful, Allen (1991) stresses the importance of the inclusion of spatial information (spatial cues). Hence in order to achieve visual literacy, not only must learners be able to navigate 3D space and construct mental models of their environment, but they must master complex rules for inferring 3D relationships from 2D cues (Allen, 1991).

Wu and Shah (2004) have suggested five principles for the design of visualisation tools that support students' visuospatial thinking and in so doing, enable them to understand chemistry concepts and develop representational skills:

- providing multiple representations and descriptions;
- making linked referential connections visible;
- presenting the dynamic and interactive nature of chemistry;
- promoting the transformation between 2D and 3D; and
- reducing cognitive load by making information explicit and integrating information for students.

Interestingly, Shepard (1978) notes that the discovery of the three-dimensional double helical structure of DNA was in large part dependent upon the integration of spatial visualisation into the deductive process. Watson's own account of the journey toward this discovery highlights the role that visualisations and the connections inferred from various incarnations of visual representations played in his cognitive processes and those of his collaborator, Francis Crick (Watson, 1968). Whilst this seminal discovery is an exemplar of visuospatial reasoning at the expert level, the processes invoked by Watson and Crick would have included those described above as the visual interpretation of new information also necessitated the reinterpretation of previous work and revision of hypotheses (see also Figure 2).

For students, the successful transition through the developmental stages of metacognitive capability described earlier in this book by Gilbert (Chapter ); i.e. *acquisition*, *retention*, *retrieval* and *amendment*; would necessitate strategic scaffolding of the cognitive processes reflective of these principles.

In the context of genomics, the amendment stage is particularly central to interpretation and knowledge construction because the field continues to evolve at a rapid pace with advances in the research technology. Traditional representations utilised by learners in genomic analysis are information-rich but comprehensively distant from the 3D world of molecular interactions and relationships they are meant to convey. Contextual relevance, authenticity, and experience in traversing these domains are therefore paramount in facilitating the transition from learner to expert. Whilst experts are able to easily navigate and make connections between multiple representations of concepts and processes, the learner is still in the process of mapping their visual and conceptual knowledge. Kozma has observed that when chemistry students are presented with a 3D representation of a molecular structure, they tend to focus on surface features without thinking about the underlying chemical principles (Kozma, 2003; Kozma and Russell, 1997).

In genomics, 2D and 3D representations are not necessarily alternate representations of the identical molecule or chemical structure. For example, a Chime-rendered 3D structure (Figure 5) can represent a coded compilation of a series of 2D alignments of sequences from multiple genomes (Figure 4A). In other words, the 3D and 2D images contain the same information, but their respective interpretation can vary significantly depending upon the context in which they are presented and the experience of the learner. When undergraduate biology students are presented with informationally equivalent 2D and 3D representations of a viral surface protein, we have observed marked differences in their descriptive explanations of concepts relevant to virus-receptor interaction (See and Takayama, unpublished). For a 3D representation to be effective in supporting student understanding, the prior 3D literacy can significantly influence successful student engagement (Richardson and Richardson, 2002; Wu and Shah, 2004). Correct perception of the 3D structure is crucial for further cognitive operations, and problem solving tasks of increasing complexity should be scaffolded strategically to achieve optimal learning outcomes (Ferk et al., 2003).

The metavisual process in interpreting the 2D alignment (Figure 4A) is distinct from that utilised in understanding the 3D protein structure (Figure 5). Pattern recognition is often the initial simplest process the researcher engages in when first examining a 2D alignment. It is important to note, however, that what may appear to the novice as a simple scanning process belies the underlying reasoning and 'visual cues' (known as 'sequence motifs') that the researcher integrates into a working model based on interpretation of functional significance. That is, certain sequence motifs, which are indicative of evolutionary and functional conservation, are

identically preserved amongst and sometimes between species. The student must learn how to link these visual cues into a conceptual framework. Boyle and Boyle (2003) have explored alternate visual representations of genomic alignment data to facilitate student engagement and inquiry. In their case, alignments are not analysed by pattern recognition but by graphical representation of alignment frequencies.

In summary, metavisual ability in genomics is inherently embedded in the learning process as the student becomes increasingly familiar with the interpretation of comparative genomic analysis. Visual literacy also necessitates the ability to not only utilise the appropriate representation to convey specific functional information, but to be cognisant of scale and relationship.

## **5. COLLABORATIVE VISUALISATION IN GENOMICS: INQUIRY-BASED TEACHING**

Edelson and Gordin (1998) have suggested the key criteria that must be fulfilled for visualisations to succeed in the construction of a successful framework for inquiry-based learning are: motivating context, learner-appropriate activities, data selection, scaffolding interfaces and support for learning. The complexity of the cognitive processes that would be required for the student to actively process multiple representations of data-rich genomic information presents a challenge to the educator. Our goal was to develop an inquiry-based project that reflected authenticity and emphasised the process of scientific inquiry rather than content and outcomes. A key strategy in this project was to facilitate collaborative learning whereby the students were instrumental in creating the visual representations of genomic analyses and in so doing, directed their knowledge construction.

### **5.1 Visualising the Science of Genomics: an international online research project**

*Visualising the Science of Genomics* (VSG) was developed as an international research project to engage students in the active process of collaborative scientific inquiry. This unique project was conducted entirely online amongst geographically distanced participants who worked in 'research teams' of five students, each from a different country. Participants represented a diversity of scientific backgrounds including: microbiology, bioinformatics, medicine, chemical engineering, biotechnology, pharmaceutical sciences, molecular biology, medical chemistry, genetics,

biochemistry, mathematics and computer science. The international and multidisciplinary composition of each research team provided the context for scientific research as a concerted global effort dependent upon contributions by scientists with specific areas of expertise.

