SIXTH FRAMEWORK PROGRAMME
PRIORITY FP6-2005-NEST-PATH
TACKLING COMPLEXITY IN SCIENCE

Proposal for:

Project full title: **Unravelling complex diseases with complexity theory: from networks to the bedside.**
Project acronym: **ComplexDis**
Date of preparation of Annex I: **14/02/2006**
Type of instrument: **STREP**
Proposal: 043241
Submission stage: FULL proposal
Activity code addressed: NEST-2005-Path-COM
Coordinator: Dr BENSON Mikael,
E-mail: Mikael.Benson@pediat.gu.se
Fax: +46 31 848952

Part. no.	Participant organisation name	Short name	Country
1 CO	Göteborg University	UGOT	Sweden
2	Rikshospitalet-Radiumhospitalet HF	Radium	Norway
3	Svabhegy National Institute of Paediatrics	SNIP	Hungary
4	Consiglio Nazionale delle Ricerche	CNR	Italy
5	University of Tennessee	UT	USA
6	Azienda Ospedaliera di Padova	AOPDIT	Italy
7	University of Navarra	UN	Spain
8	Vrije Universiteit Brussel	VUB	Belgium

Table of contents *Page*

Proposal summary

Unravelling complex diseases with complexity theory: from networks to the bedside.

ComplexDis

Activity code addressed: *NEST-2005-Path-Com*

Proposal abstract

Common diseases like allergy, autoimmunity and cancer are complex, i.e. caused by multiple interacting genes and environmental factors. Using high-throughput methods all genes and their products can be studied in common diseases. The challenge lies in understanding the complex gene and protein interactions that underlie these diseases. The Pathfinder project has shown that common principles, such as network theory can be used to understand widely different complex systems. Despite this there have been few applications in medical research. The hypothesis behind this project is that complexity theory can be used to unravel disease mechanisms and to develop predictive models of complex disease. The models will be validated experimentally and by solving a real-world clinical problem, i.e. to find biological markers for personalized medication. This involves the design, implementation and enhancement of complexity tools (e.g. network theory metrics) used to assess emergent properties from experimental studies of complex diseases, such as allergy, autoimmunity and cancer. Network theory plays a central role in our approach through the use of separators, connectivity, covers and cliques. Complex diseases are seen from a systems perspective and the focus is on the emergent properties of systems rather than individual components. This involves transfer of knowledge from complexity studies in other fields, such as language evolution. An important feature of the project is that it aims to transfer and develop techniques from complexity science to solve a real-world problem, i.e. to find markers for personalized medication in two different diseases, i.e. allergy and multiple sclerosis. We also want to disseminate analytical tools for such research. We believe that this can be achieved within a three-year period and used to promote the science of complexity in medical research. The long-term goals are to mimic and reverse pathogenic network changes in experimental models of complex disease. The project is a multi-disciplinary collaboration between clinicians, experts in complexity theory, bioinformaticians, computer scientists, molecular biologists, immunologists and geneticists in Europe and the USA. Thus, the project adheres to the objectives of the second NEST-PATHFINDER call, in that it is a cross-disciplinary effort to apply the science of complexity to a specific area where complexity is a key issue.

B 1. Project Objectives

This project is based on applying the science of complexity to a central problem; how to identify and functionally understand emergent, disease-associated gene interaction modules. We apply cutting-edge tools, many of which are novel and have not been previously applied in clinical research. For example, emergence of new disease-associated modules is analyzed with algorithms that were recently described by one of the applicants in a study of inbred mice (*Nature Genetics 2005; 37:233-42*). This is an example of transfer of complexity tools from one area to another. Another example is the application of the principles of language evolution to understand emergent properties in disease-associated modules. Together with other complexity tools this project is likely to contribute to solving a significant complex real-world problem, i.e. to find markers for personalized medication. This is a short-term goal. We also want to develop and share analytical tools for clinical research based on the science of complexity. The long-term goal is to manipulate specifically modules in order to control cellular behaviour and ultimately to reverse disease processes.

Our ongoing studies have led to the identification of theoretical and methodological *problems* that will be addressed in the project:

1. *Pleiotropic genes*

The same gene may be involved in different modules, and have different functional roles. For example, in one module a gene may be involved in cell proliferation and in another chemotaxis. In network models derived from standard clustering methods the same gene can only be part of one module.

Objective: To develop methods to form network models in which genes can be part of more than one module.

2. *Emergence of new network modules*

This is a central problem in this project and an example of how the science of complexity has direct clinical implications. Ideally, the expression of the same set of genes in a module would increase or decrease in disease compared to health. Clustering methods suggest that such modules exist. If this were true, diseases could be cured by simply blocking a key gene in a key module. Similarly, the same set of biomarkers could be used to follow a disease process. However, studies of other networks such as social systems or words in languages indicate that new modules emerge if external conditions change. Consider for example a group of friends going from high-school to university. The original friends will form new friendships and interact differently with each other.

In complex diseases high-throughput analyses show that hundreds of genes differ in expression, some of which are not even expressed in health. By inference, many new interactions and modules may be formed in disease. We are currently examining if new gene interactions and modules are formed (i.e. modules that contain genes and interactions that are not expressed in health). Preliminary data indicate that this is the case. This is highly relevant for clinical and experimental research. It may, for example, not be possible to perform experimental validation studies by blocking a gene that has a key regulatory function in modules found in health, even though that gene appears relevant and changes greatly in expression in disease. Altered gene interactions and modules could also explain why drug targets often do not show expected effects.

Objective: To develop new methods to analyse the emergence of new network modules and to assess their role in complex disease networks.

3. *The multiple network layer problem*

Multiple layers of interacting networks, e.g. DNA-, RNA- and protein networks are involved in complex disease. These networks, in turn, operate at different levels in cells, tissues and organs. They also interact with environmental networks such as microbes and pollutants. Thus, the layer and scale problem is going to be addressed in our consortia by assessing the network properties at every scale and by studying their interaction in order to obtain a given phenotype such as the homeostasis or a pathological condition.

Objective: To develop new methods to make multi-layered network models.

4. *Network dynamics*

Mathematical or computational models based on discrete conditions give a simplified but representative description of the dynamical processes involved in complex disease like allergies or cancer. Studies of biological and socio-economic networks indicate that dynamic processes can be modelled and simulated. C-ImmSim[1] is an example of simulators of this kind. This computational model and its derivates have been shown to make testable predictions of the dynamics of cells and molecules involved in complex diseases like cancer[2, 3], allergy[4] or AIDS[5] , both qualitatively and quantitatively. The versatility of computer models is the true power of such approach: fast and cheap *in-silico* experiments. Obviously, mathematics needs approximations and simplifications and therefore the results need to be interpreted with extreme caution, but the discovery of general principles and mechanisms is not completely out of reach.

Objective: To apply modelling methods to understand network dynamics in complex disease.

5. *Genetic heterogeneity of complex disease*

Genome-wide linkage studies show that different chromosomal loci may be involved in different populations and patients that appear to have the same disease. This indicates that there may be disease subtypes, in which different gene combinations cause a similar phenotype. In the context of this project, this means that there may be different disease-associated modules in patients with the same diagnosis, which complicates identification of such modules. On the other hand, addressing this problem is directly related to one of the specific goals, i.e. to find markers for personalized medication.

Objective: To find methods to identify disease subtypes.

<u>The above problems and objectives will be addressed as follows:</u>

Combinatorial algorithms to process high-throughput biological data to form network models that describe pleiotropic genes and emerging modules

We rely on our longstanding work in graph theory for the design, implementation and enhancement of combinatorial algorithms used to process high-throughput biological data. Tools born of our research in fixed-parameter tractability are used to solve huge *NP*-complete problems in order to elucidate subgraph structures of putative biological relevance to the study of allergy. Cluster, supercomputing and other high-performance computing platforms will be used to extract cliques and related dense gene sets suggestive of co-regulation, and to perform genomic data mining to highlight the most promising genesets for detailed study. A primary goal is the discovery of genetic regulatory pathways and networks. We build upon our novel algorithms as recently described (*Nature Genetics 2005; 37:233-42).* Important advantages of these algorithms are that pleiotropic genes can be included in different modules and that emergence of disease-associated modules with new interactions can be described.

Novelty: The clique-based approach to describe pleiotropic genes and emergence of modules was first described by us[6].

Competing research: Disease-associated modules have recently been reported in cancer (reviewed *in Nat Genet. 2005;37 Suppl:S38-45*).

Co-citation literature networks to form and functionally annotate network modules

PubGene is a literature co-citation network that provides a gene- or protein interaction network template based on automated mining of 16 million publications in Medline[7]. PubGene can be used to organize genes identified in DNA microarray experiments in networks and disease-associated modules. These can be functionally annotated, for example for known-disease association. An advantage of PubGene is that modules are not restricted by canonical gene interaction pathways that have been described in healthy cells. It may therefore be particularly suitable to functionally annotate emerging modules with new gene interactions. Another advantage is that PubGene is combined with gene- and protein databases in the public domain. Thus, PubGene also provides multi-layer network information. This can be used to interactively annotate modules or module components (see below).

Novelty: Functional annotation of emerging disease-associated gene interaction modules by combining clique-based algorithms and co-citation networks.

Competing research: Identification of inflammatory modules and pathways based on a manually curated gene interaction network was recently described in healthy cells[8].

Using principles of language evolution to understand the emergence of disease-associated gene interactions

Language is another complex system that has been extensively studied since the pioneering work of Chomsky. In language several interacting networks contribute to language generation such as co-occurrence, semantic and syntactic networks[9]. The complexity generated by language evolution also mimic the molecular language developed by biology[10]. Thus, we hypothesize that tools developed in the study of language can be applied to understand gene and protein semantics.

Novelty and competing research: Language evolution principles have not been previously been applied to modules in complex disease.

Multi-layered network models

Because of current technological constraints we use the transcriptomal (RNA) network as a detailed template to find modules and key regulatory transcripts. The corresponding genes are proteins are then analysed and used as projections of the main features of DNA and protein network layers.

These analyses are mainly based on experimental models of complex disease. This work involves application of PubGene to organize genes into networks. For example, analysis of disease-associated transcriptomal module may indicate the presence of an individual transcript with putative key regulatory functions. The corresponding gene can be searched for conserved regions that may contain regulatory elements or known polymorphisms. Similarly, the corresponding protein can be analysed for variants and functional annotation. It is of note that the multi-layer model may be used to interactively refine the different layers. For example, if a transcriptomal module encodes a putative key interaction, this can be examined on the protein level using public databases; are the proteins expressed in the same cell and sub-cellular location? Do they physically interact? If the answer is yes this supports the putative transcriptomal interaction.

Novelty and competing research: To our knowledge multi-layered network modules based on a genome-wide transcriptomal template have not been previously described.

Modelling and simulations to capture and mimic network changes

The computer model C-ImmSim and its derivation to simulate allergic reactions (i.e. type-I hypersensitivity) represent the perfect model candidates to perform the kind of analysis described above. In fact, knock out genes can be easily simulated at the functional level by inhibiting a certain cellular feature (e.g., cytokine production or receptor expression). By carefully designing a set of computer experiments, and by correctly applying the necessary approximations, we can verify the validity of our hypothesis on the existence of gene modules, promiscuous genes and their relative importance in the emergence and progress of the disease.

The computation model belongs to the class of “agent-based” models also called discrete event simulations, individual based simulations or micro-simulations. In very simple words this model is not constructed on ordinary or partial differential equations like the majority of mathematical models, but on the discrete representation of individual entities (e.g., cells and molecules in this case) that have their own internal representation. Entities of the same type obey the same rules, but since these rules are based on the information contained in the entity's internal representation, the history of each individual entity is unique. This kind of approach combines the power of statistical mechanical methods of modern physics with the versatility of such a detailed description, and is made possible by the availability of powerful computers.

Novelty: The novelty of this approach consists in using a detailed simulation tool to systematically study the effects of knock-out genes and to link the results to the molecular level by means of graph theory.

Competing research: A similar approach, but applied to the study of infectious diseases, is the Pathsim project (http://staff.vbi.vt.edu/pathsim/) recently used to study EBV infection. Another is the IMMSIM++ software (http://www.cs.princeton.edu/immsim/software.html) although its use is more educational.

Twin studies to compensate for genetic heterogeneity of complex disease

Genetic and environmental heterogeneity may result in different gene interaction networks despite a similar clinical phenotype. In the context of this project, this may confound experimental studies on cells from patients with complex disease. On the other hand, addressing this problem is directly related to one of the specific goals, i.e. to find markers for personalized medication. Ideally, we need clinical material that is easy to obtain, can be used for both experimental and clinical studies and can be divided into homogenous subtypes.

Allergic rhinitis (common hay fever) may meet these criteria. To identify different subtypes we study monozygous twins concordant for allergic rhinitis in different European populations

(Italy and Sweden). The rationales are that different populations are more likely to have different genetic subtypes. 60 % of monozygous twins are concordant for hay fever. Thus, if the twins are concordant for either disease or health this is more likely to represent a homogenous genetic subtype that is representative of the population that they belong to. In previous studies of inbred mice we and others have found that variations in gene expression is partially genetically determined[6, 11]. We therefore hypothesize that concordant twins will have similar transcriptomal modules due to similar genetic polymorphisms. Thus, we aim to identify disease-associated transcriptomal modules and their corresponding genetic and protein network markers. We will use those markers to classify subtypes to ensure that experimental studies are performed on homogenous materials.

Novelty: To our knowledge identification of genetically determined, disease-associated transcriptomal modules based on twin-studies is a new concept.

Competing research: Disease-associated genetically determined variation in gene expression has recently been described in a mouse model of obesity[11]. Network-based identification of single nucleotide polymorphisms has been performed in animal models of complex disease (*Nature Genetics 2005; 37: 413-19*).

Finding markers for personalized medication

The subtype specific markers described above will be tested as markers for treatment response. Allergic rhinitis is commonly treated with intranasal cortisone spray. The effects vary considerably in different patients. We will study the clinical effects of intranasal cortisone on the twins described above and relate these effects to changes in subtype markers. We hypothesize that we will be able to predict treatment response based on those markers. These markers can then be tested in larger studies of other allergic diseases (but this is beyond the scope of this application).

