6

Objective Bayesian Nets for Systems Modelling and Prognosis in Breast Cancer

Sylvia Nagl[1], Matt Williams[2], and Jon Williamson[3]

[1] Department of Oncology, University College London
s.nagl@medsch.ucl.ac.uk
[2] Advanced Computation Laboratory, Cancer Research UK, and Computer Science, University College London
m.williams@cs.ucl.ac.uk
[3] Department of Philosophy, University of Kent
j.williamson@kent.ac.uk

Summary. Cancer treatment decisions should be based on all available evidence. But this evidence is complex and varied: it includes not only the patient's symptoms and expert knowledge of the relevant causal processes, but also clinical databases relating to past patients, databases of observations made at the molecular level, and evidence encapsulated in scientific papers and medical informatics systems. Objective Bayesian nets offer a principled path to knowledge integration, and we show in this chapter how they can be applied to integrate various kinds of evidence in the cancer domain. This is important from the systems biology perspective, which needs to integrate data that concern different levels of analysis, and is also important from the point of view of medical informatics.

In this chapter we propose the use of objective Bayesian nets for knowledge integration in the context of cancer systems biology. In Part I we discuss this context in some detail. Part II introduces the machinery that is to be applied, objective Bayesian nets. Then a proof-of-principle application is presented in Part III. Finally, in Part IV, we discuss avenues for further research.

Part I: Cancer Systems Biology and Knowledge Integration

6.1 Cancer Systems Biology

Cancer systems biology seeks to elucidate complex cell and tumour behaviour through the integration of many different types of knowledge. Information is obtained from scientific and clinical measurements made across biological scale, ranging from molecular components to systems, and from the genome to the whole patient. Integration of this information into predictive computational models, and their use in research and clinical settings, is expected to improve prevention, diagnostic and prognostic prediction, and treatment of cancer.

D.E. Holmes and L.C. Jain (Eds.): Innovations in Bayesian Networks, SCI 156, pp. 131–167, 2008.
springerlink.com

Systems biology addresses the complexity of cancer by drawing on a conceptual framework based on the current understanding of complex adaptive systems.[1] Complex systems are composed of a huge number of components that can interact simultaneously in a sufficiently rich number of parallel ways so that the system shows spontaneous self-organisation and produces global, emergent structures and behaviours.[2] Self-organisation concerns the emergence of higher-level order from the local interactions of system components in the absence of external forces or a pre-programmed plan embedded in any individual component.[3]

The challenges posed by the complex-systems properties of cancer are severalfold and can be thought about in terms of a taxonomy of complexity put forward by Mitchell:[4]

- Structural complexity;
- Dynamic complexity—complexity in functional processes;
- Evolved complexity—complex systems can generate alternative evolutionary solutions to adaptive problems; these are historically contingent.

Decisions need to be made in the face of great uncertainty regarding all three aspects of the complexity that is exhibited by the cancer systems in which one seeks to intervene.[5] This is true both for therapeutic decisions for individual patients and also for design strategies leading to new anti-cancer therapies. Although our ability to collect ever more detailed quantitative molecular data on cells and cell populations in tumours is growing exponentially, and clinical investigations are becoming more and more sophisticated, our understanding of system complexity advances more slowly for the following reasons.

It is very difficult to directly observe and measure dynamic processes in complex systems, and this is particularly challenging in biomedicine where human subjects are involved. Research relies on data generated from tissue samples by high throughput technologies mainly directed at the 'omic' levels of the gen*ome*, transcript*ome* (gene transcripts) and prote*ome* (proteins).[6] However, data sampling is highly uneven and incomplete, and the data themselves are noisy and often hard to replicate. The molecular data that are being gathered typically only illuminate 'single-plane' omic slices 'dissected' out of entire systems which are characterized by highly integrated multi-scale organization and non-linear behaviour (Fig. 6.1). Furthermore, due to technological and economic constraints, current techniques can only capture a few time points out of the continuous systems dynamics, and are not yet able to address the 'complexity explosion' of control at the proteomic level. This situation is likely to persist for some time to come.

[1] (Nagl, 2006.)
[2] (Holland, 1995; Depew and Weber, 1996.)
[3] (Holland, 1995, 1998; Mitchell, 2003.)
[4] (Mitchell, 2003, pp. 4–7.)
[5] (Nagl, 2005).
[6] (Abramovitz and Leyland-Jones, 2006).

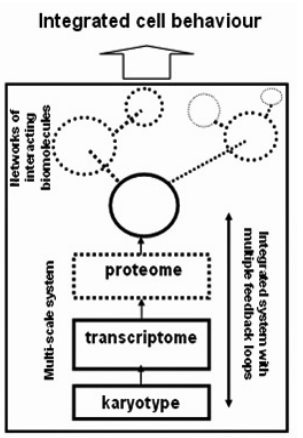

Fig. 6.1. Multi-scale integrated cell systems

A further challenge is posed by the need to relate molecular data to clinical measurements, notably those obtained in clinical trials, for identification of molecular parameters underlying physiological (dys)function. This integration task spans spatial and temporal scales of several orders of magnitude and complexity. Ultimately, cancer systems analysis needs to cut across all biological levels—genome, transcriptome, proteome, cell, and beyond to tissue, organ and patient (and the environment, but this is outside the present discussion). Toyoda and Wada (2004) have coined the term *omic space* and presented a hierarchical conceptual model linking different omic planes. They showed that this structuring of omic space helps to integrate biological findings and data comprehensively into hypotheses or models combining higher-order phenomena and lower-order mechanisms through a comprehensive ranking of correspondences among interactions in omic space. The key idea behind the concept of omic space may also serve as a general organising principle for *multi-scale systems*, and may be extended beyond cells to the tissue, organ and patient level. Below, we discuss how objective Bayesian nets can make significant contributions to the elucidation of multi-scale relationships in cancer systems (see §§6.4, 6.20).

6.2 Unstable Genomes and Complexity in Cancer

Tumours maintain their survival and proliferative potential against a wide range of anticancer therapies and immunological responses of the patient. Their robustness is seen as an emergent property arising through the interplay of genomic instability and selective pressure driven by host-tumour dynamics.[7]

[7] (Kitano, 2004).

Progression from normal tissue to malignancy is associated with the evolution of neoplastic cell lineages with multiple genomic lesions (abnormal karyotypes).[8] Most cancer cells do not have a significantly higher mutation rate at the nucleotide level compared to their normal counterparts, whereas extensive gross chromosomal changes are observed in liquid, and nearly all, solid tumours. The most common mutation class among the known cancer genes is chromosomal.

In cancer, dynamic large-scale changes of genome structure occur at dramatically increased frequency and tumour cell–microenvironment interactions drive selection of abnormal karyotypes. Copy-number changes, such as gene amplification and deletion, can affect several megabases of DNA and include many genes. These extensive changes in genome content can be advantageous to the cancer cell by simultaneous activation of oncogenes and elimination of tumour suppressors. Due to the magnitude of the observed genomic rearrangements, it is not always clear which gene, or set of genes, is the crucial target of the rearrangement on the basis of genetic evidence alone.

These changes encompass both directly cancer-causing and epiphenomenal changes (bystander mutations) which can nevertheless contribute significantly to the malignant phenotype and can modulate treatment resistance in a complex fashion. The ability of abnormal karyotypes to change autocatalytically in response to challenge, the masking of specific cancer-promoting constellations by collateral variation (any chromosomal combination that is specific for a selected function is also specific for many unselected functions), and the common phenomenon of several alternative cell pathways able to generate the same phenotype, limits the usefulness of context-independent predictions from karyotypic data. Interestingly, several researchers have put forward the theory that the control of phenotype is distributed to various extents among all the genetic components of a complex system.[9]

Mathematical modelling, adapted from Metabolic Control Analysis, suggests that it may in fact be the large *fraction* of the genome undergoing differential expression as a result of changes in gene dose (due to chromosomal rearrangements) that leads to highly non-linear changes in the physiology of cancer cells.

6.3 A Systems View of Cancer Genomes and Bayesian Networks

Genomes are dynamic molecular systems, and selection acts on cancer karyotypes as integrated wholes, not just on individual oncogenes or tumour suppressors. Given the irreversible nature of evolutionary processes, the randomness of mutations and rearrangements relative to those processes, and the modularity and redundancy of complex systems, there potentially exists a multitude of ways to 'solve' the problems of achieving a survival advantage in cancer cells.[10] Since each patient's cancer cells evolve through an independent set of genomic lesions and

[8] (Nygren and Larsson, 2003; Vogelstein and Kinzler, 2004).
[9] (Rasnick and Duesberg, 1999, and references therein).
[10] (Mitchell, 2003, p. 7).

selective environments, the resulting heterogeneity of cell populations within the same tumour, and of tumours from different patients, is a fundamental reason for differences in survival and treatment response.

Since the discovery of oncogenes and tumour suppressors, a reductionist focus on single, or a small number of, mutations has resulted in cancer being conceptualized as a 'genetic' disease. More recently, cancer has been recast as a 'genomic' or 'systems' disease.[11] In the work presented in this chapter, we apply a systems framework to karyotype evolution and employ Bayesian networks to generate models of non-independent rearrangements at chromosomal locations from comparative genome hybridisation (CGH) data.[12] Furthermore, we present a method for integration of genomic Bayesian network models with nets learnt from clinical data. The method enables the construction of multi-scale nets from Bayesian nets learnt from independent datasets, with each of the nets representing the joint probability distributions of parameter values obtained from different levels of the biological hierarchy, i.e., the genomic and tumour level in the application presented here (together with treatment and outcome data). Bayesian network integration allows one to capture 'more of the physiological system' and to study dependency relationships across scales.[13]

Some of the questions one may address by application of our approach include

- utilising genomic (karyotype) data from patients:
  - Can we identify probabilistic dependency networks in large sample sets of karyotypes from individual tumours? If so, under which conditions may these be interpreted as causal networks?
  - Can we discover key features of the 'evolutionary logic' embodied in gene copy number changes of individual karyotypes?
  - Can we characterise the evolutionary 'solution space' explored by unstable cancer genomes? Is there a discernible dependence on cancer types?
- utilising omic and other molecular data together with clinical measurements:
  - Can we identify probabilistic dependency networks involving molecular and clinical levels?
  - How may such probabilistic dependencies aid diagnostic and prognostic prediction and design of personalised therapies?

6.4 Objective Bayesianism and Knowledge Integration

Bayesian networks are well suited to problems that require integration of data from various sources or data with different temporal or spatial resolutions. They can model complex non-linear relationships, and are also very robust to missing information. Bayesian network learning has already been successfully applied to data gathered at the transcriptomic and proteomic level for predictions regarding structure and function of gene regulatory, metabolic and signalling

[11] (Khalil and Hill, 2005; Lupski and Stankiewicz, 2006).
[12] (Reis-Filho et al., 2005).
[13] (Nagl et al., 2006).

networks.[14] Bulashevska and colleagues have applied Bayesian network analysis to allelotyping data in urothelial cancer.[15] However, these studies also demonstrate persistent limits—only very partial answers have so far been obtained concerning the organization and dynamic function of whole biological systems which are by definition multi-scale and integrated by multiple feedback loops (see §6.1).

