Navigating genomes: the space in which genes happen.
By Steve Hughes, Egenis, Exeter

Abstract
I will explore contrasting conceptualisations (modellings, mappings) of genomes, either as uni-dimensional text or alternatively as activity spaces, in relation to the ways in which genomic knowledge translates into instrumental value or causal explanation. I will argue that the contrasting models are reflections of a basic and inevitable divergence in the modelling of complex systems and I will relate this to the worlds of intellectual property, invention and innovation as well as to the ways in which genomic exploration is driven.

Introduction
I base the following argument on the proposition that the way in which we see the world influences the way in which we act in it and vice versa that the way in which we choose or are persuaded to act influences the way in which we see. In particular I shall argue that common representations of genomes reflect a practice of commodification and annexation rather than providing heuristic value.

The contemporary view of genes and genomes is dominated by re-materialised conceptualisations in which the DNA component, a string of sequential nucleotide representations ACGT which we see poured out by automated chemical analysis, is accorded a special and isolated status as prime determinant of the nature of living things. Haraway (1997) has referred to this "trope" or style of representation as a fetishism sustained by the need to present constructive endeavour as "under control". We have become accustomed to a set of persuasive metaphors linked to the string conceptualisation, "the book of life", " the language of the genes", "recipe for life", "blueprint" In this form, as a material script sited at the centre of a system to which all else is peripheral and subservient it is a prime target for annexation and commercial ambition which only serves to reinforce the tendency to reify. A one dimensional string is a simple place to map and to navigate and in which to stake one's claim to what appears to be a valuable segment. Perhaps it is not surprising that the language of genomic analysis is rich with gold-rush metaphor (gene prospecting, data mining, panning (in phage genomic display)). The rush to acquire sequence and to control and centralise the deployment of the associated knowledge in a way reflects the DNA-centric view of biological organisation in which DNA is packed into the nucleus as the cellular command and control sector, enclosed by the nuclear membrane like the castle walls of the feudal/baronial control system, or the secure buildings of government ministries. Thus all other cellular components are excluded from the control function and held at the periphery (extra mural) in the role of responsive producers like feudal surfs. To have control or ownership of the controllers makes one king of the castle.

This walling-off of the genome into the DNA component of living systems supports the representation of "the Human Genome" as a computerised database of DNA sequences detached from its functional context. Indeed the formal annexation of segments of the DNA string as patents actually formally requires separation from context. A consequence of this narrow conceptualisation has been the need to invent "functional genomics", "comparative genomics", "transcriptomics" and "proteomics" as distinct subjects of investigation. Rematerialisation of genomes as the DNA sequence dataset has similarly spawned "post-genomic" for the phase of human development following the completion of major DNA sequencing projects during which we will ponder what to make of the information in our electronic files and how to learn from it. In this phase we expect to witness attempts at the reconnection of the genomic text with its functional context. The enabling process is called among other things *annotation* and amounts to the building of another text around the core, which assigns a linkage to function reflected at some level or other of the "DNA makes RNA makes protein makes phenotype" of the central dogma of molecular biology.

I find the above proliferation of –omics objectionable because it segments fields of investigation in a facile way according to the tools used to acquire datasets rather than to distinguish the nature of the questions we need to ask. It reflects a mode of thinking, or perhaps an alternative to thinking which says "lets get the data in the bag and work out what to do with it later". A quote from one of the keynote speakers at a recent conference on genomics (Genomics Momentum, Rotterdam 2004) "we got human we got mouse and rat we got the dog, we're close on bat and kangaroo", is illustrative. One could not help but reflect on earlier triumphalism emanating from the same part of the world "we got Saddam"!

The materialised concept of genes as discrete DNA strings interspersed among so called junk DNA is increasingly untenable as we learn more about the patch-working eg trans-splicing, of transcripts from dispersed loci as templates for a kind of mix and match production system for proteins. It is apparent that there is a greater scope for diversity among genes and their outcomes than can be accounted for in the DNA script taken at the face value envisaged in current annotations. Organisms exhibiting dramatically different levels of organisational complexity (eg Arabidopsis, Homo sapiens), as seen from a conventional human perspective, seem to draw upon DNA strings sets of rather similar complexity suggesting that it is the actions of deployment of that resource which underpins genomic diversity.