Students worked in teams to analyse, hypothesise, reflect, predict, visualise and formulate models based on genomic sequence data from the Human Immunodeficiency Virus-1 (HIV-1), the causative agent of AIDS. Contextual relevance was provided through the creation of case studies based on actual data. The goal of VSG was to allow students to assess and interpret available information, and to develop their own research questions and methodology. This approach contrasts significantly from the traditional university laboratory practical, in which the student learning experience is dictated by the 'aim of the experiment' and the prescribed methodology in the lab manual. The biological sciences have traditionally placed emphasis on laboratory classes to promote inquiry-based learning. In principle, the laboratory introduces the student to the *practice* of biology, whereby the learner is provided with the opportunity to apply his/her theoretical knowledge. One of the goals of the biology educator is to teach students how to 'think like scientists'; we aim to engage the student in a cognitive apprenticeship as a researcher. Paradoxically, in most laboratory courses the lab manual specifies the 'aims' or 'hypotheses', and the student follows an established protocol to conduct the experiment. A true cognitive apprenticeship, however, must include development of the thought processes that facilitate the formulation of a hypothesis, as well as the reasoning processes invoked in the development or application of appropriate methodology to test the hypothesis. Hence whilst technically the laboratory provides a tangible context, focus on content and outcome may override learning *how*. Kozma (2003) has highlighted the argument by Dunbar (1997) that naturalistic studies of scientific practice (that is, cognitive studies of how scientists think and solve problems) require the process to be examined in an authentic setting whereby scientists solve complex, extended scientific problems as they interact with colleagues and with resources in their research environment. In other words, a true cognitive apprenticeship must engage the student in what biologists would qualify as an *in vivo* (natural) setting as opposed to an *in vitro* setting (in an artificial environment). The VSG project endeavoured to foster authentic inquiry through the creation of a collaborative research community.

In view of the diverse backgrounds of the participants, preliminary information was sent to all participants prior to the start of the project. The information included background reading on HIV-1 as well as a CD-ROM tour and necessary technical information for the web-based work. The students were also encouraged to post brief introductions about themselves

in their team sites to initiate students into the social framework of their learning community. Indeed, the merits of a constructivist approach to learning in a networked community, whereby the social construction of knowledge is engendered through collaboration and opportunities for the transfer of authority to the student, have been demonstrated by others (Brown et al., 1989; Greeno, 1998; Greeno et al., 1996; Resnick, 1988).



Figure #11-8. [Homepage for VSG interface. A coloured version of this figure is available on the CDROM. The navigation buttons on the left enable students to access their individual research team areas for collaborative work, as well as the areas common to all teams. The VSG site can be visited at: [http://www.omnium.edu.au/courses/vsg\\_2003s1/base/index.php](http://www.omnium.edu.au/courses/vsg_2003s1/base/index.php)]

## 5.2 HIV-1 genomics analysis as an authentic inquiry-based approach

A two-week pilot project was conducted with 42 participants from 24 universities, based in eleven different countries. Each team was provided with a case study modelled on HIV-1 genomic sequence datasets obtained from the GenBank international public database (Benson, 2000). The sequences represented the clones (types) of HIV-1 viruses present in blood samples drawn from HIV-1-positive subjects. Students analysed sequences

derived from a region of the HIV-1 genome encoding gp120, a key viral surface protein involved in virus-host interaction. The interaction between gp120 and the host immune cell is a core concept that serves as a 'launching point' for student discussions on their ideas about significance of the data and strategies for further exploration. Online tutorials and 'seminars' were readily available, allowing students to refer to them as needed at their own pace. After assessing the available information, the teams faced the challenge of developing their own research question(s) and appropriate methodology for investigation. The greatest initial challenge was 'What is our question?' or 'What is our hypothesis?' before teams could proceed to analyse, interpret, and create visual representations of their analyses. Student discussions sought to determine appropriate strategies for comparative genomic sequence alignments (as in Figure 4A) to test their hypothesis; this became an iterative process for some teams as they revisited their initial hypothesis after performing preliminary alignments to seek patterns or trends in HIV-1 genomic mutations (which could be depicted as in Figure 4B).

VSG was modelled as a unique online collaborative research model for teaching and learning the scientific process. By shifting the emphasis from a content- and outcome-focused approach to a process-oriented approach, reflection and analysis can be facilitated. One of the greatest challenges for the educator is to teach students to apply what they have learned to a new situation; the process of transfer. The VSG project aims to identify and clarify the scientific process and situate abstract concepts in **relevant** and **authentic** contexts to aid students' understanding of the 'big picture' as well as the functional details.

The pedagogical approach focused on open-ended inquiry, so that students were not pressured to 'produce results', allowing them to explore, research and pursue approaches that were of interest to them. Open exploration can encourage intrinsically motivated learning; Vollmeyer and colleagues (1996) have demonstrated that learning from computer-based biology problems was retarded when tertiary students were directed to solve problems with a specific goal in comparison to students who were instructed to 'explore the system'. In teaching biology students about the relationship between genomic sequence analysis and an organism's evolutionary history, Parker has also argued that an open-ended question-driven approach motivates students and is a more realistic way to teach science (Parker et al., 2004).

### **5.3 Modes of visual and textual interaction online**

Collaborative learning was facilitated through a highly interactive interface originally created to teach students in graphic design (Bennett, 1999). The Web-based platform was ideally suited to facilitate visual communication as well as textual communication amongst participants. Figure 9 shows the different modes of interaction amongst research team members (A) and between research teams (B). A coloured version of this figure is available on the CDROM.