Multiple Sclerosis is treated with immunomodulatory drugs such as interferon beta. However, there is a significant percentage of non-responder patients (up to 40%). Thus we need to find new therapies that improve the control of such diseases. In a preliminary study, we have analyze the effect of interferon beta therapy in the transcriptome network of patients with MS and we identify several pathways not covered by such therapy. Thus, using network analysis we were able to identify new therapeutic targets that will not be identify by using classical molecular approaches (Palacios 2005)

Novelty: To our knowledge identification of markers for personalized medication and drug targets based on network theory is a novel approach.

Competing research: There are large-scale efforts to find markers for personalized medication based on systems biology in the USA[12], but these have not yielded any published clinical studies. In breast cancer gene expression signatures are tried to personalize medication[13].

B 2. Relevance to the objectives of NEST

The objectives of the NEST-Pathfinder call are to promote successful tackling of specific but important, complex real-world problems and the transfer of knowledge between disciplines. This project meets these objects by applying different complexity tools to a central problem; how to identify and functionally understand emergent, disease-associated modules. We apply cutting-edge tools, many of which are novel and have not been previously applied in clinical research. For example, emergence of new disease-associated modules are analyzed with clique-based methods that were recently described by one of the applicants in a study of inbred mice[6]. This is an example of transfer of complexity tools from one area to another. Another example is the application of the principles of language evolution to understand emergent properties in disease-associated modules. Together with other complexity tools this project is likely to contribute to solving a significant complex real-world problem, i.e. to find markers for personalized medication. Since there are few applications of complexity science in complex diseases, successful tackling of this problem may lead to increased interest in such applications. This project will contribute to this by publications, presentations at conferences, educational efforts, interactions with academic and industrial leaders, as well as collaborations with pharmaceutical and bioinformatics companies on the SME level. One such collaboration has already led to a patented drug target. We will also develop standardized complexity tools for clinical research via the web and EU sponsored projects such as Exystence. Together, these efforts may lead to this project serving as a "beacon" for other clinical researchers. Our long-time goal is to mimic and reverse disease processes in an experimental model of complex disease. This may lead to identification of new therapeutic principles. It is possible that this project will contribute to a significant increase of complexity projects in clinical research and that both will interactively benefit from this.

B 3. Potential Impact

Since most common diseases are complex, it is likely that the science of complexity will contribute significantly to understanding of their pathogenesis. This has been projected to reach the clinic in the next 5-10 year period, to predict and prevent disease as well as to personalize medication[12]. Personalized medication is seen as the "low-hanging fruit" in this context, since there are high-throughput methods to find candidate biomarkers and well-defined clinical phenotypes in many complex diseases. Personalized medication is likely to significantly improve health care and also to save the cost of pre-scribing sub-optimal treatment. In the case of allergic disease, which affects some 30-40 % of the population in the EU, this may contribute both to health and lowering of pharmaceutical costs. Another likely result of this project is the identification of new drug targets. Collaborations between clinicians and complexity scientists have already resulted in identification of a patented drug target in multiple sclerosis (see workpackage 8). Many complex diseases have reversibility mechanisms. For example, in allergy low-dose administration of allergen may cure the disease. Such reversibility mechanisms have not been previously addressed with complexity tools. A long-term goal of this project is to study and mimic these mechanisms. This could have significant impact on treatment of complex diseases. The participants of this effort represent a cross-disciplinary mixture of scientists from Europe and the USA. Since their expertise is not available in any single location in Europe, collaboration at the European level is imperative. The scientists in the project will promote the application of complexity to science via publications, presentations at conferences, educational efforts and collaborations with pharmaceutical companies. The project will also benefit from interactions with other EU sponsored projects, such as *Exystence, UniNet* and *ImmunoGrid*. Taken together, these efforts may have significant impact on health care in Europe, both on the academic and clinical levels.

B 4. The consortium and project resources

The consortium is composed of scientists in complexity and related fields on one hand and of clinician researchers on the other. In the description below their contributions are estimated in workmonths, but we have not tried to estimate the value of other resources. It is, for example, difficult to put a price tag on supercomputers at the Oak Ridge National Laboratory or facilities at the Artificial Intelligence Laboratory in Brussels.

The first project is based on applying clique-based methods to find modules that were first described by professor Michael Langton's group (*Nature Genetics 2005; 37:233-42).* These are unique in that they can be used to analyze pleiotropic genes and emergence of modules with new gene interactions:

Michael A. Langston, Professor of Computer Science, University of Tennessee, USA

Professor Langston received the PhD in Computer Science from Texas A&M University in 1981. At the University of Tennessee he leads a team of students, post doctoral fellows and research associates whose work is focused on efficient algorithm design, analysis and high performance implementations, with a special emphasis on applications to computational biology. He also serves as Visiting Scientist at Oak Ridge National Laboratory, where he consults in the Computer Science and Mathematics Division, the Life Sciences Division, the Chemical Sciences Division, the Joint Institute for Computational Science and the Computational Biology Institute. He is currently in the process of developing portals through which the community at large may access his team's computational tools. His work in developing ClustalXP is a well-known example. Professor Langston has authored over 180 refereed publications, including journals such as *Nature Genetics* and is perhaps best known for his long-standing work on combinatorial algorithms, complexity theory and design paradigms for sequential and parallel computation. In addition to maintaining his research program, he regularly teaches courses on algorithmic analysis, bioinformatics, discrete optimization, graph theory and related subjects. His research has been funded by the National Science Foundation, the Department of Defense, the Department of Energy, the National Institutes of Health, and a variety of other agencies. He has received numerous awards, most recently the Distinguished Service Prize from the Association for Computing Machinery Special Interest Group on Algorithms and Computation Theory.

Contribution to the project: 18 workmonths from himself and a post doc. Applications in this project involve cluster, supercomputing and other high-performance computing platforms at Oak Ridge National Laboratory.

Requested contribution from the European Commission: None

Functional annotation of emergent modules is done by professor Eivind Hovig and associates using the PubGene co-citation literature network (*Nature Genetics* 2001;28:21-8). An advantage of PubGene is that modules are not restricted by canonical gene interaction pathways that have been described in healthy cells. It may therefore be particularly suitable to functionally annotate emerging modules with new gene interactions. Another advantage is that PubGene also provides multi-layer network information.

Eivind Hovig, professor at Department of Informatics at the University of Oslo, Norway, holds positions as a group leader at the Institute for Cancer Research, and as section head at the Department of medical informatics at The Norwegian Radium Hospital. Professor Hovig received his PhD in molecular genetics in 1992. He leads two research groups of students, post doctoral fellows and research associates whose work is focused on development and implementation of high-throughput techniques for genomics and clinical bioinformatics. His has extensive teaching experience at the University of Oslo the and currently supervises three masters students, five PhD students and two post docs. He has authored some 70 articles in journals like *Nature Genetics* and *Lancet*, and received awards for scientific excellence as well as inventor prices. He is the leader of the FUGE functional genomics platform of bioinformatics in the Oslo region, member of the board of the national FUGE bioinformatics steering group, member of the FUGE national microarray steering group. He is chief scientific officer of the Norwegian based bioinformatics company, PubGene Inc., based on a patent application of Hovig, Jenssen et al. and also engaged in the formation of a company co-founded by Ideas ASA/GammaMedica and the Research Foundation of the Norwegian Radium Hospital, called Biomolex AS. This company is based on a patent application of Hovig, Skretting et al. He also serves as scientific adviser the Norwegian Lab-on-a-Chip company NorChip. Coholder of four US patents one Norwegian patent, and two patent pending applications.

Resource contribution to the project: 18 workmonths from himself and associates corresponding from the Norwegian Research Council Functional Genomics Funding. Laboratory and bioinformatics facilities at the Institute for Cancer Research *Resources requested from the European Commission:* 36 workmonths for a post doc, 2500 euro for a notebook for the post doc, 55 000 euro for DNA microarrays and other consumables, as well as 12000 travelling.

The next project involves application of principles of language evolution to understand the emergence of disease-associated gene interactions. This is done with two unique approaches by Luc Steels, professor of Computer Science at the Vrije Universiteit Brussels, and Ricard V. Solé, professor at the ICREA-Complex Systems Lab in Barcelona.

Luc Steels, professor of Computer Science at the Vrije Universiteit Brussel (VUB). He graduated in linguistics at the University of Antwerp and in computer science at the Massachusetts Institute of Technology, working in the MIT AI Laboratory. After that he worked in the domain of geophysical measurement interpretation as a project leader for geological expert systems at Schlumberger. In 1983 he founded the VUB Artificial Intelligence Laboratory, which he still directs this day. The Artificial Intelligence Lab of the Vrije Universiteit Brussel (VUB) will be the primary group involved in this project. It has a research tradition in investigating cognitive science issues using artificial intelligence, artificial life and computational modelling techniques. More than 20 ph.d's have graduated from the laboratory and many quite large projects have been executed in the past for the European Commision starting from the first framework program in 1982 until recently. Since five years the major research topics have centred on the origins and evolution of language, comprising themes such as the categorization of perception, meaning creation, communication, grammaticalisation and imitation. Currently several projects are going on financed by the Fund for Scientific Research of Flanders (FWO) and the Industrial Fund of Flanders (IWT). He was cofounder and chairman (from 1990 until 1995) of the VUB Computer Science Department (Faculty of Sciences). And also directs the Sony Computer Science Laboratories in Paris. His scientific research interests cover the whole field of artificial intelligence, including natural language, vision, robot behavior, learning, cognitive architecture, and knowledge representation. His publications are in top AI/cogscie journals such as *Behavioral and Brain Sciences, Trends in Cognitive Science, AI journal*, etc. He also edited a dozen books. At the moment his research focus in on fundamental research into the origins of language and meaning. Steels has played a significant role in the field of artificial life and organised several workshops on genome/language analogies

Financial contribution to the project: 18 workmonths from himself and associates.

Resources requested from the European Commission: 36 workmonths for a post doc, 2500 euro for a notebook for the post doc, 12000 euro for travelling and 12147 euro for consumables and additional costs.

Ricard V. Solé, professor, ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Barcelona and external Professor of the ***Santa Fe Institute*** (New Mexico, USA), senior member of the NASA-associate Center of Astrobiology ***(CAB)*** in Madrid, and member of the Council of the ***European Complex Systems Society.*** He is member of the editorial board of several international journals, including ***Advances in Complex Systems***, ***FRACTALS***, ***Ecological Complexity*** and ***Marine Ecology***. He completed a five-year degree in Physics and another 5-year degree in Biology at the University of Barcelona and received the PhD in Physics in 1991. He has supervised 12 Ph D theses and is PI of three EU projects and one american, NIH project. His main research interests go around understanding the possible presence of universal patterns of organization in complex systems, from prebiotic replicators to evolved artificial objects as well as the relevance of selection, tinkering and emergence in shaping complex networks. He has published more than 140 scientific papers and book chapters on complex systems and has written three books in the area (including Signs of Life, Basic Books, New York 2001, in collaboration with Brian Goodwin). His contributions have been featured in diverse publications, from *The New York Times, New Scientist, Nature* or *Science* to several popular and technical books. His most recent research projects involve cellular networks in development and cancer as well as language webs. Most of his recent research will be collected in a forthcoming book, Evolving Webs, to be published by Princeton U. Press within its Series on Complex Systems. His role in this project is analysis of global and local network properties in disease networks. His contribution is development of theoretical and modeling tools for the study of network diseases. Using input from both existing literature on neurodegenerative diseases and data obtained through the project, mathematical and computer models of complex networks involved in MS and AD will be constructed. A general purpose software for representing, analysing and modeling these networks will be constructed. Suitability for allocated work: The Complex Sytems Lab located at the Universitat Pompeu Fabra is recognized as one of the leading world groups in the world exploring diverse aspects of complex systems (see our website ***http://complex.upf.es***). Its members have been trained in interdisciplinary approaches to cmplexity and include a wide array of scientific perspectives, including Biology, bioinformatics, molecular biology, statistical physics, computer science and nonlinear dynamics.

Resources requested from the European Commission: Shares funding with main partner of workpackage 8, Dr Pablo Villoslada

Modelling and simulations are needed to understand the dynamics of emergent disease-associated modules. The aims of the next project are to mimic and reverse disease processes in *in silico* and experimental models of complex disease.

Filippo Castiglione, senior researcher at the Institute for Computing Applications (Rome) of the National Research Council. He received his PhD in Scientific Computing at the Univeristy of Cologne (Germany). He was a postdoc at the Institute for Medicine and Bio-Mathematics (Israel). FC currently holds Its research interests range from modelling of biological phenomena, simulation of the immune system and clinical applications, scientific and high-performance computing. He has published more than 30 scientific articles. FC has been holding visiting/research positions at the University of Bielefeld (2001) and Harvard Medical School (2001-02). FC's group at IAC has been modeling biological systems since 1998, thus gaining significant experience in the development of mathematical and computational models of biological phenomena based on stochastic cellular automata. The group collaborates with the Harvard Medical School for the study of cell signalling; the Institute for Medicine and Bio-mathematics (Israel) for the study of hypersensitive reactions; researchers and MDs of the Italian Institute for Infectious Disease "L. Spallanzani" in Rome

for the study of the HIV-1 infection; the Policlinico Monteluce in Perugia (Italy) and Dipartimento di Patologia Sperimentale all'Università di Bologna (Italy) for the study of cancer vaccines and immunotherapies; the Molecular Genetics Group of the University of Rome "Tor Vergata" for the study of protein-protein interaction networks.

Financial contributions: 18 workmonths from himself and associates

Resources requested from the European Commission: 36 workmonths for a post doc, 2500 euro for a note book for the post doc, 12000 euro for travelling and 12904 euro for consumables.