The singular sequence concept of "the" genome of "an" organism similarly becomes untenable as we learn more about patterns of methylation of the DNA component, a chemistry which significantly alters DNA and its functionality within genomes in a way which is dynamic and variable over short ranges and timescales (from cell to cell and minute to minute) and also long timescales with implications for heritability(often called genetic imprinting). I argue that this implies an increased dimensionality even for the DNA string and that we have to think of genomes rather than "the" genome within an organism's space and time coordinates. Conventional chemical deciphering of DNA sequence has conveniently expunged the methylation pattern of the DNA bases making it possible to speak of a single ,"the", genome, in doing so further distancing the concept of "the" genome from functional realities and pushing it in the direction of a clean slate. This may be seen by some as driving the genome into abstract space to which I would counter that on the contrary what we are doing is forcing it into the familiar material of hard text, the very convention through which we conduct transactions and exercise controls, title and ownership and mark identity.

An alternative and more broadly accommodating viewpoint, which I favour, regards the genome as a space of activity and sees genes as activities within that space. This configuration has been adopted by some mathematical modellers interested in the complexity of biological systems (Wolkenhaur, Kolch Cho 2003). For them as for me activity is characterised by molecular interactions between a diverse set of components of which DNA strings are just one. In fact DNA as "stuff" takes its properties from its interactions with other stuff. Molecular interactions may be regarded as transient information transfer events or transformations, which are difficult to represent as material entities and still more difficult to annex within a simply defined space. We may see them as having some similarity with abstract thoughts which collectively produce action or in our case a phenotype. At the same time the interacting components are undeniably material (chemically defined entities) as are the outcomes of the transitions they promote. There is an intrinsic dualism here akin to mind/brain which inevitably will make spatial modelling of genomes unattractive to some.

As discussed above, however, the reality of genomic space is multidimensional, dynamic and complex for the navigator, and additionally, given what we know of the significance of mobile genomic components, like siRNAs and plasmids it is increasingly necessary to maintain a flexible view of its boundaries. Spatial conceptualisation and modelling of genomes is significant and attractive in the heuristic sense as a vehicle for building explanations of genes and genomes in terms of the nature of the relationships between their interacting components . Genomes in this conceptualisation are not limited to nuclear space, in fact genomic space can be extended, dependent on the activities in question beyond the boundaries of the host organism itself as we would see for instance in the discourse on genes and animal behaviour.

Combination of heuristic intent and intrinsic complexity ensures that there will be no post-genomic era for this conceptualisation. Nor is there nor will there be disciplines of comparative genomics or functional genomics. Genomics is nothing if not intrinsically comparative or functional.

This spatial perspective of genomes does have an historical foundation just as do genes as actions. The term genome was coined by Winkler (1920)[1] in a paper dealing the complexities of explaining plant

[1] This is the 1920 definition of the term 'genome' by Hans Winkler, a professor of botany in Hamburg. Presumably, there must be a rich, as yet unwritten history which would connect this definition to his experiments on plant hybridisation and contemporary theories about parthenogenesis and apogamy. I have tried to translate as faithfully as possible but the nuances of language are of course infinite. Adam Bostanci 2003

"I propose to use the expression genome for the haploid set of chromosomes, which, together with the associated protoplasm, constitute the material basis of the systematic unit, and to call 'homogenomatisch' those nuclei, cells, and organisms, in which a genome of the same kind is present more than once in each nucleus, and 'heterogenomatisch' those which possess different kinds of genomes in the cell nucleus. Individuals, which possess