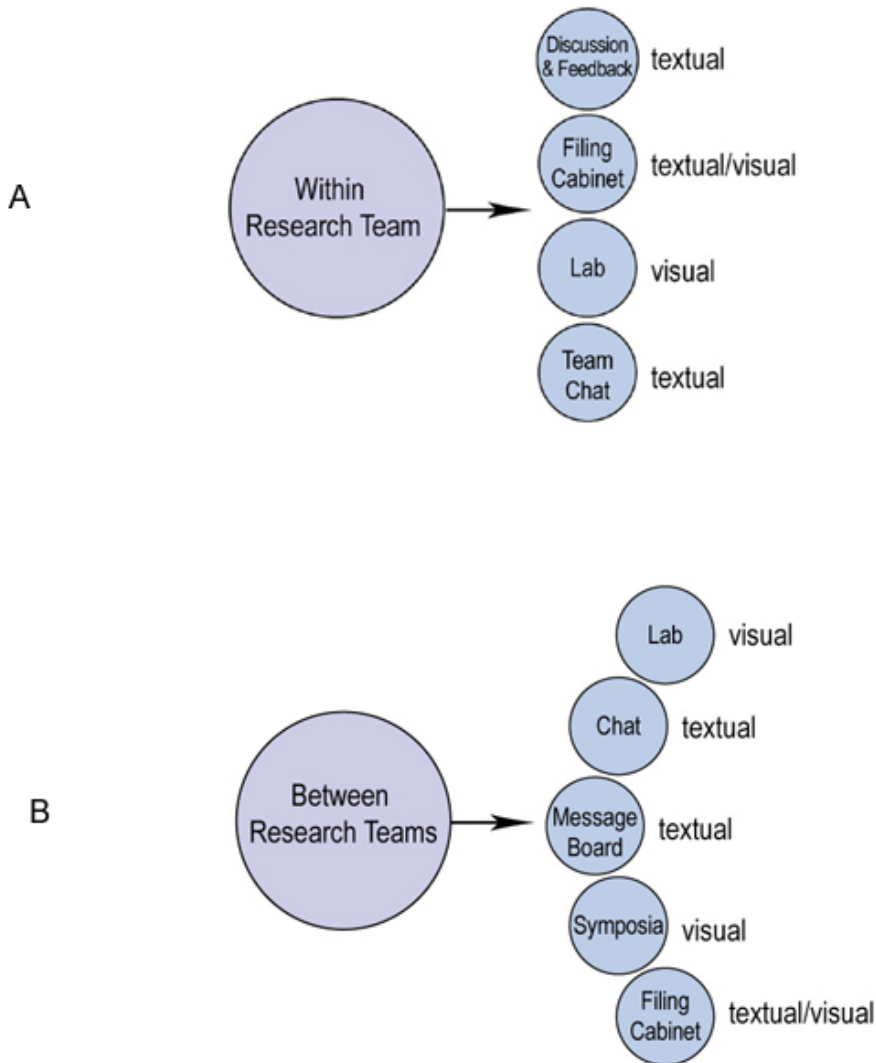


Figure #11-9. [Modes of interaction amongst the students in each VSG research team (A) and between research teams (B). A coloured version of this figure is available on the CDROM]

The availability of both visual and textual areas for communication amongst these multidisciplinary teams allows students to determine appropriate modes of interaction. The learner-centred approach also sought to promote facility in engagement of students from various disciplines with different cognitive learning styles. Studies on multiple representations have demonstrated that text information is better remembered when it is

illustrated by images than without illustrations (Levie and Lentz, 1982; Levin et al., 1987). The basis for such findings can be ascribed to Paivio's dual coding theory which proposes that cognitive processing of verbal information occurs in a separate system from cognitive processing of pictorial information (Paivio, 1986). Hence, the engagement of both systems for illustration + text enhances the memory. Furthermore, as verbal and pictorial explanations are processed in distinct cognitive systems, they are organised into mental models in different parts of the working memory (Chandler and Sweller, 1991; Mayer, 1997). Successful integration of the two processes results when components of the textual model and the corresponding components of the visual model are simultaneously activated (Chandler and Sweller, 1991; Sweller, 1999). How the text is integrated with the image is important; Mayer (1997) proposes that visual representations optimally support comprehension when text and images are explanatory, when textual and visual information are presented closely together spatially and temporally, their respective content is related to each other, and when learners have low prior knowledge about the subject but possess high spatial cognitive ability. Yet, others caution that the design of the integration of textual and visual information should be considered carefully to avoid learning interference due to task-inappropriate graphics (Kirby, 1993; Schnotz and Bannert, 2003). As the structure of graphics affects the structure of the mental model, the form of visualisation employed should support the construction of a task-appropriate mental model, and well-designed visualisations can indeed facilitate this process in learners with high prior knowledge (Schnotz and Bannert, 2003).

The VSG approach takes into consideration all of the above arguments by focusing on the *student's primary role in the construction and iterative adaptation of integrated visual and textual models*. An examination of the team discussions reveals that the students have utilised visualisations in several ways:

- To represent the process of their genomic analysis.
- To facilitate the creation and interpretation of conceptual links between HIV genomic sequences, HIV protein structure and function, HIV evolution, and clinical implications for patients.
- To provide appropriate representations (2D alignment vs 3D model) to convey specific concepts.

The following excerpt from a Team Capella discussion, in which a team member proposes a strategy for their case study, exemplifies the integration of visual cognition into the scientific process:

“Summary of the team’s reunion on Sunday 25:

Choosing the subjects and how to use the data obtained:

1. Select 2 patients to focus on: one slow progressor, one fast progressor (as defined by their CD4 profiles)
2. Either: a) select one to follow over time, and align the sequences from this clone for all visits; b) do this for several or all clones

You will be performing (2) for both patients.

3. Visualise the divergence over time by phylogenetic tree-building of your alignments. This function is easily available in Bio Workbench. This allows you to measure evolutionary distances and to see whether there are marked new ‘branches’ that form as the viral clones are mutating.
4. Model your alignments onto a 3D structure of gp120 in Protein Explorer to examine where the mutating regions are located, and where the conserved regions are located in relation to the actual shape of gp120, and their functional interactions with CD4, neutralising antibodies, etc.”

Another student from Team Avior represented his hypothesis of the presence of a converging selective pressure by linking the patterns of variability in genomic sequence at two distinct onsets of disease, and statistical analysis of evolutionary distances. His integration of visual and statistical information demonstrates his ability to integrate his cognitive processes to create his conceptual model, as shown in Figure 10 (a coloured version of this figure is available on the CDROM):