The above complexity tools are applied to clinical materials consisting of peripheral blood mononuclear cells (PBMC) from patients with allergy and multiple sclerosis. These two diseases may depend on altered function of the same cell type, the T helper cell. Moreover, both diseases have inducible reversibility mechanisms that are relevant for the long term goal, i.e. to mimic and reverse disease mechanisms. These materials are obtained by a team of clinician researchers with unique experience of applying complexity tools:

Mikael Benson, senior researcher and consultant, Queen Silvia Children's Hospital, Göteborg, Sweden. He is a senior consultant in pediatric allergology and received his PhD based on a thesis about allergic inflammation at the department of pediatrics. He has had full-time research grants from the Universities of Lund and Göteborg as well as the Swedish Research Council since 2000 and has currently received a six-year grant as a senior researcher from the Swedish Research Council. The aims are to develop predictive models of allergic inflammation based on combining complexity theory with high-throughput techniques, such as DNA microarrays. He started using these techniques in 1999, and was the first to do so in allergology. This work is based on co-ordinating a national and international network of scientists and has resulted in some 30 publications in top allergy journals like *J Allergy Clin Immunology* and the formation of a multi-disciplinary team that is involved in this project. He is a frequently invited speaker at national and international conferences. He has extensive teaching experience of medical and research students, and currently supervises two PhD students and one post doc. He is presently engaged in founding a Centre for Clinical and Genomic Systems Biology at Göteborg University, together with a fellow senior researcher who is a geneticist. This involves the Bio-X-Med research centre in Göteborg that has cutting-edge technology for high-throughput studies in genomics and proteomics. He is chairman of an interest group in functional genomics of the European Academy of Allergy and Immunology, which has 160 members in Europe and USA. His role in this project is that of co-ordinator and also responsible for workpackage 5.

Financial contributions: 24 workmonths from the Swedish Research Council and mobilization of technical resources from the Bio-X-Med research centre.

Resources requested from the European Commission: 36 workmonths for a post doc, 7.5 months for project co-ordination, 2500 euro for a note book for the post doc, 60 000 euro for DNA microarrays and other consumables, and 12 000 for travelling.

Professor Antonella Muraro, consultant Paediatric Allergist and full time member of the academic staff, of the Allergy Unit of the Department of Paediatrics of the University of Padua, in particular in the field of paediatric food allergy, rhinitis and sinusitis as well as on prevention of allergic diseases in childhood. She has been professor of Pediatric Allergology at the Allergy and Clinical Immunology School, University of Padua since 1990. In 1992 she achieved a PhD at the University of Rome. Since 2001 she has been member of the Board of the Section on Pediatrics of the European Academy of Allergology and Clinical Immunology (EAACI) and since 2003 she has been the scientific secretary of the EAACI – Section on Pediatrics Board. Since 2004 she has been member of the the Italian Society of Pediatric Allergy and Immunology. She is also Member of American College of Allergy, Italian Society of Pediatrics, Italian Society for Pediatric Respiratory diseases. She is the chairwoman of the EAACI task-force on anaphylaxis in children and she is the author of many publications on food allergy and atopic dermatitis in top allergy journals like *J Allergy and Clin Immuno*l and *Allergy.* She is responsible for workpackage 6.

Resources requested from the European Commission: 36 workmonths for a post doc, 2500 euro for a notebook for the post doc, 51 489 euro for DNA microarrays and other consumables, and 12000 travelling.

Dr. Lajos A. Réthy, senior scientific consultant and chief physician, Svábhegy" National Institute of Paediatrics, Budapest, Hungary. Dr Rethy received his PhD at 2001 Medical School of the University of Pécs, Hungary. Thereafter he was awarded a three-year grant for post doctoral research. His research is focused on large-scale population-based genomic studies and development of novel methods for DNA extraction, in collaboration with the Karolinska Institute and the Queen Silvia Children's Hospital in Sweden. One such study dealt with polymorphism in an asthma related gene, GPRA, in 700 atopic patients. This has resulted in several publications in top allergy journals like *J Allergy and Clin Immunology* and *Allergy.* He has contributed to founding a national DNA bank and natione-wide screening programs. His work also includes gender-related differences in human immune-responses. Dr. Réthy is also a senior lecturer of basic and nutritional immunology at the Dept of Nutritional Sciences /Faculty of Health Sciences of the University of Vienna, Austria since 2003 and a lecturer of nutritional genomics there since 2005 (http://data.univie.ac.at/pers?zuname=rethy).

He is also active participant of the functional genomics interest group of the European Academy of Allergology and Clinical Immunology as well as other national and international scientific societies. He is thereby in a unique position to develop and promote complexity science in medical research in central Eastern Europe. He is responsible for WP7.

Financial contributions: Mobilization of clinical, diagnostic and laboratory resources from National Institute for Pediatrics and affiliated pediatric centers in Budapest.

Resources requested from the European Commission: 36 workmonths for a post doc, 2500 euro for a notebook for the post doc, 55 776 euro for DNA microarrays and SNP analyses, and 12 000 euro for travelling.

Pablo Villoslada, director of the Multiple Sclerosis Center at the Department of Neurology and Neuroimmunology Lab (UNAV), Consultant Neurologist of the Department of Neurology, Clinica Universitaria de Navarra and Assistant Professor, University of Navarra. He received his PhD in neuroscience at the Autonomous University of Barcelona (1996). He has some 30 international publications on neuroimmunology, immunology, genetics, system biology and clinical aspects of Multiple Sclerosis (MS).

Invited speaker of national and international meetings and workshops. Member of the Spanish Neurological Society (SEN), American Academy of Neurology (AAN), International Society of Neuroimmunology (ISNI) and International Society of Computational Biology (ISCB). The University of Navarra is an experienced centre in biomedical research using animal models, cellular and molecular immunology and neurobiology. Dr Villoslada pioneered applying complexity science to neurology and is active in a bioinformatic and systems biology group composed of 3 professors of mathematics and 4 PhD students with expertise in Bayesian reverse engineering, mathematical modelling, network analysis and system biology studies. All equipment required to develop the research program is currently operative in his institute. His work is based on animal models and in human vitro models of multiple sclerosis gene and protein networks and phenotypic analysis. Important features are analysis of network topology and dynamics in MS and developing a new *in* vitro system models to identify new diagnostic markers and therapeutic targets. This work has already resulted in a patented drug target in MS. He has extensive national and international collaborations. An important part of this collaboration is that with professor Solé of the *ICREA-Complex Systems Lab* in Barcelona. He participated (2001-2005) in a NIH grant as investigator and has 3 grants from the Spanish Health and Science ministries (PI).

Resources requested from the European Commission: 36 workmonths for a post doc, 2500 euro for a notebook for the post doc, 49990 euro for DNA microarrays and other consumables, and 12 000 euro for travelling.

Other countries:

Professor Langston of the USA participates in this application because he has unique qualifications as described above.

B 5. Project management and exploitation/dissemination plans

The project management will be undertaken at the University of Göteborg, with the assistance of a part-time project manager. A **management committee** consisting of one representative elected from each of the six institutions will meet every 6 months in the context of a ComplexDis workshop, to which the European Commission’s Project Officer will be invited. These meetings will provide an opportunity to review milestones and outputs, to revise the project plan to meet changes in circumstances if necessary, to respond to new initiatives and identify emerging projects and proposals. Before the start of the project, each work-package will be reviewed at a meeting of all the investigators who will be contributing to that work-package. They will be asked to prepare a detailed project plan (including milestones and deliverables) for their Research Assistants (RA) and Research Students (RS). This will form the reference document against which to monitor the progress of the project. The progress of each RA and RS will be the responsibility of their home organisation. A **steering committee** will be formed that will meet annually to review progress and suggest further collaborations, sources of funding and potential applications. This committee will consist of senior academics and the representatives of regional and national public bodies, from within the participating organisations and beyond. This steering committee may also be called for guidance, should particular difficulties arise. The mechanisms that will be in place to maintain and support the cross-disciplinary work and international collaboration are:

- Research students will be encouraged to arrange and hold their own project workshops.
- Every member of the team will publish in journals outside their core discipline.

• Web cams will be used to link investigators in different departments and institutions quickly for short meetings, to avoid delays caused by scheduling, and to augment regular face-to-face meetings.

B 6. Implementation plan

B 6.1. General descriptions and milestones

This project revolves around a central concept, i.e. that disease-associated modules can be found in gene expression networks. These modules are identified and analysed using different computational and knowledge-based bioinformatic methods. We also transfer methods from studies of language evolution to functionally analyze the modules.

These methods are applied on data derived from DNA microarray analysis of cells from patients with allergy and multiple sclerosis. DNA microarray analysis is a technique that allows simultaneous analysis of expression of all human genes. The validity of the disease-associated modules and genes is tested experimentally, but also in clinical studies. The aim is to demonstrate that the science of complexity can be used for an important real-world problem. We have already used complexity tools to identify a patented drug target in multiple sclerosis. In this project we focus on an attainable short-term goal. Emergent modules are analysed to find markers for personalized medication in allergy and multiple sclerosis. The long-term aim of this application is to develop tools to understand reversibility mechanisms in complex diseases. Allergy and multiple sclerosis may be particularly suited for understand reversibility mechanisms in that they are both reversible and caused by different subsets of the same type of lymphocyte, the T helper cell. These two subsets have opposing functions, and it is possible that they may have opposing effects in the two diseases. In other words, the T helper subset that reverses multiple sclerosis may cause allergic disease, and vice versa. To our knowledge, cross-disciplinary research between neurologists and allergists to exploit this polarity has not been previously been performed. This project involves several other examples of transfer of knowledge between different disciplines, such as complexity science, bioinformatics, experimental biomedicine and clinician research.

Many of the complexity tools in this project are applications or new developments of cutting-edge work described by the applicants. It can be summarized as follows:

WP 1 applies clique-based methods to identify disease-associated modules and genes. Important advantages of these algorithms are that pleiotropic genes can be included in different modules and that emergence of disease-associated modules with new interactions can be described[6].

WP 2 is based on the literature co-citation network PubGene that provides a gene- or protein interaction network template based on automated mining of 16 million publications in Medline[7]. PubGene can be used to organize genes identified in DNA microarray experiments in networks and disease-associated multi-layer modules.

WP 3 is another example of the cross-disciplinary nature of this project. It is based on the hypothesis that the complexity generated by language evolution also mimics the molecular language developed by biology. Thus, we can apply tools developed in the study of language to understand gene and protein semantics in emergent disease-associated modules[9]

WP 4 aims to understand and mimic the dynamics of emergence of disease-associated modules through modelling and simulations. The effects of key regulatory genes can be simulated at the functional level by inhibiting a certain cellular feature. By carefully designing a set of computer experiments, and by correctly applying the necessary approximations, we can verify the validity of our hypothesis on the existence of gene modules, pleiotropic genes and their relative importance in the emergence and progress of the disease[2]. This project will lay the foundation for the long-term aim of the project, i.e. to mimic reversibility mechanisms in complex diseases.

WP 5-7 apply complexity tools on pan-European clinical data derived from DNA microarray 0studies of cells from patients with allergic rhinitis. The aims are to implement and interactively develop the methods in workpackages 1-4 to identify emergent, disease-associated modules and genes with key regulatory functions. These genes are examined on the DNA, RNA and protein levels for multi-layer network models. The genes will also be further examined experimentally as well as biomarkers to personalize medication.

WP 8 is based on collaboration between clinicians and complexity scientists aiming to find new therapeutic targets and biomarkers in multiple sclerosis. This has already resulted in a patented drug therapeutic target that is being developed by a SME.

The projects are described in detail below:

Workpackage 1: Combinatorial algorithms to process high-throughput biological data to form network models that describe pleiotropic genes and emerging modules

The tools of molecular biology and the evolving tools of genomics can now be exploited to study the genetic regulatory mechanisms that control cellular responses to a wide variety of stimuli. These responses are highly complex, and involve many genes and gene products. To increase our understanding of these responses, we will develop novel graph algorithms to generate highly distilled gene sets, produce scalable implementations for cutting-edge high performance computing platforms and use these implementations to extract disease-associated modules and genes, as well as *cis*-regulatory regions for analysis of single nucleotide polymorphisms. This is based on our innovative mathematical tools for gene network analysis that we recently reported in *Nature Genetics*[6, 14, 15]. Since the project revolves around module identification, WP1 has a key role.

Previous approaches. A wealth of clustering approaches has previously been proposed [16-20]. The most common are either hierarchical, in which all genes begin in their own clusters and are eventually merged into one, or centroid, in which genes are organized into a predefined number of clusters by iterative adjustments based on similarity [21]. The clusters these methods produce are typically disjoint, which places an artificial limitation on the biology under study in that many genes play important roles in multiple but distinct pathways [22]. There are recent clustering techniques, for example those employing factor analysis [23], that do not require exclusive cluster membership for single genes. Unfortunately, these tend to produce biologically uninterpretable factors without the incorporation of prior biological information [24]. Relevance networks [22, 25, 26] have been proposed as a means to overcome the limitations of traditional clustering methods. Without an algorithmic means to extract the aggregate relationships between multiple genes, however, the most interesting relationships (those with tight connections between multiple genes) remain embedded within the vast sea of correlations.

Benefits of a graph theoretical approach. It is therefore necessary to develop more powerful tools to extract subsets of coordinately regulated genes from large aggregates of gene expression data. Graph theory offers unique advantages to this problem. Our novel graph algorithms are based on decades of basic research, and constitute a class of tools that can help elucidate relationships in highly complex data structures, in our case as matrices of correlations across thousands of genes. We employ fixed parameter tractability (FPT), whose roots can be traced at least as far back as the work of Fellows and Langston to show that a variety of parameterized problems are tractable when the relevant input parameter is fixed [27, 28]. (A problem is FPT if it has an algorithm that runs in $O(f(k)n^c)$ time, where n is the problem size, k is the input parameter, and c is a constant independent of both n and k.)