Employing objective Bayesianism as our methodology, we present a multi-scale approach to knowledge integration which utilises more fully the very considerable scope of available data. Our method enables integration of 'omic' data types and quantitative physiological and clinical measurements. These data combined offer rich, and as yet largely unexplored, opportunities for the discovery of probabilistic dependencies involving system features situated at multiple levels of biological organisation.

The technique supports progressive integration of Bayesian networks learnt from independently conducted studies and diverse data types, e.g., mRNA or proteomic expression, SNP, epigenetic, tissue microarray, and clinical data. New knowledge and new data types can be integrated as they become available over time. The application of our knowledge discovery method is envisaged to be valuable in the clinical trials arena which is undergoing far-reaching changes with steadily increasing incorporation of molecular profiling. It is our aim to assess the potential of our technique for integrating different types of clinical trial datasets (with and without molecular data). The methods described here are highly complementary to ongoing research initiatives, such as the Cancergrid project (www.cancergrid.org) and caBIG (cabig.nci.nih.gov) which are already addressing pressing informatics requirements that result from these changes in clinical study design.

6.5 Complementary Data Integration Initiatives

We are currently not in a position to make maximal use of existing data sets for Bayesian network analysis, since data have not yet been standardised in terms of experimental and clinical data capture (protocols, annotation, data reproducibility and quality), and computational data management (data formats, vocabularies, ontologies, metadata, exchange standards). Basic requirements are the generation of validated high-quality datasets and the existence of the various data sources in a form that is suitable for computational analysis and data integration. This has been well recognised, as is amply demonstrated by the aims and activities of a collaborative network of several large initiatives for data integration within the cancer domain which work towards shared aims in a coordinated fashion (the initiatives mentioned below are meant to serve as example projects and do not represent the sum total of these efforts on an international scale).

The National Cancer Institute Center for Bioinformatics (NCICB) in the United States has developed caCORE which provides an open-source suite of

[14] (Xia et al., 2004).
[15] (Bulashevska et al., 2004).

common resources for cancer vocabulary, metadata and data management needs (biological and clinical), and, from Version 3.0, achieves semantic interoperability across disparate biomedical information systems (for detailed information and access to the caCORE components, see ncicb.nci.nih.gov/ NCICB/infrastructure/cacore_overview). caCORE plays an essential integrative role for the cancer Biomedical Informatics Grid (caBIG), a voluntary network connecting individuals and institutions to enable the sharing of data and tools, creating a 'World Wide Web of cancer research' whose goal is to speed up the delivery of innovative approaches for the prevention and treatment of cancer (cabig.nci.nih.gov/).

In the United Kingdom, the National Cancer Research Institute (NCRI) is developing the NCRI Strategic Framework for the Development of Cancer Research Informatics in the UK (www.cancerinformatics.org.uk). The ultimate aim is the creation of an internationally compatible informatics platform that would facilitate data access and analysis. CancerGRID develops open standards and information management systems (XML, ontologies and data objects, web services, GRID technology) for clinical cancer informatics, clinical trials, integration of molecular profiles with clinical data, and effective translation of clinical trials data to bioinformatics and genomics research (www.cancergrid.org).

Part II: Objective Bayesian Nets

6.6 Integrating Evidence Via Belief

In this information-rich age we are bombarded with evidence from a multiplicity of sources. This is evidence in a defeasible sense: items of evidence may not be true—indeed different items of evidence often contradict each other—but we take such evidence on trust until we learn that it is flawed or until something better comes along. In the case of breast cancer prognosis we have databases of molecular and clinical observations of varying reliability, current causal knowledge about the domain, knowledge encapsulated in medical informatics systems (e.g. argumentation systems, medical ontologies), and knowledge about the patient's symptoms, treatment, and medical history. The key question is how we represent this eclectic body of evidence and render it coherent.

Knowledge impinges on belief, and one way in which we try to make sense of conflicting evidence is by finding a coherent set of beliefs that best fits this knowledge. We try to find beliefs that are consistent with undefeated items of evidence where we can, and where two items conflict we try to find some compromise beliefs. But this is vaguely put, and in this Part we shall describe a way of making this idea more precise.

Objective Bayesianism offers a formalism for determining the beliefs that best fit evidence; §6.7 offers a brief introduction to this theory. While this provides a useful theoretical framework, further machinery is required in order to find these beliefs and reason with them in practice—this is the machinery of objective Bayesian nets outlined in §6.8. In §6.9 we sketch a general procedure for constructing these nets, then in §6.10 we see how objective Bayesian nets can be used to

integrate qualitative evidence with quantitative evidence. Finally §6.11 discusses objective Bayesian nets in the context of the problem at hand, breast cancer.[16]

6.7 Objective Bayesianism

According to Bayesian theory, an agent's degrees of belief should behave like probabilities. Thus you should believe that a particular patient's cancer will recur to some degree representable by a real number x between 0 and 1 inclusive; you should believe that the patient's cancer will not recur to degree $1 - x$. Many Bayesians go further by adopting empirical constraints on degrees of belief. Arguably, for instance, degrees of belief should be calibrated with known frequencies: if you know just that 40% of similar patients have cancers that recur then you should believe that this patient's cancer will recur to degree 0.4. Objective Bayesians go further still, accepting not only empirical constraints on degrees of belief but also logical constraints: in the absence of empirical evidence concerning cancer recurrence you should equivocate on the question of this patient's cancer recurring—i.e. you should believe the cancer will recur to the same degree that you should belief it will not recur, 0.5.[17]

From a formal point of view the objective Bayesian position can be summed up as follows.[18] Applying Bayesian theory, the agent's degrees of belief should be representable by a probability function p. Suppose that the agent has empirical evidence that takes the form of a set of quantitative constraints on p. Then she should adopt the probability function p, from all those that satisfy these constraints, that is maximally equivocal, i.e. that maximises entropy $H = -\sum_v p(v) \log p(v)$, where the sum is taken over all assignments $v = v_1 \cdots v_n$ to the variables $V_1, \ldots, V_n$ in the domain. This is known as the *maximum entropy principle*.[19]

Note that two items of empirical evidence may conflict—for example, the agent might be told that the frequency of recurrence is 0.4, but might also be told on another occasion that the frequency of recurrence is 0.3, with neither of the two reports known to be more reliable than the other and neither more pertinent to the patient in question. Arguably, the agent's degree of belief that the patient's cancer will recur will be constrained to lie within the closed interval $[0.3, 0.4]$. More generally, empirical constraints will constrain an agent's belief function to lie within a *closed convex* set of probability functions, and consequently there will be a unique function p that maximises entropy.[20] Thus the agent's rational belief function p is objectively determined by her evidence (hence the name *objective Bayesianism*).[21]

[16] See Williamson (2002); Williamson (2005a, §5.5–5.8) and Williamson (2005b) for more detailed descriptions of the theory behind objective Bayesian nets.
[17] (Russo and Williamson, 2007).
[18] (Williamson, 2005a, Chapter 5).
[19] (Jaynes, 1957).
[20] (Williamson, 2005a, §5.3).
[21] See Williamson (2007b) for a general justification of objective Bayesianism, and Russo and Williamson (2007) for a justification within the cancer context.

We see then that objective Bayesianism provides a way of integrating evidence. The maximum entropy probability function p commits to the extent warranted by evidence: it satisfies constraints imposed by evidence but is non-committal where there is insufficient evidence. In that respect, objective Bayesian degrees of belief can be thought of as representative of evidence.

6.8 Obnets

Finding a maximum entropy probability function by searching for the parameters $p(v)$ that maximise the entropy equation is a computationally complex process and impractical for most real applications. This is because for a domain of n two-valued variables there are 2^n parameters $p(v)$ to calculate; as n increases the calculation gets out of hand. But more efficient methods are available. Bayesian nets, in particular, can be used to reduce the complexity of representing a probability function and drawing inferences from it.

A *Bayesian net* is a graphical representation of a probability function. The variables in the domain form the nodes of the graph. The graph also contains arrows between nodes, but must contain no cycles. Moreover, to each node is attached a probability table, containing the probability distribution of that variable conditional on its parents in the graph. As long as the *Markov condition* holds—i.e. each variable is probabilistically independent of its non-descendants in the graph conditional on its parents, written $V_i \perp\!\!\!\perp ND_i \mid Par_i$—the net suffices to determine a probability function over the whole domain, via the identity

$$p(v) \stackrel{\text{df}}{=} p(v_1 \cdots v_n) = \prod_{i=1}^{n} p(v_i|par_i).$$

Thus the probability of an assignment to all the variables in the domain is the product of the probabilities of the variables conditional on their parents. These latter probabilities can be found in the probability tables. Depending on the sparsity of the graph, a Bayesian net can offer a much smaller representation of a probability function than that obtained by listing all the 2^n probability values $p(v)$. Furthermore, the Bayesian net can be used to efficiently draw inferences from the probability function and there is a wide variety of software available for handling these nets.[22] The other chapters in this volume are testament to the importance of Bayesian nets for probabilistic reasoning.

An *objective Bayesian net*, or *obnet*, is a Bayesian net that represents objective Bayesian degrees of belief, i.e. that represents an agent's entropy-maximising probability function. Because the objective Bayesian belief function is determined in a special way (via the maximum entropy principle) there are special methods for constructing an objective Bayesian net, detailed in §6.9. These methods are more efficient to carry out than the more direct maximisation of the parameters $p(v)$ in the entropy equation.

[22] (Neapolitan, 1990; Korb and Nicholson, 2003).

Given an objective Bayesian net, standard Bayesian net algorithms can be used to calculate probabilities, e.g. the probability of cancer recurrence given the characteristics of a particular patient. Thus an obnet can help with the task in hand, breast cancer prognosis. But an obnet can address other tasks too, for example the problem of knowledge discovery. An objective Bayesian net can suggest new relationships between variables: for instance if two variables are found to be strongly dependent in the obnet but there is no known connection between the variables that accounts for this dependence, then one might posit a causal connection to explain that link (§§6.16, 6.20). An obnet can also be used to determine new arguments to add to an argumentation framework: if one variable significantly raises the probability of another then the former is an argument for the latter (§6.18). Thus an obnet is a versatile beast that can assist with a range of tasks.

6.9 Constructing Obnets

One can build an objective Bayesian net by following a 3-step procedure. Given evidence, first determine conditional independencies that the entropy maximising probability function will satisfy. With this information about independencies one can then construct a directed acyclic graph for which the Markov condition holds. Finally, add the probability tables by finding the probability parameter $p(v_i|par_i)$ that maximise entropy.

The first step—finding the conditional independencies that p must satisfy—can be performed as follows. As before, we suppose that background knowledge imposes a set of quantitative constraints on p. Build an undirected *constraint graph* by taking variables as nodes and linking two variables if they occur together in some constraint. We can then read off probabilistic independencies from this graph: for sets of variables X, Y, Z, if Z separates X from Y in the constraint graph then X and Y will be probabilistically independent conditional on Z, $X \perp\!\!\!\perp Y \mid Z$, for the entropy maximising probability function p (Williamson, 2005a, Theorem 5.1).

