Dynamical properties of gene regulatory networks involved in long-term potentiation

a thesis submitted for the degree of Doctor of Philosophy

Gonzalo Sánchez Nido
Dynamical properties of gene regulatory networks involved in long-term potentiation
The research work presented on this thesis has been partially published and portions of these publications are reused. Textual portions and figures of these publications are reused with permission.


  “The IEEE does not require individuals working on a thesis to obtain a formal reuse license”


  “Under the Frontiers Terms and Conditions, authors retain the copyright to their work. All Frontiers articles are Open Access and distributed under the terms of the Creative Commons Attribution License, (CC-BY 3.0), which permits the use, distribution and reproduction of material from published articles, provided the original authors and source are credited, and subject to any copyright notices concerning any third-party content”

The template for this publication (including the front cover), The Legrand Orange Book, was originally compiled by Mathias Legrand (legrand.mathias@gmail.com) and has subsequently been modified by Vel (vel@latextemplates.com) and Gonzalo Sánchez Nido (gonzalo.s.nido@gmail.com), and is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at http://creativecommons.org/licenses/by-nc/3.0. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

First submitted, March 2015
Hardbound copy deposited, July 2015
Abstract

The significant increase in the availability of postgenomic data has stimulated the growth of hypothesis-generating strategies to unravel the molecular basis of nature. The application of systems theory to biological problems emerged in the early 1970s, and yet the computational methods developed to model biological networks and analyse their functionality have been seldom used for understanding the neurogenetic basis of cognition. The main interests of this thesis are the application of computational models to microarray expression data for the identification and analysis of biological networks related with long-term potentiation (LTP), the cellular correlate of learning and memory in mammals. The models include the analysis of co-expression and studies of dynamical stability.

The thesis starts with the application of established methods on gene expression analysis on the available expression data from LTP in order to identify networks of closely correlated genes in their patterns of expression to ultimately pinpoint putative key regulators not identified previously by classical differential expression analysis.

The thesis continues with the analysis of previously identified gene networks regulated 20 min, 5 h, and 24 h post-LTP induction. A dynamical stability analysis using weight matrices suggests that the early network has a significant sensitivity to perturbations compared with randomly generated networks of similar characteristics. In addition, using random Boolean networks, we study the differential sensitivity to perturbations of these networks and we find that our results are consistent with a model of LTP as a complex cellular switch. In such a scenario, earlier networks are dynamically more unstable than later regulatory networks, which are proposed to be responsible for the new homeostatic state reached by the stimulated neurons. Key genes responsible for the dynamic properties observed are identified and discussed. In particular, we found that Egr2, a member of the Egr family of transcription factors was crucial to the bistability observed in the early-response network. Other genes previously associated with LTP have a more modest contribution.

A functional analysis of these networks is presented and integrated with previous knowledge on the molecular basis of LTP.
A Miki y a Javi
Acknowledgements

This thesis would not have been possible without emotional support. I left family and friends behind when I started my PhD at the University of Otago. I have to thank each member of the P.B. group of WhatsApp, made up of my Biology classmates from Uni, now scattered all over the world – Jero, Djordje, Pin, Carol, Pablulu, Palo, Ainhoa, Rincho, Mery, Manolo, Isa, Paula and Paulita.

I would also like to extend my thanks to the members of the P.H. group on Telegram, set up during a quick encounter in Berlin with Molly, Pako and David Lee. They helped me to keep up with music and politics, and at the same time share desperation and joy.

I would like to express my very great appreciation to the members of the Bioinformatics Unit of the Centre for Molecular Biology Severo Ochoa, where I worked as a research assistant before starting my PhD – Ugo, Alberto, David, Klett, Antonio, Ruben, Fons and Amigo.

“Survival” in Dunedin has been possible thank to local friends, of course. Throughout these years, people have come and gone, but some of them have been there for quite a long time. I am deeply grateful to Sam and Andrew. You have made me feel at home up in the summits and back home in Dunedin and I feel a great deal of admiration for you. Also, thank you Dave for being always there when I needed a place to stay or someone to talk to. Pretty much all my friends in New Zealand are amazing climbers – Sarah, Kate, Dani, Rocío, Martin, Paul and Michelle. It has been delightful to climb mountains, ice and rock with you during these years (and the years to come).

I would also like to thank Lech for the conversations during coffee breaks and lunches. Your enthusiasm and company have definitely made a difference to my life in the Department.

Central to this thesis are my main and co-supervisors, Lubica and Joanna. I would like to express my deep gratitude for the opportunity to embark this journey. Thank you for your help, knowledge and encouragement throughout these years. I would also like to thank Margaret for her advice and assistance with the microarray data.

My family, of course, despite the distance, has always been there for me. Mum, dad, Miguel and Javi – thank you so much.

I am particularly grateful to Maud, who has been there pretty much every day during my PhD, dealing with my unbearable personality, my ups and downs, and my contradictions.
## Contents

1 **Introduction** ........................................... 15  
1.1 Neuroscience and the study of the brain 15  
1.2 Learning and memory 16  
1.3 Motivation 17  
1.4 Thesis goals 18  
1.5 Thesis outline 18  

2 **Overview of genetics and microarrays** .................. 19  
2.1 Introduction 19  
2.2 Biology of gene expression 20  
2.3 Regulation of transcription 22  
2.4 Cell homeostasis 23  
2.5 Gene expression profiling 24  
2.6 Oligonucleotide microarrays 24  
2.7 Analysis of microarray data 25  
2.7.1 Low-level analysis .................................. 25  
2.7.2 High-level analysis ................................. 27  
2.8 Microarrays in neuroscience 27  

3 **Biology of long-term potentiation** ....................... 29  
3.1 The neuron 29  
3.1.1 The resting potential ................................ 29  
3.1.2 The action potential ................................. 30  
3.1.3 The chemical synapse ............................... 32
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.8 Extracellular signalling and cell adhesion</td>
<td>88</td>
</tr>
<tr>
<td>6.3.9 Changes in membrane composition, endocytosis and exocytosis</td>
<td>90</td>
</tr>
<tr>
<td>6.3.10 Neuronal structural changes via cytoskeletal remodelling</td>
<td>92</td>
</tr>
<tr>
<td>6.3.11 Neurogenesis and cell death</td>
<td>93</td>
</tr>
<tr>
<td>6.3.12 Immunity-related genes</td>
<td>93</td>
</tr>
<tr>
<td>6.4 Discussion</td>
<td>93</td>
</tr>
<tr>
<td>6.4.1 LTP-induced transcription</td>
<td>94</td>
</tr>
<tr>
<td>6.4.2 Epigenetic control of gene expression</td>
<td>95</td>
</tr>
<tr>
<td>6.4.3 Regulation of signalling pathways via changes in gene expression</td>
<td>96</td>
</tr>
<tr>
<td>6.4.4 Other overrepresented functions</td>
<td>97</td>
</tr>
<tr>
<td>6.5 Concluding remarks</td>
<td>100</td>
</tr>
<tr>
<td>7 Stability of an early LTP network using TReMM</td>
<td>105</td>
</tr>
<tr>
<td>7.1 Introduction</td>
<td>105</td>
</tr>
<tr>
<td>7.2 Methods</td>
<td>106</td>
</tr>
<tr>
<td>7.2.1 Weight matrices dynamics</td>
<td>106</td>
</tr>
<tr>
<td>7.2.2 LTP-GRN and gene expression</td>
<td>107</td>
</tr>
<tr>
<td>7.2.3 Frequency of bistable outputs $p_B$</td>
<td>107</td>
</tr>
<tr>
<td>7.2.4 Random and knock-out networks</td>
<td>107</td>
</tr>
<tr>
<td>7.3 Results</td>
<td>108</td>
</tr>
<tr>
<td>7.3.1 LTP-GRN bistable frequency</td>
<td>108</td>
</tr>
<tr>
<td>7.3.2 Knock-out effects</td>
<td>108</td>
</tr>
<tr>
<td>7.4 Discussion</td>
<td>108</td>
</tr>
<tr>
<td>8 Stability of LTP networks using Boolean models</td>
<td>111</td>
</tr>
<tr>
<td>8.1 Introduction</td>
<td>111</td>
</tr>
<tr>
<td>8.2 Methods</td>
<td>112</td>
</tr>
<tr>
<td>8.2.1 Random Boolean Networks</td>
<td>112</td>
</tr>
<tr>
<td>8.2.2 Network topologies</td>
<td>112</td>
</tr>
<tr>
<td>8.3 Results</td>
<td>114</td>
</tr>
<tr>
<td>8.3.1 LTP-related networks are dynamically similar to the yeast network</td>
<td>114</td>
</tr>
<tr>
<td>8.3.2 Knock-out analyses</td>
<td>115</td>
</tr>
<tr>
<td>8.4 Concluding remarks</td>
<td>115</td>
</tr>
<tr>
<td>8.5 Discussion</td>
<td>117</td>
</tr>
<tr>
<td>9 Conclusions – LTP as a cellular switch</td>
<td>121</td>
</tr>
<tr>
<td>References</td>
<td>125</td>
</tr>
<tr>
<td>A Appendices</td>
<td>151</td>
</tr>
<tr>
<td>A.1 List of differentially expressed genes</td>
<td>151</td>
</tr>
<tr>
<td>A.1.1 20 min post-LTP</td>
<td>151</td>
</tr>
<tr>
<td>A.1.2 5 h post-LTP</td>
<td>152</td>
</tr>
<tr>
<td>A.1.3 24 h post-LTP</td>
<td>153</td>
</tr>
<tr>
<td>A.2 Top Gene Ontology terms for co-expression networks</td>
<td>153</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Examples of protein structures ............................................. 21
2.2 Central dogma of molecular biology ...................................... 22
2.3 Hybridization of the sample to the probes ............................... 26
2.4 Microarrays in the scientific literature .................................... 28

3.1 Structure of a neuron ......................................................... 30
3.2 Voltage-gated channels during the action potential ..................... 31
3.3 Hippocampal neural circuit .................................................. 36
3.4 LTP by HFS ................................................................. 37
3.5 Molecular basis of LTP induction .......................................... 44

4.1 Boolean network model ...................................................... 54
4.2 Types of attractors in random Boolean models .......................... 55
4.3 Derrida plots of random networks ....................................... 56
4.4 Dynamics of a set of genes using weight matrices ...................... 60

5.1 Normalisation of raw microarray data by RMA ......................... 64
5.2 Volcano plots ............................................................... 65
5.3 Differentially expressed genes ............................................. 67
5.4 LTP networks .............................................................. 68

6.1 Clustering of samples ....................................................... 71
6.2 Soft threshold selection ...................................................... 72
6.3 Modular differential connectivity example ............................... 75
6.4 Soft threshold parameter selection ....................................... 76
6.5 Co-expression modules heatmaps ....................................... 78
6.6 Module overlap ........................................................... 79
6.7 Module reconfiguration following LTP .................................... 80
6.8 Main trends in intramodular connectivity changes following LTP .... 81
6.9 Differential expression of hub genes across time ....................... 83
6.10 MAPK pathways activated by extracellular stimuli .................... 89
6.11 Lipid signalling ............................................................ 98
7.1 Distribution of bistable output frequencies ........................................ 108
7.2 LTP-related gene regulatory network 1 ............................................. 109

8.1 Types of network topologies used ..................................................... 113
8.2 Degree distribution for the LTP and yeast transcriptional networks ...... 114
8.3 Derrida plots for the yeast network .................................................... 115
8.4 Derrida plots for the LTP networks ................................................... 116
8.5 Derrida plots for the LTP in silico knock-outs, $H(0)$ vs $H(1)$ .......... 117
8.6 Derrida plots for the LTP in silico knock-outs, $H(0)$ vs $H(5)$ .......... 118
8.7 Effect of small perturbations in the LTP and yeast networks .......... 119

9.1 Nested networks ............................................................................. 123
List of Tables

3.1 Immediate early genes ................................................................. 47
4.1 Example of Boolean functions .................................................... 53
5.1 Experimental design ................................................................. 65
5.2 Top differentially expressed genes .............................................. 66
6.1 Top hubs .................................................................................. 74
6.2 List of modules .......................................................................... 76
6.3 Hubs membrane channels .......................................................... 91
6.4 Genes co-expressed 20 min post LTP involved in cytoskeletal remodelling 92
6.5 Gene modules co-expressed 5 h post LTP .................................... 100
A.4 Top Gene Ontology terms for each of the modules (p < 0.01) ........... 153
1. Introduction

If a man carries many such memories into life with him, he is saved for the rest of his days. And even if only one good memory is left in our hearts, it may also be the instrument of our salvation one day

Fyodor Dostoyevsky

1.1 Neuroscience and the study of the brain

The brain is the centre of the nervous system and is localised in close proximity to the main sensory organs. Information from the sensory organs is integrated by the brain to elicit adequate responses. The ability to learn and retain memories represents an evolutionary advantage for the success of a species in a changing environment. Whereas the capacity to retain information over long periods of time is not exclusive to the mammalian brain, some notable differences exist between their brains and the other vertebrates.

Perhaps the most striking feature of the mammalian brain is its larger size when compared to other vertebrates. But more interestingly, the brain of mammals is structurally different as well. The mammalian brain possesses a highly complex structure called neocortex, which differs from the more simple pallium found in other vertebrates (Aboitiz, Morales, and Montiel, 2003). Other areas such as the amygdala and the hippocampus are more developed. The hippocampus, in particular, is a brain region that, as such, is only present in mammals.

Differences also exist between human brain and their close relatives – the great apes. The human brain is remarkably large in size. Despite their similar overall body size, our brain is about three times heavier. The brain represents around 20% of the total body oxygen consumption and around 25% of the glucose.

The complexity of the nervous system is difficult to conceive. For example, an average human adult brain consists of around 85 billion neurons connected in a network of around $10^{14}$-$10^{15}$ connections, known as synapses (Herculano-Houzel, 2009). The brain of a rat, on the other hand, consists of 200 million neurons and between $10^{11}$ and $10^{12}$ synapses (Herculano-Houzel and Lent, 2005). The computational capabilities that such a network of processing units can hold is fascinating.
Especially considering that each of those synaptic connections possesses an extraordinary plasticity subject to fine tuning dependent on the specific requirements. Furthermore, neurons are able to create new connections with other neurons, reconfiguring the wiring of the network as we interact with the environment. At the same time, the brain withstands a disruption of the neural connectivity across the major neurodegenerative diseases. Indeed, these changes can be specific to the syndromes to some extent and track the pattern of the pathological development (Zhang, Wang, Wu, Kuang, Huang, He, and Gong, 2011; Pievani, de Haan, Wu, Seeley, and Frisoni, 2011).

While the first references to the human brain can be traced back as far as the times of the Ancient Egyptians, the consideration of the brain as a seat of intelligence dates to the times of the Ancient Greeks. The structural characterisation of the brain was a subject of study since the Hellenistic period, and knowledge of the anatomy of the brain has steadily accumulated ever since.

It was, however, during the late 19th Century that the field of neuroscience, benefiting from technical advances, broadened more greatly and incorporated different disciplines. The seminal work carried by Ramón y Cajal using the staining technique developed by Camillo Golgi represents perhaps the most influential breakthrough in the field of neuroscience. Their results lead to the neuron doctrine – the hypothesis that the functional unit of the brain is the neuron (Kandel, Schwartz, Jessell, et al., 2000). Additional support for this theory came from the field of modern electrophysiology and the introduction of electron microscopy. While these experiments were carried out using animal models, other studies were focused on language disorders. French neurologist Pierre Paul Broca identified particular regions in the brain responsible for language. These results served as starting point for a number of subsequent studies to characterise functionally the different anatomical regions of the brain.

The advent of molecular biology and the 20th Century together with the consequent movement of the “gene” to the spotlight of biology brought an interest in the role of genes in behaviour. The first large-scale experiments where conducted in flies by the group of Seymour Benzer in the early 1970s, studying the effects of mutations in the circadian rhythm. Joseph Takahashi, in the 1990s, conducted similar studies on mice. Although human genetics represent a much more challenging scenario, there has been exceptional progress in the last decades in this regard, with a variety of mutations being identified with neurological disorders. These genes have been put into the context of molecular pathways and have shed light on the genetic basis of neurological and psychiatric disorders. Aside disease-oriented research, studies at the molecular level have characterised in great detail the mechanisms underlying neuronal function. Molecular biology and genetics methods have been adopted to understand neuronal development and how genetic changes influence biological functions.

