

A SYSTEMS-LEVEL AND EPIGENETIC GENE CONTROL APPROACH TO HUNTINGTON'S DISEASE

Christiaan Henkel¹, Xiao-Xian Yang², Timo M. Breit¹ and Pernette J. Verschure²

¹⁾ Integrative Bioinformatics Unit

²⁾ Nuclear organization group

Swammerdam Institute for Life Sciences and Informatics Institute

University of Amsterdam

Kruislaan 318, 1098 SM Amsterdam

Huntington's disease (HD) is a well known model for autosomal dominant neurogenetic disorders. However, the molecular mechanisms underlying the multifaceted character of the disease are still poorly understood. We have initiated two related lines of HD research following a combined systems level approach with two perspectives, (i) an experiment based focus on epigenetic gene control to understand cellular and molecular mechanisms underlying HD (ii) a dry lab focus on integration of experimental data from disparate sources into a comprehensive HD knowledge base. There is an iterative interaction between these efforts: experimental data will be fed into the growing knowledge model, which will in turn supply novel hypotheses for laboratory validation. Understanding of the complex mechanisms underlying HD will benefit from this multi-disciplinary approach.

(i) DISREGULATION OF GENE EXPRESSION IN HUNTINGTON'S DISEASE BY EPIGENETIC CHANGES IN CHROMATIN STRUCTURE

Gene expression control is crucial to maintain differentiated cell types in a multi-cellular organism, whereas aberrant gene regulation can lead to pathological situations. The cell nucleus is the site where the genetic material is stored and where key processes take place, i.e. orchestration of many thousands of genes, faithful replication of the genome during cell division and repair of damaged DNA. Packaging of genomic DNA into higher order chromatin structures is a basic epigenetic mechanism to achieve proper control of critical cellular processes. The linear genome is partitioned into distinct functional chromatin domains that help to assure independent regulation of such domains. Posttranslational modifications of histones (i.e. acetylation, methylation, phosphorylation and ubiquitination) are critical determinants of functional changes in chromatin structure over large distances. We are still largely ignorant how levels of epigenetic gene regulation affect biological systems. Our research focuses on the functional organization of chromosomes and chromatin in the mammalian interphase cell nucleus in living cells. We investigate whether triplet repeat expansions as observed in HD are involved in altering gene expression via epigenetic gene control mechanisms, i.e. changes in chromatin structure. We will start using cells of a rat HD model, carrying varying glutamine repeat sizes that are well characterized and correlate with HD progression. In addition, we will use lymphoblastoid cell lines derived from HD patients carrying alleles with different CAG repeat lengths (i.e. > 36 CAG repeats) and from healthy controls (i.e. < 36 repeats). We compare the chromatin structure in cell lines from a rat HD model carrying several repeat sizes and from HD patients carrying long or shorter CAG repeats versus healthy controls. We will use chromatin immunoprecipitation (ChIP) analysis, nuclease accessibility analysis and Western (protein) blotting to determine chromatin compaction, and the presence of histone modifications and other chromatin associated proteins. Moreover, we will interfere with the chromatin structure by using several drugs that specifically inhibit histone deacetylation or DNA methylation. In parallel, we analyze the overall nuclear structure of these cell lines. To this end, we perform immunofluorescent labeling with antibodies against several known nuclear factors involved in gene expression to analyze the 3-dimensional distribution of such factors in the interphase nucleus using state-of-the-art microscopy techniques. Knowledge whether changes in epigenetic gene control are involved in the mechanism underlying HD are essential to elucidate the cellular and molecular basis of HD and to eventually interfere with dysregulation of such epigenetic control mechanisms.

(ii) TOWARDS A KNOWLEDGE BASE FOR HUNTINGTON'S DISEASE

The amounts of experimental data, information, and knowledge on HD and related phenomena are vast. However, these resources differ greatly in scope, format, availability and detail (literature, omics data, and so on), which prevents straightforward interpretation of the relationships between such data. To overcome this problem, we will employ a novel systems-based approach in which we will focus on creating an HD knowledge base.

Knowledge bases can be used to capture explicit knowledge of an organization, or, in this case, a life sciences problem domain. To structure a knowledge base, the entity types, relationships, and classification scheme can be specified in an ontology. Together with domain-specific instances this constitutes a domain specific knowledge base. The primary benefit of such a knowledge base is that it provides the means for computerized gathering, organization, and retrieval of knowledge of a complete domain. This task is too demanding for any individual research group working in a given field and as such the creation of a domain-specific knowledge base should be approached as a collaborative effort. However, it needs no explanation that such an (ever evolving) knowledge base will support and boost the research of a domain immensely.

The proposed HD knowledge base will consist of two parts, a resource identification (RI) model and a knowledge model. The first component, the HD-RI model, encompasses potentially all resources of HD related data, information, and knowledge. The initial step towards an HD-RI model is the creation of a generic resource identification (GRI) model that is able to capture all relevant resources. An important aspect of this GRI model is the ability to provide researchers with an intuitive and flexible overview, adequate search engine, and where possible access to the resources. The generic character of the GRI model should ensure that it can be (re)applied to a wide variety of life science problem domains. Once the GRI model is in place, we will instantiate it for HD with the help of HD experts plus publicly available information, and thus create the HD-RI model.

The second part of the HD knowledge base is an HD knowledge model in which the HD associated data, information and knowledge themselves are formalized, together with their relationships. As with the GRI model, ontologies and semantic web technology will be used for this. The HD knowledge model can be used for computational experimentation, such as mining for as yet undiscovered relationships, and machine-reasoning to generate novel hypotheses. Every HD researcher can use this model to put his own research in a systemic perspective, and subsequently contribute to the expanding HD knowledge model by feeding results back into it.

To be able to allow researchers to use the HD knowledge base, all parts will be implemented in an HD problem-solving environment of the Virtual Lab for e-Science. This environment is enabled by emerging ICT technologies, including grid computing and semantic web technology. Domain interaction will take place in a physical Laboratory for e-BioScience (e-BioLab), which is currently under construction. As such, the HD model will serve as a case study in the development of generic methodology for integrative bioinformatics and e-bioscience.