The second step—determining the directed acyclic graph to go in the objective Bayesian net—is equally straightforward. One can transform the constraint graph into a directed acyclic graph G that satisfies the Markov Condition via the following algorithm:[23]

- triangulate the constraint graph,
- re-order V according to maximum cardinality search,
- let $D_1, \ldots, D_l$ be the cliques of the triangulated constraint graph ordered according to highest labelled node,
- set $E_j = D_j \cap (\bigcup_{i=1}^{j-1} D_i)$ for $j = 1, \ldots, l$,
- set $F_j = D_j \backslash E_j$ for $j = 1, \ldots, l$,
- take variables in V as the nodes of G,

[23] See Williamson (2005a, §5.7) for an explanation of the graph-theoretic terminology.

- add an arrow from each vertex in E_j to each vertex in F_j $(j = 1, \ldots, l)$,
- ensure that there is an arrow between each pair of vertices in D_j $(j = 1, \ldots, l)$.

The final step—determining the probability tables to go in the objective Bayesian net—requires some number crunching. One needs to find the parameters $p(v_i|par_i)$ that maximise the entropy equation, which can be written as $H = \sum_{i=1}^{n} H_i$ where $H_i = -\sum_{v_1 \cdots v_n} \left(\prod_{V_j \in Anc_i} p(v_j|par_j) \right) \log p(v_i|par_i)$, ($Anc_i$ being the set of ancestors of V_i in G). This optimisation task can be carried out in a number of ways. For instance, one can use numerical techniques or Lagrange multiplier methods to find the parameters.

This gives the general method for constructing an obnet. In §6.11 we shall tailor this method to our particular problem domain, that of breast cancer. But first we shall see how the method can be extended to handle qualitative evidence.

6.10 Qualitative Evidence

In the breast cancer domain, as elsewhere, evidence can take qualitative form. As well as quantitative evidence gleaned from clinical and molecular databases, there is qualitative causal knowledge and also qualitative evidence gleaned from medical ontologies and argumentation systems. In order to apply the maximum entropy principle in this type of domain, qualitative evidence must first be converted into a set of quantitative constraints on degrees of belief. Here we shall describe how this is possible.

Consider the following qualitative relationships: A is a cause of B; A is a sub-type of B; A is an argument in favour of B. These causal, ontological and evidential relations are all examples of what might be called *influence relations*. Intuitively A influences B if bringing about A brings about B but bringing about B does not bring about A. More precisely, a relation is an *influence relation* if it satisfies the following property: learning of the existence of new variables that are not influences of the other variables should not change degrees of belief concerning those other variables.[24]

Qualitative knowledge of influence relationships can be converted into quantitative constraints on degrees of belief as follows. Suppose $V \supseteq U$ is a set of variables containing variables in U together with other variables that are known not to be influences of variables in U. As long as any other knowledge concerning variables in $V \backslash U$ does not itself warrant a change in degrees of belief on U, then $p^V_{\beta \downarrow U} = p^U_{\beta_U}$, i.e., one's belief function on the whole domain V formed on the basis of all one's background knowledge β, when restricted to U, should match the belief function one would have adopted on domain U given just the part β_U of one's knowledge involving U. These equality constraints can be used to constrain degrees of belief so that the maximum entropy principle can be applied. The equality constraints can also be fed into the procedure for constructing objective Bayesian nets: build the constraint graph as before from the

[24] (Williamson, 2005a, §11.4).

non-qualitative constraints; transform that graph into a directed acyclic graph as before; but take the qualitative constraints, converted to quantitative equality constraints, into account when determining the probability tables of the obnet (Williamson, 2005a, Theorem 5.6).

6.11 Obnets and Cancer

In the context of our project we have several sources of general (i.e., not patient-specific) evidence: databases of clinical data; databases of molecular data; a medical ontology; arguments from an argumentation framework; evidence of causal relationships from experts and also from published clinical trials. The study discussed in Part III focusses on databases of clinical and molecular data, but in this section we shall show how all these varied evidence sources might be integrated.

These sources impose a variety of constraints on a rational belief function p. Let C be the set of variables measured in a database of clinical data. Then $p_{\downarrow C} = freq_C$, the rational probability function when restricted to the variables in the clinical dataset should match the frequency distribution induced by that dataset. Similarly, if M is the set of variables measured in a molecular dataset, then $p_{\downarrow M} = freq_M$. A medical ontology determines influence relationships amongst variables. For example, knowledge that assignment a is a type (or sub-classification) of assignment b imposes the constraint $p(b|a) = 1$, as well as the influence constraint (§6.10) $p_{\downarrow A}^{\{A,B\}} = p^{\{A\}}$. An argumentation framework also determines influence relationships. An argument from a to b indicates that a and b are probabilistically dependent. This yields the constraint $p(b|a) \geq p(b) + \tau$ where τ is some threshold (which measures the minimum strength of arguments within the argumentation framework), as well as the influence constraint $p_{\downarrow A}^{\{A,B\}} = p^{\{A\}}$. Finally, causal evidence yields influence constraints of the form $p_{\downarrow A}^{\{A,B\}} = p^{\{A\}}$, and, if gleaned from a clinical trial, quantitative constraints of the form $p(b|a) \geq p(b) + \tau$.

In order to construct an objective Bayesian net from these sources, we can follow the three-step procedure outlined in §6.9.

The first task is to construct an undirected constraint graph. Following the recipe, we link all the variables in C (the clinical variables), link all the variables in M (the molecular / genomic variables), and link pairs of variables that are connected by an argument or by a clause in the ontology or by a causal relation. But we can reduce the complexity of the resulting obnet, which is roughly proportional to the density of the graph, still further as follows. One can use standard algorithms to induce a Bayesian net that represents the frequency distribution of the clinical database. Similarly, one can induce a Bayesian net from the clinical database. Then one can incorporate the independencies of these nets in the constraint graph, to render the constraint graph more sparse. This can be done as follows. Rather than linking every pair of variables in C with an edge in the constraint graph, include a link only if one is the parent of the other in frequency net, or if the two have a child in common in that net. Similarly for the variables in M. This yields a constraint graph with fewer edges, and thus a smaller obnet as a result.

The next two steps—converting the constraint graph into a directed acyclic graph, and adding the probability tables—can be carried out as detailed in §6.9.

Note that there is a particularly simple special case. If the evidence consists only of two databases which have just one variable in common, then one can construct the directed acyclic graph of the obnet thus: for each database learn a frequency net, ensuring that the variable in common is a root variable (i.e. has no parents); then just join the frequency nets at the root variable.

Having discussed the theoretical aspects of objective Bayesian nets, we now turn to a detailed description of the breast cancer application.

Part III: The Application

6.12 Obnets and Prediction in the Cancer Domain

We have applied objective Bayesian nets to the domain of breast cancer using three sources of data: one clinical, and two genomic, as well as a published study. The use of two genomic data sets was necessary as the more substantial genomic Bayesian net did not have a node in common with the clinical network, and so we used a smaller network to link the larger genomic network and the clinical one. We start by reviewing the data used (§§6.13, 6.14), and then in §6.15 describe how we constructed and merged the three separate networks. We then present some initial data on the performance of the network, and conclude the Part with a discussion of the uses of such networks in §6.16.

6.13 Breast Cancer

Breast Cancer is one of the commonest cancers in the Western World. It is the commonest non-skin cancer in women in the UK and US, and accounts for approximately a third of cancers in women, with lifetime rates of 1 in 10. Some 36000 cases are diagnosed each year in the UK, of whom about a third will die from the disease.[25] Consequently there has been a considerable amount of research focused on breast cancer, and death rates have fallen over the last 10 years.[26]

The mainstay of treatment for breast cancer remains surgery and radiotherapy,[27] with hormonal and chemotherapeutic agents often used to treat presumed micro-metastatic disease. One of the advantages of surgery is that, as well as removing any local disease, a sample can also be taken of the axillary lymph nodes. These are a common site of metastatic spread for the cancer, and their removal not only removes any spread that may have occurred, but also allows analysis of the nodes to describe the degree of spread. The two main aims of treatment are to provide local control of, and to prevent premature death from, disease.

[25] (McPherson et al., 2000).
[26] (Quinn and Allen, 1995).
[27] (Richards et al., 1994).

Examination of the primary tumour and lymph nodes lets us define certain characteristics of the disease that make local recurrence and death more likely. These characteristics are primarily the grade of the tumour, (which represents the degree of abnormality displayed by the cells, scored 1-3), the size of the tumour (as its maximum diameter, in mm) and the number of involved nodes.[28] There are also newer tests for the presence or absence of certain proteins on the cell surface that may predict tumour behaviour or response to certain drugs.[29]

The central aim of therapy planning is to match treatment with the risk of further disease. Thus those at high risk should be treated aggressively while those at low risk should be treated less aggressively. This allows more efficient use of resources, and restricts the (often considerable) side effects of intensive treatment to those patients who would benefit most.

Current Prognostic Techniques

These prognostic characteristics are currently modelled using statistical techniques to provide an estimate of the probability of survival and local recurrence. Two commonly used systems are the Nottingham Prognostic Index (NPI),[30] which uses data from large UK studies, and results derived from the American Surveillance, Epidemiology and End Results (SEER) database,[31] which are used by systems such as Adjuvant Online.[32] Both techniques rely on multivariate analyses of large volumes of data (based on over 3 million people for SEER) to calculate prognostic formulae.

These tools, and others like them, are effective at providing estimates of risk of death and local recurrence. However, they have two major weaknesses. Whilst effective, they lack explanatory power in a human-readable form. Therefore, extra knowledge that has not been captured by the statistical analysis (such as the presence and impact of other co-existing conditions) cannot be easily incorporated. Secondly, knowledge that post-dates the formation of the formulae (such as the discovery of Her-2neu, a cell-surface protein that is a marker for more aggressive disease) is very difficult to incorporate. Therefore, while they excel at providing an accurate assessment of population-based risk, they have weaknesses in the individualisation of that risk.

Humans are often poor at manipulating explicit probabilities;[33] however, clinicians have the ability to process additional knowledge that statistically-based systems often either ignore or treat on a perfunctory level. We would like to support clinical decision making by providing explicit probabilistic estimates of risk based on an integration of the variety of our knowledge sources.

[28] (Richards et al., 1994).
[29] (Veer et al., 2005; Cristofanilli et al., 2005).
[30] (Galea et al., 1992).
[31] (Ries et al., 2004).
[32] (Ravdin et al., 2001).
[33] (Kahneman and Tversky, 1973; Borak and Veilleux, 1982).

6.14 Our Knowledge Sources

Clinical Data

We used clinical data from a subset of the American Surveillance, Epidemiology and End Results (SEER) study. The total study is very large (over 3 million patients) and presents summary results on cancer diagnosis and survival in the USA between 1975 and 2003,[34] and subsets of the data are available for public use. We used a subset that initially consisted of 4878 individuals with breast cancer, which, once cases with incomplete data were removed, was reduced to 4731. The dataset consists of details of patient age (in 5 year bands, from 15–19 to 85+), the tumour size and histological grade, Oestrogen and Progesterone receptor status, the number of positive lymph nodes (if any), surgical type (mastectomy vs breast conserving), whether radiotherapy was given, the patients' survival from diagnosis (in months) and whether they had survived up until 5 years post-diagnosis. Patients in this subset were only followed up for 5 years, and so there is no data available on longer survival times.