The genomic and proteomic revolutions of the 1990s, together with the advent of other -omic fields has fired the popularisation of the discipline of systems biology in the late 1990s. The fast accumulation of genomic data has led to the development of a class of methodologies at the systems-level, with the aim to provide a more integrative understating of the nervous system function (Geschwind and Konopka, 2009; De Schutter, 2008). The adoption of functional genomics or molecular-systems methods that allow a dynamic measurement of gene products in a highly parallel manner, coupled with an underlying systems-level knowledge of the organization of these gene products, has the potential to provide a more integrative understanding of nervous system function.

1.2 Learning and memory

Somewhere within that complex network of neurons that constitute the brain, we encode, store, and retrieve information. This brain function is what we define as “memory”. Memory is, in other words, the ability of the mind to hold and recall past experiences, and represents one of the most intensively studied subjects in the field of neuroscience.
The historic case of Henry Molaison (H.M.) is perhaps the most paradigmatic in the study of the neural basis of memory. In 1953, H.M. underwent a brain surgery to alleviate a case of severe intractable epilepsy in the medial temporal lobe. The surgery included the resection of the hippocampus and the adjoining regions in both hemispheres, which resulted in a very specific and severe anterograde amnesia. H.M. was able to recall memories of the times before the operation, including his normal vocabulary. He was, however, unable to form new memories. He would be able to retain new information for short periods of time, but he would forget after some minutes. H.M. had lost the ability to form new long-term memories, even though his short-term working memory was intact.

Despite not being able to remember new episodes after the surgery, H.M. was able to form a very specific type of long-term memories. He was able to improve his motor skills upon training, yet he would not be able to remember the training itself (Squire, 2009). This form of memory is what we know today as *implicit* memory (Kandel *et al.*, 2000; Roediger, 1990).

The case of H.M. case showed that new memories that require a conscious retrieval (*explicit* or declarative memories) were formed in the medial temporal lobe. Other cases narrowed the anatomical location for the formation of new explicit memories down to the hippocampus and more specifically highlighted the importance of the CA1 region. Long-term memories, however, are thought to be ultimately stored in the cerebral cortex. While the capacity of neurons to fix short-term events into long-term memories has been attributed to activity-dependent changes in the efficacy of synaptic connections throughout the 20th century, the experimental evidence did not come until 1966 by the hand of Terje Lømo. Lømo observed in the rabbit hippocampus that repetitive activation of excitatory synapses resulted in an increase in synaptic strength that could last for long periods of time (Lømo, 1966, 2003). This form of synaptic plasticity is known as LTP.

As it shall be discussed in more depth in Chapter 3, LTP represents the cellular mechanism responsible for long-term memories in mammals. Although some of their basic molecular mechanisms have been described, the molecular basis of LTP is an active field of research. Perhaps one of the most active discussions on LTP lies on the role of gene expression in the mechanistically different phases, and is the central topic of this thesis.

### 1.3 Motivation

It is known that the maintenance of LTP depends on gene expression (Abraham, Mason, Demmer, Williams, Richardson, Tate, Lawlor, and Dragunow, 1993; Abraham and Williams, 2003; Goelet, Castellucci, Schacher, and Kandel, 1986; Bramham, Alme, Bittins, Kuipers, Nair, Pai, Panja, Schubert, Soule, Tiron, *et al.*, 2010), but the mechanisms are still far from being well understood. Efforts towards a deeper understanding of the gene regulatory networks underlying LTP have been undertaken by Ryan, Mason-Parker, Tate, Abraham, and Williams (2011); Ryan, Ryan, Kyrke-Smith, Logan, Tate, Abraham, and Williams (2012) among many others. The new high-throughput methods developed in recent times have rendered available a great amount of valuable data on gene expression that is yet to be fully studied.

As we have mentioned, LTP is the change in synaptic efficiency most likely to be one of the cellular mechanisms responsible for long-term memory storage in the mammalian brain. From an experimental perspective, LTP behaves as a cellular switch whose output on/off value can be measured by the means of recording electrodes. Namely, cells that do not receive a stimulus sufficient to induce LTP will not show a significant change in their field excitatory post-synaptic potential. Figure 3.4 in Chapter 3 represents the effect of a high-frequency stimulus on the field excitatory post-synaptic potential (see also Chapter 3). As any other biological switch, a compromise between (a) robustness to genetic and environmental noise and (b) sensitivity to discriminate meaningful signals, is likely to be embedded in its structure. Furthermore, these
characteristics are expected to be distributed at different levels of biological organisation, requiring an integrative, systems-level approach to its study.

From a genetic control perspective, attractors have been postulated to represent a specific pattern of protein expression that defines the cell’s character (Kauffman, 1969b), and LTP offers an attractive model of biological switch – LTP can be experimentally measured and at the same time its long-lasting effect overlaps through different spatial and temporal scales of organisation.

1.4 Thesis goals

This thesis aims to:

1. Identify additional regulators expressed at different times post-LTP by the use of available tools for the analysis of gene expression data. In particular the classical differential expression methodology is compared with the co-expression networks (Chapters 5 and 6).

2. Study the modules of co-expressed genes to identify the functions that are significantly enriched. Discuss the relevance of central genes and functions within the context of LTP (Chapter 6).

3. Characterize from a dynamical stability perspective the biological networks previously identified (Chapters 7 and 8).

1.5 Thesis outline

The present Chapter 1 offers a brief introduction to the thesis and the motivation behind the experimental work carried out. Aims and problem statement are described. An overview on cellular biology focused on gene expression and microarray technology is presented throughout Chapter 2. LTP is then reviewed in Chapter 3, together with the basic biology of neurons and neural systems. Chapter 4 describes the methodology for gene regulatory networks. The models used in this thesis – random Boolean networks and transcription regulation modelling using weight matrices are described in particular depth. These chapters should be considered introductory since they mostly cover a literature review of the field. Note, however, that Chapter 4 reviews methodologies in gene regulatory networks and hence provides a methodological ground for subsequent chapters. A description of the microarray data used is given in Chapter 5, which also replicates the classical differential expression analysis. The results presented in this chapter do not contain novel findings and are not thoroughly discussed in the thesis although they are referred to when relevant. The aim of the chapter is to describe the data processing conducted with the microarray datasets used as a starting point for the analyses performed subsequently.

Differential expression analysis introduced in Chapter 5 is extended and complemented by a co-expression analysis using the weighted gene co-expression network analysis tool at different times post-LTP induction in Chapter 6. Chapter 7 focuses on a previously identified early network (20 min after LTP induction) by Ryan et al. (2011) and characterises its dynamics by implementing transcription regulation modelling using weight matrices. Results of the stability analysis of the network are described. Chapter 8 further extends the analysis of gene expression underlying LTP by characterising the dynamic regime of the networks identified by Ryan et al. (2011) and Ryan et al. (2012) at different times post-LTP (20 min, 5 h, and 24 h). Conclusions are outlined in Chapter 9.
2. Overview of genetics and microarrays

2.1 Introduction

Just as the vast majority of cell types in mammals, neurons possess a nucleus within their cell body. In the nucleus, the whole genome encodes the blueprint of life. The molecule in charge of storing the inheritable genetic information is the DNA. According to the more classical view, inserted along the linear DNA molecule there are “coding regions” that encode functional proteins. These segments are transcribed to an intermediate molecule (RNA) which is subsequently translated into proteins by complex molecular machineries. Proteins are the actual effectors and in this regard, responsible for the maintenance of the cell homeostasis. Protein functions range from the formation of the structural scaffolding of cells and tissues to the processing and implementation of complex intra- and extra-cellular signals.

The regulation of gene transcription becomes of fundamental importance to understand the different cellular mechanisms. The key role of the regulation of gene expression is exemplified by the differentiation programs that cells undergo, which constitutes the basis of the difference in complexity between humans and lower organisms. Distinct cell types of the same organism differ thoroughly structurally and functionally, yet they possess the same exact copy of DNA. It is owing to the differential gene expression driven by key regulators that cells undergo differentiation to liver cells or neurons during the developmental stages of an individual.

In this chapter an overview on the biology of gene expression is given, together with a brief review on the methods for measuring gene expression levels. The emphasis will be on the DNA microarray technology owing to the scope of the present thesis.
2.2 Biology of gene expression

The necessary information for life is encoded in the DNA, and the complete complement of DNA for a species is referred to as its genome. The genome represents a blueprint for the development and maintenance of a living organism. The molecule of DNA consists of subunits known as nucleotides of which there are four different types: adenine, cytosine, guanine and thymine. The nucleotides are combined linearly in a single long string. With few exceptions, every human cell contains a copy of the whole genome in a cellular compartment called nucleus. Prior to cell division, the DNA replicates so that each of the daughter cells holds a complete copy of the genome.

Proteins, on the other hand, represent the structural and functional mediators between the living organism and the environment. Proteins, like DNA, are constituted of similar units bound in a linear fashion. The units of proteins are called amino acids, and there are 20 different of them. The particular combination of different amino acids in the chain will be reflected in the 3-dimensional structure of the protein. In fact, the different amino acids possess different physicochemical properties and it is owing to their interactions that a particular sequence folds to a particular protein structure. The 3-dimensional structure is responsible for the particular cellular function of the protein.

The concept of gene has changed significantly throughout history. The notion of “one gene-one protein” was originally proposed by Beadle and Tatum in 1941, referring to the fact that one strand of DNA coding for a protein would only make one protein – a one-to-one mapping. However, as a result of alternative splicing, trans-splicing, and other protein modifications that take place within the cell, we know today that rarely that is the case. Furthermore, DNA is not the only unit of inheritance, and some important RNA molecules are in fact passed to daughter cells and function as inheritance units. To make the picture even more complicated, proteins can be built up from different genomic sequences located far apart in the genome. The patterns of regulation at the different levels of biological organisation make difficult to physically delineate the gene. And yet, the operational definitions adopted by the scientific organisations still rely on the sequence – the Sequence Ontology Consortium defines the gene as the “locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions” (Pearson, 2006).

A more recent definition (or reinterpretation) of the notion of gene comes as a computational metaphor. Genes can be thought of as subroutines in the transcription/translation operating system (Gerstein, Bruce, Rozowsky, Zheng, Du, Korbel, Emanuelsson, Zhang, Weissman, and Snyder, 2007). For the sake of clarity, in this thesis we adopt a definition which, although overly simplistic, is still deeply rooted in contemporary narrative. According to such a definition, a gene represents a blueprint for a molecule (generally for a protein) which operates as a functional unit, either interacting with other biological units (DNA, RNA, or protein) or by having an effect on the cell and/or its environment.

During replication, an exact copy of the genome is synthesised by a dedicated cellular machinery. Proteins, however, do not replicate. Based on the cellular needs, they are synthesised using the corresponding DNA segment as a template. From a reductionist point of view, a gene is said to encode a protein, since the information in the DNA strand is translated from the 4-letter alphabet of nucleotides to the 20-letter alphabet of amino acids (the one-to-one mapping). The set of rules that map these alphabets is known as the genetic code. Each consecutive set of three nucleotides (also known as codon) is translated to one of the 20 amino acids. As there are four different nucleotides, there are 64 possible codons mapping to a total of only 20 amino acids. Hence, some of the codons map for the same amino acid (the genetic code is said to be redundant) but a given codon maps to only one amino acid (the genetic code is non ambiguous). Three of the codons are stop codons and signal a termination of translation.

In reality, the synthesis of a protein sequence from DNA passes through another intermediate
2.2 Biology of gene expression

Figure 2.1: The protein structures depicted correspond to collagen (left) and haemoglobin A (right). The corresponding sequences are also represented. Collagen is the most abundant protein in mammals. It is of a fibrous nature and represents the main component of connective tissue. Its function is of structural support. Haemoglobin A is a protein found in the red blood cells of vertebrates and responsible for carrying oxygen from the respiratory organs to the tissues. It is made up of four subunits that bind together forming a pocket that accommodates the so-called heme group. The latter contains an atom of iron that bonds to the protein subunit. Oxygen binds to the heme component during respiration. Sequences and coordinates obtained from PDB database (IDs 1hhb and 1cag, Fermi et al., 1984; Bella et al., 1994).

molecule known as RNA, constituted by similar units (ribonucleotides) and depicted in Figure 2.2. RNA molecules are said to be transcribed from the gene with a complementary mapping – adenosine, cytosine, guanine, and thymine residues of the DNA segment are transcribed into uracil, guanine, cytosine, and adenine respectively to form the RNA string. In contrast to DNA, the messenger RNA (mRNA) molecules transcribed from a gene travel outside of the nucleus. The protein will be synthesised by yet another cellular machinery, the ribosome, by using the aforementioned genetic code mapping table to translate the sequence of mRNA. Although the whole picture is hardly as simple as summarised so far, the fundamental concepts of replication, transcription and translation suffice to give a general intuition of the basic protein synthesis mechanisms needed to understand the following sections.

In summary, the gene expression taking place in the nucleus leads to the synthesis of mRNA which is in turn translated into the functional effectors, proteins. The rate of gene expression of the different genes of the genome is responsible for the cellular levels of mRNA, and hence for the amount of each protein. Morphological and functional differences between cell types, such as liver cells and neurons, are in fact a result of different rates of expression (transcription) of
different key genes. While the expression of the haemoglobin genes is enhanced in the precursor cells that will eventually become mature red blood cells, cells that form the connective tissue will upregulate the expression of collagen. Such different expression profiles in different cell types is the result of a differentiation process from less differentiated cells guided by specific signals. Gene regulation during the differentiation process is crucial. In this regard, the traditional picture of the genotype-phenotype relationship, in which the observable characteristics of an individual are mapped to the particular sequence in his DNA, may be further complemented with the consideration of the expression profiles.

### 2.3 Regulation of transcription

There are a number of mechanisms adopted by living cells to regulate the specific needs in terms of gene expression. These mechanisms can modulate the final concentration levels of a protein at different levels:

- Transcription
- mRNA processing
- mRNA transport
- Translation

In general terms, the control at these different steps is mediated by proteins whose expression levels are, in turn, regulated by other up-regulated genes. Some RNA molecules such as microRNAs and small interfering RNAs are not translated into proteins but can still regulate gene output at
2.4 Cell homeostasis

The internal conditions of multicellular organisms, such as blood temperature in endothermic animals, blood pH and water content, are under strict control. The regulation of these variables is commonly known as homeostasis, and it operates at different hierarchical levels – organs, tissues, and cells, including unicellular organisms. This hierarchical and overlapping organisation of the different control levels poses serious difficulties to the systematic study of the different contribution of each component.

From a cellular perspective, the maintenance of the intracellular variables within their respective ranges includes the concentration of proteins needed to maintain a normal cellular physiology. Protein turnover is compensated by a tightly regulated gene expression profile, which ensures cell survival and, ultimately, organism survival. Regulation of gene expression serves as a homeostatic control mechanism: changes in the environment will be sensed by the appropriate receptor systems and subsequently processed. A response to correct the deviation of the particular variable will be then triggered resulting in changes in gene expression and protein levels. Some authors have

the transcriptional, translational and mRNA-stability level. However, we will focus here on the regulation that occurs at the transcriptional level by a direct effect of regulatory proteins acting on the genome. Later on in the discussion we will introduce the function of the RNA regulators.

At the transcriptional level, the regulation of the cellular concentration of a given protein happens indirectly through the control of RNA synthesis. Firstly, chemical modifications on the DNA, such as methylation and acetylation, are used to unpack the DNA molecules and allow access of the transcriptional machinery. This represents a first coarse level of regulation. Mediating this local uncoiling of the DNA structure are two classes of proteins, the histone acetyltransferases and methyltransferases (Kouzarides, 2007). A perhaps finer control of transcription is carried by the binding of proteins to specific sequences of the DNA, which are generally outside the coding regions. These DNA regions can be at different distances from the actual coding region upon they have an effect, both upstream and downstream of the direction of translation. A protein bound to its corresponding DNA sequence can have an effect on the rate of transcription of only one gene or act on a number of targets. The ultimate expression level of a given gene will therefore be the result of the combined effect of the different regulatory elements acting on it and their strength. The whole expanse of DNA involved in regulating the transcription of a gene is known as gene control region.