The challenge, once the graph is created, is not to study the graph in its entirety but rather to extract its embedded subgraphs, or small, tightly connected regions of the graph that represent subsets of genes with strong correlations between every pair of its members and thus likely to represent biologically significant interactions. In the most extreme case, in which a subgraph contains all possible edges between its vertices, this structure is called a *clique*. Each and every pair of vertices is joined by an edge, from which we can infer some form of co-regulation among the corresponding genes. Clique is *NP*-complete, and widely known for its application in a variety of combinatorial settings, a great number of which are relevant to computational molecular biology [29]. It is particularly useful in microarray analysis, because it addresses the previously-noted shortcomings of traditional clustering algorithms. Cliques need not be disjoint. A vertex can reside in more than one clique, just as a gene can operate in more than one regulatory network. In terms of gene expression, clique represents the most trusted potential for identifying a set of interacting genes [22]. Clique is so difficult, however, that guaranteeing solutions even to within only n^{ε} cannot be accomplished within polynomial time for any $\varepsilon>0$ unless $P=NP$ [30]. In fact clique is not even FPT unless the *W* hierarchy collapses [31]. (The *W* hierarchy, whose lowest level is FPT, can be viewed as a fixed-parameter analog of the polynomial hierarchy, whose lowest level is *P*.) Such a collapse is widely viewed as an exceedingly unlikely event, roughly on a par with the likelihood of the collapse of the polynomial hierarchy [32]. Thus we focus instead on clique's complementary dual, the *vertex cover* problem. Like clique, vertex cover is *NP*-complete. Unlike clique, however, vertex cover is FPT. We search for a minimum vertex cover in a graph, thereby finding the desired maximum clique in the graph's complement. Currently, the fastest known vertex cover algorithm runs in $O(1.2759^k k^{1.5}+kn)$ time [33]. Contrast this with the $O(n^k)$ time needed in a straightforward approach. The requisite exponential growth (assuming $P\neq NP$) is therefore reduced to a mere *additive* term. Our recent work on this subject is described in [34-40]

Workpackage 2: Co-citation literature networks to form and functionally annotate network modules

Cellular entities interact in modular units or functional components of the cells complex machinery. Depending on the biological process, these modules function in isolation or in a network of relationships with each other, The aim of this workpackage is to extract, catalogue and identify modules from the literature all cellular components and events that compose molecular networks of relevance. The network relationships of these entities will be characterized from the text by applying computational approaches that capture the semantic relationships that govern their network topology in healthy and diseases cellular systems and knowledge extraction from the literature that reflects the modular organization of these cellular components. These modules will be used as templates to functionally annotate emergent modules identified using clique-based methods.

Gene literature networks have high levels of connectivity (data from PubGene) due to high levels of co-citation of cellular components in more than 16 million biomedical articles. Latent in the complexity of these literature networks are various types of cellular directed networks (including protein–protein interaction, metabolic, signaling and transcription-regulatory networks), as well as specific networks of disease and normality. In this project PubGene will develop systems to allow for the functional annotation of high-throughput experimentation from the other partners. Furthermore novel algorithms will be developed that discover putative critical hubs of gene nodes and capture the modular organization of cellular components that is also latent in the vast deluge of medical text in the Public domain.

PubGene has unique capabilities in fulfilling these objectives. The group was established in 2001 by researchers who pioneered the field in biomedical text mining (Jenssen et al. *Nature Gen.* 2001). The PubGene solution explores information contained in millions of articles and citation records and structures this information into an organized knowledge representation. The PubGene algorithms and solutions are presently utilized in a highly reputable text mining product in the bioinformatics community, known as PubGene 2.5**™.** In this product, the Medline database of citation records is the main textual source for mining information about biomedical entities and concepts such as genes, proteins, mutations, diseases, and keywords from structured vocabularies.

Workpackage 3: Using principles of language evolution to understand the emergence of disease-associated gene interactions

The aims are to examine if phenomena such as lexicalization and grammaticalisation can be observed in a pre-defined set of genes that are expressed in health and disease.

From a systems perspective, gene interaction changes in disease may be analogous to word interaction changes during language evolution. It is well known that words are polysemous. They have on average 30 different meanings that are 'expressed' in specific contexts, similar to the way a gene may be associated with a new meaning in disease because it is expressed with other genes than in health. The language context is either established through the communicative situation or through associations between the words which are partially activated due to priming effects. Moreover studies of language change and dialogue have shown that language is not a static system but constantly changing. The meaning of a word is not fixed forever but the subject of constant negotiation and change for the purposes of dialogues. New grammatical constructions arise in a bottom up manner or they get stretched in order to deal with new circumstances. The re-arrangement and adjustments are typically the result of miscommunications, which clearly trigger many additional cognitive processes that are not active in normal routine language processing. Once a new innovation has emerged, it may spread in the population through cultural transmission. We call the processes by which words and grammatical constructions arise, change, and spread in a population semiotic dynamics. Although these language forming processes are far from understood, there is a lot of empirical data on language change which has suggested a number of basic operators that are responsible for maintaining the adaptability and evolvability of language[41]. Second, there is a growing set of software simulations and robotic experiments in which these language formation processes are simulated. These experiments rely on complex mechanisms for representing hierarchical structure, matching and merging of structures to build new ones, analogical processing, reinforcement learning, a.o.[42]. Third, there is a lot of work recently to understand the processes by which the context-sensitivity of language can be achieved. Some of this work relies on grounding language communications in embodied interactions[43], whereas other work relies on the acquisition of large-scale associative networks based on latent semantic analysis[44] and its use during the on-line parsing and interpretation of language. Another important recent development that helps both the understanding of language evolution and disease-associated gene interactions is based on recent advances in complex systems, particularly in network analysis[45]. Whereas traditional graph theory analyses static networks, new approaches focus on the dynamics of networks and their interconnectivity properties. These tools have already lead to major results for studying language networks in populations of agents and they are being actively applied to genomics as shown in the work of another partner[46]. We therefore intend to apply these techniques in interaction with the other partner.

There is already a long tradition to find analogies between genes and language (starting with Jerne's 1984 Nobel lecture, The Generative Grammar of the Immune System) and more recently several workshops and research papers have attempted to find more concrete and realistic mappings[47]. This project is however a first in using the latest insights from gene-interactions and the latest results from language evolution studies to become very concrete and realistic. Rather than following a structuralist approach as in generative grammar which takes a static view on language, we will explore formalisms that have come out of viewing language as a complex adaptive system. Specifically, the aim of this project is to examine if phenomena such as lexicalization and grammaticalisation can be observed in a pre-defined set of genes that are expressed in health and disease. We will focus on "hub" genes with many interactions. These genes and their nearest neighbours will be examined in the most significant modules. Lexicalization and grammaticalisation will be examined in relation to well-defined phenotypic changes. The hypotheses are that we can observe that in certain gene interaction contexts a study gene has specific meanings. For example, in combination with genes A and B the study gene causes a cell to proliferate, but in combination with C and D the gene causes the cell to die. These effects will be reversed by experimental blockade of the gene. By focusing on a small number of disease-associated genes and modules the aim is to lie to grounds for a "disease grammar".

Workpackage 4: Modelling and simulations to capture and mimic network changes

The aim of this workpackage is to implement a computer code to simulate the phenomena of the emergence of allergic rhinitis. This will be made on the base of the C-ImmSim simulator and in particular of its customised version used to simulate type-I sensitivity reactions[4]. The computational model will represent cells and molecules as unities and for this reason is said to be working at the extracellular level. However, since we are interested in the effects of gene expressions, this might lead to the incorrect interpretation that the results produced by the simulator will not be relevant to our research. This turns out to be not true, indeed. The model has already shown to be well suited to simulate the “effects” of a certain change in the function of the cellular and molecular entities and therefore we can simulate knock-out scenarios by simply modelling the effect of under/over expression of certain genes (taken alone or in combination). For the sake of clarity, it should be said that whenever the causal relation between gene expression and relative effect is unknown and cannot even be guessed, this approach is unfruitful. However, in the majority of cases one can at least make a guess and this tool can cheaply calculate the consequences in terms of cellular dynamics, disease progression, etc.

The development of the agent-based microsimulation model for allergic rhinitis will consists of the following stages.

1. Analysis of the phenomena of the emergence of allergic rhinitis by the identification of the cellular and molecular entities that are likely to be involved. This will be done by strictly collaborating with our partner experts on the topic and by careful collecting information on recent literature.
2. Selection of the most important features in terms of “who are the actors” (i.e., the agents) of the phenomena.
3. Identification of the rules of interaction/cooperation among the cellular and molecular entities that we choose to represent in the computer model. These rules will be coded in computer language to be executed during the simulation to determine the agent's behaviour. The choice of the entities and of the rules governing their behaviour is the most important modelling stage.
4. Various other modelling decisions are required, as for example, the definition of a time step (e.g., second, minute, fraction of a day, ...), the definition of the simulation space (e.g., blood, lymphnode, ...). These choices are likely to influence the performance of the simulator in terms of CPU and memory requirements. Various optimization techniques might become necessary during the programming phase of the development of the simulator.
5. Coding and debugging the computer simulator. In this stage a careful analysis of the computational requirements of the simulator are in order. Although modern computers are equipped with extremely fast processors and very large memory space there is nothing available on the market to represent in a one to one correspondence the “numbers” of the immune system, not just in the number of cells but rather in the potential clonotypic repertoire, just to mention an example. To overcome these limitations, various abstractions and simplifications are needed. The coding goes hand in hand with debugging to find and solve the inevitable errors in computer programming and/or technical modelling.
6. Validation of the simulator. This part is very important and is done in two stages:
   - 6.1 Correctly reproduction of the system behaviour *qualitatively*. By observing the output (e.g., cell/molecular populations/concentrations) one expects to recognize the typical temporal patterns obtained by real laboratory experiments or in clinical practice.
   - 6.2 Second, a parameter tuning is necessary to make the output as much adherent to the reality as possible. This is a *quantitative* assessment of the fidelity of the simulator.

In order to perform the quantitative assessment of the ability of the simulator to reproduce the reality, experimental data is necessary. This data will be provided in the other workpackages.

Once the validation of the simulator is done one can proceed with different computer experiment to test or verify the validity of hypothesis. The computer model described in the previous section is very well suited to reproduce the experimental settings of knock-out animal models. In few words, it is extremely easy, provided a careful modelling design (items 2 and 3), to activate/deactivate a certain dynamical feature of a cell (e.g., the production of a certain cytokine). This will allow to observe the dynamics of the cellular and molecular concentrations in knock-out experiments and to eventually determine the network of relationships between cause (genes expressions) and effects (the macroscopic observables).

Workpackage 5: Implementation studies on experimental models of allergic disease

The short-term aim of this workpackage is to identify disease-associated modules and key regulatory genes using the methods described above. The long-term aims are to mimic and reverse disease processes in a cellular model of allergic disease.

Allergic rhinitis is a unique example of complex diseases in that the cause, i.e. allergen is known and can both induce and reverse the disease. In large doses during the season allergen causes disease, but in low doses given at intervals during allergen vaccination patients are cured. Moreover, the key regulatory cell, the T helper (Th) cell, is known and has a well-defined phenotype in the disease (the Th2 cell). The presence of this cell can be determined by flow cytometry or analysis of the cytokines in cell supernatants. *It is thus possible to study induction and reversal of the disease process in detail.*

There is increasing interest in using high-throughput methods in allergy research, but there are relatively few publications. We have used DNA microarrays since 1999 and this has resulted in several publications describing new genes and pathways, but also the realization that there is considerable individual variation, even in tightly controlled experimental settings[48-63].

Recently, we have performed a network-based meta-analysis of a large number of microarray experiments of cells and tissues from healthy controls and patients with allergy. This was as done as described in Workpackage 1 and showed the emergence of new gene interactions and modules in disease. We are currently analysing how these modules relate to phenotypic traits in allergen-stimulated T cells. We have also performed pilot studies to make multi-layer network models; we identified putative key genes whose relevance have been validated on the RNA and protein level in allergen-stimulated T cells. We are currently performing analyses of single nucleotide polymorphisms (SNP) in *cis*-regulatory regions of these genes in a large material of patients and controls. Thus, rather than performing genome-wide searches for SNPs we use a hypothesis-based approach, that in turn, is founded on network theory. To our knowledge no other groups have made a network-based analysis of allergic disease.

Our focus in this project is to study Th cells from patients and controls using DNA microarrays. Based on the results of deliverable 1 below and previously performed meta-analysis we will select 6.000 genes for a customized microarray that will be used in subsequent studies. This is done to reduce cost and allow studies of some 180 individuals from three European countries. The lymphocytes are challenged with allergen *in vitro.* Gene interaction networks are defined and the networks further analysed to find disease-associated modules of functionally related genes. This is done as described in workpackages 1-4. These modules are studied in various conditions, such as treatment with drugs or after allergen vaccination. The aim is to find modules that reliably predict and explain responses to the different conditions. We also want to find key regulatory genes in the modules, in other words genes that control the modules. As an example, a switch may be turned on during the pollen season, and turned off following medical treatment or allergen vaccination. We are particularly interested in the modules that cause disease following high dose allergen-stimulation (like in the pollen season) and reverse the disease following repeated low-dose stimulation (like during allergen vaccination). These modules may be relevant for the long-time goal, i.e. to mimic and reverse the disease processes in an experimental model of complex disease. The switches are found and validated using combinations of different bioinformatic methods described in workpackages 1-4. Manipulation of switches is done using blocking antibodies or RNA interference and the effects studied in relation to phenotypic changes of Th cells. Because of genetic heterogeneity the detailed studies of the modules depend on characterization of transcriptome disease subtypes in twin-studies in different European populations and their genetic correlates (see workpackages 6-7). This information will also be used for the clinical studies, i.e. to find markers to personalize medication.

Workpackages 6-7: Identification of transcriptome subtype modules and multi-layer network modules in allergic disease

The short-term aim of these workpackages is to find protein markers for response to treatment with cortisone in allergic rhinitis. The long-term aim is to find markers for other medications in other allergic diseases. Such markers may be useful to personalize medication. Thus, these workpackages aim to show a clinically useful application of the science of complexity in medical research. This entails development of a standardized analytical protocol that can be used by other researchers in the field. The material from these workpackages is also used for experimental studies in workpackage 5. Workpackage 6 that is performed in Italy and includes treatment with immunotherapy to find modules and genes that are associated with reversibility. Workpackage 7 is performed in Hungary and includes analysis of polymorphisms in key regulatory genes and how these relate to protein expression. This information is used for multi-layer network models.

Allergic rhinitis is commonly treated with intranasal cortisone spray. This treatment has variable effects. Previous DNA microarray studies by us and others indicate that cortisone alters expression of hundreds of genes and that there are considerable individual variations. Detailed studies of one gene showed that cortisone decreased its expression in some patients and not in others, and that this was linked to a polymorphism in the *cis*-regulatory region of the gene. This led to the hypothesis that there may be genetic variations in gene expression that explain variable effects of cortisone treatment in allergic rhinitis. In these studies we aim to identify protein markers that can be measured in nasal fluids from patients with allergic rhinitis that can be used to predict treatment response to cortisone.