Initial inspection of the clinical data was carried out using standard spreadsheet software (OpenOffice.org 2, 2005). Initial work concentrated on regrouping some of the data as follows. Oestrogen receptors are produced as part of the same intra-cellular pathway as Progesterone receptors, and as a result there is a very close correlation between ER & PR status. Since they are regarded as being one entity for most clinical purposes, we combined them into a single 'Hormone Receptor' variable. The Lymph Node status was converted from a number of positive lymph nodes (from 0–15) into a binary variable (True/ False), patient age was converted from 5 year age bands into 15–50, 50–70, 70–90, and Tumour size was converted from size in millimetres to sizes 0–20, 20–50, and 50–150 (these corresponding to clinical T Stages 1, 2, and 3+4). Patients with incomplete data (for example missing number of involved lymph nodes) were deleted from the dataset. A sample of the dataset is depicted in Table 6.1.

Table 6.1. A sample of the clinical dataset

Age	T Size	Grade	HR_status	Positive LN	Surgery	Radiotherapy	Survival	Status
70-74	22	2	1	1	1	1	37	1
45-49	8	1	1	0	2	1	41	1

The variables of the clinical database—i.e., the column headings of Table 6.1 are as follows:

Age: Age in years;
T Size: size of the primary tumour, in millimetres;
Grade: Histological grade of the tumour, from 1–3; 3 being most abnormal;

[34] (Ries et al., 2004).

HR_Status: Positive if the sample was positive for *either* Oestrogen *or* Progesterone receptors;
Positive LN: 1 if the patient had any lymph nodes involved by tumour, 0 otherwise;
Surgery: 1 if the patient had surgery for the tumour;
Radiotherapy: 1 if the patient received radiotherapy for their tumour;
Survival: Recorded survival in months;
Status: Status at final follow-up, 1 = alive, 0 = died.

Genomic Data

We used two karyotype datasets from the progenetix database (www.progenetix. de).[35] Progenetix contains discretised data on band-specific chromosomal rearrangements of cancer and leukemia cases (loss -1, gain +1, no change 0). It consists of a compilation of published data from comparative genome hybridisation (CGH), array CGH, and matrix CGH experiments, as well as some studies using metaphase analysis. Progenetix is, with 12320 CGH experiments, by far the largest public CGH database. We had available:

(i) a breast cancer CGH dataset of 502 cases which lacked consistent clinical annotations, which we used to learn the genomic Bayesian net from band data only,
(ii) a second CGH data set of 119 cases with clinical annotation, including lymph node status (an additional 12 individual cases with clinical annotation were set aside as a validation set), and
(iii) a recent study, Fridlyand et al. (2006), which contains quantitative information concerning the probabilistic dependence between the variables HR_status and 22q12—this provided a further bridge between clinical and genomic variables.

From the total number of chromosomal bands in the human genome, we selected 28 bands in this proof-of-principle application. The chosen bands were hypothesised to be closely associated with tumour characteristics, progression and outcome (as represented by variables in the clinical net) based on genes with known function present on the bands. Genes were evaluated according to the biological processes they participate in, using their Gene Ontology annotations, e.g., cell cycle regulation, DNA damage repair and cancer-related signal pathways. An additional selection criterion was the presence of at least 3 relevant genes on the band.

The larger dataset consisted of 116 separate data fields. For reasons of space, 12 sample fields are reproduced in Table 6.2. The code of the form Np/qn indicates:

- N: which chromosome (1–22, X or Y);
- p/q: the short (p)/ long (q) arm of the chromosome;
- n: the band on the chromosome arm (0–40, depending on chromosome).

[35] (Baudis and Cleary, 2001).

Table 6.2. A sample of the larger of the genomic datasets

1p31	1p32	1p34	2q32	3q26	4q35	5q14	7p11	8q23	20p13	Xp11	Xq13
0	0	0	1	-1	0	0	1	0	0	0	-1
0	0	1	1	0	0	0	-1	-1	0	0	0

Table 6.3. A sample of the smaller of the genomic datasets

Lymph Nodes	1q22	1q25	1q32	1q42	7q36	8p21	8p23	8q13	8q21	8q24
0	1	1	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0

The level of each band could either be unchanged (0) decreased (-1) or increased (1).

The smaller dataset was similar to the larger one, except for the fact that it included data on whether the patient's lymph nodes were also involved. 11 of the 26 data fields are reproduced in Table 6.3.

6.15 Constructing the Network

Our knowledge takes the form of a clinical database, two molecular databases, and information about the relationship between two variables derived from a research paper (§6.14). The clinical database determines a probability distribution $freq_c$ over the clinical variables and imposes the constraint $p_{\downharpoonright C} = freq_c$, i.e., the agent's probability function, when restricted to the clinical variables, should match the distribution determined by the clinical dataset. Similarly the first molecular database imposes the constraint $p_{\downharpoonright M} = freq_m$. The additional molecular dataset and the paper contain the observations that define the probability distribution $freq_s$ of three variables $S = \{$ HR_status, Positive LN, 22q12$\}$, the first two of which occur in the clinical dataset and the other of which occurs in the molecular dataset; it imposes the constraint $p_{\downharpoonright S} = freq_s$.[36]

Given constraints of this form, an obnet on the variables in C, M and S can be constructed in the following way. First use standard methods, such as Hugin software, to learn a Bayesian net from the clinical dataset that represents $freq_c$, subject to the condition that the linking variables (positive LN, HR status) are root variables (Fig. 6.2). Similarly learn a Bayesian net from the larger genomic

[36] Fridlyand et al. (2006) report frequency data on gain and loss of 22q12 in breast cancer dependent on oestrogen hormone receptor status. Interestingly, loss of 22q12 is far more frequent in ER positive tumours; in their study, 45% of ER positive cases showed loss of 22q12, and 5% exhibited gain. In contrast, in ER negative tumours, loss or gain occurs with equal frequency of 20%. This information was used to add an additional arrow between HR status and 22q12, and the conditional probability table for 22q12 was amended to reflect this dependence using the frequency data.

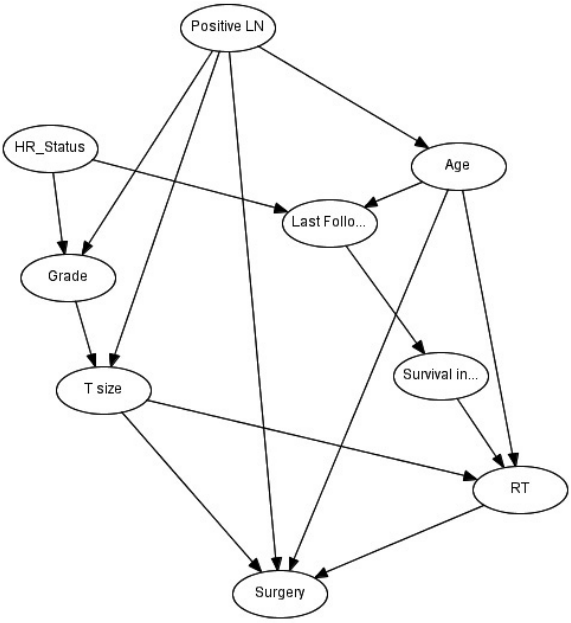

Fig. 6.2. The graph of the Bayesian net constructed from the clinical dataset

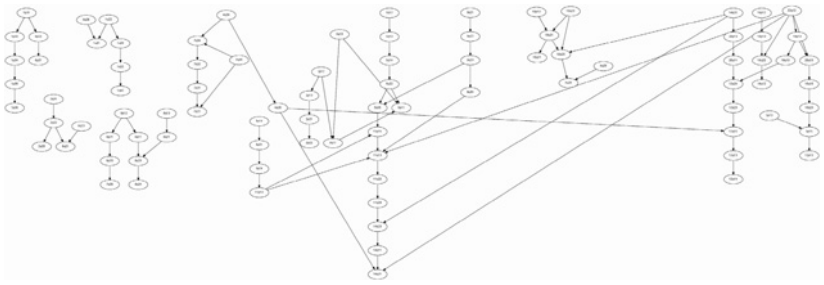

Fig. 6.3. The graph of the Bayesian net constructed from the large genomic dataset

dataset that represents $freq_m$, ensuring that the linked variable (22q12) is a root of the net (Fig. 6.3). Finally learn a bridging network, Fig. 6.4, from the smaller genomic dataset and the study, merge the three graphs to form one graph by merging identical variables, and integrate the conditional probability tables. Fig. 6.5 shows the graph of the resulting integrated obnet.

Here a conflict arose between the probability distribution determined by the clinical dataset and the probability distribution determined by the genomic dataset used to bridge the genomic and clinical variables: these gave different values to the probability of Positive LN. In §6.7, we pointed out that if neither dataset were to be preferred over the other, then the conflicting datasets

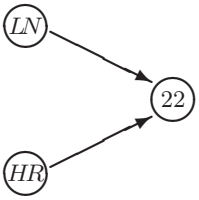

Fig. 6.4. The graph of the Bayesian net constructed from the smaller genomic dataset and the published study. The variables are Positive LN, HR status and 22q12.

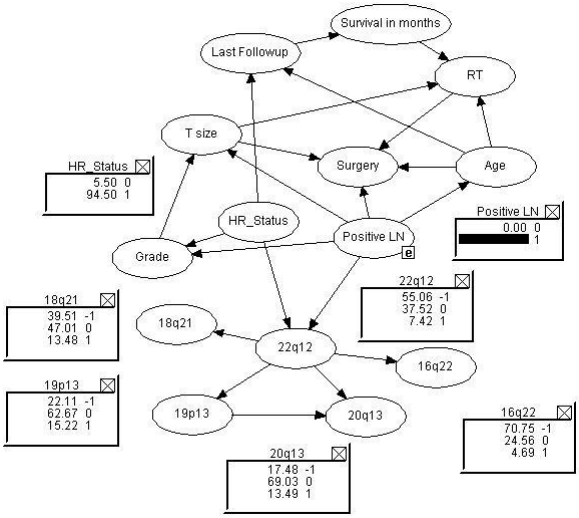

Fig. 6.5. The objective Bayesian net. Probability for positive lymph node status is set to 1 (black bar), and the calculated probability distributions for selected nodes are shown (HR status: 0 negative, 1 positive; chromosomal bands: -1 loss, +1 gain, 0 no rearrangement; RT radiotherapy).

constrain the probability of the assignment in question to lie within the closed interval bounded by the conflicting values; one should then take the least committal value in that interval. But in this case there are reasons to prefer one dataset over the other. First, the clinical dataset is based on a much larger sample than the bridging genomic dataset—for example, the clinical dataset has 2935 people with $LN = 0$, while the bridging genomic dataset has 56. Second, the molecular dataset has a clear sample bias: overall a frequency bias in favour of loss of 22q12 in breast cancer has been observed (20% loss vs. 7% gain in 800 cases in the progenetix database; accessed on the 14th of June 2006); furthermore, one may hypothesise that the presence of the KREMEN1 gene on 22q12 suggests that band loss, rather than gain, is more likely to be correlated with positive lymph node status, at least in certain karyotype contexts (§6.16). Thus the clinical data should trump the genomic data over the conflict on LN, i.e.,

Table 6.4. Probability table for Positive LN in the obnet

Positive LN	p
0	0.62
1	0.38

Table 6.5. Conditional probability table for 22q12 in the obnet

	Positive LN	0	0	1	1
22q12	HR_Status	0	1	0	1
-1		0.082	0.205	0.266	0.567
0		0.835	0.772	0.468	0.370
1		0.082	0.023	0.266	0.063

the probability table of *LN* in the obnet, Table 6.4, is simply inherited from the clinical net. The conditional probability table for 22q12 is depicted in Table 6.5.