variation as observed in hybrids. It was founded on the concept of a package (of traits) possessing a defined provenance and a self defining distinct continuity denoting that provenance (my interpretation). This allowed for the description of a hybrid in terms of the ancestral genomes it contains, though as we now know , dependent upon the difference between the parents of the hybrid (closely related , or distantly related (the "wide" hybrid)) the constituent genomes may physically blend conserving the diploid chromosome number (for example: parentA 4 pairs crossed with parent b 4 pairs = hybrid 4 pairs, or alternatively the sum of the two contributed genomes ; parentA 4 pairs x parent b 4 pairs = hybrid 8 pairs. Winkler's concept, early as it was, set the landscape for spatial conceptualisation and is still a challenge for uni-dimensional string thinking. Take for example hexaploid bread wheat the genomic space of which contains three diverse ancestral genome complements reflecting its ancestry as a natural hybrid. However, as illustrated by ribosomal RNA transcription units (nucleolus organising centres), structurally equivalent loci (equivalent DNA strings) among the three genomes exhibit diverse behaviours dependent upon which part of genomic space they occupy, for instance whether they lie within diploid or hexaploid space. I will cite this as a sufficient example of the realistic and the operational value of the spatial concept and will not use this paper to recite the collection of other genomic observations which support the argument that spatially distributed activity is key to a generally valuable way of conceptualising genomes, that is the stuff of a sister paper. Suffice it to remind ourselves that the language of genetics has stabilised around terms like "position effect", "translocation", "inversion" "ectopic expression" "scaffold-association" and "syntenic" to support explanation of the relationship of position and action.

Reconciliation of the string and spatial conceptualisations

As I have suggested above in a tangential way the instrumental versus heuristic divergence of the contrasting conceptualisations reflects a divergent motive or set of motives or imperatives for dealing with the world and its complexity. The former aligns with the need for command, control and containment and the latter with the urge to explore and to know about the deeper nature of things. As such they reflect a long established pattern of behaviour witnessed in man's artefacts since earliest times. For me the divergence is most attractively portrayed in the cave paintings of Southern Europe where the dramatic and vivid contrast between naturalistic and formalised representations has been cited (Witkin 1995 and authors cited therein) as illustrative of continuing strands in art in society linked to their aristocratic (socially codified, instructional, un-negotiated mores and rules,) or individualistic nature. So perhaps we have to accept the two conceptualisations will be retained by those who demand control, "completeness" and uniformity in data sets or by those who are content with fragmentary but realistic (spatial) visions of activity in the genomic space.

I have drawn on the term "happen", as in "the space in which genes happen" in order to support coexistence but at the same time to bolster the neglected spatial view. Concepts can "happen to be", i.e. they can be re-materialised in some form, or they can happen by being the basis of actions or events. But "happen" in common use is most strongly pinned to events or transformations.

Perhaps in time the so called "systems biology" can provide an alternative route to reconciliation. Systems biology recognises all the fragmented approaches to data collection in genomics (xxxomics) as contributing entities and provides an institutional banner for their re-aggregation.

Implications for knowledge management

---

the same genomes, shall be referred to as 'isogenomatisch', those which possess different kinds of genomes as 'anisogenomatisch'. Depending on the number of sets of chromosomes present in the nuclei, we shall distinguish between organisms that are 'monogenomatisch', 'digenomatisch', 'trigenomatisch' and, in general, 'polygenomatisch', wherein it is initially without consequence, whether the genomes are of the same kind and comprise the same number of chromosomes, or whether they are made up from different chromosome numbers; merely the number of genomes present shall be designated.

I will consider the implications of these concepts in more detail elsewhere. Here I will merely consider briefly to what extent the terms of 'Polygenomatie' and 'Homogenomatie' differ from 'Polyplodie' and 'Homozygotie'. (...)"

Hans Winkler, "Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche", Verlag von Gustav Fischer, Jena, 1920. (pp. 165-166)

I have argued above that the dominant uni-dimensional textural configuration lends itself to facile annexation of apparently valuable sectors. It may be further argued that some of the very strategies and methodologies of translating genomes into text (eg the so-called shotgun sequencing approach) are optimised around capturing the location of valuable regions (Adam Bostanci ……). The other requirement of annexation in relation to formalised intellectual property is the demonstration of utility or instrumental value. Classically utility is demonstrated by association of the DNA sector with a trait, which concomitantly supports the common notion that there are genes for this and genes for that, bypassing the need for any causal explanation . Let me illustrate this by reference to the well documented anthropocentric[2] utility of forecasting expectations of breast cancer. Statistical studies of families and occurrences of breast cancer have supported the association of DNA sequence differences (single nucleotide polymorphisms SNPs see- http://snp.cshl.org/) among humans at a locus denoted as Brca1 with the probability of occurrence of a particular type of breast cancer among carriers of the SNPs. The Brca1 DNA SNPs are accordingly annexed for diagnostic purposes in a patent. Nothing in what is described above portrays an explanation of the causal relationships involved but we do have to recognise a predictive test of instrumental value. Nevertheless it is not uncommon to hear Brca1 described as "the gene for breast cancer" rather than what it really is, simply a predictive marker and one which at best gives a probability estimate rather than defining a discrete category of persons. The broader utility of brca1-based diagnosis is contested by diverse actors. Health care institutions are under pressure to provide the test as part of breast cancer surveillance but have problems with the associated royalties if widespread screening is undertaken, while critics of genetic screening (Nelkin and Tankredi Dangerous Diagnostics) argue that widespread screening will enable an unreasonable categorisation of people as objects of risk and social burden. This form of innovation in health care is thus contested on several fronts for lack of negotiation of its meaning, its cost constraints and its control.