Note that the presence of some DNA-binding proteins will influence the target gene by effectively increasing its expression (activators) while some others will lead to the opposite effect, inhibiting gene expression (repressors). In the literature, these proteins are often referred to as transcription factors. Genomic responses to intra- and extracellular signals are represented by the complexity in the regulation of expression, rather than by the absolute number or genes. This regulation is, in fact, a good indicator of the anatomical and developmental complexity of an organism (Pilpel, Sudarsanam, and Church, 2001; Markstein, Markstein, Markstein, and Levine, 2002).

As mentioned earlier in the discussion, the complexity of the regulation further increases due to other regulators acting on the subsequent steps of gene expression. This large class of non-coding RNAs, (the microRNAs) act as repressors on the already synthesised mRNA by binding to it and promoting its degradation (Mattick and Makunin, 2006). The last stage of control happens after the mRNA has been translated into protein. Chemical modifications to the proteins themselves, mediated by other enzymes, can modulate their levels of activity, without directly affecting their cellular concentrations. This post-translational control is carried by kinases and phosphatases, which chemically add or remove, respectively, a phosphate group. These modifications change the activity of the protein by enhancing or inhibiting its function. Other post-translational modifications have an effect on the actual level of protein by promoting their degradation.
proposed the concept of proteostasis to refer to this cellular homeostatic control system (Powers, Morimoto, Dillin, Kelly, and Balch, 2009; Balch, Morimoto, Dillin, and Kelly, 2008).

Whereas housekeeping proteins tend to show similar expression profiles throughout different cell types, the homeostatic expression profiles for different tissues will differ according to the specific functional and structural needs of the tissue. Similarly, some genetic cellular subsystems will be activated only under particular circumstances, such as changes in metabolism. These changes in gene expression profiles between different cell states can be transient or last for long periods of time. Long-lasting changes such as the one triggered by the high frequency stimulus leading to LTP require the gene expression machinery to shift from one homeostatic state to another in a semi-permanent fashion. In this respect, an interesting hypothesis was raised by Kauffman (1969a) in 1969 – If we consider gene expression as a dynamic system, cell fates can be seen as high-dimensional attractor states in the gene expression space. The long-lasting changes following LTP can be interpreted from such a perspective.

The evident complexity of the circuitry integrated by the different cellular mechanisms remains a major technical difficulty to approach by classic experimental means. There exist a range of methods to infer, model, and analyse gene regulatory networks, and some will be described further on in Chapter 4.

2.5 Gene expression profiling

According to the scheme depicted in Figure 2.2, gene expression consists in the conversion of the information stored in the coding DNA segments into the actual protein, via the intermediate mRNA. The cellular concentration of a given protein is hence dependent on the rate of gene transcription and protein turnover. Mass spectrometry and two-dimensional gel electrophoresis are the main methods to quantify protein level. However, compared to measuring protein levels, it is technically easier to profile the amount of RNA. While in some cases mRNA levels are considered as an indirect measurement of protein levels, this is an oversimplification which, especially in the case of neurons, has to be taken into consideration as will become clear in Chapter 3.

The experimental techniques available nowadays to measure RNA levels in a cell can be classified according to different criteria. In terms of the overall strategy, some techniques rely on sequencing the RNA output, such as the serial analysis of gene expression (SAGE; Velculescu, Zhang, Vogelstein, Kinzler, et al., 1995) and its variants (Matsumura, Ito, Saitoh, Winter, Kahl, Reuter, Krüger, and Terrauchi, 2005), RNA-seq (Khan, Wei, Ringner, Saal, Ladanyi, Westermann, Berthold, Schwab, Antonescu, Peterson, et al., 2001), and massively parallel signature sequencing (MPSS; Brenner, Johnson, Bridgham, Golda, Lloyd, Johnson, Luo, McCurdy, Foy, Ewan, et al., 2000). Other techniques are based on hybridization of RNA to probes. This is the case of microarrays (Schena, Shalon, Davis, and Brown, 1995; Schulze and Downward, 2001), Northern Blot (Alberts, 2008), and quantitative polymerase chain reaction (qPCR; VanGuilder, Vrana, and Freeman, 2008). Microarrays are significantly cheaper to perform on large-scale studies and hence generally preferred in genome-wide experiments.

In terms of the number of genes that can be studied in each assay, the different techniques can be classified into low- or high-throughput. The former allows the study of a limited number of genes at the same time, while the second group embodies those technologies with the potential to encompass the whole transcriptome in the same assay. From the high-throughput technologies, microarrays and RNA-seq are by far the most widely used nowadays.

2.6 Oligonucleotide microarrays

A microarray consists of a set of DNA spots attached to a solid surface which can hybridise in a specific manner to complementary sequences in the sample. The hybridisation can be detected and
2.7 Analysis of microarray data

2.7.1 Low-level analysis

Low-level data analysis deals with extraction of the signal from the high level of noise and experimental artefacts, and includes image analysis, gene filtering, background correction, probe level analysis and gene summarisation. Although the present thesis focuses on the high-level analysis of microarray data, we will briefly summarize some basic low-level analysis procedures.

Following the experimental procedure for hybridization, the array is scanned and the resulting image (the DAT file) is inspected for defects. In general terms, if these defective areas represent less than 10% of the total probes, the areas can be masked as outliers (Li and Wong, 2001). The probe array image is then analysed to obtain a set of expression indices associated to each probeset. The probe intensity data for the chip is recorded in a text file (the CEL file), of which there will be one quantified to measure the expression levels of large numbers of genes (Heller, 2002). While it is possible to prepare custom microarrays in regular research facilities, the popularity of microarrays owes to the availability of commercial platforms, mass-produced by industrial companies. The precision of the commercially available microarrays allows the comparison between two (or more) identical arrays hybridized with different samples. This parallel microarray experiments can be used to measure the relative abundance of RNA in different tissues or different conditions, such as developmental stages or carcinogenic versus healthy tissues. The gene expression profiles constructed using parallel experiments are represented in a matrix where the rows represent genes and the columns represent samples. The cells of the matrix contain the values of expression of a single gene (Schulze and Downward, 2001).

The term “oligonucleotide microarrays” is often used to refer to the *GeneChip* class of microarrays developed by *Affymetrix*. They consist of a solid surface onto which a collection of DNA spots are attached. Each of these spots contains molecules of DNA with a specific sequence that corresponds to a section of a particular DNA element. Among those elements, many of them are complementary gene-coding sequences. Typically, microarray platforms may have tens of thousands of spots in a single array, and this size makes it possible to detect and quantify the expression of every known gene in the genome.

In an Affymetrix oligonucleotide microarray experiment, the RNA from the sample is extracted, purified, and subsequently transcribed back into its complementary DNA sequence (cDNA). Next, the second-strand of the cDNA is synthesised and the purified double-stranded cDNA is used as a template for an in vitro transcription reaction in the presence of labelled ribonucleotides. After filtering out the double-stranded cDNA, the resulting labelled RNA is fragmented and hybridized onto the microarray. The molecules will bind to their complementary probes attached onto the surface of the array. The strength of the signal from a spot will be proportional to the amount of cDNA present in the sample (Schena *et al.*, 1995; Brazma and Vilo, 2000). A summary is provided in Figure 2.3.

Each probeset is actually comprised of two signals – a set of 11-20 “PM” (perfect match) probes with a length of around 25 nucleotides, and a “MM” (mismatch) signal, resulting from another set of probes with a single nucleotide substitution in the centre of the probe, which can be used to account for non-specific probe binding. In fact, the average of the PM-MM differences for all probe pairs in a probe set (“average difference”) is generally used as an expression index. Indeed, this is often the starting point for the so-called high-level analysis (see Section 2.7.2).

These platforms capture an additional layer of complexity which cannot be grasped solely by gene sequencing. The capacity to carry a systematic and comprehensive analysis of gene expression in a fast and reproducible manner, together with its relative low costs, makes of microarray technologies a remarkably valuable tool.
Chapter 2. Overview of genetics and microarrays

Figure 2.3: The microarray (a) contains the probes which will hybridise the complementary labelled samples (b). After washing off the non-specific bonding sequences, the strength of the signal of a spot will be proportional to the amount of sample binding to the probes on that spot.

for each sample. An additional Chip Description File (the CDF file) summarises the information for each probe cell on the chip.

The preprocessing of the CEL file is accomplished in three subsequent steps:

- Background correction (adjust for random noise)
- Normalisation (calibrate measurements of different arrays)
- Summarisation (summarise the PM/MM levels for each probeset to a single expression value)

A number of available tools perform this analysis in an automated fashion requiring minimal user input. The most significant difference between the tools resides in the normalisation procedure used. Some methods, such as MAS 5.0, the default measure available from Affymetrix, and dChip (Li and Wong, 2001), are based on PM-MM. In contrast, another set of methods base the normalisation on PM values only. This is the case of robust multi-array average (RMA) (Irizarry, Hobbs, Collin, Beazer-Barclay, Antonellis, Scherf, and Speed, 2003) and GCRMA (Wu, Irizarry, Gentleman, Murillo, and Spencer, 2004). The choice of normalisation procedure is guided by the microarray platform and array design, the experimental design, and the underlying biological assumptions. The R package affcomp (Cope, Irizarry, Jaffee, Wu, and Speed, 2004) facilitates the comparison of expression measures and the selection of methods.

RMA is one of the most widely used methods. The RMA background normalisation method only considers PM probe intensities as the sum of signal and noise. The normalisation on the PM probe intensities is carried out by quantile normalization, and the summarisation of the values into expression values is done using the median polish method. The implementation of the RMA
2.8 Microarrays in neuroscience  

method is included in the *affy* R package of the Bioconductor project (Gentleman, Carey, Bates, Bolstad, Dettling, Dudoit, Ellis, Gautier, Ge, Gentry, et al., 2004).

In Chapter 5 we will describe the microarray datasets in which part of the present thesis is based. Additionally, we will discuss the results of the RMA analysis in some depth. The high-level analysis discussed in Chapter 6 uses the microarray expression values normalised and summarised using RMA as a starting point.

### 2.7.2 High-level analysis

Following the summarisation step, the intensity values of the probesets in the array have been translated into single gene expression values. As mentioned earlier, this often represents the starting point for further high-level analyses concerned with extracting knowledge about the underlying biological processes. These analyses are often based on the identification of differentially expressed genes. While there exist several tools that perform this function, LIMMA (Smyth, 2005) is one of the most widely used. Results of this algorithm with an microarray dataset are shown in Chapter 5.

The LIMMA analysis takes a linear modelling approach to handle the experimental design, fitting a linear model to the expression data for each gene. The output is a set of $\log_2$ fold change values which indicate the degree of change between the two samples. It uses a moderated t-test to estimate the p-values and calculates the false discovery rate (FDR). Both measures take into account multiple testing bias (an indication of the frequency of false positives among the significant hypotheses). The implementation of the LIMMA analysis can be found as an R package within de Bioconductor project (Smyth, 2005; Gentleman et al., 2004).

Differential expression analysis can be used as a starting point for further studies, since it allows to rank the genes according to their putative relevance in the biological processes linked with the different conditions of the samples studied. In this regard, differential expression analysis can be seen as a middle-level analysis. A rather large number of methods exist to study the gene expression matrix obtained from microarrays. In general terms, the strategies aim to compare the expression profiles either in rows or columns of the expression matrix, although both methods can be combined (Brazma and Vilo, 2000). Commonly, the studies that compare microarrays of different conditions correspond to different types of tissues. For instance, to understand the gene expression mechanisms of a particular type of cancer, the disease and non-disease samples will be used to hybridise different arrays (or use different fluorescent dyes). The differential expression analysis will offer valuable information on the physiological differences between carcinogenic tissue and healthy tissue. These types of parallel analyses are not limited to different tissues. For example, investigations on temporal changes in gene expression can also be tackled.

The comparisons between genes are based on the similarity of their profiles. The Euclidean distance or correlation measures are often used to measure the distances, although other measures could have been used as well (e.g. rank correlation coefficient, mutual information; Smith, 2007). The data can subsequently be analysed by unsupervised methods to cluster genes and/or samples. Supervised methods are often used to build classifiers. The underlying assumption is that genes with similar expression profiles, may be co-regulated.

Other high-level analyses of microarray data include the integration of data from other sources, such as the one from function annotation databases.

### 2.8 Microarrays in neuroscience

The brain functions as the central processing unit in vertebrates, integrating the information collected from the environment. The complexity of the gene networks involved in the brain processes adds an additional layer of organisation that exchanges information with the highly organised synaptic networks. The study of gene expression in the brain aims to shed light on the mechanisms of brain
functioning as well as to gain a deeper understanding on the processes that lead to pathological states.

The capability of analysing a whole genome in parallel allows us to tackle the study of these complex systems. In this regard, the study of the transcriptional regulatory networks associated with mechanisms involved in learning and memory represents a step towards bridging the classic methodologies applied by neuroscientists with the molecular and systems neuroscience (Geschwind and Konopka, 2009). In fact, since the very first microarray experiments described by Schena et al. (1995), this technology has gained an increasingly important role in neuroscience (see Figure 2.4).

Figure 2.4: The main plot represents the number of scientific papers returned by a PubMed database search with the word “microarray” on the title or abstract from 1995 to the current day. The orange bars in the inset plot represent the subset of those papers in the field of neuroscience.
3. Biology of long-term potentiation

“For all that let me tell thee, brother Panza,” said Don Quixote, “that there is no recollection which time does not put an end to, and no pain which death does not remove”

Miguel de Cervantes

3.1 The neuron

Neurons represent the unit of transmission of information in the nervous system. Information travels along their axons as electric potential in their membranes. From a morphological point of view, neurons are made of a soma (cell body), dendrites, and the axon. Brain tissue is constituted of two cell types – neurons and glial cells. There are several classes of each of these cell types, and have different functions within the nervous system. The glial cells provide physical and nutritional support to the neural cells, and in addition they can play active roles in synaptic plasticity. In particular, glial cells of the hippocampus and cerebellum have been shown to actively participate in synaptic transmission (Newman, 2003; Koob, 2009).

Although mitochondria – the cellular “powerhouses” – are found in axon terminals and dendrites, most of the metabolic processes take place in the soma. The soma accommodates the cell nucleus, which contains the genetic information of the neuron. The dendrites are protrusions of the cell membrane that receive the inputs from incoming connections from other neurons (Figure 3.1).

In the nervous system, neural cells are connected forming neural circuits – nets of interconnected neurons that allow the organism to integrate the information gathered from the environment and generate specific responses. The nervous system, in addition, allows for the storage and retrieval of information, both implicit and explicitly.

3.1.1 The resting potential

Neurons keep a gradient of electrical charges between the cytoplasm and the exterior. This resting membrane potential is around $-70mV$ (negative inside) and it is actively maintained by the cell.
Chapter 3. Biology of long-term potentiation

Figure 3.1: Pyramidal neuron from the hippocampus. In blue, the tree-like structures known as dendrites carry information from other neurons. The information is then integrated eventually leading to an action potential which travels along the axon (in red) to other neurons. The arrows indicate the direction of information flow.

The resting membrane potential is the result of an equilibrium in the passive flow of ions through the membrane. Neurons are permeable to $K^+$, $Na^+$, and $Cl^-$, but the permeability differs for each of these ions. Protein channels that allow the flow of the ions are embedded in the membrane of the cell, and their different concentration leads to different permeabilities. The flow of ions across the membrane is dependent on two factors – the unequal concentration of the ions on each side (chemical driving force) and the difference in potential across the membrane (electrical driving force). The flow of a particular ion depends not only on the sum of the chemical and electrical driving forces, but also on the membrane permeability to that particular ion.

Given the concentration gradients in the cell, the total flow of ions across the membrane would tend to reach an equilibrium in which $K^+$ ions leave the cell (the concentration of $K^+$ is significantly higher inside the cell) and $Na^+$ ions enter the cell. To compensate this concentration gradient leak, neurons actively transport these ions against their concentration gradients using a special channel known as sodium-potassium pump. In contrast to the channels described previously, the sodium-potassium pump is not a passive channel and needs cellular energy in the form of ATP.

$Ca^{2+}$ is a fundamental ion in the neurons. Its concentration is considerably inferior in the cytoplasm than in the extracellular fluid. In order to maintain its the gradient across the membrane, there are several types of channels that transport the $Ca^{2+}$ out of the cell or hoards it in intracellular stores. Similarly, the $Cl^-$ anion is actively transported to the extracellular fluid.