Validation

Validation of our merged network is difficult; almost by definition, a suitable dataset involving the whole domain does not exist (if it did, we would not need to use this technique; we would simply use the dataset to learn a Bayesian net). Because of this, the best we were able to do was to use a small validation set with 11 test cases; validation showed reasonable agreement between the test cases and the obnet, but we must be careful not to over-interpret the results.

We approached the validation from two sides; the first was to set the 22q12 status and observe the effect on lymph node (LN) status; the second was to set the status of the lymph node variable, and observe the effect on 22q12 status. Unfortunately, the test set is both small and contains few cases of of 22q12 alteration.

Table 6.6. Setting 22q12 and observing LN status

22q12	LN status: Predicted from net	Actual	No.
1	0.62	0	1
0	0.5	0.55	9
-1	1	1	1

As can be seen from Table 6.6, only the linkage between no change of 22q12 and LN status predicted by the net is reflected in the validation set. It is difficult to interpret the results for the other values of 22q12.

Table 6.7. Setting LN status and observing 22q12

LN status	22q12: Predicted from net	Actual	No.
+	0: 0.84, 1: 0.08, -1:0.08	0: 1.00	5
-	0: 0.95, 1: 0.05	0: 0.66, 1: 0.16, -1:0.16	6

Entering evidence in LN status, we have the results of Table 6.7. As we can see, both cases of LN status agree reasonably well with the observed cases, but again we must be careful not to over-interpret this relationship.

6.16 Interpretation of the Integrated Obnet

We have presented a general method for merging Bayesian networks which model knowledge in different areas. This method was applied to an example application linking knowledge about breast cancer genomics and clinical data. As a result, we have been able to examine the influence of karyotype pattern on clinical parameters (e.g., tumour size, grade, receptor status, likelihood of lymph node involvement) and vice versa (Fig. 6.5).

In the post-genome era, prognostic prediction and prediction of targets for new anti-cancer treatments from omic and clinical data are becoming ever more closely related—both need to relate molecular parameters to disease type and outcome. This correspondence is very clearly reflected in the uses to which the integrated obnet may be put. Obnet analysis may facilitate (i) discovery of genomic markers and signatures, and (ii) translation of clinical data to genomic research and discovery of novel therapeutic targets.

Discovery of Genomic Markers

Since hormone receptor status and lymph node involvement are well-known prognostic factors for survival and disease recurrence in patients with breast cancer, the ability to link karyotype patterns to this is clearly of great potential significance. Previous tumour genotyping in breast cancer has already shown the usefulness of genomic rearrangements as prognostic indicators.[37]

For clinical decision making, this technique may also be useful when applied to integrate karyotype or other molecular data with parameters that cannot be observed in routine clinical practice, but are of clinical significance. An example might be the presence of distant metastasis on PET-CT, an imaging modality that may be present in the research setting but is not widely available in the clinic, but which may have prognostic significance for breast cancer recurrence. The use of such a net would then allow practitioners, where PET-CT is not available, to use genomic data to estimate the likelihood of a positive scan. There are of course, many different possible options for such networks, and it

[37] See, e.g., Al-Kuraya et al. (2004).

remains an open question as to which will, in clinical terms, prove to be the most useful.

Large clinical datasets are extremely expensive and difficult to collect. This is particularly true in diseases such as breast cancer, where the risk of recurrence extends up to at least 10 years, and hence requires long-term follow-up for accurate estimation. However, the generation of potential new predictive markers, such as genomic information or cell surface proteins, for exploration is currently a significant area of research. The correlation of such markers with better known clinical markers is (relatively) simple, in that it does not require long-term follow-up, and can be estimated following standard surgical treatment. However, for such information to be useful, it must be integrated with the existing databases on long-term outcomes, and it is this that we have demonstrated here.

Translation of Clinical Data to Genomic Research

The probabilistic dependence between 22q12 status and lymph node involvement was followed up by analysis of the genes with known function on this chromosomal band. This strongly suggested a causal interpretation of the dependency relationship based on knowledge of cellular pathways which regulate biological processes (mechanistic causation). KREMEN1 encodes a high-affinity dickkopf homolog 1 (DKK1) transmembrane receptor that functionally cooperates with DKK1 to block wingless (WNT)/beta-catenin signalling, a pathway which promotes cell motility.[38] Loss of 22q12 may therefore contribute to cancer cell migration through loss of the inhibiting KREMEN1 protein. The probability distribution for 22q12 is consistent with this hypothesis (Fig. 6.5).

In total, twelve genes implicated in cell migration and metastatic potential were identified on 22q12 and the other bands shown in Fig. 6.5. Like KREMEN1, the protein products of the other eleven genes can also be placed in the context of the metastatic pathways they participate in. Provided that appropriate kinetic interaction data are available, computational pathway modelling[39] can be employed to predict changes in pathway function resulting from the probabilistically dependent band gains and losses and concomitant changes in gene copy number. Molecularly targeted intervention strategies aimed at bringing about a therapeutic response in the cells so affected can be explored by running simulations using such pathway models. Simulation may be seen as being motivated by an agency-oriented notion of causation (see also §6.20).

Part IV: Further Development of the Method

6.17 Qualitative Knowledge and Hypotheses

There are various ways in which we intend to develop the method presented here.

[38] (Mao et al., 2002).
[39] (Alves et al., 2006).

First, as discussed in §6.11, there are a variety of knowledge sources that we hope to integrate. These include argumentation systems, medical ontologies, and causal relationships, as well as the clinical and molecular datasets which have been the focus of this chapter. In this Part, we shall discuss some of these other knowledge sources.

Second, as indicated in §6.16, we intend to exploit the objective Bayesian net that integrates these knowledge sources by using it not only for prognosis but also as a device for hypothesising new qualitative relationships amongst the variables under consideration. If the obnet reveals that two variables are probabilistically dependent, and that dependence is not explained by background knowledge, then we may hypothesise some new connection between the variables that accounts for their dependence. For example, we may hypothesise that the variables are causally (§6.20) or ontologically (§6.19) related. Furthermore, any such dependence can be used to generate qualitative arguments (§6.18): each variable will be an argument for or against the other, according to the direction of the dependence.

Third, we can increase the complexity of the formalism, in order to model temporal change or different levels of interaction, for instance. We shall discuss such extensions in §6.21.

These are avenues for future research. In this Part it will suffice to make some remarks on the likely directions that such research will take.

6.18 Argumentation

So far we have described how the network was developed, and analysed its performance as a Bayesian network. However, as suggested in §6.10 above, we are interested in more than just the probabilistic interpretation of the network—we are also interested in what the new network says about the world. Whereas in that section we suggested moving from qualitative to quantitative knowledge, here we shall discuss the opposite.

Bayesian Networks as Arguments

Bayesian networks are a useful tool for providing estimates of the probability for a variable of interest based on the evidence. Of course, Bayesian networks are not the only method of doing so, and there has been much work over the years on different formal methods to support decision making (rule-based systems, support vector machines, regression models, etc.). More generally, humans often use the notion of weighing up 'the arguments' for a belief or action to help them come to a conclusion. The argumentative method goes back at least 2500 years, and extends beyond the Graeco-Roman tradition.[40] Arguments have the advantage that they can present not only a conclusion, but also its justification. The idea of trying to base decision-making on arguments has a long history. The first clear example of an (informal) procedure for doing so was described by Benjamin Franklin.[41]

[40] (Gard, 1961).
[41] (Franklin, 1887).

Of course, this informal notion of an argument can be neither implemented nor assessed in any rigorous fashion, but over the last 10-15 years there has been some work on developing competing formal (and computable) models of argument.[42] More recent work has drawn together developments in non-monotonic logic and insights from cognitive science to produce a number of different argumentation frameworks.[43] We do not intend to present a review of the field here, but suffice to say that there are two general themes, *argument formation* and *argument resolution*. Each competing formalism defines these slightly differently, but in general an argument is a set of premises that allow one to deduce, via a set of rules, some set of conclusions. Resolution of competing arguments varies considerably between formalisms, is difficult to summarise in general terms, and matters less for our discussion here. However, what interests us is how one can interpret our new Bayesian network in terms of arguments. In other words, given a Bayesian network, what can we say about the arguments for and against a set of propositions, and given a new Bayesian network (formed from two or more existing ones) what new arguments can we make? Two of the authors of this paper have previously presented a simple technique for developing arguments from a Bayesian network,[44] basing the arguments on a relatively simple argumentation formalism,[45] and we use the method outlined there to develop our arguments from the Bayesian network. For reasons of space, we do not present the details of our method here; they can be found in Williams and Williamson (2006). Instead, let us consider what it might mean, in general terms, for a probabilistic statement to be interpreted as an argument. Firstly, therefore, let us consider what we mean by an argument.

Rules, Arguments and Probability

Intuitively, an argument is a line of reasoning that proceeds from some premise(s) via a set of deductions to some conclusion. As we all know, arguments are defeasible—that is, their conclusions may at some point be challenged and what was at some point held to be 'true' by argument may later be found to be untrue. We can formalise this 'method of argument' in various different ways (as mentioned above) but in general we have a quartet of premises, rules (for deduction), conclusions, and conflict between arguments. In order to map probabilistic statements into an argumentative framework, therefore, we need to consider how different aspects of a Bayesian system map into the quartet of argumentation, and what effects this has. We shall do this below, but first we need to establish a (fairly trivial) mapping between Bayesian notation and argumentative notation. We do this by considering all variables to be binary in nature, and each node in the network to represent a single binary-valued variable. The mapping between such a network and a truth-valued logic (say, propositional logic) should be clear: for any variable X, $p(X = 1)$ is interpreted as x and $p(X = 0)$ is interpreted as $\neg x$.

[42] (Fox and Parsons, 1997).
[43] (Amgoud et al., 2004; Hunter and Besnard, 2001; Krause et al., 1995).
[44] (Williams and Williamson, 2006).
[45] (Prakken and Sartor, 1996).

Given the correspondence above, mapping the premises is fairly simple: they are the inputs to the network, i.e., the variables in the net that are instantiated. Similarly, the conclusions are also relatively simple—they are the values of other nodes in the network. Given any two nodes in a Bayesian network, the absence of any connection between them implies probabilistic independence, and therefore precludes one being an argument for the other. The presence of a connection suggests that there may be a relationship between them. It is this relationship that we interpret as forming the 'rules'[46] for an argumentation system, and in its most basic form, if the truth of one variable A increases the probability of another variable B being true, then we might write $a \Rightarrow b$. An argument is the association of a set of premises and rules that lead to a conclusion[47]—e.g., $< \{a, a \Rightarrow b\}, b >$, where $<>$ denotes the argument, the first element $\{a, a \Rightarrow b\}$ is the support and b is the conclusion. Arguments are in conflict if they argue for conclusions which are mutually exclusive—so if we had another argument $< \{c, c \Rightarrow \neg b\}, \neg b >$, this would be in conflict with our first argument.