To identify different subtypes of patients that may respond differently to cortisone we study monozygous twins concordant for allergic rhinitis in two different European populations (Italy and Sweden). The rationales are that different populations are more likely to have different genetic subtypes. 60 % of monozygous twins are concordant for hay fever, compared to less than 10 % in dizygous twins. Thus, if the twins are concordant for either disease or health this is more likely to represent a homogenous genetic subtype that is representative of the population that they belong to. We also hypothesize that such genetic subtypes will have distinct transcriptome (RNA) subtypes. These transcriptome subtypes are identified by DNA microarray analysis of allergen-challenged Th cells. Using methods described in WP1 and WP2 we identify the corresponding DNA polymorphisms and variation in protein levels. This information will be used for multi-layer network models and to identify proteinmarkers for those subtypes. We will then examine if the selected protein markers can be used to predict treatment response.

Workpackage 8: Network-based and dynamical analysis of Multiple Sclerosis pathogenesis and therapy

The short-term aim of this workpackage is to assess the topology and dynamics of the immune system network in multiple sclerosis (MS) and to identify new modules and pathways involved in the pathogenesis of the diseases using this analysis. The long-term aims are to improve our understanding of MS pathogenesis but more importantly to develop new biological markers to assess response to therapy and to identify new therapeutic targets. In order to obtain such objective we are going to apply several complexity tools such as network theory, non-linear dynamics and language theory. This subproject will benefit from the long-term collaboration between a neuroscientist and medical group with special interest in complexity in biology and cognitive sciences (University of Navarra, Spain) with a group with great expertise in studying complexity in biology and language (University Pompeu Fabra, Spain). Finally, we are going to compare our findings with the network analysis done in allergy by other members of the consortia. Current biological data supports the concept that Th1 diseases such as MS and Th2 diseases such as allergy are the opposite phenotypes of the same phenomena such as autoimmunity. Indeed, we will take advantage of the approaches and tools developed in the other WP of the project.

1. <u>Analysis of network topology in MS:</u>

Our aim is to identify the differences in the network topology of the immune system and brain transcriptome networks, and protein interaction (IP) networks in multiple sclerosis patients compared with healthy controls.

1.1. Transcriptome network reverse engineering: we will assess the transcriptome in PBMCs from MS patients and controls using DNA microarrays in a already collected prospective cohort (N=60) at different stages of the disease. Reverse engineering will be performed using a Bayesian network inference approach. The dataset will be curate using gene expression and IP information at Ingenuity Database. IP networks will be created using public databases and our transcriptome analysis.

1.2. Transcriptome network topology analysis: Network metrics will be analyzed using Pajek and Matlab software. New interactions identified in our analysis will be validated using IP databases, such as String (http://string.embl.de/). In addition we are going to address how robustness in the transcriptome network is affected by the disease and by the effect of immunomodulatory therapy. To this end, we are going to develop custom software that will be freely available in the web. In addition we are going to apply language theory tools to analyze our networks, including the problem of multi-layered networks and generation of new modules along disease evolution.

2. <u>Analysis of network dynamics from health to autoimmunity:</u>

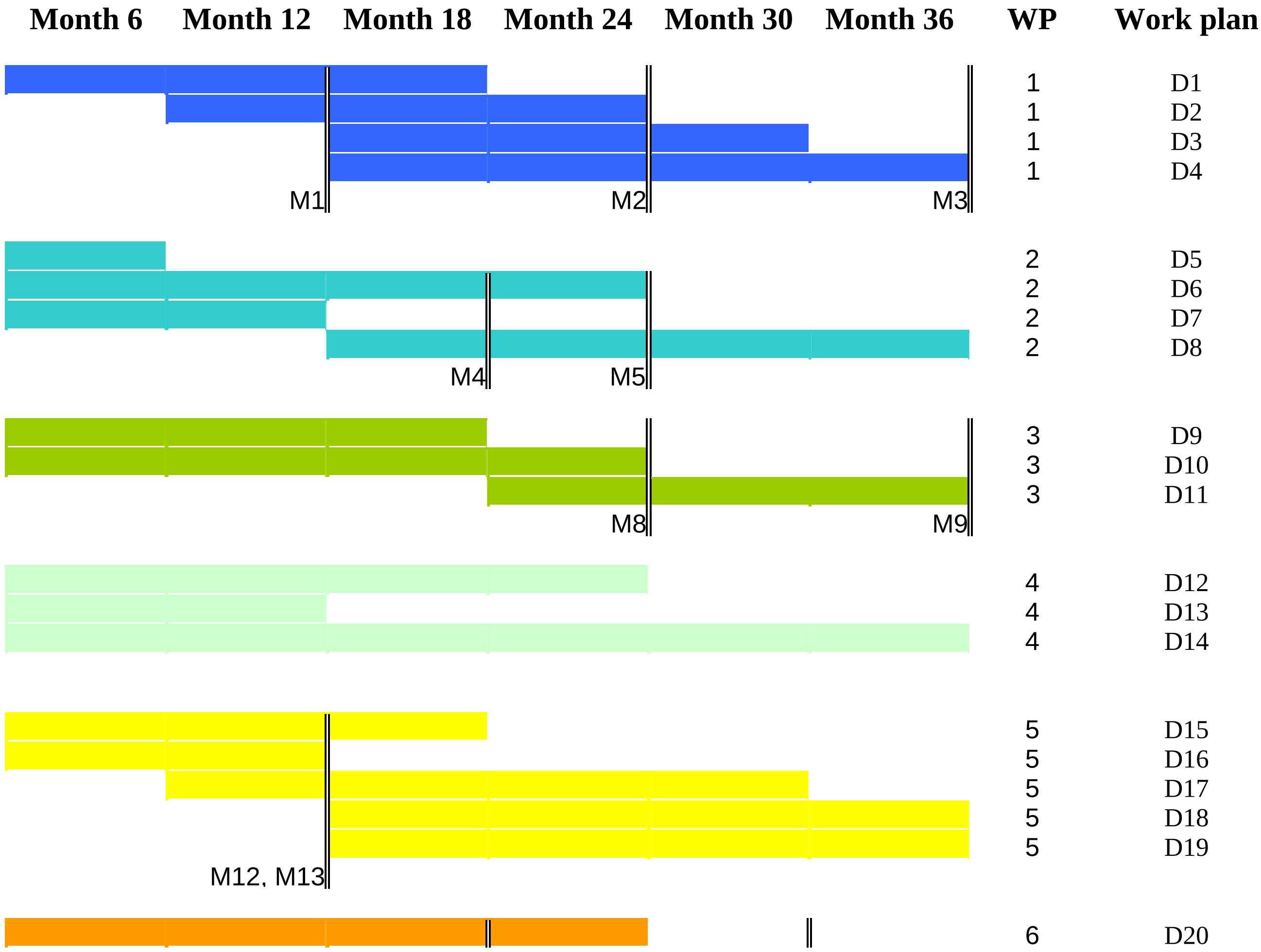
Our second aim is to assess the dynamics of MS transcriptome and IP networks during the development of the disease compared with healthy controls and in the response to immunomodulatory therapy. Our working hypothesis is that autoimmune disease are dynamical diseases in which long-term dynamics in the genetic and protein network that control the immune response evolve to a disease state or attractor. In order to assess the dynamics of the transcriptome network we are going to assess longitudinally the gene expression levels (real time PCR) of a set of genes from the core of the transcriptome network (between 50-200 genes) identified in the Aim 1 in our prospective collected samples. Dynamics of gene expression levels will be modelled with ordinary differential equations in order to model the different developmental stage of MS and the changes induced by immunomodulatory therapy. In order to assess the changes in the dynamics we will use standard statistics and metrics but new methods are need. In this sense we are going to use new tools generated in the analysis of language networks.

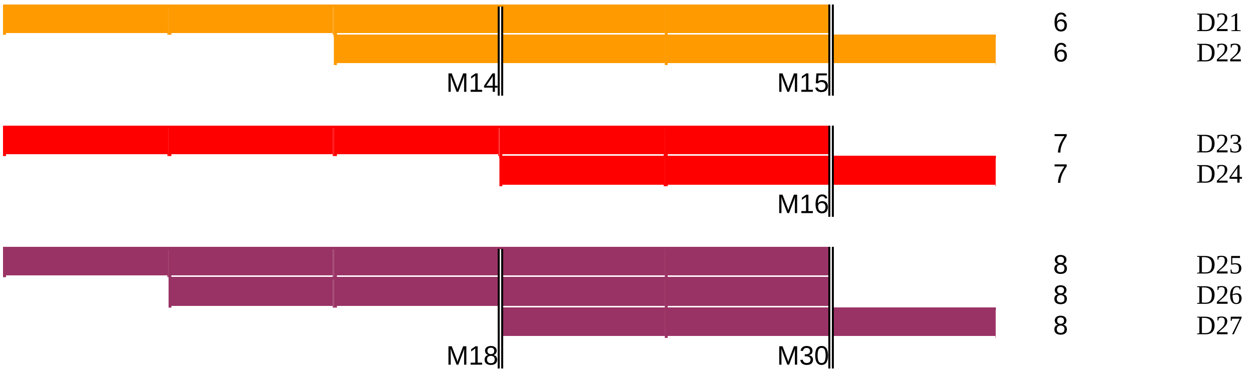
3. <u>Validation of the modules and pathways identified</u>:

The differences in transcriptome and IP networks identified in the topological and dynamical analysis between MS and healthy controls will be validated using *in vitro* experiments. Special focus will be dedicated to the network modules associated with response to immunomodulatory therapy or the identification of new pathways that can become a therapeutic target. In this proposal we aim to identify pathways that might suitable to target for therapy using network analysis. However, full molecular and clinical validation of therapeutic targets is out of our aims in the present proposal schedule for timing reasons, but we have planed to perform such studies after completing the present study. Our group has expertise in performing molecular and *in vivo* validation, and preclinical studies in order to move basic research to the bed side. These efforts also include protection of the intellectual property generated from our research and collaboration with biotech companies to complete such efforts. As an example, using a systems biology approach we have recently identified a small molecule that ameliorates the animal model of MS and we have submitted an international patent protection (PCT/ES2005/000139). We have started the preclinical studies for validating this compound for MS therapy in collaboration with a spin off of the University of Navarra: Digna Biotech (http://www.dignabiotech.com/eng/index.asp).

B 6.2. Planning and timetable

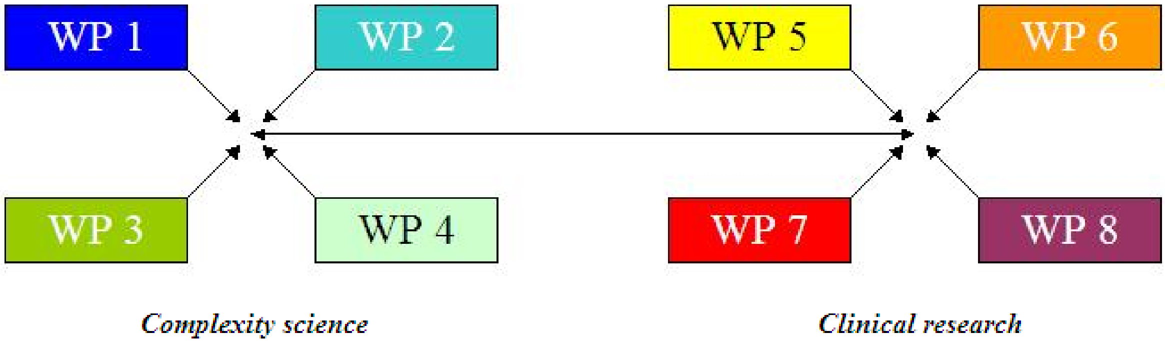
Details of the timing of the different phases of each work package are provided in the work package descriptions. The chart shown below gives a global overview of all deliverables that form the ComplexDis project.





B 6.3. Graphical presentation of work packages

The graph below shows the interactions between the different work packages. The work packages can be associated with two broadly defines categories: complexity science and clinical research.



B 6.4. Workpackage list

Work-package No	Workpackage title	Lead contractor No	Person-months	Start month	End month	Deliverable No
1	Combinatorial algorithms to process high-throughput biological data to form network models that describe pleiotropic genes and emerging modules		20	1	36	1 – 4
2	Co-citation literature networks to form and functionally annotate network modules		42	1	36	5 – 8
3	Using principles of language evolution to understand the emergence of disease-associated gene interactions		40	1	36	9 – 11
4	Modelling and simulations to capture and mimic network changes		40	1	36	12 – 14
5	Implementation studies on experimental models of allergic		48	1	36	15 – 19
6	Identification of transcriptome subtype modules in allergic disease		48	1	36	20 – 22
7	Identification of transcriptome multi-layer network modules in allergic disease		40			23-24
8	Network-based and dynamical analysis of Multiple Sclerosis pathogenesis and therapy		36	1	36	25 – 27
	TOTAL		314			

Only shows posts funded by this project

B 6.5. Deliverables list

WP no	Deliverable No	Deliverable title	Delivery date	Nature	Dissemination level
1	D1.	Annotated sets of putatively co-regulated allergy-specific genes	18	R	PU
1	D2.	Beta versions of graph analysis software	24	D	PP
1	D3.	User-ready graph analysis software package	30	D	PU
1	D4.	Comprehensive paper on computational tools	36	R	PU
2	D5.	Language evolution to understand disease-associated gene interactions	6	R	PU
2	D6.	Topology robustness of literature networks	24	R	PU
2	D7.	Software and database solution	12	D	PU
2	D8.	Publication on a study that validates the text-mining and NLP algorithms	36	R	PU
3	D9.	Vision paper on using recent work in language evolution	18	R	PU
3	D10.	Software that performs computer simulations	24	D	PU
3	D11.	Comprehensive overview of how simulation results can be analysed	36	R	PU
4	D12.	The simulator	24	D	PU
4	D13.	Technical reports describing the architectural issues	12	R	PU
4	D14.	Visualization tools and various statistical analysis utilities	36	D	PU
5	D15.	Description of emergence of new network modules	18	R	PU
5	D16.	Description of customized DNA microarray gene chip	12	R, D	PU
5	D17.	Detailed mechanistic description of subtypes	30	R	PU
5	D18.	Description and validation of key modules and genes	36	R	PU
5	D19.	Description of a multi-layer model of a network module	36	R	PU

6	D20.	Identification of markers for treatment response	24	R	PU
6	D21	Identification of markers for treatment respose in nasal fluids from twins	30	R	PU
6	D22.	Analysis of predictive value of markers for treatment respose in nasal fluids from twins	36	R	PU
7	D23.	Identification of combinations of disease-associated SNPs	24	R	PU
7	D24.	Multi-layer network models	36	R	PU
8	D25.	A set of biological pathways involved in MS pathogenesis	30	R	PP
8	D26.	New tools and software for assessing dynamic changes and robustness in networks	30	P	PU
8	D27.	A set of new biological markers and candidate therapeutic pathways	36	R	CO