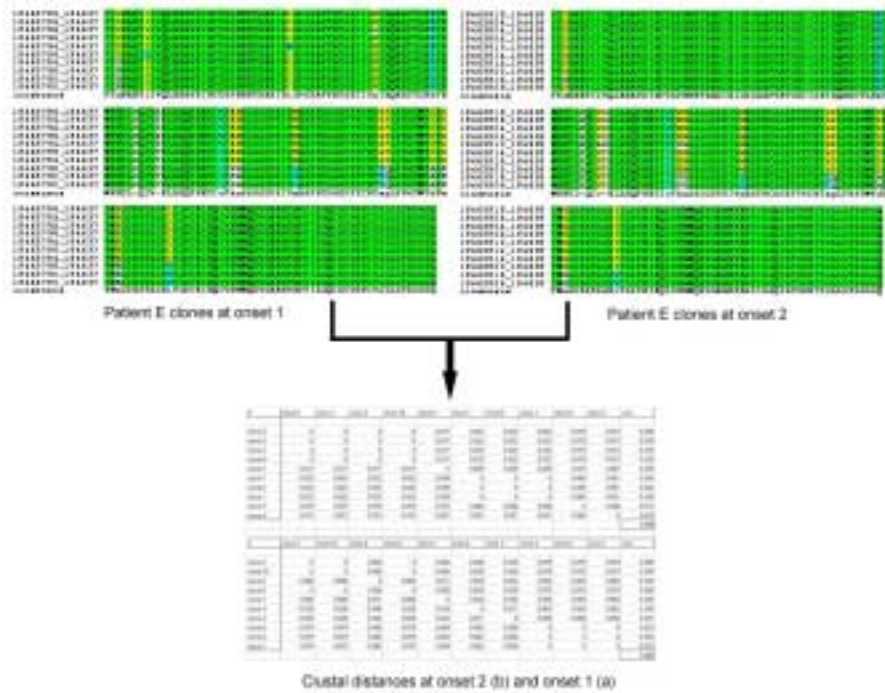


Figure #11-10. [Conceptual model by a Team Avior member (Samiran) integrating patterns of variability in genomic sequence at two distinct onsets of disease, and statistical analysis of evolutionary distances based on HIV-1 gp120 sequence information derived from clinical blood samples. A coloured version of this figure is available on the CDROM.]

The collaborative outcomes of the student teams also extended to 3D representations of mutational patterns of HIV-1 viruses. HIV-1 protein sequence alignments were mapped onto three-dimensional (3D) structural models of gp120. The 3D visualisation facilitates the understanding of how mutations in the HIV-1 gp120 genome may ultimately affect protein structure and functional interaction with the CD4<sup>+</sup> T cell receptor. The open source web-based program, Protein Explorer (Martz, 2002), was utilised to model protein sequence alignments and test predictions. Protein Explorer is a versatile program that enables 3D modelling, visualisation, and manipulation of protein structures. Multiple sequence alignments were mapped using Protein Explorer and ConSurf (Glaser et al., 2003) to identify amino acid residue positions that remained highly conserved in comparison to those that mutated significantly.

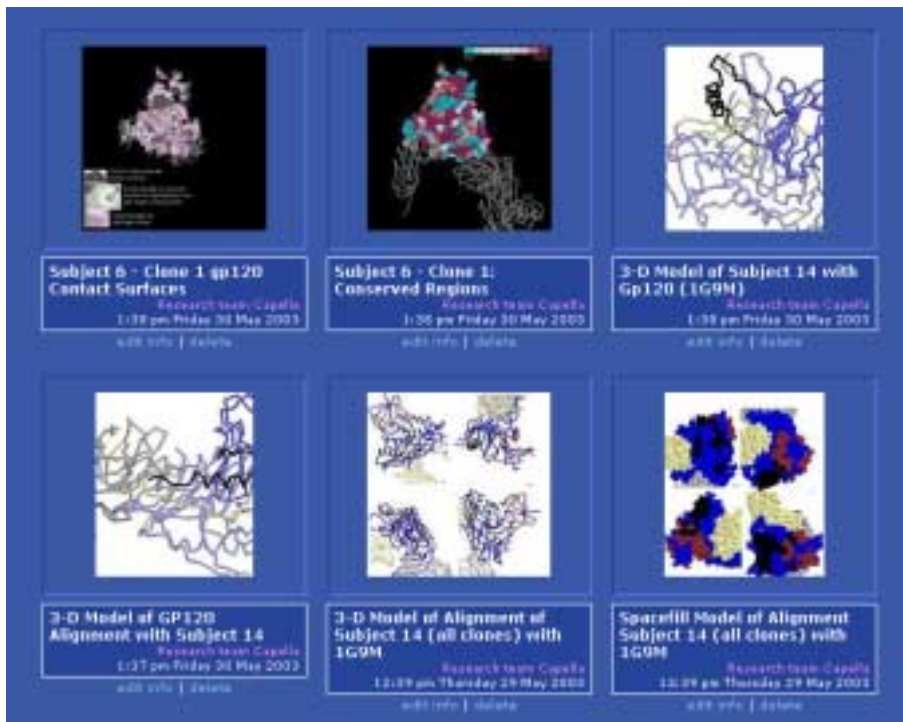


Figure #11-11. [Screen capture from the Symposium entitled ‘Week 2: 3D Models of HIV-1 GP120’. Each panel represents a team submission of a 3D model representing comparative analysis of protein sequences derived from HIV-1 gp120 genomic information. A coloured version of this figure is available on the CDROM.]

The images in Figure 11 represent screen captures of the students’ 3D models of the HIV-1 gp120 protein structure (a coloured version of this figure is available on the CDROM). The models represent variations in scale, depth, perspective, and context; all of these aspects are outcomes of conceptual frameworks created through team collaboration. Kozma’s observations of students using a chemical modelling package have suggested that materials features of representations can support student learning if they correspond to certain characteristics of abstract, scientific entities that do not otherwise have a concrete, visible character (Kozma, 2000). Similarly, these computer-generated 3D models are visual representations of genomic data analyses which provide structural insight into the consequences of the high frequency of mutations in the HIV-1 genome. The case study further provides students with a biological and clinical context for HIV-1 genomic mutations, and the colour-coded comparisons demonstrate how the consequent changes in viral protein structure makes the development of vaccines and therapeutics extremely difficult.

## 5.4 The VSG learning community

According to Shaffer and Anundsen (1993), 'community' is defined as a dynamic entity that emerges when a group of individuals share common practices, are interdependent, make decisions jointly, identify themselves with something larger than the sum of their individual relationships, and make a long-term commitment to well-being (their own, one another's, and the group's). The online community is dependent upon these same attributes in the absence of face-to-face contact or a voice. Indeed, the initial challenge of a project like VSG is the development of the community itself, as learning goals are concurrently being frameworked.