Our Network as Arguments

We are now in a position to return to the first question we posed above—'what can we say about the arguments for and against a set of propositions?' The first thing to observe is that our approach will only allow us to develop arguments about literals that correspond to nodes in the network. Secondly, we can only develop rules between literals that are linked in the network. Thus, while *we* might know of some connection, that connection will not appear unless it also appears as a conditional dependence (and hence a link) in our network. This is why the procedure outlined in §6.10 is so important: all background knowledge must be taken into account in the construction of the objective Bayesian net. Of course, in general we might add some additional rules (from other sources), but the rules (and hence the arguments) developed from the network will only be concerned with those literals that appear and are associated in the network. Thirdly, given the dichotomised nature of the variables, we have a tendency to develop arguments both for and against literals. We can see this from the following example. The CPT for the Tumour size node from our clinical network is shown in Table 6.8. From these values we can calculate that $p(T\_Size = 0$–$20)$ is 0.657, whilst $p(T\_Size = 0$–$20 \mid LN = 0)$ is 0.753, and $p(T\_Size = 0$–$20 \mid LN = 1)$ is 0.5. Therefore, following the method outlined above, we can see that we should develop the following rules:

- $(LN = 0) \Rightarrow (T\_Size = (0$–$20))$
- $(LN = 1) \Rightarrow \neg(T\_Size = (0$–$20))$

On the one hand, this may seem to problematic—the generation of pairs of opposing rules (and hence arguments) might lead us to some sort of deadlock.

[46] A 'rule' in this context is a defeasible piece of knowledge that allows one to infer the value of one variable from another.

[47] Most systems, including ours, also impose further restrictions to ensure arguments that are consistent and non-circular.

Table 6.8. Conditional probability table for Tumour Size

	Grade	1	1	2	2	3	3
T_Size(mm)	Positive LN	0	1	0	1	0	1
0—20		0.85	0.66	0.76	0.5	0.62	0.42
20—50		0.14	0.33	0.21	0.5	0.35	0.58
50—150		0.005	0	0.02	0	0.03	0
No.		689	271	1668	990	578	535

However, this is not a bad as it may seem. Firstly, it seems intuitively correct to develop rules for both options—after all, the whole point of the Bayesian net is that it contains information about both options. Secondly, the 'deadlock' between the rules can be resolved in a variety of ways (for example, we could encode the likelihood of the different options as 'weights' given to the rules). Thirdly, the point of the rules, and arguments, is to allow us to help integrate human decision-making with our Bayesian techniques, and to allow us to do this we need to display arguments for both options, even if they have different weights. Finally, in this case it would of course be impossible to have a measurement for both $LN = 0$ and $LN = 1$ at the same time (although we might have other *arguments* for both at the same time, as we shall see below).

New Arguments from New Networks

The second question we asked above was 'given a new Bayesian network. . .what new arguments can we make?' In a sense, the answer is (almost) 'none'. After all, as we said above, all the rules are developed from existing literals and relationships in a Bayesian network. Since our new network is only a combination of the existing networks, there should not be anything new. However, this answer misses one of the aspects that is crucial to the difference between argumentation and Bayesian networks.

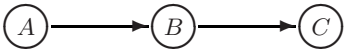

Fig. 6.6. The graph of a Bayesian network

One of the key features of Bayesian networks, as mentioned in §6.8 is that each variable is probabilistically independent of its non-descendants conditional on its parents, and as is noted above, this has some very desirable properties from a computational aspect. However, once we consider developing arguments from our network, we see that this relationship comes out differently when applied to arguments. For example, consider a (very simple) Bayesian net, whose graph is shown in Fig. 6.6. Depending on the probabilities in the net, we may develop the rules:

- $(A = 1) \Rightarrow (B = 1)$ which we write as $a \Rightarrow b$
- $(B = 1) \Rightarrow (C = 1)$ which we write as $b \Rightarrow c$

- $(A = 0) \Rightarrow (B = 0)$ which we write as $\neg a \Rightarrow \neg b$
- $(B = 0) \Rightarrow (C = 0)$ which we write as $\neg b \Rightarrow \neg c$

Now, according to the Bayesian network, B screens C off from A. However, under our argumentation formalism, we might have the following arguments:

- A_1: $< \{a, a \Rightarrow b, b \Rightarrow c\}, c >$ as an argument for c
- A_2: $< \{\neg a, \neg a \Rightarrow \neg b, \neg b \Rightarrow \neg c\}, \neg c >$ as an argument for $\neg c$

which seems to make explicit the dependence of c on a. Now, if we do not know the status of b, then in both formalisms, we understand that in fact the best guide to the state of c is the state of a, and so the two approaches are in agreement. If we know both a and b, and they are 'concordant' (e.g. a and b or $\neg a$ and $\neg b$) then we will find that indeed a is 'redundant', and c is entirely determined by b. However, our work is motivated by the fact that our knowledge is often partial and conflicting. For example, we might have one piece of information about a and another about b, and they may conflict—for example, we may believe both $\neg a$ and b. In such a situation, the Bayesian net approach would typically discard the information about a, as it would be over-ridden by the information about b. Under an argumentative approach, however, we are able to construct arguments for *both* c and $\neg c$, as shown below:

- A_1: $< \{b, b \Rightarrow c\}, c >$
- A_2: $< \{\neg a, \neg a \Rightarrow \neg b, \neg b \Rightarrow \neg c\}, \neg c >$

Thus the argumentation differs from the Bayesian net in that it does not follow the probabilistic independencies of the net. Obviously at some point we will need to resolve this disagreement, but we can at least start by considering both cases. The Bayesian net approach retains probabilistic validity, but only by enforcing a set of strict rules, one of which is committing to a particular value of certain variables (b in our example); the argumentative approach loses this precision, but has the advantage that it can handle conflicting premises and generate arguments based on them, which it then resolves, rather than losing this information (b and $\neg a$ in our example). Such differences are not unique to our particular brand of argumentation (for example, they are seen in Parsons' qualitative probability framework,[48] which devotes a considerable amount of space to discussing the problem).

We are now finally in a position to answer our second question. When we add a new network, we can still only develop the same rules that we developed in each network. However, because we can use the rules to form arguments, we can form arguments that are 'bigger' than those formed in either network alone. For example, consider our merged network, Fig. 6.5. In this case, we can see that 22q12 is connected to lymph node status, and we would have been able to form an argument linking the two from the genomic net alone. However, given the links in the network, we can form an argument (but not a valid probabilistic relationship) which would link lymph node status and 19p13 status, even if we

[48] (Parsons, 2003, 2004).

know 22q12 status. Such an argument is not interpreted probabilistically but may still be useful for explaining the links to human users.

6.19 Ontologies

In §6.11 we mentioned some of the different types of knowledge that we might try and integrate into our Bayesian network; one important category is ontological knowledge. Ontologies (formal representations of vocabularies and concept relationships) and common data elements support automated reasoning including artificial intelligence methods such as Bayesian networks. Various standard vocabularies and object models have already been developed for genomics, molecular profiles, certain molecular targeted agents, mouse models of human cancer, clinical trials and oncology-relevant medical terms and concepts (SNOMED-RT/CT, ICD-O-3, MeSH, CDISC, NCI Health Thesaurus, caCORE, HUGO). There are also existing ontologies describing histopathology (standards and minimum datasets for reporting cancers, Royal College of Pathologists; caIMAGE, NCI). The European Bioinformatics Institute (EBI) is developing standards for the representation of biological function (Gene Ontology) and the Microarray Gene Expression Data (MGED) Society is developing MIAME, MAGE, and the MAGE ontology, a suite of standards for microarrays (transcriptomic, array CGH, proteomic, SNP). However, significant gaps still exist and eventually, all cancer-relevant data types (see the NCRI Planning Matrix, www.cancerinformatics.org.uk/planning_matrix.htm) will need to be formalised in ontologies. These efforts are ongoing and pursued by a large community of researchers (see above, and ftp1.nci.nih.gov/pub/cacore/ ExternalStds/ for further details on available standards). Clearly, it would be desirable to incorporate the fruits of these efforts in our scheme for knowledge integration; the potential for using ontologies as a knowledge source will increase with the maturation of these other initiatives.

While we would like to base our objective Bayesian net on ontological knowledge as well as our other knowledge sources, we also believe that we could establish some possible ontological relationships from our Bayesian network. The most obvious of these is in establishing a sub-/ super-class relationship between two variables. For example, imagine a dataset which recorded both whether someone had had breast cancer and if they had had each individual subtype of breast cancer (and also included those without breast cancer). Such a network would contain several nodes, but in each case, if any subtype of cancer was positive, then the 'has cancer' variable would also be positive. Such patterns of conditional dependence may be complex—in our example, there would be several different nodes linking to the 'has cancer' node, but in general are indicative of the presence of a sub-/super-class relationship (where the dependent variable is the superclass). We may take this idea further by suggesting that if there are certain combinations of variables that (together) are highly predictive of another variable, we might regard those individuals as acting as a 'definition' of the outcome variable. For example, consider a dataset which records whether

individuals (of different species) are human or not, whether they are women or men, and if they have XX or XY chromosomes. Now, those individuals with XX chromosomes are of course women, and so all individuals who are human and have XX chromosomes will also be women. From this, we might deduce that in fact, the two are equivalent, and thus being human and having XX chromosomes is the same as being a woman. This approach is, to say the least, prone to error, but provides a way to start learning such definitions from data, something that currently has to be done by hand.

In an ideal world, we would expect these relationships to be absolute, but in reality, we should allow for there being some 'noise' in the data, and may well be willing to accept a near-absolute (say 95% or 98%) as being suggestive of such a link. However, this is *not* the same as saying that the ontological relationship is probabilistic, as some authors do. Instead, it is based upon a supposition that the ontological relationship is absolute, but that the data may imperfectly reflect this. Interestingly, however, we seem to rarely see this sort of relationship in our networks. The reason for this is that the resolution of ontological relationships is one of the things that we tend to do at either the data collection or pre-processing stage. For example, as a part of our pre-processing of the data we combined the Oestrogen and Progesterone receptor status into a new variable, Hormone Receptor status, where the class of Oestrogen and Progesterone receptors are subclasses of Hormone Receptors. However, this is not to say that such relationships will never be important. One of the aims of the semantic web, and science on the semantic web, is to enable large amounts of data to be shared; such sharing will necessitate automatic handling of data (as manual processing of larger and larger databases becomes harder and harder), and tools for the handling of data may be able to use such strong probabilistic relationships to highlight potential ontological issues to the user.