3.1.2 The action potential

The transmission of information is encoded in electrical pulses that travel throughout the length of the axons. This depolarisation is known as “action potential” and propagates along the axon without attenuation, self-regenerating until reaching the synaptic terminals. There, the axon transmits the
information to other neurons.

For the action potential to happen, the depolarisation has to reach a threshold. Small depolarising currents can increase the resting potential from $-70mV$ to that threshold, which lies around $-50mV$. The action potential differs from other graded biological responses in that it is an *all-or-none* event – once the threshold is reached, the signal is transmitted without decrement.

The action potential is the result of the presence of Na$^+$ and K$^+$ channels in the membrane, whose opening is regulated by the local potential of the membrane surrounding them. These type of channels are known as voltage-gated channels. The depolarisation of the membrane above the threshold increases the permeability to Na$^+$ by opening these voltage-gated channels, leading to a further depolarisation. This feedback cycle is responsible for initiating the action potential. The new membrane potential acquired after the action potential takes place is restored by the flow outwards of K$^+$, eventually reaching the initial charge separation. This flow is driven by voltage-gated K$^+$ channels, whose opening times are slower than the voltage-gated Na$^+$ channels.

In addition to the open and closed states, the voltage-gated Na$^+$ channels have an inactivated state. Following a sufficiently large depolarisation, these channels become unable to open until the membrane restores its resting potential. This three-state switch is responsible for the refractory period, which limits the frequency of action potentials (Figure 3.2).

**Figure 3.2:** *(a)* Action potential propagates along the axon reaching the synaptic terminals, where the signal is passed to postsynaptic neurons. *(b)* Temporal membrane permeability changes with the onset of the action potential. Positive feedback by the entry of Na$^+$ and the resulting Na$^+$ voltage-gated channel opening generates a fast depolarisation of the membrane. Eventually, the slow-opening voltage-gated K$^+$ channels increase the permeability of the membrane to K$^+$ restoring the resting potential. Figure *(b)* based on Hodgkin and Huxley (1952)

In addition to the channels briefly described here, there are other channels of similar properties which are permeable to other ions such as Ca$^{2+}$ or Cl$^-$. There are multiple variants that bestow specific properties to the action potential in different types of neurons.
3.1.3 The chemical synapse

The information transmission between different neurons takes place in the synaptic terminals, where the action potential reaches from the presynaptic neuron and is passed over to the postsynaptic neuron. The action potential can cross directly to the other membrane (electrical synapse) or cross by chemical intermediates (chemical synapse). The electrical synapses communicate very fast, but chemical synapses are more common and have the fundamental property of signal amplification (Kandel et al., 2000).

The chemical synapses are dependent on the release of a neurotransmitter molecule from the axonal ending of the presynaptic cell to the extracellular space between the cells. The molecule diffuses reaching specific receptors in the postsynaptic membrane. The release of neurotransmitter depends on the opening of Ca$^{2+}$ channels in the presynaptic membrane, which in turn lead to the fusion of vesicles stored in the axon terminal and loaded with the neurotransmitter molecules. The eventual binding of the neurotransmitter to the receptors in the postsynaptic cell drives the opening of other transmembrane receptors, changing the postsynaptic membrane potential. There are two types of neurotransmitter receptors, ionotropic receptors and metabotropic receptors (Nicoll, Malenka, and Kauer, 1990). Ionotropic receptors form an ion channel pore, while metabotropic receptors act through secondary messengers, activating intracellular signalling cascades that can lead to changes in membrane permeability or other cellular effects.

Ionotropic receptors (Wisden and Seeburg, 1993) are made of several subunits that form a ion channel. The extracellular domain of the channel functions as a binding site to the neurotransmitter. Upon binding, the channel undergoes a conformational change that permits the transit of ions through its transmembrane domain. Ionotropic receptors act faster, since they combine transmitter-binding and channel functions in the same molecular entity.

Ionotropic receptors are generally subdivided into 3 superfamilies:

1. Receptors that resemble the nicotinic acetylcholine receptors
   - Glycine receptors
   - Gamma-aminobutyric acid A (GABA_A) receptors
   - Nicotinic acetylcholine receptors
   - Some serotonin receptors

2. Ionotropic glutamate receptors (GluR)

3. ATP-gated purine receptors

The other family of neurotransmitter receptors are known as metabotropic receptors (Traynelis, Wollmuth, McBain, Menniti, Vance, Ogden, Hansen, Yuan, Myers, and DINGLEDINE 2010). Their contribution to the membrane permeability to ions is indirect, through series of metabolic steps. Upon binding of the neurotransmitter in the extracellular domain, the coupled G-proteins become active and dissociate from the receptor. Activated G-proteins can directly activate transmembrane channels or activate other effector proteins which will, in turn, regulate other postsynaptic channels. The activation of metabotropic receptors can trigger complex cellular responses through the production of second-messengers, diffusible molecules able to activate other proteins.

Neurons in the central neural system receive excitatory and inhibitory inputs depending on the type of receptors located in the postsynaptic membrane. In the brain of vertebrates, presynaptic neurons that release the glutamate neurotransmitter act on neurons that carry excitatory receptors in their membranes, while neurons that release GABA or glycine act on neurons that carry receptors that cause inhibition. The information coming from a number of neurons is integrated to generate an action potential in the postsynaptic neuron.
3.1.4 The excitatory synaptic transmission

The major excitatory transmitter in the brain is L-glutamate which binds to specific ionotropic and metabotropic receptors. There are three subtypes of ionotropic receptors for glutamate:

1. α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor
2. N-methyl-D-aspartate (NMDA) receptor
3. Kainate

While open, all these allow the flow of Na\(^+\) and K\(^+\). NMDA receptors, in addition, are permeable to Ca\(^{2+}\). Their active structure in the membrane is in the form of a tetramer, an oligomer formed from four protein subunits. Among the three types of receptors, kainate receptors are the less understood, but they may have an effect on the amount of neurotransmitter released presynaptically (Schmitz, Mellor, and Nicoll, 2001).

There are four different types of subunits of AMPA receptor (GluR1, GluR2, GluR3, and GluR4) that combine in two dimers (an odimer of GluR1 and either a dimer of GluR2, GluR3 or GluR4). The differences in the subunits determine their interactions with other proteins (Greger, Ziff, and Penn, 2007). In the case of the NMDA receptor, the tetramer is formed by two GluN1 subunits and two GluN2 subunits. There are eight variants of GluN1 and 4 of GluN2, and the different combinations confer the channel different kinetic properties (see Teng, Cai, Zhou, Zhang, Liu, Wang, Dai, Zhao, and Sun (2010) for a recent review). NMDA receptor is considered the predominant molecular device for controlling synaptic plasticity and memory function (Li and Tsien, 2009).

While the non-NMDA receptors open upon neurotransmitter binding, the NMDA receptors require both the binding of glutamate and the postsynaptic membrane depolarisation to open. The postsynaptic depolarisation when at resting potential is, hence, due mainly to the action of the AMPA receptors. The NMDA receptors act as detectors of the co-occurrence of glutamate release and postsynaptic depolarisation. When NMDA receptors open, they allow the flux of Ca\(^{2+}\) from the extracellular space to the cytoplasm. This rise in Ca\(^{2+}\) activates a set of signalling cascades that can lead to long-lasting changes in the synaptic efficacy. The ability of modulating the efficacy of synaptic transmission is involved in the formation of some forms of memory. In most neural pathways in the mammalian hippocampus, LTP is NMDA receptor-dependent (see Section 3.2.3).

3.2 Long-term potentiation

3.2.1 The discovery of long-term potentiation

The archetypical case of H.M. in the 1950s commented in Chapter 1 was influential for its contribution to the understanding of human memory organisation in the brain, pinpointing the anatomical location of explicit memories in the hippocampus. About 20 years later, in 1973, Timothy Bliss and Terje Lømo, discovered what today is considered the cellular mechanism underlying long-term memory – LTP. Their experiments demonstrated that while a single pulse of electrical stimulation to the presynaptic cells caused an excitatory postsynaptic potential (EPSP), a high-frequency train of stimuli enhanced the subsequent EPSPs to single stimuli. This effect lasted for long periods of time. Effectively, a long-lasting change in the synaptic strength was being induced by the high-frequency stimulus (HFS).

Although the discovery of LTP is placed within a scientific context where several research groups participated actively to the development of such a cellular model, Lømo has been credited as the “discoverer” of LTP. Scientific literature often refers to the seminal papers of Bliss and Lømo (1973) and Bliss and Gardner-Medwin (1973) as the keystone, although the discovery began earlier in Oslo in 1964. According to Lømo himself (Lømo, 2003), the first findings related with what
would be an intense area of research later on were presented in a meeting of the Scandinavian Physiological Society in 1966 (Lømo, 1966), and the first mention to the “possible relationship between frequency potentiation and learning processes” came in a paper prepared for a meeting in 1965 by Andersen and Lømo (1967).

It is noteworthy that the results of the refined experiments after the 1966 meeting would not have been published until 1973 and, more interestingly, that there was a “relative lack of enthusiasm expressed by most people upon hearing the results”. Apparently, on Lømo’s own opinion, both the scientific community and himself were overcome by the complexity of the system, and the lack of understanding of what was behind the findings (Lømo, 2003).

By the late 1970s, LTP-related research was being conducted worldwide, and in 1983 the central role of NMDA receptor was revealed by the work of Collingridge and coworkers (Collingridge, Kehl, and McLennan, 1983; Coan and Collingridge, 1985; Collingridge and Bliss, 1987). At the same time, the hippocampus was gaining attention as a crucial structure for spatial learning Morris, Garrud, Rawlins, and O’Keefe (1982); Morris (2003).

Despite the early lack of interest shown by the scientific community, the results published in 1973 in the aforementioned paper represent the first detailed description of LTP, and 40 years later it has been referenced in more than 3,000 publications. Indeed, LTP is still today the focus of intense research. The modern studies fall either into a mechanistic perspective, seeking to clarify the cellular biology of the phenomenon – or into a functional perspective, aiming to shed some light on the link between LTP and behavioural learning. With the advent of the computer era, new lines of research have ventured into the study of LTP by means of computational models (Benuskova, 2000; Benuskova and Kasabov, 2007). Many of the fundamental questions, however, remain yet unanswered.

### 3.2.2 Definition and properties

In a review by Malenka and Bear (2004), the LTP definition is generalised as the “long-lasting enhancement of synaptic effectiveness that follows certain types of tetanic electrical stimulation”. This enhancement of the synaptic strength is measured as the percent change in the EPSPs.

Although LTP was discovered in the hippocampus – more precisely in the synapses formed from the inputs of the perforant path fibres to the granule cells of the dentate area – other structures in the brain have been shown to possess the ability to form LTP, including the cerebral cortex, cerebellum, and amygdala. This has raised the possibility that LTP may potentially occur in all excitatory synapses in the mammalian brain (Malenka and Bear, 2004). However, the best studied form of LTP in terms of the amount of experimental work has been performed in the synapses between the Schaffer collateral and commissural axons and the apical dendrites of CA1 pyramidal cells of the hippocampus – an associative, NMDA receptor-dependent LTP (Reymann and Frey (2007), see Section 3.2.3).

The majority of the biochemical data available comes from LTP induction in the CA1 region of the hippocampus, and is therefore considered as the prototypic form of synaptic plasticity NMDA receptor-dependant. In such context, LTP displays some properties that have been classically considered as “suggestive of an information storage device” (Martin, Grimwood, and Morris, 2000):

1. **Input specificity**
2. **Associativity**
3. **Persistence**

*Input-specificity* refers to the fact that when LTP is generated at one particular set of synapses by
repetitive activation, the corresponding increase in synaptic strength does not necessarily occur in other synapses, even in the same cell. This property may be important in terms of storage capacity. 

**Associativity** refers to the observation that when stimulation in a set of synapses is not enough to trigger LTP, the latter can be facilitated by a strong activation of an independent set if both sets are activated simultaneously.

**Persistence** is the ability of LTP to last from several minutes to many months, and is this persistence that separates LTP from other short-term plasticity mechanisms.

### 3.2.3 Types of LTP

The input to the hippocampus comes from neurons of the entorhinal cortex through the so-called perforant pathway, which connects with the granule cells of the dentate gyrus. The latter make synaptic contacts with the pyramidal cells of the CA3 region (the mossy fiber pathway). In turn, the axons of the pyramidal cells of the CA3 region are connected to the pyramidal cells of the CA1 region via the Schaffer collateral pathway (see Figure 3.3). The importance of each of these pathways is highlighted by the fact that damaging any of them is sufficient to produce memory disruption in humans (Zola-Morgan, Squire, and Amaral, 1986).

The dependence on the frequency of the presynaptic activity encodes the long-term changes in synaptic efficacy at excitatory glutamatergic synapses. The frequency of the local $\text{Ca}^{2+}$ concentration at the postsynaptic spines will vary according to the presynaptic activity, and consequently distinct forms of potentiation can be triggered in the synapse. This is particularly evident in the extreme case of low-frequency stimulation, which triggers a completely different form of long-lasting change in synaptic strength known as long-term depression (LTD). While out of the scope of the present thesis, it is worth mentioning that while both LTP and LTD are triggered by increased $\text{Ca}^{2+}$, the distinct frequencies will be decoded to initiate the pertinent cellular and molecular changes. The cell, hence, requires a mechanism to discriminate these different temporal patterns.

The mechanism underlying the LTD/LTP decision relies on a different activation of NMDA receptors, which in turn leads to different increases in postsynaptic $\text{Ca}^{2+}$. A small increase in intracellular $\text{Ca}^{2+}$ triggers LTD, while a stronger activation of NMDA receptors and the resultant rise in $\text{Ca}^{2+}$ leads to LTP (Bear and Malenka, 1994).

The forms of LTP in the different pathways possess different cellular and molecular characteristics. Moreover, LTP seems to be a family of different processes that alter the synaptic strength in the different areas of the hippocampus. The two main classes of LTP in the hippocampus are **associative** and **nonassociative** LTP.

In the mossy fiber pathway (see Figure 3.3), an example of nonassociative LTP takes place. In this class of LTP, only AMPA receptors have a significant role in synaptic transmission – experimentally blocking NMDA receptors has no effect on the onset of LTP (Zalutsky and Nicoll, 1990). On the contrary, LTP induced in the Schaffer collateral pathway is associative, requiring activity in the postsynaptic cell to activate the NMDA receptors (Lynch, 2004; Nicoll and Malenka, 1995; Larkman and Jack, 1995). Hence, questions regarding the functional roles of LTP must be placed into very specific contexts.

It is important (and reassuring to a certain extent) to realise that despite the repertoire of long-lasting modifications and the complexity of the underlying molecular mechanisms, the synaptic changes undergone by LTP are considered to be a fundamental property of the majority of excitatory synapses in the mammalian brain (Malenka and Nicoll, 1999). A deeper understanding of a prototypic LTP will definitely help to uncover not only the common mechanisms among the different subtypes, but also point towards the differences and their functional roles in different regions of the brain. As pointed out by Malenka and Bear, “just as different neurons express different complements of ion channels to control their firing properties, neurons can vary in terms of the specific forms of LTP and LTD they express” (Malenka and Bear, 2004).
Chapter 3. Biology of long-term potentiation

3.2.4 Experimental induction

The experimental induction of LTP consists in the implantation of stimulating electrodes in the input pathway and similar recording electrodes in the postsynaptic area. The position of the post-synaptic recording electrodes should produce a maximal positive-going field EPSP (fEPSP). If the procedure is being conducted in vivo, the animals are given a recovery time, after which sets of baseline recordings be made (Abraham, Logan, Greenwood, and Dragunow, 2002).

The induction of LTP follows a previous baseline recording period and is triggered by the HFS. Different areas require different patterns of HFS protocols, and even in the same areas, it is possible to elicit different types of LTP by different HFS. In general terms, the stimulation needed to elicit LTP consists of a number of trains of pulses applied in sets. A train is defined as a series of pulses at a particular frequency and for a certain amount of time. As an example, the microarray data used in this thesis was originally obtained by Ryan et al. using 50 trains of 400Hz stimulation (250µs pulse duration), 10 pulses/train, in sets of five trains 1s apart, with 60s between sets, delivered to the ipsilateral perforant pathway, which is the main excitatory input of granule cells in the dentate gyrus. This protocol was developed based on electroencephalography signals observed in the hippocampus of freely moving rats, and has a physiological meaning.