The social interactions in the VSG community were integral to the collaborative learning efforts of its members. Palloff and Pratt (1999) stress the importance of the development of shared goals that are related to the learning process in an online community. The integration of these goals into the social dialogue amongst the students was indeed reflective of this effort. VSG reflects a situative approach to learning (Brown et al., 1989; Greeno, 1998; Resnick, 1988), whereby participation in processes shaped by interactions in the VSG community construct students' conceptual understanding of genomics. Such situative approaches focus on the construction of knowledge within the context of student interactions with others, with the learning material, and with their learning environment (Brown et al., 1989; Greeno, 1998; Greeno et al., 1996; Resnick, 1988).

Social interactions in the VSG community were assessed using a modified rubric based on Sringam and Geer's Cognitive Development Interactive Analysis Model (Sringam and Geer, 2000), with the inclusion of an additional category, 'socialisation'. The seven categories of interaction in the rubric were as follows:

- Socialisation
- Planning
- Sharing/comparing/contributing information
- Identifying or clarifying inconsistency of ideas, concepts or statements
- Negotiation of meaning/co-construction of knowledge
- Testing and modification of proposed synthesis or co-construction of knowledge
- Agreement statement(s) and application of newly constructed knowledge

Figure 12 summarises the distributions of the categories of social interactions that occurred amongst students in the Discussion and Feedback interface (a coloured version of this figure is available on the CDROM).

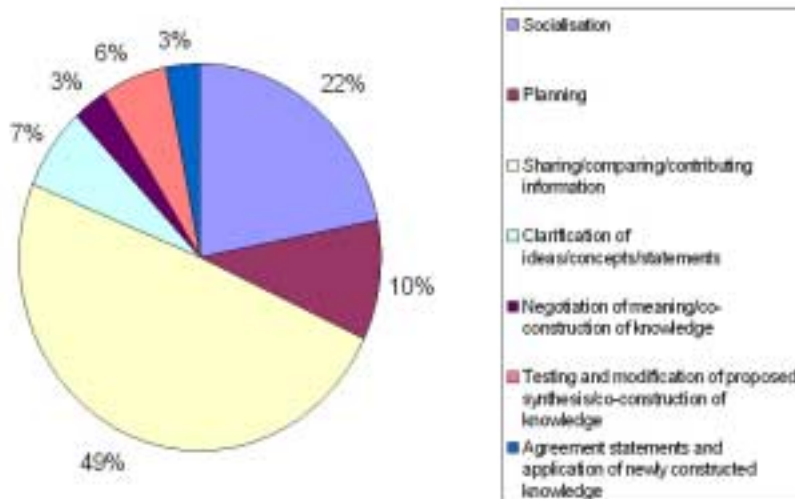


Figure #11-12. [Categories of social interaction amongst all students in the VSG Discussion and Feedback area. Interactions were scored using a modified rubric based on Sringam and Geer's Cognitive Development Interactive Analysis Model (Sringam and Geer, 2000). Data analysis was performed by J. Ang (2003). A coloured version of this figure is available on the CDROM.]

The two main categories of social interaction were i) socialisation and ii) sharing, comparing and contribution of information. Whilst socialisation (22% of interactions) plays a significant role in facilitating collaborative learning, task-related interactions (78%) were predominant in the VSG community. This is indicative of student commitment and strong involvement in the investigation. Furthermore, a common observation amongst the groups was the negotiation and development of a shared understanding re: 'the problem' (task). While this sometimes took the majority of the students' time during the 2-week project, this process was in itself a conduit toward cognitive development. This is evidenced by the dialogic progression of group discussions which fostered continual reflection and process-oriented critical analysis integrated with visual cognition.

The diversity of experience and background of the students were instrumental in creating the rich tapestry of this international online research community. The VSG community was further enhanced through the participation of five world-renowned HIV researchers, who appeared as 'guest tutors'. The guest tutors interacted in synchronous chat sessions to provide students with the opportunity to interact with 'real scientists' in the field, each of whom represented a distinct specialisation. This experience

underscored the students' appreciation for the multi-disciplinary nature of research.

### 5.5 Analysis of cognitive interactions in the VSG community

One of the key criteria for an authentic learning experience is that of fidelity of context (Herrington and Herrington, 1998; Meyer, 1992; Reeves and Okey, 1996; Wiggins, 1993). For many of the students, it was the first time in their academic careers that they found themselves immersed in collaborative authentic inquiry, whereupon they were driven by intrinsic motivation.

A rubric for assessing cognitive interactions was developed, based on Biggs' and Collis' Structure of Observed Learning Outcomes (SOLO taxonomy) (Biggs and Collis, 1982). The majority of the interactions assessed via SOLO taxonomy were indicative of higher levels of cognitive ability (relational and extended abstract). The cognitive interactions are summarised in Figure 13 (a coloured version of this figure is available on the CDROM).

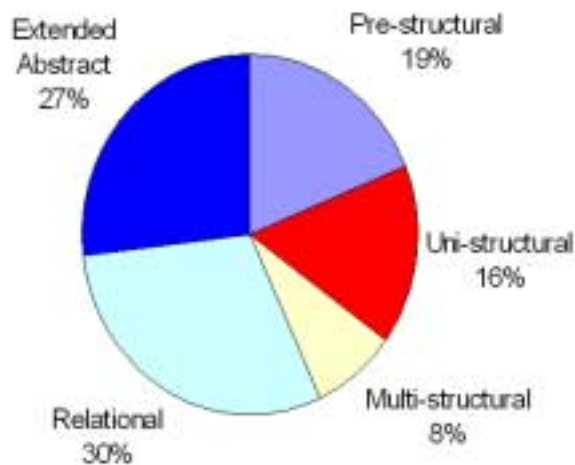


Figure #11-13. [An example of cognitive assessment of student discussions. Transcripts from the VSG Discussion and Feedback area were analysed by SOLO taxonomy. Data analysis was performed by J. Ang (2003). A coloured version of this figure is available on the CDROM]

Student discussions were characterised by analytical, contextual, and social dialogue. The instructor provided feedback and facilitation when appropriate. Whilst most teams initially performed similar preliminary ‘experiments’, the defined goals, specific datasets chosen, and strategies developed by the teams varied. This diversity of approaches exhibited by teams to develop their ‘question’ and ‘process’ may be reflective of the different perspectives, analyses, and expertise provided by the members of each research team. It is possible also that students may exhibit variations in metavisual competence, but this was not rigorously assessed.