I would argue from this example that the DNA string concept as a basis for modelling or mapping genomes, besides providing ease of annexation also provides worrying shortcuts to instrumentalism though a statistically-based predictive value coupled to obviousness in utility. This construction equates broadly with the concept of genes-P articulated by Lennie Moss (Moss ….) to explain preformationist justifications of apparent purposiveness in genetic systems. In contrast Genes –D where the gene-trait association is based on chains of causal explanation rather than statistical association unlike genes-P have little quantitative predictive value and need more collective work to acquire utility. As such Moss regards genes-D has having value as resource (latent utility).

Playing the game according to the shortcutting norm we run the risk of extending prediction into technologies for which we have scant causal explanation. This can become a serious deficit for other types of utility such as alterations to physiological function (gene therapy, GM-crops (Calvert 2004)) when we look at broader issues of the innovation process particularly those which relate to the practical and social justification for changes we implement. An interesting example in this arena is the use of gene therapy in the treatment of children with severe combined immune deficiency (SCID). Such children are obliged to live in physical isolation for lack of the enzyme adenosine deaminase (ADA). Restoration of the missing enzyme in patients requires the reinsertion of a piece of DNA corresponding to the protein sequence of the missing enzyme via a virus-derived vector. An early treatment of this sort was successful. Later treatments in France similarly alleviated the SCID condition but some patients were found to suffer leukaemia subsequently. The explanation provided for this was that the particular virus vector used in France had inserted into a site within the genome which promoted the onset of leukaemia. I would argue that in this case of gene therapy the instrumental value of gene replacement had not been negotiated beyond the DNA string deficit/restoration concept and that consideration of the genomic space that the DNA string was to be added back to needed to be considered in developing a utility for medical treatment (Hacein-Bey-Abina et al., 2003) (Misteli 2004)…

Another example is herbicide tolerant crops. While the causal relationships between DNA sequences and the manifestation of herbicide tolerance in plants was satisfactorily worked out by the inventors and patent holders of this technology prior to its deployment, the relationships between test-plants field-crops and agronomic practice and environmental outcomes were not. At least if they were the information did not enter agronomics-informed discourse in the UK. I do not recall any discussion for instance of how herbicide tolerance should be deployed within crop rotations or of how engineered tolerance might be deployed logistically to reduce the impact of herbicide use until SCIMAC [3]was set up in 1998 in the wake

[2] See Calvert 2004 for a categorisation of utilities

[3] SCIMAC Supply Chain Initiative for Modified Agricultural Crops see: http://www.agindustries.org.uk/scimac/default.html

of NGO-inspired consumer reaction. Regardless of what we see when we look over the hedge, even at the extremes of monoculture, crops in the field represent a complex genomic space characterised by contributions from a large number of species including root associated fungi and bacteria, nematodes, weeds, insects, birds and mammals. These all play a role in agricultural productivity in the negative, positive long- or short-term senses. Despite recent attempts to re-engage and draw these considerations into the debate (UK farm scale evaluations) the deployment of herbicide tolerant crops is stalled supporting my contention that what's good for patenting and its interpretation of utility is not necessarily good for innovation and established utility.

Short-cuts to instrumentality can also raise issues for the scope of patenting itself in relation to the breadth of claims which should be allowed in the absence of a causal teaching as well as with respect to the authenticity of the inventiveness. In their study of the patenting of genes the Nuffield Council on Bioethics were critical of the inventiveness portrayed in many patents. After all the methods for making associations between loci and proneness to disease are well worked out and obvious as is the utility and method of practice of a diagnostic test once one has establish the association.