The HFS results in the induction of LTP. The recording electrodes show an increase in the efficiency of synaptic transmission together with an increase in the excitability of the cell population. Figure 3.4 shows how the HFS affects the size of the fEPSP as measured by the recording electrodes.
3.2 Long-term potentiation

3.2.5 Phases of LTP

There are at least three mechanistically and temporally distinct phases following HFS in hippocampal neurons. Short-term potentiation (STP), lasting 15-30 min, early-LTP (E-LTP), stable for up to 2-3 h, and late-LTP (L-LTP), which in some cases can last up to 6-8 h in hippocampal slices, months in vivo. While these mechanisms are initiated by the same extracellular neurotransmitter, the underlying molecular mechanisms differ. Proteins and mRNAs critical for earlier phases are likely to be pre-existing and turn over slowly. As noted from the earlier studies on LTP, the molecular species responsible for L-LTP and long-term memory must be either induced or, if constitutively expressed, only transiently accessible to modification (Goelet et al., 1986).

STP is typically restricted to the first 20-30 minutes, with a gradual decay that uncovers the overlapping E-LTP. This is, however, dependent on the stimulation protocol and varies notably from experiment to experiment. The magnitude of STP depends on the frequency of the HFS, and the duration depends on the number of stimuli used during HFS and the frequency of low-frequency stimuli that are delivered following its induction, generally applied to monitor the level of synaptic plasticity during experiments. This implies that STP can be stored (Volianskis and Jensen, 2003). In a recent review (Park, Volianskis, Sanderson, Bortolotto, Jane, Zhuo, Kaang, and Collingridge, 2014), the authors speculate that STP is expressed by an alteration in the probability of transmitter release, while LTP is expressed by a postsynaptic component.

Although STP has classically been characterised as a single mechanism and sometimes even considered inconsequential, an in vitro study that employed a range of different NMDA receptor antagonists differing on their preferring type of NMDA receptor subunits, distinguished a total of three pharmacologically different components of synaptic potentiation – STP1, STP2, and LTP (Volianskis, Bannister, Collett, Irvine, Monaghan, Fitzjohn, Jensen, Jane, and Collingridge, 2013).

Similarly, LTP has at least two distinct components, which seem to be the result of different
stimulation patterns – single-episode or compressed HFS versus spaced HFS (Woo, Duffy, Abel, and Nguyen, 2003). These patterns elicit E-LTP and L-LTP respectively, which classically have been interpreted as temporally distinct phases. E-LTP is independent on PKA activity and protein synthesis contrarily to L-LTP, that depends on PKA activation and protein synthesis.

The inherent complexity of memory storage processes at the different levels of organisation result in a plethora of different methodologies and experimental approaches applied to the study of LTP. The extensive literature on LTP available as per today reflects these difficulties. It is well known that there are different subtypes of LTP that can be induced in the same synapses by distinct induction protocols, brain regions, or developmental stages. While modulatory molecules can play fundamental roles in some aspects of LTP, such as long-term morphological changes or processes specific to each subtype of LTP, it is of a practical and theoretical interest to pinpoint the core elements necessary and sufficient for the establishment LTP. The following sections represent a starting point for further improvement and do not preclude the possibility that other forms of LTP may differ in some of its key components or transduction mechanisms. However, as stated previously, our study is restricted to the NMDA receptor dependent LTP induced at perforant path synapses using multiple, spaced trains of stimuli and analysing molecular changes in the dentate gyrus. The following discussion is articulated around these core molecules and transduction pathways. The most studied downstream elements will also be discussed.

3.3 Molecular mechanisms of LTP

It is well accepted that E-LTP requires synaptic activation of postsynaptic NMDA receptors. Repetitive tetanic stimulation leads to the depolarisation of the postsynaptic cell and the subsequent activation of the NMDA receptor (Huang and Kandel, 1994). The mechanism underlying the triggering of the early phase requires both AMPA receptor and NMDA receptor activation.

While the depolarisation in the postsynaptic cell caused by AMPA receptors upon glutamate binding is short-lived, the summation of EPSPs will eventually lead to the dissociation of the Mg\(^{2+}\) cation blocking the NMDA receptor allowing Ca\(^{2+}\) as well as Na\(^{+}\) to enter the postsynaptic cell. This rapid rise in intracellular Ca\(^{2+}\) concentration activates the enzymes mediating E-LTP induction. It is thought that a short-lasting threshold level of Ca\(^{2+}\) must be reached to trigger LTP (Svoboda and Mainen, 1999; Malenka and Nicoll, 1999). Note, however, that whereas the increase in postsynaptic Ca\(^{2+}\) concentration is a necessary element for the induction of LTP, NMDA receptor activation can be experimentally bypassed (Kauer, Malenka, and Nicoll, 1988), and therefore may not be necessary for all forms of LTP.

L-LTP, as well as long-term memory, is dependent on changes in gene expression and protein synthesis (Nguyen, Abel, and Kandel, 1994; Abraham and Williams, 2003, 2008; Alberini, 2009). Some authors further subdivide L-LTP into two mechanistically different phases according to the dependence on new gene transcription. The first phase (some times referred to as LTP2) requires of the regulation of translation of mRNA already present at the synapses while the second (LTP3) depends both on new gene transcription and protein synthesis (Raymond, 2007; Abraham and Williams, 2008).

The new transcripts and proteins that characterise LTP3 could be targeted to the specific synapses from the nucleus by means of a synaptic “tag” generated at the potentiated synapses to maintain their specificity (Frey and Morris, 1997). The rise in intracellular Ca\(^{2+}\) activates second messenger systems that converge (via kinase activity) to phosphorylate in the nucleus a set of constitutive transcription factors such as NF-\(\kappa\)B, CREB, Egr1, and serum response factor (SRF). The activation of constitutive transcription factors (CTFs) brings a first wave of de novo gene transcription consisting of effector genes and inducible transcription factors (ITFs). In turn, ITFs are responsible for a second wave of transcription (Ryan et al., 2012; Abraham and Williams, 2003).
3.3 Molecular mechanisms of LTP

3.3.1 The CaMKII pathway and early LTP

Temporal decoding of intracellular Ca\(^{2+}\)

The concentration of Ca\(^{2+}\) in the postsynaptic cell encodes both temporal and spatial information which has to be decoded reliably. In addition to being decoded, the information contained in the transient intracellular elevation of Ca\(^{2+}\) has to be translated into a prolonged physiological response. Experimental evidence seems to indicate that CaMKII may act as the decoding molecule. CaMKII is a multifunctional kinase with a catalytic domain, an autoinhibitory domain, a variable segment, and a self-association domain. In the basal state, the protein is inactive – its catalytic domain is blocked by the autoinhibitory domain. CaMKII is a large holoenzyme comprising 12 homologous subunits associated via the self-association domain forming large multimers.

Intracellular Ca\(^{2+}\) binds to Calmodulin (CaM), a small protein that serves as an intermediate messenger. The Ca\(^{2+}/CaM\) complex is able to bind and activate transiently CaMKII subunits. A long-lasting activation of CaMKII only takes place by phosphorylation of the Thr 286 residue by neighbouring subunits via the catalytic domain. The ability to phosphorylate neighbouring subunits (autophosphorylation) represents the model by which a local transient increase in Ca\(^{2+}\) can be decoded into a longer-lasting form of molecular memory (Miller and Kennedy, 1986). In fact, once phosphorylated at Thr 286, CaMKII is independent of Ca\(^{2+}/CaM\). The levels of Ca\(^{2+}\) and CaM, however, must reach a threshold – autophosphorylation does not occur if neighbouring subunits are not activated simultaneously (see Figure 3.5c). Several computational studies have been recently proposed to model the complex dynamic interactions of Ca\(^{2+}\), CaM, and CaMKII (e.g. Pepke, Kinzer-Ursem, Mihalas, and Kennedy, 2010; Byrne, Putkey, Waxham, and Kubota, 2009). The computational models for postsynaptic signal transduction in LTP have been recently reviewed by Manninen, Hituri, Kotaleski, Blackwell, and Linne (2010).

Direct evidence for the requirement of CaMKII autophosphorylation in hippocampal LTP was provided by the work of Giese, Fedorov, Filipkowski, and Silva (1998). A mutation that precluded autophosphorylation was introduced in CaMKII. While the enzyme was still activated by Ca\(^{2+}/CaM\), the inability to lock off the transient increase in Ca\(^{2+}\) prevented the transgenic mice from undergoing LTP. The Ca\(^{2+}\)-independent activation of CaMKII by autophosphorylation is stable for at least 1 hour (Fukunaga, Stoppini, Miyamoto, and Muller, 1993), during which the enzyme slowly phosphorylates other substrates. Furthermore, injecting active CaMKII enzymes into CA1 hippocampal neurons, increases the size of excitatory postsynaptic currents (EPSCs) and prevents induction of LTP by synaptic stimulation (Lledo, Hjelmstad, Mukherji, Soderling, Malenka, and Nicoll, 1995; Lisman, Yasuda, and Raghavachari, 2012). These studies seem to imply that activation of CaMKII is sufficient and necessary for LTP induction.

Spatial decoding

Consistent with a local specificity in sensing the Ca\(^{2+}\) signal, CaMKII is localised nearby the postsynaptic density (PSD), Sheng and Kim (2011), a protein-rich area attached to the postsynaptic membrane in close proximity to the presynaptic active zone. Furthermore, in hippocampal cultured neurons it has been shown that from an initial localisation at F-actin in the dendrite, CaMKII translocates to the PSD (Shen and Meyer, 1999). A number of substrates of CaMKII have been identified. In the PSD, a subunit of the NMDA receptor, GluN2B, has been shown to bind activated CaMKII (Leonard, Lim, Hemsworth, Horne, and Hell, 1999; Strack and Colbran, 1998). Accumulating evidence suggests that the CaMKII/GluN2B system is required for activation and translocation of CaMKII (Bayer, Paul De Koninck, Hell, and Schulman, 2001; Bayer, LeBel, McDonald, O’Leary, Schulman, and De Koninck, 2006). CaMKII Ca\(^{2+}\)-independent persistent activation, LTP induction (Halt, Dallapiazza, Zhou, Stein, Qian, Junitti, Wojcik, Brose, Silva, and Hell, 2012; Barria and Malinow, 2005), and LTP maintenance (Sanhueza, Fernandez-Villalobos, Stein, Kasumova, Zhang, Bayer, Otmakov, Hell, and Lisman, 2011).

Other mechanisms outside of the CaMKII pathway contribute to the spatial decoding of the Ca\(^{2+}\)
Chapter 3. Biology of long-term potentiation

signal. For example, the dendritic spine heads physically segregate the concentration dynamics of intracellular species (Svoboda, Tank, and Denk, 1996), although interactions between neighbouring synapses can take place (Harvey, Yasuda, Zhong, and Svoboda, 2008).

The molecular switch hypothesis

During the Ca\(^{2+}\)-independent CaMKII activated phase, the activity of phosphatases acting on the Thr 286 residue has to be suppressed. While a number of phosphatases are able to dephosphorylate CaMKII in vitro (Strack, Barban, Wadzinski, and Colbran, 1997), the activity of protein phosphatase 1 (PP1) has been shown to be inhibited during LTP by the phosphatase inhibitor 1 (I1, see Cohen (1989) for a review). Protein kinase A (PKA, discussed in Section 3.3.3) directly phosphorylates I1 which, in turn, inhibits the phosphatase activity of PP1 (Blitzer, Connor, Brown, Wong, Shenolikar, Iyengar, and Landau, 1998). Another phosphatase, Calcineurin (PP2B), has been directly linked to LTP (see Figure 3.5c-d). It seems that low concentrations of Ca\(^{2+}\)/CaM are enough to inhibit PP2B in hippocampal neurons after LTP induction (Wang and Kelly, 1996), preventing the phosphatase activity upon I1. The CaMKII-phosphatase coupling is consistent with a bistable switch, with low and high CaMKII activities both stable, originally proposed by Lisman (1985).

Yet, a study conducted in vitro found no evidence of bistability in the CaMKII-PP1 system (Bradshaw, Kubota, Meyer, and Schulman, 2003). In addition, a detailed model of CaMKII activation published by Michalski in 2013 predicted that the CaMKII system would never be bistable at resting Ca\(^{2+}\) concentrations. Specifically, the concentrations of intracellular Ca\(^{2+}\) needed for a bistable regime should lie above the observed physiological levels. Taken together, these studies call into question the role of CaMKII as the sole responsible for the biochemical switch underlying LTP and memory. Furthermore, a study by Lee, Escobedo-Lozoya, Szatmari, and Yasuda (2009) using two-photon fluorescence on single spines found that CaMKII activation is brief, and returns to baseline in about 1 min. It has been proposed that the process responsible for maintaining LTP must be downstream of CaMKII (Nicoll and Roche, 2013).

To further clutter the picture, a recent in vitro study published in 2014 by Urakubo, Sato, Ishii, and Kuroda (2014) demonstrated how a single holoenzyme of CaMKII can function as a molecular switch in the presence of an NMDA receptor-derived peptide. The intracellular Ca\(^{2+}\) transient was sufficient to trigger the phosphorylated ON state. In order for the system to switch from the ON state to a dephosphorylated OFF state, an increase in PP1 concentration was needed. Interestingly, reversal of LTP by phosphatases PP1 and PP2A has been documented in the literature (Kang-Park, Sarda, Jones, Moore, Shenolikar, Clark, and Wilson, 2003).

Note, however, that while the Ca\(^{2+}\) concentrations needed to reach bistability were found to be above 200nM (physiological levels are in the 80 – 150nM range), it is conceivable that the synergistic interplay between other proteins, perhaps in conjunction with other, less studied CaMKII phosphorylation sites (Thr 305 and Thr 306), may shift or extend the bistability phase to basal Ca\(^{2+}\) levels. The complexity of CaMKII regulation is remarkable, and a study seems to point out that phosphorylation at Thr 305 and Thr 306 may promote LTD rather than LTP (Pi, Otmakhov, Lemelin, De Koninck, and Lisman, 2010). The active involvement of other proteins would also account for the high concentrations of CaMKII observed in the PSD that cannot be accounted for solely by the binding to NMDA receptor (Petersen, Chen, Vinade, Dosemeci, Lisman, and Reese, 2003).

In the light of evidence from these studies, the molecular switch that CaMKII may constitute does not seem to be sufficient to elicit LTP. As we have emphasised, LTP is a complex process which may encompass a number of different overlapping and interrelated mechanisms acting at different levels of biological organisation. The putative binary switch that CaMKII presents provides an attractive hypothesis consistent with the robustness to noise and integrative capabilities of a molecular gating system. At different levels, however, it is conceivable that synaptic changes in strength may behave as multi-level mnemonic devices. In fact, graded storage devices can
3.3 Molecular mechanisms of LTP

Theoretically, have longer retention times, more noise tolerance, and greater storage capacity (Fusi, Drew, and Abbott, 2005; Satel, Trappenberg, and Fine, 2009).

AMPA receptor phosphorylation and trafficking

The glutamate receptor 1 (GluR1) subunit of the AMPA receptor represents another crucial substrate of CaMKII in the PSD (see Soderling and Derkach (2000) for a review and Barria, Muller, Derkach, Griffith, and Soderling (1997) for a paradigmatic study carried out by inducing LTP in region CA1 of hippocampus). Among the different phosphorylation sites in GluR1, Ser 831 is likely to be the responsible for the increase in conductance observed during E-LTP (Barria et al., 1997). The Ser 831 residue is phosphorylated by CaMKII and PKC (see Section 3.3.2) and results in an increase on the average single channel conductance. However, this is only one mechanism by which AMPA receptor function is potentiated during E-LTP. The early recruitment of other AMPA receptor channels already present in the plasma membrane, together with the temporally subsequent exocytosis of AMPA receptor-enriched vehicles are the other two fundamental mechanisms known to contribute. The phosphorylation-induced increment of AMPA receptor conductance appears to be a fundamental component of LTP (Giese et al., 1998).