The cognitive levels of interaction in VSG were higher than that observed in the instructor’s classroom teaching. Whilst this may have been due in part to the calibre of students that had volunteered to participate in the project, the approach utilised was characterised by several qualities that may also have strongly contributed to the learning outcomes:

- Student-centred collaborative approach
- Open-ended scientific inquiry process
- The creation of a strong online community of students and instructor
- Contextual visualisations

Charlin and colleagues (1998) emphasise a learner-centred approach towards problem-solving as being of key importance, and define four principles related to their effect on learning:

1. Learners are active processors of information;
2. Prior knowledge is activated and new knowledge is built on it;
3. Knowledge is acquired in a meaningful context;
4. Learners have opportunities for elaboration and organisation of knowledge.

These principles are indeed reflective of the importance of contextual relevance for students who are presented with an abstract concept like genomics (Chinn and Malhotra, 2002; Tobias and Hake, 1988). The dialogue amongst students and between student and instructor revealed that the principles were indeed effectively utilised.

## **5.6 Analysis of the instructor’s role in VSG**

The instructor in a student-centred learning environment takes on new roles that are crucial in maintaining an interaction and collaboration amongst students. Technology-based learning communities like VSG where learning is dependent upon a socially interactive and collaborative experience are

guided by a social constructivist approach to teaching and learning (Blanton et al., 1998; Duffy and Cunningham, 1996; Jonassen and Reeves, 1996; Maor and Taylor, 1995; Tobin, 1993). Student cognition via such an approach takes place within a social context and collaboration is an essential component. In this environment, the instructor functions in several capacities: pedagogy, social interaction, management and technology (Bonk et al., 2001). Initial analysis of the instructor's contributions toward student team discussions revealed that the role as facilitator/motivator was nearly as prevalent as the pedagogical role. This is markedly distinct from what occurs in face-to-face teaching. The transfer of 'pedagogical authority' during discussions was facilitated through the collaborative nature of the modes of interaction between students and instructor. Figure 14 illustrates the modes of interaction between the instructor and the students (a coloured version of this figure is available on the CDROM).

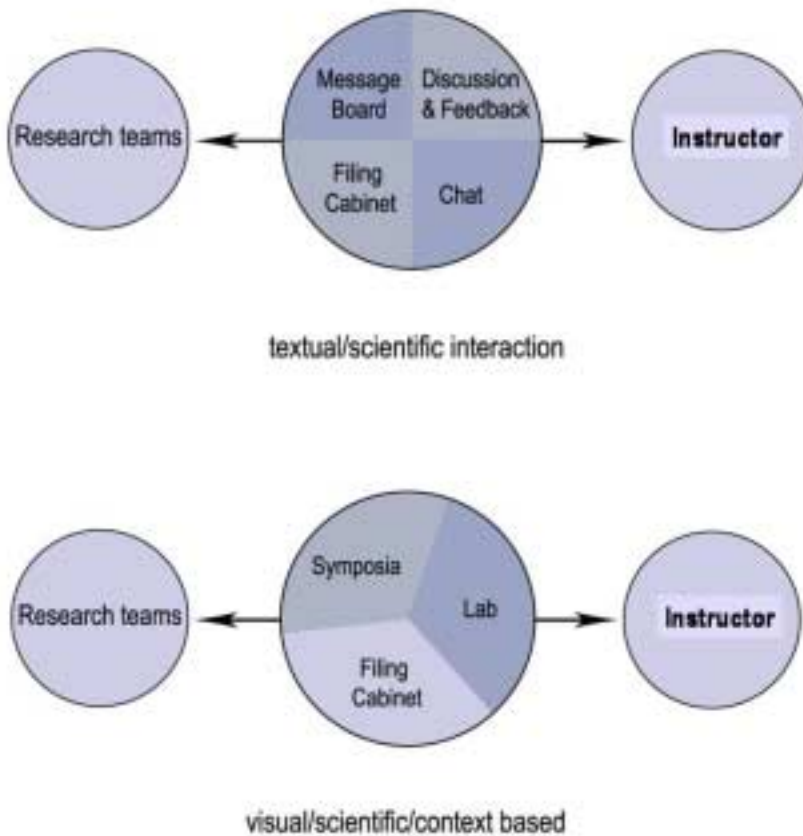


Figure #11-14. [Modes of interaction between the students and the instructor in VSG. A coloured version of this figure is available on the CDROM.]

The facilitative nature of these interactive spaces was reflected in the dialogue between the instructor and the students, as visual cues were integrated with conceptual discussions. The situative learning approach also enabled the instructor to adapt to the individual needs of the teams in order to promote their metacognition of the scientific process.

## 6. SUMMARY AND CONCLUSIONS

As we continue to witness the rapid progression of post-genomic technology and the consequential exponential increase in the quantity and diversity of information, educators must effectively integrate highly networked concepts into the curriculum without overloading the cognitive capabilities of the learner. Wandersee (2000) describes ‘meaning-making’ in biology as the understanding of a concept, which often involves visualisation. Visual representations can indeed enhance the learning process if they are strategically designed and effectively utilised with both metacognitive and metavisual scaffolding. Whether the student is viewing a static visualisation in the textbook or a 3D interactive representation on a computer, the learning goals must be appropriately supported by the mode of representation. To achieve these pedagogical goals, I would venture to propose that university researcher-educators develop an awareness of their own metavisual thinking to engage in a distillation of the layers of ‘assumed visual knowledge’ necessitated by their practice. Such a process can facilitate the generation of cognitively more intuitive representations for not only students but the research audience as well.