6.20 Causal Relationships

Concepts of Causation in Complex Systems

Since each patient's cells evolve through an independent set of mutations and selective environments, the resulting population of cancer cells in each patient is likely to be unique in terms of the sum total of the mutational changes they have undergone. Inter-personal variability has given rise to the new field of 'pharmacogenomics' (in cancer and other diseases) which has as its ultimate aim diagnostic and prognostic prediction, and design of individualised treatments based on patient-specific molecular characteristics. Given the prevailing high degree of uncertainty in the face of biological complexity, pharmacogenomics offers great promise, but is also 'high risk'. Risk has, for example, been highlighted by recent findings of the nonreproducibility of classification based on gene expression profiling (expression microarrays, transcriptomics).[49] In this situation, diagnosis and prognosis based on biomarkers or profiles of multiple molecular indicators

[49] See, for example, Michielsa et al. (2005).

may lead to mis-classification, and may identify patients as likely non-responders to a given treatment when in fact they would derive benefit or, conversely, may falsely predict efficacy in patients for whom none can be achieved. One may argue that this uncertainty, at least in part, is compounded by prevailing notions of biological causality which is still preoccupied with the search for single (or a small number of) physical causes, and a failure to take into account the characteristics of complex systems.

Different views on the nature of causality lead to different suggestions for discovering causal relationships.[50] Medicine has been, due to its very nature, particularly focused on an agency-oriented account of causality which seeks to analyse causal relations in terms of the ability of agents (doctors, health professionals and scientists) to achieve goals (cures, amelioration of symptoms) by manipulating their causes. According to this conception of causality, C causes E if and only if bringing about C would be an effective way of bringing about E. Or conversely, for example, in the context of therapeutic intervention, C is seen as a cause of E if by inhibiting C one can stop E from happening. In this intervention-oriented stance, the agent would also seek to ground this view of causality in a mechanistic account of physical processes, as, for example, in the mechanistic mode of action of a drug. In diagnostic and prognostic prediction from patient data, a causal framework is also implied; here, causation may be conceptualised as agency-based, mechanistic or in terms of a probabilistic relationship between variables. However, the extensive literature on the subject reveals a number of problems associated with all three approaches.[51]

An alternative view of causality, termed *epistemic causality* by Williamson (2005a), overcomes the strict compartmentalisation of current theories of causation, and focuses on causal beliefs and the role that *all* of these indicators (mechanistic, probabilistic, agency-based) have in forming them. It takes causality as an objective notion yet primarily a mental construct, and offers a formal account of how we ought to determine causal beliefs.[52] This approach will be applied to glean causal hypotheses from an obnet, as outlined below.

We are faced with a profound challenge regarding causation in complex biological systems. In her discussion of developmental systems and evolution, Susan Oyama observes 'what a cause causes is contingent and is thus itself caused'.[53] The influence of a gene, or a genetic mutation, depends on the context, such as availability of other molecular agents and the state of the biological system, including the rest of the genome. Oyama argues for a view of causality which gives weight to all operative influences, since no single influence is sufficient for a biological phenomenon or for any of its properties. Variation in any one influence, or many of them, may or may not bring about variation in the result, depending on the configuration of the whole. The *mutual dependence of (physical) causes* leads to a situation where an entire ensemble of factors contribute to any given

[50] (Williamson, 2007a).
[51] (Williamson, 2005a; Williamson, 2007a, and references therein).
[52] (Williamson, 2007a).
[53] (Oyama, 2000, pp. 17–18 and references therein).

phenomenon, and the effect of any one factor depends both on its own properties and on those of the others, often in complex combinations. This gives rise to the concept of organised *causal networks* and is a central insight of systems thinking. The biological relevance of any factor, and therefore the information it conveys, is jointly determined, typically in a statistically interactive fashion, by that factor and the entire system's state.[54]

Whilst 'systems thinking' is likely to be fundamental for biomedicine and for cancer, in particular, due to its overwhelming complexity, we still lack a principled methodology for addressing these questions. Methodology development is a pressing need, and it is with this major objective in mind that our research is undertaken. The work presented here combines a multidisciplinary framework of biological systems theory and objective Bayesian network modelling; our next step will be to integrate epistemic causality into this framework.

Here, it may be helpful, or even necessary, to draw a distinction between fundamental science and applied biomedical research. In systems biology, the ultimate goal may be to gain a complete mechanistic explanation of the system complexity underlying causal networks. The achievement of this aim still lies in the, possibly far distant, future. In contrast, in biomedical research and clinical practice, we tend to be more immediately interested in discovering molecular predictors for diagnostic and prognostic purposes, and in developing effective strategies for intervention in malfunctioning body systems and disease processes.

Perhaps surprisingly, an applied focus of this kind may work in our favour vis-à-vis biological complexity, as progress is not as severely constrained by a requirement for an exhaustive mechanistic elucidation of the complete system. In this chapter we sketch a discovery strategy which is based on the epistemic view of causality (see below and §6.16). The strategy integrates *probabilistic* dependency networks (Bayesian networks) with expert knowledge of biological *mechanisms*, where available, to hypothesise causal networks inherent in the system. This approach enables one to predict probabilistic biomarker profiles and targets for intervention based on the identified dependencies between system components. Interventions may then be tested by computational modelling and experimental validation which may be seen as foregrounding '*agency-based*' causation. This is a pragmatic strategy which can yield insights into system function which are attainable now and are valuable from a biomedical point of view.

Gleaning Causal Relationships from an Obnet

The epistemic theory of causality maintains the following.[55] Just as, under the objective Bayesian account, an agent's rational degrees of belief should take the form of a probability function objectively determined by her evidence, so too her rational causal beliefs, represented by a directed acyclic graph (*dag*), are objectively determined by her evidence. An agent should adopt, as her causal belief graph, the most non-committal graph (i.e., the dag with fewest arrows) that satisfies constraints imposed by her evidence.

[54] (Oyama, 2000, p. 38).
[55] (Williamson, 2005a, Chapter 9).

Now, evidence imposes constraints in several ways. (i) The agent may already know of causal connections, in which case her causal graph should contain the corresponding arrows. (ii) She may know that A only occurs after B, in which case her causal graph should not contain an arrow from A to B. (iii) Or a causal connection from A to B may be incompatible with her scientific knowledge inasmuch as her scientific knowledge implies that there is no physical mechanism from A to B and hence no possible physical explanation of B that involves A; then there should be no arrow from A to B in her causal belief graph. (iv) Or there may be a *strategic dependence* from A to B (i.e., A and B may be probabilistically dependent when intervening to fix A and controlling for B's other causes) for which the agent has no explanation in her background knowledge; she should then have an arrow from A to B in her causal graph to explain the dependence, as long as other knowledge does not rule out such an arrow.

One can determine the agent's causal belief graph by running through all dags and taking a minimal graph that satisfies the constraints (i–iv) imposed by background knowledge; but such a method is clearly computationally intractable. Two other, more feasible methods are worth investigating. The agent's causal belief graph can be approximated by a minimal dag that satisfies the Markov condition and constraints of type (i–iii); standard Bayesian net software can be used to construct such a graph. Or one can generate an approximation to the causal belief graph by constructing a graph that satisfies constraints (i–iii) and incrementally adding further arrows (also satisfying these constraints) that correspond to strategic dependences in the obnet. These extra arrows are causal hypotheses generated by the objective Bayesian net.

6.21 Object-Oriented, Recursive and Dynamic Obnets

Another avenue for future research concerns extensions of the objective Bayesian net framework to cope with object orientation, recursion and temporal change.

Object-oriented and dynamic Bayesian networks possess certain advantages for the modelling of complex biological systems. For example, Dawid, Mortera and Vicard have applied OOBNs to the domain of genetics and complex forensic DNA profiling[56] and Bangsø and Olesen showed how OOBNs can be adapted to dynamically model processes over time, such as glucose metabolism in humans.[57]

Object-oriented Bayesian networks (OOBNs) allow one to represent complex probabilistic models.[58] Objects can be modelled as composed of lower-level objects, and an OOBN can have nodes that are themselves instances of other networks, in addition to regular nodes. In an OOBN, the internal parts of an object can be encapsulated within the object. Probabilistically, this implies that the encapsulated attributes are d-separated from the rest of the network by the object's inputs and outputs. This separation property can be utilised to locally

[56] (Dawid et al., 2007).
[57] (Bangsøand Olesen, 2003).
[58] (Koller and Pfeffer, 1997; Laskey and Mahoney, 1997).

constrain probabilistic computation within objects, with only limited interaction between them.

By representing a hierarchy of inter-related objects, an OOBN makes organisational structure explicit. OOBN 'is-a' and 'part-of' hierarchical structuring mirrors the organisation of ontologies in the biomedical knowledge domain (see §6.19), and ontologies can therefore be used as background knowledge to structure an OOBN.

Recursive Bayesian networks offer a means of modelling a different kind of hierarchical structure—the case in which variables may themselves take Bayesian networks as values.[59] This extra structure is required, for instance, to cope with situations in which causal relationships themselves act as causes and effects. This is often the case with policy decisions: e.g., the fact that smoking causes cancer causes governments to restrict tobacco advertising.

The timing of observations (e.g., symptoms, measurements, tests, events) plays a major role in diagnosis, prognosis and prediction. Temporal modelling can be performed by a formalism called Temporal Bayesian Network of Events (TBNE).[60] In a TBNE each node represents an event or state change of a variable, and an arc corresponds to a causal-temporal relationship. A temporal node represents the time that a variable changes state, including an option of no-change. The temporal intervals can differ in number and size for each temporal node, so this allows multiple granularity. The formalism of dynamic Bayesian nets can also be applied.[61]

OOBN properties allow one to exploit the modular organisation of biological systems for the generation of complex models. To our knowledge, OOBNs have not been applied to systems-oriented cancer modelling. We aim to assess the usefulness of OOBN methods for multi-scale models of cancer systems, especially to represent variables associated with heterogeneity in tumours. Our research will also evaluate the uses of the TBNE formalism and dynamic Bayesian nets for temporal models of karyotype evolution (§§6.2, 6.3) and evolving therapeutic systems (patient/tumour-therapy-response).

In sum, then, there are a variety of situations which call for a richer formalism. Since an obnet is a Bayesian net, one can enrich an obnet using all the techniques available for enriching Bayesian nets: one can render an obnet object-oriented, recursive or dynamic. The details of these extensions are questions for further work.

6.22 Conclusion

In this chapter we have presented a scheme for systems modelling and prognosis in breast cancer. A multiplicity of knowledge sources can be integrated by forming the objective Bayesian net generated by this evidence. This obnet represents the probabilistic beliefs that should adopted by an agent with that evidence;

[59] (Williamson and Gabbay, 2005).
[60] (Arroyo-Figueroa and Sucar, 2005).
[61] (Neapolitan, 2003).

Table 6.9. The interplay between evidence and belief

Evidence	⟷	Belief
Clinical data		Probabilistic (obnet)
Genomic data		Argumentative
Published studies		Ontological
Argumentation systems		Causal
Medical ontologies		
Causal knowledge		
Biological theory		

it can be used to assist prognosis of cancer patients. The obnet together with evidence can, in turn, be used to generate sets of argumentative, ontological and causal beliefs. These are just hypotheses and require testing; more data must be collected to confirm or disconfirm these hypotheses. These new data increase the base of evidence and consequently new beliefs (probabilistic, causal and so on) must be formulated. We thus have a dialectical back-and-forth between evidence and belief, as depicted in Table 6.9.

This iterative approach to knowledge discovery facilitates novel insights and hypotheses regarding the organisation and dynamic functioning of complex biological systems, and can lead to fruitful discovery from limited data. Objective Bayesian nets thus provide a principled and practical way of integrating domain knowledge, and of using it for inference and discovery.