Furthermore, the different subunits that form the AMPA receptor can play a role in the redistribution of the channels in the membrane. While the primary subtype of AMPA receptor inserted during LTP is the heteromeric GluA1/GluA2 (Adesnik and Nicoll, 2007), it has been suggested that a GluA2-lacking AMPA receptor is required for LTP maintenance. This type of AMPA receptor seems to appear during the first 10-20 min post-LTP induction (Plant, Pelkey, Bortolotto, Morita, Terashima, McBain, Collingridge, and Isaac, 2006). A study in vivo in rat dentate gyrus suggested that trafficking of GluA1-containing AMPA receptors takes place in the early-phase of LTP, although these studies are not conclusive (Adesnik and Nicoll, 2007).

AMPA receptor activity is also dependent on auxiliary proteins. The anchoring of the AMPA receptor at the synapse is mediated by the phosphorylation of the protein stargazin, which allows more receptors to bind to PSD-95 (Tomita, Stein, Stocker, Nicoll, and Bredt, 2005; Opazo, Labrecque, Tigaret, Frouin, Wiseman, De Koninck, and Choquet, 2010). Stargazin acts as an auxiliary subunit of most AMPA receptors influencing their conductance and kinetics – the expression of a stargazin construct unable to undergo phosphorylation prevented LTP in hippocampal neurons (Tomita et al., 2005). The TARP family have been shown to have an important role in membrane insertion and lateral redistribution of AMPA receptors (see Blakely and Edwards (2012) for a review). A member of the TARP family, γ-8, appears to participate in the formation of the pool of extrasynaptic AMPA receptors. Mice lacking γ-8 show a severe defect in LTP (Rouach, Byrd, Petralia, Elias, Adesnik, Tomita, Karimzadegan, Kealey, Bredt, and Nicoll, 2005) due to the lack of extrasynaptic AMPA receptors.

CaMKII role in LTP maintenance

The increase in the number of channels by exocytosis is also mediated indirectly by CaMKII, and it is essential for the later stages of LTP. Activated CaMKII stimulates the fusion of AMPA receptor-containing vesicles with the cell membrane via the Ras-Erk1/2 pathway (Patterson, Szatmari, and Yasuda, 2010), Rab-GTPase proteins, SNARE proteins, syntaxin 4 and 13, and the motor protein myosin V (see Kennedy and Ehlers (2011) for a review, pathway represented with a black dashed arrow in Figure 3.5). These components, although not necessary for the early phase of LTP, strongly reduce the potentiation at later times (Lisman et al., 2012), which suggest that during E-LTP it suffices to recruit and anchor the AMPA receptors readily available in the cell membrane. On the other hand, L-LTP relies on the exocytosis of AMPA receptor vesicles. Note that the Ras-Erk1/2 pathway is likely not to be the only one mediating the vesicle fusion (Cullen and Lockyer, 2002; Patterson et al., 2010).

Nevertheless, conflicting views exist regarding the persistence of the CaMKII/NMDA receptor
complexes and the multiple roles of CaMKII in LTP maintenance (Lee et al., 2009; Bayer et al., 2006).

### 3.3.2 The PKC pathway

The induction of LTP is not only dependent on CaMKII. In 1989, Malinow, Schulman, and Tsien reported that the inhibition of either CaMKII or protein kinase C (PKC) blocked the induction of LTP, but had no effect on its expression after its establishment. In fact, the role of PKC in LTP was the subject of a number of earlier studies. While the exact molecular mechanisms by which PKC is activated after increases in intracellular Ca$^{2+}$ are unclear, the information carried by its activation converges with the flux of information carried by the activated CaMKII – both molecules phosphorylate the identical site on GluR1 subunit of AMPA receptor, increasing its conductance (Mammen, Kameyama, Roche, and Huganir, 1997; Roche, O’Brien, Mammen, Bernhardt, and Huganir, 1996). This view is consistent with the fact that inhibition of both PKC and CaMKII is necessary to reverse the expression of E-LTP (Wang and Feng, 1992).

PKC may also be associated with later stages of LTP. PKMζ, an autonomously active isoform of PKC, has been proposed to be responsible for the maintenance of the initial cellular changes established following LTP induction. The kinase acts to preserve both LTP maintenance and the long-term memory trace by a continuous phosphorylation that enhances AMPA receptor (Sacktor, Osten, Valsamis, Jiang, Naik, and Sublette, 1993). PKMζ mRNA contains a long untranslated region that is crucial for the regulation of the PKMζ protein synthesis at the synapse (Hernandez, Blace, Crary, Serrano, Leitges, Libien, Weinstein, Tcherapanov, and Sacktor, 2003). The translational block is mediated possibly via microRNA, allowing stores of PKMζ mRNA that can be activated promptly following synaptic stimulation (Sacktor, 2008). A number of intracellular signalling cascades converge in the activation of the repressed mRNA, including CaMKII, PKA, and PI3K-Akt-mTOR pathways. For reviews, see Sacktor (2010) and Kwapis and Helmstetter (2014).

### 3.3.3 The PKA pathway

The protein kinase A (PKA) cascade is often referred to as the cAMP signalling pathway (Figure 3.5d). Activation of NMDA receptor increases cAMP levels via Ca$^{2+}$-sensitive adenylyl cyclases AC1 and AC8, which activate PKA by promoting the detachment of its inhibitory subunit. This process is Ca$^{2+}$/CaM-dependent. The activated PKA has the ability to enhance voltage-gated Ca$^{2+}$ channel and AMPA receptor conductance, while simultaneously attenuating voltage-dependent K channel function. By this mechanism, an initial Ca$^{2+}$ influx can be amplified (Chetkovich, Gray, Johnston, and Sweatt, 1991; Roche et al., 1996). This feedback is complemented by the crosstalk between the PKA cascade and the CaMKII switch. Specifically, PKA activates I1 by phosphorylation, which will inhibit the phosphatase activity of PP1 and PP2A upon CaMKII (Blitzer, Wong, Nouranifar, Iyengar, and Landau, 1995; Blitzer et al., 1998). As the dephosphorylation of I1 can be regulated by PP2B, it is probable that the antagonistic regulation of PKA and PP2B on I1 regulates the level of synaptic output (Winder, Mansuy, Osman, Moallem, and Kandel, 1998). In this scenario, the cAMP pathway can be interpreted as a gate acting on the CaMKII cascade rather than carrying the signal for LTP.

Yet, the role of PKA has been regarded as having a fundamental importance in L-LTP (Abel, Nguyen, Barad, Deuel, Kandel, and Bourouchcouladze, 1997). Activated PKA translocates to the nucleus and activates CREB, which mediates the transcription of genes responsible for L-LTP maintenance (Thomas and Huganir (2004), see Section 3.3.6). While E-LTP and L-LTP have been classically regarded as temporally and mechanistically distinct phases, they may well overlap in time. In fact, the PKA- and protein synthesis-independent E-LTP can last for periods of time that overlap with L-LTP which, in the other hand, can develop quickly (Park et al., 2014).
3.3 Molecular mechanisms of LTP

3.3.4 The Ras-Erk1/2 pathway

The Ras-Erk1/2 signalling pathway has been shown to be of crucial importance for induction and maintenance of LTP (Impey, Obrietan, and Storm, 1999), and for some types of memory (Atkins, Selcher, Petraitis, Trzaskos, and Sweatt, 1998; Blum, Moore, Adams, and Dash, 1999; Runyan, Moore, and Dash, 2004). Erk is a member of the MAPK family of protein kinases. Upon activation by an Erk kinase (MEK), Erk can phosphorylate the protein kinase Rsk2, which in turn translocates to the CREB and the SRF (Xing, Ginty, and Greenberg (1996), see Section 3.3.6 and Figure 3.5b). Erk can also dimerize and translocate to the nucleus to activate other transcription factors such as Elk-1 (Gille, Sharrocks, and Shaw, 1992), and SMAD3 (Gille, Kortenjann, Thomae, Moomaw, Slaughter, Cobb, and Shaw, 1995).

The activation of MEK is accomplished by another class of protein kinases, the Raf kinase family. Raf activation pathway is not clear, but seems that at least two mechanisms must converge – Ras interaction plus phosphorylation mediated by an Src kinase or PKC (Morrison and Cutler Jr, 1997).

A relevant feature of the Ras-Erk1/2 cascade is the capacity for serial amplification through the successive protein kinases, which could have an important role in the induction phase of LTP.

3.3.5 The presynaptic component

The mechanisms described so far are all located within the postsynaptic cell. A long dispute has taken place since the discovery of LTP regarding the locus of expression. Early studies showed an increase in glutamate release following LTP (Dolphin, Errington, and Bliss, 1982; Errington, Galley, and Bliss, 2003; Bliss, Errington, Galley, and Lynch, 1987; Lynch, Errington, Clements, Bliss, Redini-Del Negro, and Laroche, 1990), but contradicting reports also exist (Lüscher, Malenka, and Nicoll, 1998; Manabe and Nicoll, 1994; Mainen, Jia, Roder, and Malinow, 1998). While the models of postsynaptic LTP expression have become widely accepted, the presynaptic component is still subject of research. A recent study reported an increase in the reliability of transmitter release in active synapses (Enoki, Hu, Hamilton, and Fine, 2009). The authors hypothesised that the pre- and postsynaptic components of LTP are differentially regulated during development – immature synapses (deficient in AMPA receptors) recruit AMPA receptors upon potentiation as the brain matures via the post-synaptic mechanism, while already matured synapses express subsequent plasticity via presynaptic changes in transmitter release probability.

3.3.6 Gene expression in LTP

Signalling crosstalk between synapses and nucleus

We have so far discussed the molecular effectors and processes underlying LTP induction. Consolidation and maintenance of LTP, however, may actually require additional molecular mechanisms that involve new mRNA and protein synthesis (Kandel, 2001; Bito, Deisseroth, and Tsien, 1997; Morgan and Curran, 1989; Okuno, 2011). The crosstalk between the nucleus and the active synapse is believed to involve a tagging of the activated synapse. The “synaptic tagging and capture” hypothesis (Frey and Morris, 1997) argues that activated synapses are marked with a long-lasting tag to direct the delivery of the new synthesized proteins towards them (for a recent review, see Redondo and Morris, 2010).

The cAMP-responsive element

Another Ca\textsuperscript{2+}/CaM-dependent protein kinase, CaMKIV, has been shown to be involved in LTP. In contrast to CaMKII, CaMKIV is prominently located in the nucleus, where it can be activated upon phosphorylation by CaMKK. Activated CaMKIV is capable of phosphorylating CREB at Ser 133, which leads to the recruitment of an additional protein, the CREB binding protein (CBP). The CREB/CBP complex is ultimately responsible for the cAMP-responsive element (CRE)-dependent
Figure 3.5: (a) Co-occurrence of membrane depolarisation and glutamate binding opens NMDA receptors. The intracellular raise in Ca^{2+} is sensed by at least three pathways. The Ras-Erk1/2 pathway (b); the CaMKII molecular switch (c); the PKA pathway (d). The pathways converge (including CaMKIV) in the activation of CREB (e). A “retrograde messenger” travels to the presynaptic cell leading to an increase in the presynaptic response to subsequent stimuli (f).

gene expression (Braun and Schulman, 1995; Bito, Deisseroth, and Tsien, 1996). It has been demonstrated that CBP activity is, in addition, positively regulated directly by CaMKIV (Chawla, Hardingham, Quinn, and Bading, 1998). It appears that the main role for CaMKIV consists in activating CPB while the Ras-Erk1/2 pathway is responsible for ensuring a prolonged CREB
3.3 Molecular mechanisms of LTP


Another pathway that has been studied in reference to CREB activation is triggered by a type of membrane channels called voltage-dependent Ca\(^{2+}\) channels (VDCC). A particular class, the L-type VDCC, has been shown to participate actively in some forms of LTP (Grover and Teyler, 1990; Morgan and Teyler, 1999). In fact, the L-type VDCC-mediated Ca\(^{2+}\) increase is likely to be a trigger for CRE-mediated transcription (Bading, Ginty, and Greenberg, 1993).

The downstream CRE-regulated genes are initiators and modulators of long-term structural and functional changes elicited by LTP and memory (Deisseroth, Bito, and Tsien, 1996; Bourchuladze, Frenguelli, Blendy, Cioffi, Schutz, and Silva, 1994; Bito et al., 1997; Okuno, 2011). Other signalling cascades, such as the PKA pathway and the Ras-Erk1/2 pathway may converge in CREB phosphorylation (Flavell and Greenberg, 2008). While the phosphorylation of CREB by the Ras-Erk1/2 cascade seems to work at a slower time scale, its role in LTP-induced gene expression is still unclear.

A large number of target genes for CREB have been proposed, and CREs exist in the promoter regions of c-fos, BDNF, Egr1, homer1a/vesl1s, cpg15, and Arc (Benito, Valor, Jimenez-Minchan, Huber, and Barco, 2011). The regulation of these genes will be discussed in Section 3.3.6.

NF\(\kappa\)B

NF\(\kappa\)B is a CTF ubiquitously expressed with a strong presence in the brain. There are a total of seven members in the Rel-NF\(\kappa\)B family – p50, p52, p65 (RelA), p100, p105, RelA, c-Rel, and RelB. They form homo- or heterodimers, and the prototypical dimer in the nervous system is p50/p65 (May and Ghosh, 1997). NF\(\kappa\)B is localised in the cytoplasm bound to an inhibitor protein, I\(\kappa\)B. Upon stimulation, the phosphorylation of I\(\kappa\)B by the I\(\kappa\)B kinase (IKK) leads to its ubiquitination and subsequent degradation by the proteasomal machinery, exposing the DNA binding site of NF\(\kappa\)B. Following translocation to the nucleus, the active dimer is able to promote transcription of its target genes (Karin and Ben-Neriah, 2000). Neuronal stimulation may activate NF\(\kappa\)B through several signalling cascades, including PI3K-Akt-mTOR, PKC and CaMKII (Lilienbaum and Israël, 2003; Rojo, Salinas, Martín, Perona, and Cuadrado, 2004).

Freudenthal, Romano, and Routtenberg (2004) documented the activation of NF\(\kappa\)B during LTP in the mouse hippocampus. An increment in p50 has also been observed (Meberg, Kinney, Valcourt, and Routtenberg, 1996). In addition, Albensi and Mattson (2000) showed that a DNA decoy containing the consensus sequence recognised by NF\(\kappa\)B reduced the magnitude of LTP. In particular, c-Rel has been shown to be necessary for LTP in vitro and for hippocampus-dependent memory formation in vivo (Ahn, Hernandez, Levenson, Lubin, Liou, and Sweatt, 2008).

In addition to the transcription factor activity following translocation to the nucleus, NF\(\kappa\)B may act locally at the synapse Salles, Romano, and Freudenthal (2014). In this regard, the fruitfly homologous gene, Dorsal, was shown to regulate glutamate receptor density in an transcriptionally independent way (Heckscher, Fetter, Marek, Albin, and Davis, 2007).

The serum response element

The serum response element (SRE) is the binding site for another inducible transcription factor activated by phosphorylation – SRF. While SRF and CREB may jointly contribute to some neuronal functions (such as the control of Arc expression), the functions of the SRF-induced immediate early genes (IEGs) seems to be restricted to plasticity-associated structural changes in neuronal connectivities via the control of actin dynamics (Knöll and Nordheim, 2009). A gene profiling study using transfected hippocampal cultures indicated that the effect of overexpression of SRF only had effect on around 200 transcripts, while CREB affected up to 10% of the transcriptome (around 30,000 transcripts, Benito et al., 2011).

Within the SRE, an additional binding site to accommodate a ternary complex factor (TCF),
a family of proteins with an Ets domain such as Sap-1, Sap-2, and Elk-1 (Price, Rogers, and Treisman, 1995). These proteins are capable of binding the SRE/SRF complex. The Ras-Erk1/2 pathway mediates the activation of the SRE pathway, initiated by the Ca\(^{2+}\) transient (Hardingham, Arnold, and Bading, 2001).