Several exemplars are indeed indicative of this intuitive direction taken by biologists. SequenceJuxtaposer (Slack et al., 2004) was developed to facilitate visualisation of large-scale genomic sequence comparisons in context. An information visualisation technique called ‘accordion drawing’ was applied to enable ‘stretching out’ of genomic sequence regions when detailed views were required, and ‘contraction’ for a ‘global view’ of the entire genome. The creation of the visualisation takes into account the cognitive load required to maintain a mental model of sequence navigation history; the accordion approach reduces this load by allowing various scales of visualisation within one computer browser frame (Slack et al., 2004). The creation of an entirely new ‘language’ for the visualisation of complex data found its basis in collaboration between artist and scientist (Keefe et al., 2004). To distill the enormous amount of data obtained from magnetic resonance imaging (MRI) scans of mouse spinal cords, concepts from oil painting were applied to represent multivalued data with multiple layers of varying ‘brush strokes’. The metavisually cognisant integration of these borrowed expressive art techniques facilitated the simultaneous representation and comprehension of enormous amounts of data.

The ‘classroom’ is no longer a spatially and temporally fixed venue for teaching, which provides both challenges and opportunities for the educator. Similarly, as the biological playing field of ‘omics’ research continues to expand and evolve, the educator faces the challenge of integrating complex networks of information into multidisciplinary contexts. The potential of

visual representations in facilitating the understanding of the elegance of these newly discovered systems can indeed be realised not merely through the utilisation of the visualisation itself, but by the careful process behind its creation, to communicate our pedagogical goals.

## ACKNOWLEDGMENTS

I am grateful for the support and opportunity provided through an Innovative Teaching and Educational Technology (ITET) Fellowship from the Office of the Pro Vice Chancellor (Education) of The University of New South Wales, which launched me on this multi-dimensional journey. I am indebted to the UNSW Faculty of Science for its funding and support of the VSG project, and I wish to acknowledge the kind support of my colleagues at the UNSW College of Fine Arts. I would like to acknowledge the valuable input of Jessica Ang, whose honours thesis under my supervision examined the learning assessments of the VSG project. Discussions with my colleagues Barbara Tversky and Jan Plass, and my honours student Karen See have indeed been inspiring during the writing of this chapter. I am also grateful to the Carnegie Foundation for the Advancement of Teaching, at which I had the opportunity to engage in many stimulating discussions with colleagues during my tenure as a Carnegie Scholar with the Carnegie Academy for the Scholarship of Teaching and Learning. My appreciation and kudos go to the editor, John Gilbert, for his infinite patience and his insightful vision. Finally, I express my heartfelt thanks to all of my students, past and present, and especially to the VSG participants, who have challenged and enhanced my thinking in embracing new uncharted territories.

## REFERENCES

- Allen, B. S. (1991). Virtualities. *Educational Media and Technology: The Year in Review*, 17, 47-53.
- Ang, J. (2003). *Visualising the Science of Genomics*. Unpublished BSc. Honours, The University of New South Wales, Sydney.
- Bennett, R. (1999). *Omnium*, from [www.omnium.unsw.edu.au](http://www.omnium.unsw.edu.au)
- Ben-Zvi, R., Eylon, B., & Silberstein, J. (1988). Theories, principles and laws. *Education in Chemistry (May)*, 89-92.
- Benson, D.A., Karsch-Mizrachi, I., Lipman D.J., Ostell, J., Rapp, B.A., Wheeler, D.L. (2000). GenBank. *Nucleic Acids Research*, 28, 15-18
- Bernal, A., Ear, U., & Kyrpides, N. (2001). Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Research*, 29, 126-127.

- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the Quality of Learning: the SOLO Taxonomy*. New York: Academic Press.
- Blanton, W. E., Moorman, G., & Trathern, W. (1998). Telecommunications and teacher education: a social constructivist review. *Review of Educational Research*, 23, 235-275.
- Bonk, C., Kirkley, J., Hara, N., & Dennen, V. (2001). Finding the instructor in post secondary online learning: pedagogical, social, managerial, and technological location. In J. Stephenson (Ed.), *Teaching & learning online: Pedagogies for new technologies*. (pp. 76-98). London: Kogan Page.
- Boyle, A. P., & Boyle, J. A. (2003). Visualization of aligned genomic open reading frame data. *Biochemistry and Molecular Biology Education*, 31(1), 64-68.
- Brooks Jr., F. P. (1996). The computer scientist as a toolsmith II. *Communications of the ACM*, 39(3), 61-68.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18, 32-42.
- Campbell, A. M. (2004, 2004). *Genome consortium for active teaching*. Retrieved May 23, 2004, from <http://www.bio.davidson.edu/projects/GCAT/gcat.html>
- Campbell, A. M., & Heyer, L. J. (2003). *Discovering genomics, proteomics, and bioinformatics*. San Francisco: Pearson Education, Inc. (Benjamin Cummings).
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293-332.
- Charlin, B., Mann, K., & Hansen, P. (1998). The many faces of problem-based learning: A framework for understanding and comparison. *Medical Teacher*, 20(4), 323-330.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175-218.
- Christopherson, J. T. (1997). *The growing need for visual literacy at the university*. Paper presented at the Visionquest: Journeys toward visual literacy; 28th Annual Conference of the International Visual Literacy Association, Cheyenne, WY.
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: making thinking visible. *American Educator*(Winter), 6 - 46.
- Consortium, T. G. O. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25, 25-29.
- Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied*, 7(1), 68-82.
- Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology*. (pp. 170-198). New York: Macmillan.
- Dunbar, K. (1997). How scientists really reason: scientific reasoning in real-world laboratories. In R. Sternberg & J. Davidson (Eds.), *The nature of insight*. (pp. 365-396). Cambridge, MA: MIT Press.
- Edelson, D. C., & Gordin, D. (1998). Visualization for learners: a framework for adapting scientists' tools. *Computers and Geosciences*, 24(7), 607-616.
- Ferk, V., Vrtacnik, M., Blejec, A., & Gril, A. (2003). Students' understanding of molecular structure representations. *International Journal of Science Education*, 25(10), 1227-1245.
- Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223), 496 - 512.
- Gilbert, J. K. (2002). *Moving between the modes of representation of a model in science education: some theoretical and pedagogic implications*. Paper presented at the