In this light I find it slightly ironic to hear senior directors of DNA sequence projects bemoan the wholesale patenting and annexation of tracts of DNA sequence while at the same time promoting the material conceptualisation (I would say mis-conceptualisation) of genomes on which the ownership/control paradigm is built. They even contribute to the language of instrumentality "we have ….. so we can…"

I conclude that while DNA string concepts of the genome are maintained for their facility in annexation of genomic territory and for supporting shortcuts to utility, they have limitations with respect to innovation[4]

Knowledge within the spatial/action conceptualisation presents a diverse route to invention which engages knowledge from many actors operating at different levels within a space or in different spaces. This increases the scope for genuine novelty and specific utility[5]. However, at the same time it increases complexity both of constructing the invention itself and of managing the diverse group of actors who have to collaborate and share their knowledge. There is experience from the software development sector which suggests that freed from the imperatives of proprietary practices such as competitor exclusion, hierarchical control, and market-driven short-termism, groups of actors are likely to self-organise and collaborate in complex spaces (Benkler 2002). Inclusion within such groups is dependent solely on the possession of appropriate knowledge (both tacit and specific) and skills and is effectively self regulating. Whilst not "democratic" in the conventional sense, there is scope within such a principle if extrapolated to the genomics world to construct broadly accessible technologies coupled to more broadly negotiated utilities. Let us illustrate this with an example from plant genomics and the application of molecular markers in plant breeding. The scenario is broadly similar to gene-based diagnostics in human disease except that in this case the predictive markers representing DNA polymorphisms (RFLP, RAPD, AFLP, microsatellite, some of which have been discussed by others in this seminar,) are associated (linked by proximity on the DNA string) with loci representing traits of positive utility in agriculture. Awareness of genetic traits of potential value is invested at several levels in the community. It is invested with farmers who are aware of the variation within the crop varieties (especially the so-called land races) they grow. It is invested in those who collect and maintain germ-plasm. It is invested with crop and plant physiologists and also with those in institutions who evaluate new varieties for their distinctness uniformity and value for cultivation and use. It is invested also in plant breeders and their technical affiliates though they are dependent on knowledge fed to them by the other knowledge holders. Associations between DNA polymorphisms and traits is established is based on the relative frequencies of co-segregation of marker and trait among progeny of a cross between parents which differ with respect to the trait. This of course can be confounded by environmental variation and requires close physiological surveillance. Even then, only rarely does a trait co-segregate with a single marker locus and so what we tend to find are a number

[4] It is important to recognise that "an" innovation refers to a new thing or practice which has been adopted by society and that innovation (verb) is the process of getting to that point. Innovation is often mistakenly used to refer to the process of producing inventions which is but a small part of innovation.

[5] Sharing of data, constructivist approach combining modelling with diverse data sets (some are calling this "systems biology" but it is really just biology)

of loci each of which singly or interactively with other loci contribute to the variation observed in the trait These are termed quantitative trait loci each of which through allelic differences contribute something to the variation in the given trait. Only rarely are their contributions simply additive reflecting the complex and interactive nature of genomic space. QTL are thus defined in two ways, one by the nature of their activity (their contribution to a particular trait, and two by their position in the DNA component of the genome relative to a linked DNA sequence polymorphism which is itself entirely neutral relative to the trait. Thus there is not a thing or "gene for x" which can be annexed
This form of analysis is significant to our argument as an example: where there is a multiplicity of actors contributing skills and knowledge, where utility is broadly negotiated upstream of genomic discovery, where there is recognition of the genome as complex and interactive space, there is no sequence-based definition of a thing (DNA string gene) which can be annexed or from which others can be excluded. Instrumentality lies in the use of the linked polymorphic DNA as markers to track the linked trait (QTL) in breeding activities (the screening of progenies, identification of valuable germ-plasm) and the choices made for the trajectories of that activity. Relevant to the social negotiation of this utility, the product of marker-assisted breeding is indistinguishable in terms of genetic intervention from that which has been in the long term discourse on the value of crop improvement through breeding. Some argued against this utility in the context of the high yielding varieties of cereal crops in the Green Revolution on the basis of loss of the crop diversity previously managed by the recipient farmers