R = report

P = Prototype

D = Demostrator

O = Other

PU = Public

PP = restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only members of the consortium (including the Commission Services)

B 6.6. STREP Project Effort Form

Full duration of project

(insert person-months for activities in which participants are involved)

Project acronym -

	UGOT	Radium	SNIP	CNR	UT	AOPDIT	UN	VUB	TOTAL PARTICIPANTS
Research/innovation activities									
WP1: Combinatorial algorithms	4		4		8	4			20
WP2: Co-citation	6	36							42
WP3: Language evolution	4							36	40
WP4: Modelling and simulations	4			36					40
WP5: Implementation studies	36	4		4	4				48
WP6: Subtype modules	4	4			4	36			48
WP7: Multi-layer modules	4		36						40
WP8: Analysis of MS							36		36
Total research/innovation	62	44	40	40	16	40	36	36	314
Consortium management activities									
Management	7,5								
TOTAL ACTIVITIES	69,5								

This table only shows posts requested in this grant application (and not the participant's contributions that have been funded by other sources such as NIH or national research councils. Such contributions are described in B4)

B 7. Other issues

B 7.1. Ethical issues

In summary this project is based on research performed on human biological samples, i.e. blood and nasal fluid samples. All clinical researchers in the project have long-standing experience of performing clinical studies and handling data in accordance with ethical guidelines and instructions from Ethics Committees. The project involves no risk for patients and controls. The potential benefits are increased knowledge of disease processes and finding markers for personalized medication. All research is performed after obtaining informed consent following written and oral information after approval from the ethical committees of the involved universities. No research is performed on stem cells, embryonic cells or animals. The project involves no cloning for reproductive purposes, experiments on humans or genetic manipulation. The use of the data will conform to the applicable national and international regulations and codes of conduct, including biobank laws. The data also has an expiration date, so that it will be impossible to retain the data after the end of the project. Since the project is dependent on this application, ethical applications have not yet been obtained, but will be sought following a positive outcome. An itemized description of ethical issues based on the guidelines for proposers is given below:

1. Potential ethical aspects of the proposed research regarding its objectives, the methodology and the possible implications of the results; Explain and justify the research design;

The objectives and methods of the project involve no risks for the patients. The only samples taken are blood and nasal fluid samples. This is done according to routine methods that at most involve some discomfort to the patients. Possible implications of the results are increased knowledge of disease processes and finding markers for personalized medication. The research designs of the projects are described in the workpackages. Briefly, blood samples are taken to obtain mononuclear cells. These are challenged with allergen *in vitro.* In this way the challenges can be standardized without risk to the subjects. The cells are analyzed with DNA microarrays. This is a method that allows simultaneous analysis of RNA expression of all human genes (please note that RNA and not DNA is analyzed). Using bioinformatic methods described in WP 1-4 disease-associated gene interactions modules and key transcripts are identified. In WP 7 blood samples are taken to seek gene polymorphisms that can explain altered expression of key transcripts. The combined information from these studies is used to find protein markers that are analyzed in nasal fluids from patients with allergic rhinitis. Nasal fluids are obtained in a standardized way that involves no risk, but some discomfort. This is justified by the possibility to identify markers for personalized medication.

2. Indicate the relevant national legislation or requirements of the Member State(s) where the research takes place.

The clinical parts of the project is performed in hospital settings by senior clinicians in strict accordance with local legislation and after approval of ethics committees.

3. Specification and justification of the type, amount and source of human biological samples to be used

In WP 5-8 15 mL of peripheral blood is taken/patient and in WP 5-7 5 mL of nasal fluids/patient. This in done using routine methods in clinical practice and research so as to allow standardization and no risk for the subjects. The samples are stored in freezers in the participating hospitals for the duration of the project. Samples are owned by the research institutions and can only be accessed by clinical and research staff involved in the projects. There is no commercial exploitation of the samples.

4. Description of the procedure for obtaining informed consent of the persons from whom the material is obtained

Informed consent is obtained after written and oral information to the subjects. Contact information to the investigators is given for questions or additional information. It is stressed that the subjects can decline participation in the studies at any time and that this involves no obligations and in no way affects treatment. The written information text can only be given after approval from the Ethics Committees and is therefore included in the applications to these committees.

5. Description of the arrangements for protecting the confidentiality of donors' personal data.

The donor’s personal data is stored in one computer/institute that can only be accessed by the principal investigator (PI) who is a clinician. The samples are coded. The code is linked to information about diagnosis, age and gender but not personal data. The samples are thus anonymized and handled according to national biobank legislation.

Ethical issues checklist

Does your proposed research raise sensitive ethical questions related to:	**YES**	**NO**
Human beings		X
Human biological samples	X	
Personal data (whether identified by name or not)	X	
Genetic information	X	
Animals		X

Please indicate whether the proposal involves	**Yes**	**No**	**Uncertain**
· **Research on human beings**			
Persons not able to give consent		X	
Children		X	

Adult healthy volunteers	X		
· **Human biological samples**			
Human foetal tissue/cells		X	
Human embryonic stem cells		X	
· **Human embryos**		X	
· **Human genetic information**	X		
· **Other personal data**			
Sensitive data about health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction		X	
· **Animals (any species)**			
Non- human primates		X	
Transgenic small laboratory animals		X	
Transgenic farm animals		X	

B 7.2. Gender issues

The gender dimension is integrated in this project both in terms of organization and research direction. On the organizational and management level, one out of nine established researchers in this project is female (Professor Antonella Muraro). She plays a significant role in the management and research direction of this project. However, where it does not contravene local legislation, we will actively promote advertisement of research assistant and research student posts so as to encourage applications from female researchers who have the equivalent qualifications as male applicants. All project participants, male and female, will be encouraged to take advantage of training and courses available within their environments to acquire additional career skills. However, in addition, we will promote participation in mentor projects directed towards the career advancement of female researchers. On the research level, it is of note that immunological responses have gender-related differences. This is already a research subject of one of the researchers in this project (Dr Lajos Rethy) and will be integrated into this project. This part of the project has direct clinical implications in that we aim to find markers for treatment response. There is limited knowledge about how gender relates to treatment response in allergic dimension, so this an important gender-related aspect.

B 7.3. Policy issues

To promote the impact of clinical applications of complexity science the results of this research will be disseminated widely to health care administrators and workers, as well as clinical and experimental researchers, industry, business and the public services. Standardized analytical programs, courses, workbooks, articles and seminars will be organized at national and international levels, in order to encourage exchange of experience an ideas by collaborators at all levels. Articles on television and radio will also be created through our network of contacts in the print and electronic media. Professional development courses, and undergraduate and graduate university courses will also be created and tested as part of the outreach of this course.

Workpackage description

Workpackage number	1	**Start date or starting event:**		Month 1			
Activity Type	RTD/Innovation						
Participant id	UT	UGOT	AOPDIT	SNIP			
Person-months per participant	8	4	4	4			

Objectives

To identify emergent disease-associated gene expression modules and key regulatory cells in DNA microarray studies of allergen-stimulated lymphocytes from allergic patients.

Description of work

Tools of molecular biology and the evolving tools of genomics can now be exploited to study the genetic regulatory mechanisms that control cellular responses to a wide variety of stimuli. These responses are highly complex, and involve many genes and gene products. To increase our understanding of these responses we will

1. develop novel graph algorithms to generate highly distilled gene sets,
2. produce scalable implementations for cutting-edge high performance computing platforms,
3. use these implementations to extract gene sets suggestive of co-regulation, and
4. performing genomic data mining to highlight the most promising gene sets for detailed scrutiny.

Our primary target is the elucidation of genetic regulatory mechanisms relevant to allergy. The driving motivation is that knowledge of these mechanisms will help clarify and interpret the physiological response to allergens, and advance our understanding of how allergic susceptibility is related to overall human health. We build upon our innovative mathematical tools for gene network analysis we recently reported in *Nature Genetics* [64-66].

Workpackage risk: Low.
Contingency plan: The project involves several alternative or complementary methods that can be applied in case one method should fail

Deliverables

D1. Annotated sets of putatively co-regulated allergy-specific genes. Paper on graph algorithms toolkits and applications (month 18)

D2. Beta versions of graph analysis software. Refined genesets. Paper on differential correlation and differential cliquification (month 24)

D3. User-ready graph analysis software package. Paracliques and other dense correlation subgraphs (month 30)

D4. Comprehensive paper on computational tools (month 36).

Milestones and expected results

M1: Evaluate ontological, statistical, and spectral thresholding methods required to convert gene correlation matrices to unweighted correlation graphs [67]. Perform differential analysis on cliques obtained from patient and control data (year 1)

M2: Implement and apply the paraclique algorithm [65] in order to handle noise inherent in microarray data. Investigate the feasibility of graph coarsening strategies (year 2)

M3: Refine and evaluate clique and cluster management tools. Apply differential clustering for multiple conditions and treatments (year 3)

Workpackage description

Workpackage number	2	**Start date or starting event:**	Month 1				
Activity Type	RTD/Innovation						
Participant id	Radium	UGOT					
Person-months per participant	36	6					

Objectives

To extract, catalogue and identify modules from the literature of all cellular components and events that compose molecular networks of relevance to complex disease. The network relationships of these entities will be characterized from the text by applying computational approaches that capture the semantic relationships that govern their network topology in healthy and diseases cellular systems. Methods will be developed to extract knowledge from the literature that reflects the modular organization of these cellular components. These modules will be used as templates to functionally annotate emergent modules identified using clique-based methods and used in the platform of high throughput data from the other partners.

Description of work

The first phase of PubGene's approach will be to represent complex cellular systems as literature networks with the text mining emphasis on complex diseases. Gene interaction networks will be extracted from the literature where connectivity is a measure of their co-citation weight to each other. These networks shall be further analyzed to find disease-associations in the modules of interest. The disease association that will be emphasized in this project shall be allergy and cancer.

This system will then allow for the functional annotation of gene networks from the literature, where connectivity will be developed as a measure of their functional relationship to each other. Given that there are currently over 16 million articles in the Medline database with the addition of over 60,000 articles per month, this effectively amounts to searching for a few nuggets of information in an overwhelming sea of background clutter. Automatic systems for feature extraction will be developed. We will also investigate the network robustness of the ensuing literature networks to identify putative critical hubs in experimental models of complex diseases. It has been well established that cellular components and function operate and interact with each other in a highly modular manner (Hartwell et al. Nature. 1999), and PubGene will develop text mining with natural language processing approaches that mines any modular organization from the literature. The organization of information extracted from the literature into its biological modular organization coupled with the identification of critical hubs will allow for the comparison of diseases and healthy literature association from experimental models using high throughput data. The system developed from this phase of the study will offer the capability of processing high-throughput biological data from the other partners to form network modules by creating superimposition networks from the literature.

Workpackage risk: Medium-high

Contingency plan: Should functional annotation of emergent modules fail or be complicated, alternative or complementary genomic and bioinformatic resources will be used.

Deliverables

D5. Software and database that is customized for the functional annotation of literature networks in complex diseases. Hubs of biological entities will be clustered from the literature, and classified by their relationships to the complex diseases of focus in the project. The tool will allow for the curation of the knowledge extracted, and create a database tuneable to annotate the putative modules from the various partners. Delivered month 6.

D6. Publication on the study of the topology robustness of literature networks and the analysis of such networks to leading to the identification of putative critical hubs responsible for the topology of the network of relationships in the complex diseases. Delivered month 24.

D7. Software and database solution that allows for the knowledge discovery and analysis of high-throughput biological data from the modular organization of cellular functions, as extracted from the literature. This modular organization of literature networks will be a novel application that complies with the needs of the other partners to annotate putative modules. Delivered month 12.

D8. Publication on a study that validates the text-mining and NLP algorithms to capture the modular organization of cellular components from the literature and the analysis of high-throughput biological data using the modular organization database. Delivered month 36.

Milestones and expected results

M4: Automatic systems for feature extraction of biological entities from the literature. A software and database solution shall result in the classification of network hubs, extracted from the literature, and responsible for the topology of complex disease networks (month 18).

M5: Algorithm development that investigates the network robustness of literature networks to identify putative critical hubs in experimental models of complex diseases. The resulting database will be tuneable for annotation by the partners (month 24).

M6: This phase of the project will concentrate on the information extraction (IE) of the modular organization of cellular functions in complex diseases from the vast deluge of the scientific literature. The resulting software and database will offer the capability of process high-throughput biological data to form network modules by creating superimposition networks and their modular organization from the literature. First iteration reached month 6, second iteration month X?

Workpackage description

Workpackage number	3	**Start date or starting event:**		Month 1			
Activity Type	RTD/Innovation						
Participant id	VUB	UGOT					
Person-months per participant	36	4					

Objectives

To examine if the principles and tools developed to study language evolution can be applied to understand gene and protein semantics in emergent disease-associated modules.

Description of work

We will examine if phenomena such as lexicalization and grammaticalisation can be observed in a pre-defined set of genes that are expressed in health and disease. We will focus on "hub" genes with many interactions. These genes and their nearest neighbours will be examined in the most significant modules. Lexicalization and grammaticalisation will be examined in relation to well-defined phenotypic changes. The hypotheses are that we can observe that in certain gene interaction contexts a study gene has specific meanings. For example, in combination with genes A and B the study gene causes a cell to proliferate, but in combination with C and D the gene causes the cell to die. These effects will be reversed by experimental blockade of the gene. By focusing on a small number of disease-associated genes and modules the aim is to lie to grounds for a "disease grammar"

Workpackage risk: Medium-high
Contingency plan: Should language evolution principles not be directly applicable to predict and understand emergent properties in disease-modules this will be of interest in itself, and a starting point for development of new analogous principles.

Deliverables

D9. Vision paper on using recent work in language evolution to understand disease-associated gene interactions. This paper will be targeted to a high profile journal (month 18)

D10. Software that performs computer simulations of disease-associated gene interactions based on adapting existing software for modeling the emergence of new meanings and context-based selection of meaning (month 24)

D11. Comprehensive overview of how simulation results can be analysed using the tools of semiotic dynamics, which includes network analysis (month 36)

Milestones and expected results

M7. Investigation of analogy between disease-related gene interactions and language evolution.