**Immediate early genes**

CREB and SRF are CTFs – the protein is present regardless of the activation state of the neuron, and its transcription factor activity is switched on upon phosphorylation. Other CTFs such as CREM and ATF-2, have unclear functional roles in LTP (Herdegen and Leah, 1998). Together with the activation of CTFs, a wave of newly transcribed genes are synthesised by the transcriptional machinery following LTP induction. These so-called IEGs can be ITFs, which will in turn regulate a second wave of gene transcription (delayed response genes). Transcription of IEGs requires only pre-existing transcription factors and is independent on *de novo* protein synthesis (Abraham and Williams, 2008). Among the IEGs, some do not act as transcription factors and exhibit other structural or functional roles associated with the synaptic changes necessary for LTP (Abraham and Williams, 2003, 2008).

The characterisation and isolation of IEGs started as early as in 1987 with the description of a rapid induction of c-Fos following stimulation (Morgan, Cohen, Hempstead, and Curran, 1987). A number of successive studies have succeeded at isolating other IEGs with the aim of defining central genes involved in learning and memory.

**c-Fos** is a proto-oncogene, part of the Fos family of transcription factors. Together with c-jun it forms an heterodimer (AP-1) which binds DNA at AP-1 specific sites to regulate transcription of other genes (Chiu, Boyle, Meek, Smeal, Hunter, and Karin, 1988). It has been used as a marker of neural activity and mutant mice with a disrupted c-Fos results in an impairment of hippocampal LTP (Fleischmann, Hvalby, Jensen, Strekalova, Zacher, Layer, Kvello, Reschke, Spanagel, Sprengel, *et al.*, 2003). The c-Fos gene has a simple proximal promoter sequence which contains a CRE and an SRE (Robertson, Kerppola, Vendrell, Luk, Smeyne, Bocchiaro, Morgan, and Curran, 1995).

The brain-derived neurotrophic factor (BDNF) is a secreted protein member of the neurotrophin family of growth factors involved in neural survival and growth. In addition, BDNF seems to play a fundamental role in the maintenance of LTP and its synaptic specificity (Barco, Patterson, Alarcon, Gromova, Mata-Roig, Morozov, and Kandel, 2005; Panja and Bramham, 2014). It contains a number of different promoters with distinct activity dependencies including CRE and SRE (Greer and Greenberg, 2008).

**Arc/Arg3.1** is probably the most studied ITF in neurons. Its expression pattern correlates with neuronal activity (Lyford, Yamagata, Kaufmann, Barnes, Sanders, Copeland, Gilbert, Jenkins, Lanahan, and Worley, 1995; Link, Konietzko, Kaeselmann, Krug, Schwanke, Frey, and Kuhl, 1995), and its disruption impairs the maintenance of LTP without affecting the induction phase (Guzowski, Lyford, Stevenson, Houston, McGaugh, Worley, and Barnes, 2000; Plath, Ohana, Dammermann, Errington, Schmitz, Gross, Mao, Engelsberg, Mahlke, Welzl, *et al.*, 2006). Arc mRNA moves to the activated dendrites after induction by synaptic activation at a rate of 300\(\mu\)m/h (Wallace, Lyford, Worley, and Steward, 1998). The means by which Arc mRNA is recruited selectively to the activated synapses was attributed to the synaptic tagging and capture hypothesis. In this regard, a study of Steward, Wallace, Lyford, and Worley (1998) suggested that the selective localisation of newly synthesised Arc mRNA was not caused by a general reorganisation of the dendritic cytoplasm or cytoskeleton nor by selective degradation of mRNA in non-activated synapses, but rather by the presence of some “tag” generated following synaptic activation that recognises a signal in the mRNA molecule itself. Contrarily to this belief, however, a study has shown that Arc targets inactive synapses, by interacting with an inactive, CaM-unbound form of CaMKII (Okuno, Akashi, Ishii, Yagishita-Kyo, Suzuki, Nonaka, Kawashima, Fujii, Takemoto-Kimura, Abe, *et al.*, 2012). This study suggests that the localisation of Arc is inverse to the activated synapses, observation
consistent with the results published earlier by Shepherd, Rumbaugh, Wu, Chowdhury, Plath, Kuhl, Huganir, and Worley (2006). They showed that in the absence of Arc, AMPA receptors endocytosis is low, leading to a shifted equilibrium towards AMPA receptor membrane insertion whereas on the contrary, high levels of Arc promote AMPA receptors endocytosis. Sustained Arc expression acts by counteracting saturation of synaptic strength.

The Arc promoter contains a region named synaptic activity-responsive element (SARE) (Inoue, Yagishita-Kyo, Nonaka, Kawashima, Okuno, and Bito, 2010), which requires the simultaneous binding of CREB, SRF and MEF2. Both CaMK- and Ras-Erk1/2-dependent pathways are involved in SARE activation (Kawashima, Okuno, Nonaka, Adachi-Morishima, Kyo, Okamura, Takemoto-Kimura, Worley, and Bito, 2009). In the promoter, there are also SREs and AP-1 (the binding motif for Fos/Jun complex) sites (Waltereit, Dammermann, Wulff, Scafidi, Staubli, Kauselmann, Bundman, and Kuhl, 2001).

The early growth response protein 1 (Egr1, also known as zif268, krox24, and NGFI-A), contains a zinc finger domain and is coregulated with c-Fos during some cellular processes (Sukhatme, Cao, Chang, Tsai-Morris, Stamenkovich, Ferreira, Cohen, Edwards, Shows, Curran, et al., 1988). Egr1 is an ITF, and only low levels of protein are detected in the absence of a stimulus. Activation of NMDA receptor converges on the phosphorylation of CREB and Elk-1, inducing Egr1 transcription. Raise in Egr1 protein has been linked with learning and memory. Its role as a signalling molecule downstream from the NMDA receptor in the hippocampal synaptic plasticity has been documented in a number of studies (Cole, Saffen, Baraban, and Worley, 1989; Wisden, Errington, Williams, Dunnett, Waters, Hitchcock, Evan, Bliss, and Hunt, 1990; Wei, Xu, Qu, Milbrandt, and Zhuo, 2000). In particular, the disruption of the cyclic AMP response element (CRE) in the promoter region of the Egr1 gene leads to impairment of in vivo dentate gyrus L-LTP (Jones, Errington, French, Fine, Bliss, Garel, Charnay, Bozon, Laroche, and Davis, 2001). For a review, see Thiel, Mayer, Müller, Stefano, and Rössler (2010).

Other members of the Egr family have been identified – Egr2, Egr3, and Egr4. In neurons, these Egr proteins are expressed following synaptic activity and have been associated with LTP (Murphy, Worley, and Baraban, 1991; O’Donovan, Tourtellotte, Millbrandt, and Baraban, 1999; Ryan et al., 2011).

Table 3.3.6 reports the most studied IEGs in the literature.

Table 3.1: **List of the some of the most studied IEGs.** Note that the temporal expression of the genes reported in the Table do not always fulfil the definition of IEGs since their expression falls within the first 2 h post-LTP induction. IEGs are typically induced within 15-45 minutes of the stimulus, returning to basal levels in few hours

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene</th>
<th>Structure/function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription factors</td>
<td>c-Fos</td>
<td>Component of AP-1</td>
<td>Greenberg et al., 1983; Morgan et al., 1987</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hope et al., 1992; Draganow et al., 1992</td>
</tr>
<tr>
<td></td>
<td>fosB</td>
<td>Component of AP-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c-Jun</td>
<td>Component of AP-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Saff et al., 1988; Cole et al., 1990</td>
</tr>
<tr>
<td></td>
<td>Jun-B</td>
<td>Component of AP-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Saff et al., 1988; Cole et al., 1990; Demmer et al., 1993</td>
</tr>
<tr>
<td></td>
<td>Jun-D</td>
<td>Component of AP-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Egr1</td>
<td>Component of AP-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zinc finger</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Egr2</td>
<td>Zinc finger</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Egr3</td>
<td>Zinc finger</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenberg et al., 1983; Morgan et al., 1987 (continued on next page)</td>
</tr>
</tbody>
</table>
Late response genes

A second wave of transcription is believed to occur following the IEGs. This set of genes are typically induced with a slower time course (2-3 h after the stimulus). While some late-response genes are transcription factors, most of them function as effector proteins and carry a wide spectrum of functions, including intracellular signalling, cell adhesion, cytoskeleton and cellular scaffolding regulation, secretory factors and membrane channels Hong, Li, Becker, Dawson, and Dawson (2004); Davies (2004).

While several efforts have been made towards defining a wider picture of the biochemical networks underlying the maintenance of LTP (Lee, Cohen, Becker, and Fields, 2005; Park, Gong, Stuart, and Tang, 2006; Håvik, Røkke, Dagyte, Bramham, and Steen, 2007a; Valor and Barco, 2010; Ryan et al., 2011), it is unclear how IEGs and late response genes (LRGs) coordinate in order to lead to the persistence of LTP.

The work by Ryan et al. (2012) represents one of the most recent approach towards understanding the role of the new gene transcription in LTP maintenance. They identified a set of differentially expressed genes whose interactions served to identify biologically relevant networks and pathways related with LTP. Genes identified as major hubs in the LTP-regulated gene networks confirmed
previous findings regarding major key regulators. In addition, using databases of cellular pathways with the most differentially expressed genes allowed to pinpoint other molecular species that may play central roles during LTP without showing a significant alteration in their expression levels (e.g. becoming active by phosphorylation), such as the CTF NFκB.

These results, taken together with previous knowledge, stress the importance of plasticity-associated gene transcription in LTP. Interestingly, some of the overrepresented functions across the LTP-related genes are development, proliferation and neurogenesis, suggesting that new gene transcription may have an active role in the morphological restructuring of the synapses following LTP. Note that these reported genes have been linked to long-term memory, specially the CREB family and its associated pathway, AP-1, Egr1, and NFκB (for a review, see Alberini, 2009).

It is worth commenting that despite the evidence of a protein-dependent phase in LTP maintenance, other lines of research have shown that LTP can be maintained in the absence of protein synthesis (e.g. Pang, Teng, Zaitsev, Woo, Sakata, Zhen, Teng, Yung, Hempstead, and Lu, 2004; Fonseca, Nägerl, and Bonhoeffer, 2006). While it is now clear that it is possible to maintain LTP in the absence of protein synthesis using particular induction protocols (BNDF-LTP, avoiding test pulses, inhibiting proteasome degradation; Abraham and Williams, 2008) these results do not rule out the requirement for protein synthesis in vivo. These studies, however, suggest that there may be a stimulation-dependent depotentiation caused by protein degradation, which raises the possibility that the role of de novo synthesised proteins is to replenish the protein pools at the synapses to avoid an LTP decay due to protein turnover.

On the other hand, de novo gene expression has been shown to be involved in L-LTP consolidation and maintenance, and constitutes a distinct phase of L-LTP. Gene expression following LTP induction is thought to mediate the structural remodelling of the stimulated synapses by neural growth of new dendritic spines and enlargement of pre-existing (Matsuzaki, Honkura, Ellis-Davies, and Kasai, 2004; Krucker, Siggins, and Halpain, 2000; Bramham, 2008). However, the integrated genomic response elicited by LTP is far from being well understood.
4. Models of gene regulatory networks

4.1 Introduction

Biological phenomena are inherently complex and difficult to study, and the application of mathematical and computational tools are strongly limited to a certain spatio-temporal scale. In contrast, biological systems have the capacity to receive signals from a surprisingly broad spatio-temporal range and process them in a highly reliable manner. The unifying principles, however, seem to be difficult to comprehend and their predictive ability limited to the available computational power.

Living cells are equipped with a robust and yet plastic analog system which equips them with the ability to respond to the environmental inputs. In order to sense and respond to this changing environment, organisms implement networks of interacting elements capable of integrating and processing information in an interdependent manner. Despite the complexity of the underlying networks, some aspects of the cellular behaviour are of apparent Boolean nature, namely, the cellular output represents a switch. A stereotypical example is the case of the cell cycle – there are several checkpoints in which the cell senses the critical variables (e.g. DNA damage) and decides whether to proceed to the next cell cycle phase. The decision of transitioning to a subsequent phase is controlled by a biochemical switch (Santos and Ferrell, 2008). Cell differentiation and apoptosis are other examples of multi-component switches that generate robust (and potentially irreversible) transitions (Siegal-Gaskins, Mejia-Guerra, Smith, and Grotewold, 2011).

An operative definition of a biological network represents the molecules as nodes and their interactions as edges. Ideally, a network representation of a biological system aims to capture its essential characteristics. Within the cell, there are a number of possible interactions between different types of molecules (e.g. protein-protein, protein-DNA, RNA-RNA, protein-metabolite interactions) and the information can flow without direct interactions (transcription from DNA to RNA and translation from RNA to protein). In addition, the interactions between the molecules can...
have different strengths. Despite the complexity that these models can entail, biological networks can often be simplified depending on the analysis undertaken. Furthermore, the network description allows the application of graph theory formalities, a well-developed field of mathematics (Strogatz, 2001).

While few general formalisms and methods to study the structure and dynamics of cellular signalling and regulatory networks have been proposed (see for example Klamt, Saez-Rodriguez, Lindquist, Simeoni, and Gilles (2006)) structural analysis of networks is a well-established field. Signal transduction research is still in its first stages, and its standards are still being settled (Saez-Rodriguez, Alexopoulos, and Stolovitzky, 2011).

The validity of network modelling arises from the assumption that digital computed steady states correspond to the actual analog steady states of the cellular circuitry. When constructing a dynamical mathematical model, the choice of variable types is one aspect, the second being the character of their dynamics in time. A prominent feature of molecular concentration changes is their often rapid change, when compared with the typical metastable character in between changes.

Computational models for regulatory network analysis fall into three broad categories (Kарлеbach and Shamir, 2008). In a first class of continuous models, explicit molecular concentrations are considered. These kind of models are generally used to understand behaviours that depend on fine time and space scales.

The second main class of models are the logical models, generally used to study networks qualitatively to understand their functionality under different conditions. Despite not being able to explain quantitative questions, they became very useful owing to their flexibility and simplicity.

A third class of models study the influence of noise at a single-molecule level, explaining the relationship between stochasticity and gene regulation.

In this chapter we overview some of the main logic-based computational methods in gene regulatory networks. Methods applied in this thesis will be covered in particular detail – random Boolean networks (Section 4.2.1) and weight matrices dynamics (Section 4.2.3).

### 4.2 Logic-based models

Logic-based network modelling was introduced in 1973 in the papers of Kauffman and Thomas (Glass and Kauffman, 1973; Thomas, 1973), and it represents the simplest modelling methodology. In these models, the state of each element in the network (genes, RNA, and/or proteins) is represented at any time as a discrete value. The update of these values follows a set of rules and takes place in discrete time steps. The simplicity of this methodology allows the study of networks when only qualitative data is available. Extensive reviews can be found in Watterson, Marshall, and Ghazal (2008); Morris, Saez-Rodriguez, Sorger, and Lauffenburger (2010).

#### 4.2.1 Boolean models

Boolean regulatory networks (Glass and Kauffman, 1973) are a particular case of logical models in which the states of the components of the network can only attain two levels (active/inactive, or overexpressed/underexpressed). In a Boolean network, there are \( N \) components (nodes) connected by a set of \( E \) edges, which represent biological interactions that modulate their expression. The values which each of the \( N \) nodes can take at any time are only two, \( x_i(t) \in \{0, 1\} \). The rules by which these states are updated during the subsequent time steps are defined by a set of \( N \) Boolean functions \( B_1, \ldots, B_N \). Note that each \( B_i \) function has \( k_i \) number of inputs, corresponding to the in-degree of the node \( i \), and an output whose value is defined by the combination of the input values (truth table, for an example see Table 4.1). The input values correspond to the values \( x_{i1}, \ldots, x_{ik_i} \) of the genes connected to that node. Hence, a bit vector \( x(t) = (x_1, \ldots, x_N) \) defines the state of the
Table 4.1: Example of two Boolean functions with three arguments. $B_{\text{rand}}$ is a fully random function, whereas $B_{\text{can}}$ is a canalizing function. In the latter the output is canalized on the input $x_1$: whenever its input value is $x_1 = 1$, the output of the function is always the same (0) regardless of the values of the other inputs. This particular rule it is also said to be nested, since the non-canalising value of $x_1$ is in turn canalized on the input $x_2 = 1$. This means that if $x_1 = 0$ and $x_2 = 1$, the output is independent of the value of $x_3$

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$B_{\text{rand}}(x_1, x_2, x_3)$</th>
<th>$B_{\text{can}}(x_1, x_2, x_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

network at a particular time step and the update at $t + 1$ of each $i$-th gene is calculated according to

$$x_i(t + 1) = B_i(x_1(t), ..., x_{ik_i}).$$

(4.1)

Figure 4.1 depicts a simple network of $N = 3$.