The principle of exclusion articulated above, expressed in the dominant IP practices of the commercial landscape as artificially elevated technology entry price, when combined with the shear complexity of the fragmented and impenetrable patent field of genomic knowledge, can and does impede innovation. The term anti-commons was coined by Heller and Eisenberg (1998) to illustrate this form of impediment which like its more widely recognised partner the Tragedy of the Commons is particularly poignant for resource-poor communities and the potential for decentralised partnerships in innovation. The sharing of knowledge of technological enablement and of local scenarios for deployment is vital to the latter endeavours and a number of initiatives have been put forward to reverse the deficit. These consist either of proposed special derogations to the exclusionary principle (illustrated by the study case of golden rice), technology bundling or brokering , or else patent informatics provisions which at least explain to the resource-poor why decentralised deployment is not an option.
I promote what I see as a more promising approach which may reside in verisimilitudes of the peer production collaborative working practices which have emerged from the open source software movement (Benkler 2002). These may serve to draw inventors from the clutches of futile technology transfer departments and back into a more satisfying and creative space of peer production networks or the commons. They may also convince commercial bioscience practitioners that, for them just as for the software producers, there is more to be gained through open and distributed deployment than through the traditional practices of competitor exclusion particularly when innovation depends more on fitting technology to consumer markets than on force fixing markets to high technology products.
This approach has much in common with the collaborative approaches required for navigating genomics spaces if we can once escape in our utilitarian ambitions from the dominant string conceptualisation.

References

http://bios.net
Y. Benkler, Coase's Penguin, or Linux and the Nature of the Firm , Yale L.J. 112 winter 2002-2003

Bostanci,A. "Two Drafts, One genome? Textual-Territorial mataphores in human genome research in press science in Culture (2005)

Calvert, J (2004) 'Genomic patenting and the utility requirement' New Genetics and Society Vol.23, No.3, pp.301-312

Cleveland, D.A., D. Soleri, and S.E. Smith. 1999. *Farmer Plant Breeding from a Biological Perspective: Implications for Collaborative Plant Breedin*g. CIMMYT Economics Working Paper No.10. Mexico, D.F.:

CIMMYT.
http://www.cimmyt.cgiar.org/Research/economics/map/research_results/working_papers/pdf/EWP%2099_10.pdf

Haraway D.J., modest Witness@Second_Millenium_FemaleMan_Meets_Ocncomouse, Routledge,New york ISBN0-415-91245-8, 1997

S. Hacein-Bey-Abina et al., "LMO2-associated clonal T-cell proliferation in two patients after gene therapy for SCID-X1," *Science,* 302:415-419, October 17, 2003.

Heller, M.,and Eisenberg, R. Can Patents Deter Innovation, The Anticommons in Biomedical Research. Science, 280 698-701 (1998)

L.Moss, From Representational Preformationism to the Epigenesis of Openness to the world, Ann. N.Y. Acad. Sci. 981, 219-230 (2002)

Mistel, T. the genome in 3-dimensions: from basics to applications, bioforum Europe 6/2004 50-51 (2004)

Nelkin, D. ; Tancredi, L., Dangerous Diagnostics: The Social Power of Biological Information. Harper Collins, New York: Basic Books; ISBN 0-226-57129-7 1989.

PHILLIPS, R. L. and VASIL, I. K. (eds). DNA Markers in Plants Kluwer Acad. P Dordrecht the Netherlands

Richards, P. 1996. Culture and community values in the selection and maintenance of African rice. In *Valuing Local Knowledg*e, pp. 209-229 (Eds. S.B. Brush and D. Stabinsky). Washington, DC: Island Press.

R.W.Witkin, chapter 3:Perceptual Sytsems and Presentational Codes, and chapter 5: Burgois Society and Realist Styles of Art , in Art and Social Structure, Polity Press, Cambridge UK, ISBN 0-7456-1134-6, 1995

O. Wolkenhauer, W. Kolch, and **K.-H. Cho**, "Dynamic Pathway Modelling: From Models to Cells", *Proc. of 3rd Workshop on Computation of Biochemical Pathways and Genetic Networks*, Heidelberg, Germany, pp. 79-89, 2003.