Acknowledgements

This research was carried out as part of the caOBNET project (www.kent.ac.uk/secl/philosophy/jw/2006/caOBNET.htm). We are very grateful to Nadjet El-Mehidi and Vivek Patkar for their assistance with this work. For financial support we are grateful to Cancer Research UK, the Colyer-Fergusson Awards of the Kent Institute for Advanced Study in the Humanities and the Leverhulme Trust.

References

Abramovitz, M., Leyland-Jones, B.: A systems approach to clinical oncology: Focus on breast cancer. BMC Proteome Science 4, 5 (2006)

Al-Kuraya, K., Schraml, P., Torhorst, J., Tapia, C., Zaharieva, B., Novotny, H., Spichtin, H., Maurer, R., Mirlacher, M., Kochl, O., Zuber, M., Dieterich, H., Mross, F., Wilber, K., Simon, R., Sauter, G.: Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer Research 64, 8534–8540 (2004)

Alves, R., Antunes, F., Salvador, A.: Tools for kinetic modeling of biochemical networks. Nature Biotechnology 24, 667–672 (2006)

Amgoud, L., Cayrol, C., Lagasquie-Schiex, M.-C.: On bipolarity in argumentation frameworks. In: NMR, pp. 1–9 (2004)

Arroyo-Figueroa, G., Sucar, L.: Temporal Bayesian network of events for diagnosis and prediction in dynamic domains. Applied Intelligence 23, 77–86 (2005)

Bangsø, O., Olesen, K.: Applying object oriented Bayesian networks to large (medical) decision support systems. In: Proceedings of the Eighth Scandinavian Conference on Artificial Intelligence. IOS Press, Amsterdam (2003)

Baudis, M., Cleary, M.: Progenetix.net: an online repository for molecular cytogenetic aberration data. Bioinformatics 17, 1228–1229 (2001)

Borak, J., Veilleux, S.: Errors of intuitive logic among physicians. Soc. Sci. Med. 16, 1939–1947 (1982)

Bulashevska, S., Szakacs, O., Brors, B., Eils, R., Kovacs, G.: Pathways of urothelial cancer progression suggested by Bayesian network analysis of allelotyping data. International Journal of Cancer 110, 850–856 (2004)

Cristofanilli, M., Hayes, D., Budd, G., Ellis, M., Stopeck, A., Reuben, J., Doyle, G., Matera, J., Allard, W., Miller, M., Fritsche, H., Hortobagyi, G., Terstappen, L.: Circulating tumor cells: A novel prognostic factor for newly diagnosed metastatic breast cancer. J. Clin. Oncol. 23, 1420–1430 (2005)

Dawid, A., Mortera, J., Vicard, P.: Object-oriented Bayesian networks for complex forensic DNA profiling problems. Forensic Science International 169(256), 195–205 (2007)

Depew, D., Weber, B.: Darwinism evolving: systems dynamics and the genealogy of natural selection. MIT Press, Cambridge (1996)

Fox, J., Parsons, S.: On using arguments for reasoning about actions and values. In: Proc. AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning, Stanford (1997)

Franklin, B.: Collected Letters, Putnam, New York (1887)

Fridlyand, J., Snijders, A., Ylstra, B., Li, H., Olshen, A., Segraves, R., Dairkee, S., Tokuyasu, T., Ljung, B., Jain, A., McLennan, J., Ziegler, J., Chin, K., Devries, S., Feiler, H., Gray, J., Waldman, F., Pinkel, D., Albertson, D.: Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer 6, 96 (2006)

Galea, M., Blamey, R., Elston, C., Ellis, I.: The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Research and Treatment 3, 207–219 (1992)

Gard, R.: Buddhism. George Braziller Inc., New York (1961)

Holland, J.: Hidden order: how adaptation builds complexity. Helix Books, New York (1995)

Holland, J.: Emergence: from chaos to order. Addison-Wesley, Redwood City (1998)

Hunter, A., Besnard, P.: A logic-based theory of deductive arguments. Artificial Intelligence 128, 203–235 (2001)

Jaynes, E.T.: Information theory and statistical mechanics. The Physical Review 106(4), 620–630 (1957)

Kahneman, D., Tversky, A.: On the psychology of prediction. Psychol. Rev. 80, 237–251 (1973)

Khalil, I., Hill, C.: Systems biology for cancer. Curr. Opin. Oncol. 17, 44–48 (2005)

Kitano, H.: Biological robustness. Nat. Rev. Genet. 5, 826–837 (2004)

Koller, D., Pfeffer, A.: Object-oriented Bayesian networks. In: Geiger, D., Shenoy, P. (eds.) Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence, pp. 302–313. Morgan Kaufmann Publishers, San Francisco (1997)

Korb, K.B., Nicholson, A.E.: Bayesian artificial intelligence. Chapman and Hall / CRC Press, London (2003)

Krause, P., Ambler, S., Elvang-Goranssan, M., Fox, J.: A logic of argumentation for reasoning under uncertainty. Computational Intelligence 11, 113–131 (1995)

Laskey, K., Mahoney, S.: Network fragments: Representing knowledge for constructing probabilistic models. In: Geiger, D., Shenoy, P. (eds.) Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence, pp. 334–341. Morgan Kaufmann Publishers, San Francisco (1997)

Lupski, J., Stankiewicz, P.: Genomic disorders: The genomic basis of disease. Humana Press, Totowa (2006)

Mao, B., Wu, W., Davidson, G., Marhold, J., Li, M., Mechler, B., Delius, H., Hoppe, D., Stannek, P., Walter, C., Glinka, A., Niehrs, C.: Kremen proteins are Dickkopf receptors that regulate Wnt/beta-catenin signalling. Nature 417, 664–667 (2002)

McPherson, K., Steel, C., Dixon, J.: Breast cancer: Epidemiology, risk factors and genetics. BMJ 321, 624–628 (2000)

Michielsa, S., Koscielnya, S., Hill, C.: Prediction of cancer outcome with microarrays: a multiple random validation strategy. The Lancet 365(9458), 488–492 (2005)

Mitchell, S.: Biological complexity and integrative pluralism. Cambrige University Press, Cambridge (2003)

Nagl, S.: Objective Bayesian approaches to biological complexity in cancer. In: Williamson, J. (eds.) Proceedings of the Second Workshop on Combining Probability and Logic. (2005),
`www.kent.ac.uk/secl/philosophy/jw/2005/progic/`

Nagl, S.: A path to knowledge: from data to complex systems models of cancer. In: Nagl, S. (ed.) Cancer Bioinformatics, pp. 3–27. John Wiley & Sons, London (2006)

Nagl, S., Williams, M., El-Mehidi, N., Patkar, V., Williamson, J.: Objective Bayesian nets for integrating cancer knowledge: a systems biology approach. In: Rouso, J., Kaski, S., Ukkonen, E. (eds.) Proceedings of the Workshop on Probabilistic Modelling and Machine Learning in Structural and Systems Biology, Tuusula, June 17-18 2006, vol. B-2006-4, pp. 44–49. Helsinki University Printing House, Finland (2006)

Neapolitan, R.E.: Probabilistic reasoning in expert systems: theory and algorithms. Wiley, New York (1990)

Neapolitan, R.E.: Learning Bayesian networks. Pearson / Prentice Hall, Upper Saddle River (2003)

Nygren, P., Larsson, R.: Overview of the clinical efficacy of investigational anticancer drugs. Journal of Internal Medicine 253, 46–75 (2003)

Oyama, S.: The ontogeny of information: developmental systems and evolution, 2nd edn. Duke University Press, Durham (2000)

Parsons, S.: Order of magnitude reasoning and qualitative probability. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 11(3), 373–390 (2003)

Parsons, S.: On precise and correct qualitative probabilistic reasoning. International Journal of Approximate Reasoning 35, 111–135 (2004)

Prakken, H., Sartor, G.: Argument-based extended logic programming with defeasible priorities. In: Schobbens, P.-Y. (ed.) Working Notes of 3rd Model Age Workshop: Formal Models of Agents, Sesimbra, Portugal (1996)

Quinn, M., Allen, E.: Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. BMJ 311, 1391–1395 (1995)

Rasnick, D., Duesberg, P.: How aneuploidy affects metabolic control and causes cancer. Biochemical Journal 340, 621–630 (1999)

Ravdin, Siminoff, Davis.: A computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J. Clin. Oncol. 19, 980–991 (2001)

Reis-Filho, J., Simpson, P., Gale, T., Lakhan, S.: The molecular genetics of breast cancer: the contribution of comparative genomic hybridization. Pathol. Res. Pract. 201, 713–725 (2005)

Richards, M., Smith, I., Dixon, J.: Role of systemic treatment for primary operable breast cancer. BMJ 309, 1263–1366 (1994)

Ries, L., Eisner, M., Kosary, C., Hankey, B., Miller, B., Clegg, L., Mariotto, A., Feuer, E., Edwards, B.: SEER Cancer Statistics Review 1975-2001. National Cancer Institute (2004)

Russo, F., Williamson, J.: Interpreting probability in causal models for cancer. In: Russo, F., Williamson, J. (eds.) Causality and probability in the sciences. Texts in Philosophy, pp. 217–241. College Publications, London (2007)

Toyoda, T., Wada, A.: 'omic space': coordinate-based integration and analysis of genomic phenomic interactions. Bioinformatics 20, 1759–1765 (2004)

Veer, L., Paik, S., Hayes, D.: Gene expression profiling of breast cancer: a new tumor marker. J. Clin. Oncol. 23, 1631–1635 (2005)

Vogelstein, B., Kinzler, K.: Cancer genes and the pathways they control. Nature Medicine 10, 789–799 (2004)

Williams, M., Williamson, J.: Combining argumentation and Bayesian nets for breast cancer prognosis. Journal of Logic, Language and Information 15, 155–178 (2006)

Williamson, J.: Maximising entropy efficiently. Electronic Transactions in Artificial Intelligence Journal, 6 (2002), `www.etaij.org`

Williamson, J.: Bayesian nets and causality: philosophical and computational foundations. Oxford University Press, Oxford (2005a)

Williamson, J.: Objective Bayesian nets. In: Artemov, S., Barringer, H., d'Avila Garcez, A.S., Lamb, L.C., Woods, J. (eds.) We Will Show Them! Essays in Honour of Dov Gabbay, vol. 2, pp. 713–730. College Publications, London (2005b)

Williamson, J.: Causality. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, vol. 14, pp. 89–120. Springer, Heidelberg (2007a)

Williamson, J.: Motivating objective Bayesianism: from empirical constraints to objective probabilities. In: Harper, W.L., Wheeler, G.R. (eds.) Probability and Inference: Essays in Honour of Henry E. Kyburg Jr., pp. 151–179. College Publications, London (2007b)

Williamson, J., Gabbay, D.: Recursive causality in Bayesian networks and self-fibring networks. In: Gillies, D. (ed.) Laws and models in the sciences, pp. 173–221. With comments, pp. 223–245. King's College Publications, London (2005)

Xia, Y., Yu, H., Jansen, R., Seringhaus, M., Baxter, S., Greenbaum, D., Zhao, H., Gerstein, M.: Analyzing cellular biochemistry in terms of molecular networks. Annu. Rev. Biochem. 73, 1051–1087 (2004)