The goal of this work theme is to establish a detailed and workable analogy between disease-related gene interaction changes and language evolution. This will be done through prior exchanges of the key literature between language experts and geneticists, an in-depth workshop, and a joint paper that map out the analogy. Our goal is to publicise as much as possible this interdisciplinary dialogue so that as many researchers as possible can participate.

M8. Year 2. Case study
We will focus on a case study that is both based in biology and realistic in terms of complexity and relevance and based in computational modeling of language evolution. The case study will draw directly on work of the other partners. This case study will require a substantial effort on the biological side to identify the case study and collect and map out the necessary data. On the computational side, highly labour-intensive work will be required to implement the processes involved in the emergence and adaptation of context-sensitive gene expression by analogy with the flexibility in word usage and grammatical constructions. This computational work will rely on a battery of software tools that have already been developed in language evolution research, particularly on 'Fluid Construction Grammars'.

M9. Year 3. Analysis
There is no doubt that the behaviour emerging from simulations and experiments will be hard to follow unless we develop adequate tools for tracking the network structure and the properties of how these networks change. Moreover we need to develop a theoretical analysis, inspired by recent tools coming from complex systems research. Work during the last year will focus on the analysis part using the computer simulations from year 2.

Workpackage description

<table>
<tr><td><b>Workpackage number</b></td><td>4</td><td colspan="4"><b>Start date or starting event:</b></td><td colspan="2">Month 1</td></tr>
<tr><td><b>Activity Type</b></td><td colspan="7">RTD/Innovation</td></tr>
<tr><td><b>Participant id</b></td><td>CNR</td><td>UGOT</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td><b>Person-months per participant</b></td><td>36</td><td>4</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Objectives

To implement a computer code to simulate the phenomena of the emergence of allergic rhinitis.

Description of work

The development of the agent-based microsimulation model for allergic rhinitis will consist of the following stages.

1. Analysis of the phenomena of the emergence of allergic rhinitis by the identification of the cellular and molecular entities that are likely to be involved. This is done by strictly collaborating with our partner experts on the topic and by careful collecting information on recent literature.
2. Selection of the most important features in terms of “who are the actors” (i.e., the agents) of the phenomena and identification of the rules of interaction/cooperation among the cellular and molecular entities that we choose to represent in the computer model. The choice of the entities and of the rules governing their behaviour is the most important modelling stage.
3. Various other modelling decisions are required, as for example, the definition of a time step (e.g., second, minute, fraction of a day, ...), the definition of the simulation space (e.g., blood, lymph node, ...). These choices are likely to influence the performance of the simulator in terms of CPU and memory requirements. Various optimization techniques might become necessary during the programming phase of the development of the simulator.
4. Coding and debugging the computer simulator. In this stage a careful analysis of the computational requirements of the simulator are in order. The coding goes hand in hand with debugging to find and solve the inevitable errors in computer programming and/or technical modelling.
5. Validation of the simulator. This part is very important and is done in two stages: (i) Correctly reproduction of the system behaviour *qualitatively*. By observing the output (e.g., cell/molecular populations/concentrations) one expects to recognize the typical temporal patterns obtained by real laboratory experiments or in clinical practice; (ii) A parameter tuning is necessary to make the output as much adherent to the reality as possible. This is a *quantitative* assessment of the fidelity of the simulator. In order to perform the quantitative assessment of the ability of the simulator to reproduce the reality, experimental data is necessary. This data will be provided in the other workpackages.
6. Once the validation of the simulator is done one can proceed with different computer experiment to test or verify the validity of hypothesis. The computer model described in the previous section is very well suited to reproduce the experimental settings of knock-out animal models. In few words, it is extremely easy, provided a careful modelling design (item 2 and 3), to activate/deactivate a certain dynamical feature of a cell (e.g., the production of a certain cytokine). This allows to observe the dynamics of the cellular and molecular concentrations in knock-out experiments and to eventually determine the network of relationships between causes (genes expressions) and effects (the macroscopic observables).

Workpackage risk: Low

Contingency plan: The main risk is related to the difficulty of interpreting the results of the numerical experiment. For this aim a set of statistical tools need to be developed.

Deliverables

D12. The simulator itself is the major deliverable in this project (month 24).

D13. Technical reports describing the architectural issues (month 12).

D14. Visualization tools and various statistical analysis utilities. Results in term of scientific achievements could likely be reached within year three. (month 36)

Milestones and expected results

M10: In terms of the time required to accomplish each of the phases described above we can roughly assume that the accomplishment of phases 1 to 4 can be reached in year 1, phase 5 in year 2, and phase 6 in year 3.

Workpackage description

Workpackage number	5	**Start date or starting event:**		Month 1			
Activity Type	RTD/Innovation						
Participant id	UGOT	UT	Radium	CNR			
Person-months per participant	36	4	4	4			

Objectives

To identify emergent disease-associated gene expression modules and key regulatory cells in DNA microarray studies of allergen-stimulated lymphocytes from allergic patients.

Description of work

Allergen-stimulated Th cells from patients with allergic rhinitis and controls will be analysed using DNA microarrays. Gene interaction networks are defined and the networks further analysed to find disease-associated modules of functionally related genes. This is done as described in workpackages 1-4. These modules are studied in various conditions, such as treatment with drugs or after allergen vaccination. The aim is to find modules that reliably predict and explain responses to the different conditions. We also want to find key regulatory genes in the modules, in other words genes that control the modules. As an example, a switch may be turned on during the pollen season, and turned off following medical treatment or allergen vaccination. We are particularly interested in the modules that cause disease following high dose allergen-stimulation (like in the pollen season) and reverse the disease following repeated low-dose stimulation (like during allergen vaccination). These modules may be relevant for the long-time goal, i.e. to mimic and reverse the disease processes in an experimental model of complex disease. The switches are found and validated using combinations of different bioinformatic methods as described in workpackages 1-4. Manipulation of switches is done using blocking antibodies or RNA interference and the effects studied in relation to phenotypic changes of Th cells. Because of genetic heterogeneity the detailed studies of the modules depend on characterization of disease subtypes in twin-studies in different European populations (see workpackages 6-7). This information will also be used for the clinical study, i.e. to find markers to personalize medication.

Risk: Medium-high
Contingency plans: The risks in this project are mainly external factors such as patient compliance or if the pollen season should be unstable. To compensate for this, extra patients will be recruited and allergen-provocation studies be included.

Deliverables

A. Identification of modules and key regulatory genes

D15. Description of emergence of new network modules in allergen-stimulated CD4 + T cells from unselected patients with allergic rhinitis- journal publication and workshop (month 18)

D16. Description of customized DNA microarray gene chip for allergy research- journal publication, description in the public domain (month 12).

D17. Detailed mechanistic description of subtypes of transcripton modules that may be specific for different European populations. Identification of protein markers for these modules- journal publication and workshop. The protein markers will also be used for clinical studies in workpackages 6-7 (month 30)

D18. Description and validation of key modules and genes in a specific subtype studied under different conditions-journal publication and workshop (month 36).

B. Development of multi-layer model of a network module

D19. Description of a multi-layer model of a network module that encompasses the DNA, mRNA and protein level (month 36)

Risk: Medium-high

Contingency plans: The risks in this project are mainly external factors such as patient compliance or if the pollen season should be unstable. To compensate for this, extra patients will be recruited and allergen-provocation studies be included.

Milestones and expected results

M11. Identification of modules and key regulatory genes

M12. Analysis of emergence of new network modules based on DNA microarray analysis of allergen-challenged T lymphocytes from unselected patients with allergic rhinitis (months 1-12).

M13. Selection of genes for a customized 6,000 gene DNA microarray based on study 1 and previously performed meta-analysis (months 1-12)

Workpackage description

Workpackage number	6	**Start date or starting event:**		Month 1		
Activity Type	RTD/Innovation					
Participant id	AOPDIT	UGOT	UT	Radium		
Person-months per participant	36	4	4	4		

Objectives

To identify transcripton subtype modules in allergic rhinitis and to find subtype-specific markers for response to treatment with cortisone.

Description of work

To identify different subtypes of patients that may respond differently to cortisone we study monozygous twins concordant for allergic rhinitis in different European populations (Italy and Sweden). The rationales are that different populations are more likely to have different genetic subtypes. 60 % of monozygous twins are concordant for hay fever, compared to less than 10 % in dizygous twins. Thus, if the twins are concordant for either disease or health this is more likely to represent a homogenous genetic subtype that is representative of the population that they belong to. We will identify protein markers for those subtypes and examine if the markers can be used to predict treatment response:

Workplan:

Year 1: Monozygous twins are identified using simple questions that have shown 95 % correlation with genetic tests in previous twin studies. Blood samples are obtained out of season from ten twin pairs with allergic rhinitis and ten healthy twin pairs. Peripheral blood mononuclear cells (PBMC) are challenged with allergen and either treated or not treated with cortisone. The PBMC are analysed with DNA microarrays to find markers for treatment response.

Year 2: The protein markers are analysed in nasal fluids during the pollen season, before and after treatment with cortisone. After the season the twins are started on immunotherapy

Year 3: a) Studies of nasal fluid proteins in unselected patients with allergic rhinitis during the pollen season, before and after treatment with cortisone to examine if the markers can predict treatment response.

b)Studies of nasal fluid proteins in unselected patients with allergic rhinitis during the pollen season, before and after treatment with cortisone to examine if the markers can predict treatment response.

Risk: Medium-high

Contingency plans: The risks in this project are mainly external factors such as patient compliance or if the pollen season should be unstable. To compensate for this, extra patients will be recruited and allergen-provocation studies be included.

Deliverables

D20. Identification of markers for treatment response in allergen-challenged Th cells – journal publication

D21. Identification of markers for treatment response in nasal fluids from twins with allergic rhinitis – journal publication (month 30)

D22. Analysis of predictive value of markers for treatment response in nasal fluids from twins with allergic rhinitis – journal publication (month 36)

Milestones and expected results

M14. Identification of markers for treatment response in allergen-challenged Th cells (month 18)

M15. Identification of markers for treatment response in nasal fluids from patients with allergic rhinitis (month 30)

Workpackage description

Workpackage number	7	**Start date or starting event:**	Month 1				
Activity Type	RTD/Innovation						
Participant id	SNIP	UGOT					
Person-months per participant	36	4					

Objectives

To describe multi-layer network models based on combined DNA, RNA and protein analysis.

Description of work

This WP is based on searching for single-nucleotide polymorphisms (SNPs) in cis-regulatory regions of putative key genes identified by DNA microarray analysis of allergen-challenged T helper cells, as well as variations in the levels of the corresponding proteins. This information is used for multi-layer network models. Transcriptome modules, putative key regulatory genes and SNPs are identified as described in WP1, 5 and 6. SNP analysis is performed on material from 400 patients with allergic rhinitis and 400 healthy controls. The aim is to describe multi-layer network models. The hypothesis is that such models can be used to identify subtypes with distinct SNP combinations, which in turn have corresponding transcriptome and proteomic subtypes. First small groups of patients with distinct genetic subtypes are identified. Then, DNA microarray analysis of allergen-challenged T helper cells from these patients is performed. The relation between transcriptome modules and SNPs are studied to examine if SNPs correlate with gene expression. Finally, a small number of proteins that correspond to both genetic and transcriptome variations is analysed in nasal fluids from the patients. Algorithms designed to analyse differences in relations between the protein concentrations rather than absolute values are used so that the confounding factor of variable nasal fluid dilution is eliminated.

Workplan:

Year 1 and 2: Blood samples for SNP analysis is obtained from 400 patients with allergic rhinitis and 400 healthy controls.

Year 2: SNP analysis is performed. Material from selected patients is obtained for DNA microarrays and analysis of nasal fluid proteins.

Year 3. Construction of multi-layer network models. Identification of combinations of proteins that can serve as markers for treatment response.

Risk: Medium-high
Contingency plans: The risks in this project are mainly external factors such as obtaining large patient materials. Based on our previous experiences we believe this can be done within the stipulated time.

Deliverables

D23. Identification of combinations of disease-associated SNPs-journal publication (month 30)

D24. Multi-layer network models describing the association between cis-regulatory SNPs and altered RNA and protein expression of key genes in emergent modules (month 36)

Milestones and expected results

M16. Identification of disease-associated SNPs (month 30)

.

Workpackage description

Workpackage number	8	**Start date or starting event:**		Month 1			
Activity Type	RTD/Innovation						
Participant id	UN						
Person-months per participant	36						

Objectives

Analysis of network topology and dynamics in multiple sclerosis (MS) and the development of new *in vitro* system models to identify new diagnostic markers and therapeutic targets.

Description of work

1. Analysis of network topology in MS

Our aim is to identify the differences in the network topology, such as new modules, of the immune system transcriptome and protein interaction (IP) networks from multiple sclerosis patients compared with healthy controls. In addition, we will perform comparisons with allergy networks analyzed in the other WP, considering allergy and MS the two opposite extremes in autoimmunity.

1.1. Transcriptome network reverse engineering: Based in our preliminary studies, we plan to assess the transcriptome in PBMC from MS patients and controls using DNA microarrays. In the last 3 years we have collected RNA samples every 3 months for 2 years from a prospective cohort of MS patients (N=60) at different stages of the disease (onset, early, middle and late stage disease). This dataset reflects the different developmental stages of MS. Reverse engineering will be performed using a Bayesian network inference approach. The dataset will be curate using gene expression information available at Ingenuity Inc Database (months 1-12)

1.2. Transcriptome network topology analysis: Network metrics will be analyzed using Pajek and Matlab software. New interactions identified in our analysis will be validated using IP databases, such as String (http://string.embl.de/). (months 12-18). In addition we are going to address how robustness in the transcriptome network is affected by the disease and by the effect of immunomodulatory therapy. To this end, we are going to develop custom software that will be freely available in the web.

2. Analysis of network dynamics from health to autoimmunity

In order to assess the dynamics of the transcriptome network we are going to assess longitudinally the gene expression levels of a set of genes from the core of the transcriptome network (50 to 200 genes) identified in the Aim 1 using real time PCR. Gene expression levels and inferred IP networks will be assessed at different developmental stage of MS in our already available samples. Dynamics of individual gene expression levels will be modelled using ordinary differential equations. In order to assess the changes in the dynamics we will use standard statistics and metrics but new methods are need. In this sense we are going to use new tools generated in the analysis of language networks (months 18-24).