- Philosophical, Psychological, Linguistic Foundations for Language and Society Literacy, University of Victoria, Canada, September 12 - 15, 2002.
- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., et al. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, *19*, 163-164.
- Greeno, J. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*(1), 5-26.
- Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology*. (pp. 15-46). New York: Macmillan.
- Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: how university students respond to a model of authentic assessment. *Higher Education Research and Development*, *17*(3), 305 - 322.
- Jonassen, D. H., & Reeves, T. C. (1996). Learning with technology: using computers as cognitive tools. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology*. (pp. 693-720). New York: Macmillan.
- Keefe, D. F., Kirby, R. M., & Laidlaw, D. H. (2004). Painting and Visualization. In C. R. Johnson & C. Hansen (Eds.), *Visualization Handbook 2004*: Elsevier.
- Keig, P. F., & Rubba, P. A. (1993). Translation of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge. *Journal of Research in Science Teaching*, *30*(8), 883-903.
- Kirby, J. (1993). Collaborative and competitive effects of verbal and spatial processes. *Learning and Instruction*, *3*, 201-214.
- Kolb, D. A. (1984). *Experiential Learning: experience as the source of learning and development*. New Jersey: Prentice-Hall.
- Kozma, R. (2000). Students collaborating with computer models and physical experiments. In C. Hoadley (Ed.), *Computer support for collaborative learning*. (pp. 314-322). Mahwah, NJ: Erlbaum.
- Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction*, *13*, 205-226.
- Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching*, *34*, 949-968.
- Krypides, N. (August 31, 2004). *Genomes OnLine Database*. Retrieved September 1, 2004, from <http://www.genomesonline.org/>
- Krypides, N. (1999). Genomes OnLine Database (GOLD): a monitor of complete and ongoing genome projects worldwide. *Bioinformatics*, *15*, 773-774.
- Lander, E., Linton, L., Birren, B., Nusbaum, C., Zody, M., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860 - 921.
- Levie, H. W., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal*, *30*, 195-232.
- Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration*. (Vol. 1, pp. 51-86). New York: Springer.
- Maor, D., & Taylor, P. C. (1995). Teacher epistemology and scientific inquiry in a computerised classroom environment. *Journal of Research in Science Teaching*, *32*, 839-854.
- Martz, E. (2002). Protein Explorer: Easy yet powerful macromolecular visualization. *Trends in Biochemical Sciences*, *27*(February), 107-109.

- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32, 1-19.
- Meyer, C. A. (1992). What's the difference between authentic and performance assessment? *Educational Leadership*, 49, 39 - 40.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.
- Palloff, R. M., & Pratt, K. (1999). *Building Learning Communities in Cyberspace*. San Francisco: Jossey-Bass.
- Parker, J. D., Ziembra, R. E., Cahan, S. H., & Rissing, S. W. (2004). An hypothesis-driven, molecular phylogenetics exercise for college biology students. *Biochemistry and Molecular Biology Education*, 32(2), 108-114.
- Reeves, T. C., & Okey, J. R. (1996). Alternative assessment for constructivist learning environments. In B. G. Wilson (Ed.), *Constructivist learning environments: case studies in instructional design* (pp. 191 - 202). Englewood Cliffs, NJ: Educational Technology Publications.
- Resnick, L. (1988). Learning in school and out. *Educational Researcher*, 16(9), 13-20.
- Richardson, D. C., & Richardson, J. S. (2002). Teaching molecular 3-D literacy. *Biochemistry and Molecular Biology Education*, 30(1), 21-26.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13, 141-156.
- Seddon, G. M., Eniayiju, P. A., & Chia, L. H. L. (1985). The factor structure for mental rotations of three-dimensional structures represented in diagrams. *Research in Science and Technological Education*, 3(1), 29-42.
- See, K., & Takayama, K. (unpublished).
- Shaffer, C., & Anundsen, K. (1993). *Creating Community Anywhere*. New York: Jeremy P. Tarcher/Perigee Books.
- Shepard, R. N. (1978). Externalization of mental images and the act of creation. In B. S. Randhawa & W. E. Coffman (Eds.), *Visual learning, thinking, and communication* (pp. 133-189). New York: Academic Press.
- Shi, L. (1998, January 7, 2002). *DNA microarray (genome chip)- monitoring the genome on a chip*. Retrieved May 23, 2004, 2004, from <http://www.gene-chips.com/>
- Slack, J., Hildebrand, K., Munzner, T., & St. John, K. (2004). *SequenceJuxtaposer: Fluid navigation for large-scale sequence comparison in context*. Paper presented at the German Conference on Bioinformatics, October 4 - 6, 2004, Bielefeld, Germany.
- Sringam, C., & Geer, R. (2000). *An investigation of an instrument for analysis of student-led electronic discussions*. Paper presented at the Learning to Choose, ASCILITE 2000 Conference, Coffs Harbour, NSW, Australia.
- Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Victoria: ACER Press.
- Tobias, S., & Hake, R. R. (1988). Professors as physics students: What can they teach us? *American Journal of Physics*, 56, 786-794.
- Tobin, K. (1993). Constructivist perspectives on teacher learning. In K. Tobin (Ed.), *The practice of constructivism in science education*. Hillsdale, NJ: Lawrence.
- Tuckey, H., Selvaratnam, M., & Bradley, J. (1991). Identification and rectification of student difficulties concerning three-dimensional structures, rotations, and reflection. *Journal of Chemical Education*, 68(6), 460-464.
- Vollmeyer, R., Burns, B., & Holyoak, K. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science*, 20, 75-100.

- Vrtacnik, M., Ferik, V., Dolnicar, D., Zupancic-Brouwer, N., & Sajovec, M. (2000a). The impact of visualisation on the quality of chemistry knowledge. *Informatica*, 24, 497-503.
- Vrtacnik, M., Sajovec, M., Dolnicar, D., Pucko-Razdevsek, C., Glazar, S. A., & Zupancic-Brouwer, N. (2000b). An interactive multimedia tutorial teaching unit and its effects on the students perception and understanding of chemical concepts. *Westminster Studies in Education*, 23, 91-105.
- Wandersee, J. H. (2000). Language, analogy, and biology. In K. M. Fisher, J. H. Wandersee & D. E. Moody (Eds.), *Mapping Biology Knowledge*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Watson, J. D. (1968). *The double helix*. New York: New American Library.
- Watson, J. D., & Crick, F. H. C. (1953). A structure for Deoxyribose Nucleic Acid. *Nature*, 171, 737-738.
- Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.
- Wu, H.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, 88, 465-492.