3. Validation of the identified pathways and modules

The differences in networks topology and dynamics between MS and healthy controls will be validated using *in vitro* experiments. Special focus will be dedicated to the patterns associated with response to interferon-beta (IFNB) therapy or the identification of new pathways that can become a therapeutic target (months 24-36). In this proposal we aim to identify pathways that might be suitable as targets for therapy using network analysis. However, full molecular and clinical validation of therapeutic targets is beyond the three-year scope of this proposal, but is planned to pursuit it in extension projects. Our group have expertise in performing molecular and *in vivo* validation, as well as preclinical studies to move basic research to the bed-side.

Risk: Medium-high

Contingency plans: The risks in this project are mainly those encountered in experimental studies, for example methodological problems and unexpected results. Methodological problems will be addressed when they occur and unexpected results often yield important new information

Deliverables

D25. A set of biological pathways involved in MS pathogenesis:

1. Description of a set gene expression patterns and IP modules associated with MS. Submission to public databases and journal publication (month 18)

2. Description of new dynamics associated with the development of MS journal publication and workshop (month 24)

D26. New mathematical and computational tools for analyzing the dynamics and robustness of evolving networks. Workshops and consortia web diffusion (month 30)

D27. A set of biological markers to monitor response to immunomodulatory therapy (IFNB) and candidate therapeutic pathways (which will require further validation). Journal publication and workshop (month 36)

Milestones and expected results

M1. Description of a set gene expression patterns and IP modules associated with MS. Submission to public databases and journal publication (month 18)

M2. A set of biological markers to monitor response to immunomodulatory therapy (IFNB) and candidate therapeutic pathways. Journal publication and workshop (month 30)

REFERENCES

1. Bernaschi M, Castiglione F. Design and implementation of an immune system simulator. Comput Biol Med 2001; 31:303-31.
2. Pappalardo F, Lollini PL, Castiglione F, Motta S. Modeling and simulation of cancer immunoprevention vaccine. Bioinformatics 2005; 21:2891-7.
3. Castiglione F, Toschi F, Bernaschi M, Succi S, Benedetti R, Falini B, et al. Computational modeling of the immune response to tumor antigens. J Theor Biol 2005; 237:390-400.
4. Castiglione F, Sleitser V, Agur Z. Analyzing hypersensitivity to chemotherapy in a Cellular Automata model of the immune system. In: Cancer Modeling and Simulation. London: Chapman & Hall/CRC Press 2003. p. 333-65.
5. Castiglione F, Poccia F, D'Offizi G, Bernaschi M. Mutation, fitness, viral diversity, and predictive markers of disease progression in a computational model of HIV type 1 infection. AIDS Res Hum Retroviruses 2004; 20:1314-23.
6. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet 2005; 37:233-42.
7. Jenssen TK, Laegreid A, Komorowski J, Hovig E. A literature network of human genes for high-throughput analysis of gene expression. Nat Genet 2001; 28:21-8.
8. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, et al. A network-based analysis of systemic inflammation in humans. Nature 2005; 437:1032-7.
9. Sole R. Language: syntax for free? Nature 2005; 434:289.
10. Searls DB. The language of genes. Nature 2002; 420:211-7.
11. Mehrabian M, Allayee H, Stockton J, Lum PY, Drake TA, Castellani LW, et al. Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. Nat Genet 2005; 37:1224-33.
12. Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science 2004; 306:640-3.
13. Pawitan Y, Bjohle J, Amler L, Borg AL, Egyhazi S, Hall P, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 2005; 7:R953-R64.
14. Chesler EJ, Langston MA. Combinatorial Genetic Regulatory Network Analysis Tools for High Throughput Transcriptomic Data. RECOMB Satellite Workshop on Systems Biology and Regulatory Genomics. San Diego, 2005.
15. Baldwin NE, Chesler EJ, Kirov S, Langston MA, Snoddy JR, Williams RW, et al. Computational, integrative, and comparative methods for the elucidation of genetic coexpression networks. J Biomed Biotechnol 2005; 2005:172-80.
16. Bellaachia A, Portnoy D, Chen Y, Elkahloun AG. E-CAST: A Data Mining Algorithm for Gene Expression Data. Proceedings,

Workshop on Data Mining in Bioinformatics, 2002:49–54.

17. Ben-Dor A, Bruhn L, Friedman N, Nachman I, Schummer M, Yakhini Z. Tissue classification with gene expression profiles. J Comput Biol 2000; 7:559-83.

18. Ben-Dor A, Shamir R, Yakhini Z. Clustering gene expression patterns. J Comput Biol 1999; 6:281-97.

19. Hansen P, Jaumard B. Cluster Analysis and Mathematical Programming. Mathematical Programming 1997; 79:191–215.

20. Hartuv E, Schmitt A, Lange J, Meier-Ewert S, Lehrachs H, Shamir R. An Algorithm for Clustering cDNAs for Gene Expression Analysis. Proceedings, RECOMB, 1999:188–97.

21. Slonim DK. From patterns to pathways: gene expression data analysis comes of age. Nat Genet 2002; 32 Suppl:502-8.

22. Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. Proc Natl Acad Sci U S A 2000; 97:12182-6.

23. Alter O, Brown PO, Botstein D. Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling. Proceedings of the National Academy of Sciences,, 2000:10101–6.

24. Girolami M, Breitling R. Biologically valid linear factor models of gene expression. Bioinformatics 2004; 20:3021-33.

25. Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, et al. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. Proc Natl Acad Sci U S A 2003; 100:8466-71.

26. Allocco DJ, Kohane IS, Butte AJ. Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 2004; 5:18.

27. Fellows MR, Langston MA. Nonconstructive Tools for Proving Polynomial-Time Decidability. Journal of the ACM 1988; 35:727-39.

28. Fellows MR, Langston MA. On Search, Decision and the Efficiency of Polynomial-Time Algorithms. Journal of Computer and Systems Sciences 1994; 49:769-79.

29. Setubal JC, Meidanis J. Introduction to Computational Molecular Biology. Boston: PWS Publishing Company; 1997.

30. Feige U, Goldwasser S, Lovasz L, Safra S, Szegedy. M. Approximating the maximum clique is almost NP-complete. Proceedings, IEEE Symposium on the Foundations of Computer Science, 1991:2-12.

31. Downey RG, Fellows MR. Parameterized Complexity. New York: Springer; 1999.

32. Garey MR, Johnson DS. Computers and Intractability; A Guide to the Theory of *NP*-Completeness: W. H. Freeman and Company; 1990.

33. Chandran LS, Grandoni F. Refined Memorisation for Vertex Cover. Proceedings, International Workshop on Parameterized and Exact Computation. Bergen, Norway, 2004

34. Abu-Khzam FN, Langston MA, Shanbhag P, Symons CT. Scalable Parallel Algorithms for FPT Problems. Algorithmica accepted for publication, 2005.

35. Abu-Khzam FN, Langston MA, Suters WH. Effective Vertex Cover

Kernelization: A Tale of Two Algorithms. Proceedings, ACS/IEEE International Conference on Computer Systems and Applications. Cairo, Egypt, 2005.

36. Abu-Khzam FN, Collins RL, Fellows MR, Langston MA, Suters WH, Symons CT. Kernelization Algorithms for the Vertex Cover Problem: Theory and Experiments. Proceedings, Workshop on Algorithm Engineering and Experiments (ALENEX). New Orleans, Louisiana, 2004.

37. Langston MA. Practical FPT Implementations and Applications (Plenary Talk). Proceedings, International Workshop on Parameterized and Exact Computation. Bergen, Norway, 2004.

38. Zhang Y, Abu-Khzam FN, Baldwin NE, Chesler EJ, Langston MA, Samatova NF. Genome-Scale Computational Approaches to Memory-Intensive Applications in Systems Biology. Supercomputing. Seattle, Washington, 2005.

39. Zhang Y, Abu-Khzam FN, Baldwin NE, Chesler EJ, Langston MA, Samatova NF. Genome-Scale Computational Approaches to Memory-Intensive Applications in Systems Biology. Proceedings, Supercomputing. Seattle, Washington, 2005.

40. Dehne F, Fellows MR, Langston MA, Rosamond FA, Stevens K. An $O^*(2^{O(k)})$ FPT Algorithm for the Undirected Feedback Vertex Set Problem. Proceedings, International Computing and Combinatorics Conference. Kunming, China, 2005.

41. Hopper PJ, Traugott EC. Grammaticalization. Cambridge: Cambridge University Press; 2003.

42. Steels L. The Emergence and Evolution of Linguistic Structure: From Lexical to Grammatical Communication Systems. Connection Science 2005; 17:213-30.

43. Steels L. Evolving grounded communication for robots. Trends Cogn Sci 2003; 7:308-12.

44. Landauer TK, Foltz PW, Laham D. Introduction to Latent Semantic Analysis. Discourse Processes 1998; 25:259-84.

45. Albert R, Barabasi AL. Statistical mechanics of complex networks. Reviews of Modern Physics 2001; 74:47-97.

46. Sole R, Pastor-Satorras R. Complex Networks in Genomics and Proteomics. In: Handbook of Graphs and Networks. Berlin; 2002.

47. Steels L. Analogies between Genome and Language Evolution. Artificial Life IX. Boston, 2004.

48. Benson M, Carlsson L, Adner M, Jernas M, Rudemo M, Sjogren A, et al. Gene profiling reveals increased expression of uteroglobin and other anti-inflammatory genes in glucocorticoid-treated nasal polyps. J Allergy Clin Immunol 2004; 113:1137-43.

49. Benson M, Jansson L, Adner M, Luts A, Uddman R, Cardell LO. Gene profiling reveals decreased expression of uteroglobin and other anti-inflammatory genes in nasal fluid cells from patients with intermittent allergic rhinitis. Clin Exp Allergy 2005; 35:473-8.

50. Johansson S, Keen C, Stahl A, Wennergren G, Benson M. Low levels of CC16 in nasal fluid of children with birch pollen-induced rhinitis. Allergy 2005; 60:638-42.

51. Benson M, Carlsson B, Carlsson LM, Mostad P, Svensson PA, Cardell LO. DNA microarray analysis of transforming growth factor-beta and related transcripts in nasal biopsies from patients with

allergic rhinitis. Cytokine 2002; 18:20-5.

52. Benson M, Carlsson B, Carlsson LM, Wennergren G, Cardell LO. Increased expression of Vascular Endothelial Growth Factor-A in seasonal allergic rhinitis. Cytokine 2002; 20:268-73.

53. Benson M, Svensson PA, Carlsson B, Jernas M, Reinholdt J, Cardell LO, et al. DNA microarrays to study gene expression in allergic airways. Clin Exp Allergy 2002; 32:301-8.

54. Benson M, Svensson PA, Adner M, Caren H, Carlsson B, Carlsson LM, et al. DNA microarray analysis of chromosomal susceptibility regions to identify candidate genes for allergic disease: a pilot study. Acta Otolaryngol 2004; 124:813-9.

55. Benson M, Wennergren G, Fransson M, Cardell LO. Altered levels of the soluble IL-1, IL-4 and TNF receptors, as well as the IL-1 receptor antagonist, in intermittent allergic rhinitis. Int Arch Allergy Immunol 2004; 134:227-32.

56. Kinhult J, Egesten A, Benson M, Uddman R, Cardell LO. Increased expression of surface activation markers on neutrophils following migration into the nasal lumen. Clin Exp Allergy 2003; 33:1141-6.

57. Benson M, Uddman R, Cardell LO. Epithelial cells in nasal fluids from patients with allergic rhinitis: how do they relate to epidermal growth factor, eosinophils and eosinophil cationic protein? Acta Otolaryngol 2002; 122:202-5.

58. Adner M, Rose AC, Zhang Y, Sward K, Benson M, Uddman R, et al. An assay to evaluate the long-term effects of inflammatory mediators on murine airway smooth muscle: evidence that TNFalpha up-regulates 5-HT(2A)-mediated contraction. Br J Pharmacol 2002; 137:971-82.

59. Benson M, Reinholdt J, Cardell LO. Allergen-reactive antibodies are found in nasal fluids from patients with birch pollen-induced intermittent allergic rhinitis, but not in healthy controls. Allergy 2003; 58:386-92.

60. Benson M, Reinholdt J, Cardell LO. No decrease of birch pollen specific IgA and IgG in nasal fluids from cortisone treated patients with intermittent allergic rhinitis. Allergy 2004; 59:365-7.

61. Keen C, Johansson S, Reinholdt J, Benson M, Wennergren G. Bet v 1-specific IgA increases during the pollen season but not after a single allergen challenge in children with birch pollen-induced intermittent allergic rhinitis. Pediatr Allergy Immunol 2005; 16:209-16.

62. Fransson M, Benson M, Wennergren G, Cardell LO. A role for neutrophils in intermittent allergic rhinitis. Acta Otolaryngol 2004; 124:616-20.

63. Benson M, Strannegard IL, Strannegard O, Wennergren G. Topical steroid treatment of allergic rhinitis decreases nasal fluid TH2 cytokines, eosinophils, eosinophil cationic protein, and IgE but has no significant effect on IFN-gamma, IL-1beta, TNF-alpha, or neutrophils. J Allergy Clin Immunol 2000; 106:307-12.

64. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, et al. Complex Trait Analysis of Gene Expression Uncovers Polygenic and Pleiotropic Networks that Modulate Nervous System Function. Nature Genetics 2005; 37:233-42.

65. Chesler EJ, Langston MA. Combinatorial Genetic Regulatory Network Analysis Tools for High Throughput Transcriptomic Data. Proceedings, RECOMB Satellite Workshop on Systems Biology and

Regulatory Genomics. San Diego, 2005.

66. Baldwin NE, Chesler EJ, Kirov S, Langston MA, Snoddy JR, Williams RW, et al. Computational, Integrative and Comparative Methods for the Elucidation of Genetic Co-Expression Networks. Journal of Biomedicine and Biotechnology 2005; 2:172-80.

67. Langston MA, Perkins AD, Saxton AM, Scharff JA, Voy BH. Innovative Computational Methods for Transcriptomic Data Analysis. Proceedings, ACM Symposium on Applied Computing. Dijon, France, 2006.