In a Boolean dynamic model, the values of $x(t)$ can be updated synchronously based on the vector at time $t - 1$ and according to the Boolean rules. Hence, given an initialised vector at $t = 0$ and the set of truth tables, this type of dynamics are deterministic. The set of rules $B$ will stay the same over time during the simulation, whereas the state vector $x(t)$ will evolve. The finite number of states, $2^N$, and the deterministic nature of the synchronous update scheme implies that a state will be inevitably revisited. Alternative asynchronous updating schemes are sometimes used. In those updating schemes, the values of the inputs for the Boolean rules can be current, $x(t)$, or prior, $x(t - 1)$. The main asynchronous algorithms are the random order asynchronous (Chaves, Albert, and Sontag, 2005), the general asynchronous (Harvey and Bossoamaier, 1997) and the deterministic asynchronous (Chaves, Sontag, and Albert, 2006) update methods. A comparative study on the dynamic behaviour of synchronous and asynchronous updating schemes can be found in Saadatpour, Albert, and Albert (2010).

The complete state space is described by all the potential values of the vector $V$. The possible trajectories can be represented by a state transition graph in which the nodes represent the states of the system (values of $V(t) = (x_1, ..., x_N)$), and the edges represent the allowed transitions among states based on the set of Boolean rules and the updating scheme (see Figure 4.1d). Eventually, the updates of the state vector $V$ will converge into a steady state or a set of recurrent states. These states are known as attractors. Any initial value for the state vector that leads to a particular attractor is part of its basin of attraction. In the toy example shown in Figure 4.1, there are two basins of attraction in the state space. Coincidentally, these basins of attraction are made up with the same states than their corresponding attractors. This is not always the case, and often the states that make up a basin of attraction include other states aside from the attractor itself. In the asynchronous models, each state in the transition graph can have several successors, and hence the basins of attraction for different attractors may overlap. This is never the case for synchronous deterministic update schemes, where each node in the transition graph has one and only one successor.
An important measure in Boolean networks quantifies how distant two states are in the state space. Between any two states $x^A$ and $x^B$, the normalised number of genes whose values differ between them is known as Hamming distance. It is defined as

$$h(x^A, x^B) = \frac{1}{N} \sum_{i=1}^{N} (x^{A}_i - x^{B}_i)^2.$$  \hfill (4.2)

Figure 4.1: Boolean network of $N = 3$. (a) The transitions between the 8 possible states are represented. The Boolean values for the nodes correspond to the two different colours (green and orange). Gray arrows represent biological interactions that influence protein expression, black arrows represent transitions from a time step to the next. (b) The set of Boolean rules $B$ are fully described. The rules determine the values that each of the nodes will have in the following time step. (c) A complete representation of the state space is depicted, including all the transitions as shown in a. The state space is characterised by two attractors, which in this minimal example overlap entirely with their respective basins of attraction. Based on Karlebach and Shamir (2008)

Whereas the choice of the set of Boolean functions $B$ can be guided by experimental evidence, often the available data does not suffice to model the truth tables. A methodology used to overcome this limitation is known as random Boolean networks (RBNs) (Kauffman, 1969b). The Boolean function $B_i(x_1, \ldots, x_k)$ of $k_i$ inputs is randomly chosen from the collection of all possible Boolean functions (Kauffman, 1993). RBN models often include a bias $p$ for the fraction of output values
4.2 Logic-based models

that are 1, \( p = P(B_i = 1) \), in the Boolean functions. To generate these functions suffices to flip a \( p \)-biased coin \( 2^k \) times. Additionally, in the original formulation networks were constructed with a constant in-degree \( k_i = \bar{k} \) for each node. Many other alternatives have been studied, mostly consisting in random sampling from various distributions with a \( \bar{k} \) mean. Notice that the bias \( p \) does not, however, preclude \textit{per se} any local structure in the function (for broad overviews of RBNs, see Drossel, 2008; Kauffman, 1993).

**Phase transitions in Boolean models**

A fundamental work in RBN modelling was performed by Derrida and Stauffer (1986), who introduced the concept of \textit{phase transition} for Boolean networks based on the Hamming distance evolution. They studied the time evolution of the state vector \( x \) in the limit \( N \to \infty \) with different initial conditions, \( x^A(t = 0) \) \( x^B(t = 0) \), comparing their Hamming distance with the Hamming distance between \( x^A(t \to \infty) \) and \( x^B(t \to \infty) \) (we denote with \( H \) an averaged \( h \), Equation 4.2, over a large number of initial conditions for \( x^A(t = 0) \) and \( x^B(t = 0) \)). They noted the existence of two possible phases for different values of the average in-degree \( \bar{k} \). The first corresponding to \( \bar{k} \leq 2 \), with \( H(t \to \infty) = 0 \) (ordered regime), and the second corresponding to \( \bar{k} > 2 \), with \( H(t \to \infty) \neq 0 \) (chaotic regime).

The initial states \( x^A(t = 0) \) \( x^B(t = 0) \) can be thought as the result of a final \( H(t \to \infty) = 0 \) (ordered regime) or, on the contrary, propagate across the network (chaotic regime).

The results of the temporal evolution of the Hamming distance studied in the seminal work of Derrida and Stauffer correspond to a mean-field theory approach, and hence do not necessarily hold for real biological networks, that are far from being statistically equivalent (Aldana, 2003). However, a work by Balleza, Alvarez-Buylla, Chaos, Kauffman, Shmulevich, and Aldana published in 2008 demonstrated that the phase transitions observed under the mean-field theory are equivalent those observed in networks of realistic topologies constructed with the same statistical properties than real biological networks.

![Figure 4.2](image-url): Types of attractors in random Boolean models. Each node corresponds to a particular state of the system. Two basins of attraction are represented in the figure, (a) is a fixed point attractor, which corresponds to a cycle of period 1, and (b) is a cycle attractor of period 10. Arrows show the direction of the dynamics as in Figure 4.1d. Adapted from Aldana et al. (2003)
Derrida plots

In order to visually represent the average effect of perturbations in a particular network, we can plot $H(0)$ (size of the perturbation at the beginning of the parallel runs, x-axis) against $H(\tau)$ (y-axis), where $\tau$ is a small number of discrete steps. Plotting different values of $H(0)$ (sampling different perturbation sizes) results in a Derrida plot, a popular tool used in RBNs (Derrida and Weisbuch, 1986). While some network architectures tend to absorb small perturbations and the final Hamming distances $H(\tau)$ are on average smaller than the initial perturbation, some topologies tend to amplify them – few genes with different values of expression lead to dramatically different network states. Crucially, in the Derrida plots these different behaviours fall in the opposite halves of the plot, with robust architectures represented by curves underneath the diagonal, $H(0) > H(\tau)$, and sensitive architectures above, $H(0) < H(\tau)$ (Fox and Hill, 2001), and a network whose Derrida mapping appears tangent to the diagonal is said to exhibit criticality. Note that choosing small integer values for $\tau$ (shorter dynamics) captures the effects of the network’s local geometry, while larger values reflect the general characteristics of the structure of the network, since information has more time to spread across the network (Aldana et al., 2003). In practice, the slope for the small $H(0)$ region reveals the average outcome of a small perturbation. Curves below the diagonal indicate a tendency towards stability (ordered regime) whereas curves above imply instability (chaotic regime). The diagonal $H(0) = H(\tau)$ represents the transition from order to chaos (see Figure 4.3).

To construct the Derrida plots, a large number of Boolean rules are constructed (for a given network topology) and for each set of randomly generated rules a large number of parallel simulations are run with random initial values, uniformly sampling different values of $H(0)$.

![Derrida plots](image)

**Figure 4.3:** Derrida plots of random Boolean networks with a Poisson distribution of in-degree. The different curves correspond to the plot of $H(0)$ vs $H(\tau)$ for networks which operate in different dynamic phases. Ordered, critical, and chaotic regimes are represented by blue, orange, and green colours respectively. The shaded area under the diagonal corresponds to the ordered regime. Networks exhibiting criticality appear as tangent to the diagonal, which represents the transition between the chaotic and ordered phases. Adapted from Balleza et al. (2008)

Network sensitivity

The tangent to the Derrida mapping at $H(\tau) = 0$ represents the average sensitivity $S$ of a Boolean network. The sensitivity $S$ of a RBN is heavily dependent both on the bias $p$ (Derrida and Pomeau,
4.2 Logic-based models

1986) as well as on the average in-degree of the network $\bar{k}$ (Luque and Solé, 1997; Aldana and Cluzel, 2003). In the original formulation (Kauffman, 1969b) the phase transition is governed by

$$\bar{k}_c = \frac{1}{2p(1-p)},$$

(4.3)

where $\bar{k}_c$ is the critical average in-degree.

If the bias is set to $p = 0.5$, the critical in-degree is $\bar{k}_c = 2$, with larger values corresponding to chaotic behaviour (higher sensitivity) and smaller values leading to ordered behaviour (lower sensitivity). Likewise, heavily biased Boolean functions (values of $p$ away from 0.5) lead to ordered dynamics. These characteristics are observed in cases where the in-degree is randomly chosen from a Poissonian or from a power-law distribution (Barabási and Albert, 1999; Aldana and Cluzel, 2003; Aldana-Gonzalez, Coppersmith, and Kadanoff, 2002).

Robustness and adaptability in the critical regime

Networks with dynamic stability are by definition more robust to small perturbations. When a Derrida mapping falls under the diagonal, perturbations tend to die out with time instead of propagating through the system. It is expected from complex biological networks a certain level of stability against environmental perturbations. Some classically studied mechanisms such as gene redundancy and epistasis (Sanjuán, Moya, and Elena, 2004; Moore, 2005) suggest that the dynamical properties of biological networks are restricted to small sets of genes. However, other studies point in the direction of a “distributed robustness” scenario. All the regulatory interactions among genes play a role in the dynamical characterisation of the network (Shmulevich, Kauffman, and Aldana, 2005; Wagner, 2005).

More specifically, it has been pointed out by Kauffman and others that biological systems may lay in a dynamical region close to the border between order and chaos, often referred to as the “critical” zone (Langton et al., 1992; Graudenzi, Serra, Villani, Damiani, Colacci, and Kauffman, 2011). Some studies, in fact, suggest that the optimal trade-off between robustness and readiness to change under certain conditions (evolvability) is found in the critical zone of the space of parameters (e.g. Kauffman, 1995; Shmulevich et al., 2005). In addition, the information flow though the networks that are in a critical regime is optimal (Ribeiro, Kauffman, Lloyd-Price, Samuelsson, and Socolar, 2008). The capacity of complex dynamical systems to respond without reaching saturation to broad dynamic ranges of external stimuli requires a collective integration of the incoming information. This emergent property is maximised in the critical regime (Kinouchi and Copelli, 2006) and is the consequence of long-range correlations that arise in the critical point (Balleza et al., 2008). This kind of responses show a remarkable sensitivity.

In addition to what we pointed out earlier in reference to the “distributed” nature of robustness across a large number of genes interacting within a network, it is likely that robustness is manifested at different levels, likewise, in a distributed fashion – from the physicochemical structural properties of macromolecules to the higher levels of multicellular organization and ecosystems.

While some biological systems are designed to elicit a graded response to an input, other systems show behaviours that resemble multistable transitions. These systems can range from relatively simple switch-like responses to more complex multistable switches (e.g. Ferrell and Machleder, 1998; Pomerening, Sontag, and Ferrell, 2003; Yao, Tan, West, Nevins, and You, 2011). The complex environmental signals integrated by the living systems generate responses at the different levels of their biological organisation, with these responses being interdependent. The interpretation of any result obtained requires, thus, a critical scrutiny.

Relevance of Boolean modelling to biological problems

During LTP, the cell state needs to transit from a stable point to another stable point. If this assumption holds for the gene expression profile, LTP induction can be seen as the perturbation
needed to push the gene expression equilibrium of the neuron to the post-LTP attractor. The discrimination of this perturbation from the environmental noise is of a fundamental importance and represents the global objective of the present thesis. In this regard, gene regulatory network analysis can help to better understanding the mechanisms of gene expression underlying LTP maintenance and consolidation, as well as generating experimentally testable hypotheses regarding (a) intermediate phases and (b) effects of gene knock-outs and mutations.

The relevance of network analysis and particularly the Boolean modelling approach, has been put forward by a number of studies that show that the Boolean approach captures the main dynamical features of gene regulatory networks (see for example Bornholdt, 2005; Espinosa-Soto, Padilla-Longoria, and Alvarez-Buylla, 2004; Davidich and Bornholdt, 2008; Serra, Villani, Barbieri, Kauffman, and Colacci, 2010; Rämö, Kesseli, and Yli-Harja, 2006; Christensen, Oliveira, and Nielsen, 2009; Thakar, Piliome, Kirimanjeswara, Harvill, and Albert, 2007; Saez-Rodriguez, Simeoni, Lindquist, Hemenway, Bommhardt, Arndt, Haus, Weismantel, Gilles, Klamt, et al., 2007).

At the same time, the inherent simplicity of the models allows to analyse complex behaviours in large-scale systems that would be otherwise inconceivable to tackle with the incomplete qualitative data available.

4.2.2 Other logic-based models

Probabilistic Boolean networks
Experimental data is very often insufficient to fully describe the regulatory functions of a biological network. An approach proposed by Shmulevich, Dougherty, Kim, and Zhang (2002) consists in allowing each node of the network to have a number of regulatory functions. Each function will have a probability based on the available experimental data. Among the possible functions, one is chosen randomly at each step according to its probability.

Petri nets
Petri nets (Petri, 1966) represent another type of stochastic model in which the nodes can be “places” (biological species) or “transitions” (regulation functions). The places have discrete values, which represent the number of “tokens”. In Petri nets, the input places are connected to transitions, and these are in turn connected to output places. At any time step, every transition can be fired if their input places have enough “tokens” (a fired transition will reduce the number of tokens in each of its input places and increase the number of tokens in each of its output places).

Multi-logic models
The MetaReg model developed by Gat-Viks, Tanay, and Shamir (2004) is a logic-based model with the entities of the network not being restricted to Boolean values. Regulation functions are also discrete. A probabilistic version also exists (Gat-Viks, Tanay, Raijman, and Shamir, 2006) and it has been implemented in the MetaReg software (Ulitsky, Gat-Viks, and Shamir, 2008). By predicting the steady states after constructing the biological model, the methodology allows to predict component values under any experimental treatment. If there are discrepancies between the predictions and the experimental results, these can be used to automatically refine the model.

Another multi-valued logic model was proposed by Huang and Hahn (2009) and applied to infer networks based on experimental data (Morris, Saez-Rodriguez, Clarke, Sorger, and Lauffenburger, 2011).

4.2.3 Continuous and single-molecule models

In contrast to the models described so far, the class of continuous models do not discretise the expression values. While from a theoretical standpoint real-valued models are more accurate at representing experimental data, it has been pointed out that they exhibit similar dynamical properties under very general conditions (Chaves et al., 2006). We will briefly review the different continuous
models with particular depth in the case of the weight matrices methodology, which is applied to characterise the dynamical properties of an LTP-related network in Chapter 7.

**Linear models: weight matrices dynamics**

An advantage of linear models is that they can be used to gain understanding of a qualitative nature without requiring extensive knowledge on the regulatory mechanisms. In this class of models, each regulatory function corresponds to a weighted linear sum of the inputs for each of the network units.

The simple model proposed by Weaver and coworkers (Weaver, Workman, and Stormo, 1999) (also referred to as *TReMM*, Transcription Regulation Modelled with Matrices) represents a compromise between the simplifying assumptions about biological systems of the Boolean networks, and the detailed models described by ordinary differential equations. The number of samples required to successfully model with continuous data are so large that continuous time models could only be based on theoretical data.

The gene expression state of a network is represented by a vector \( u(t) \), where \( u_i \) corresponds to the expression level of the \( i \)-th gene. The regulatory interactions between genes are modelled with a weight matrix, \( W \) so that the effect of gene \( j \) on gene \( i \) is the expression level of \( j \), \( u(t) \) times its regulatory influence on \( i \), \( w_{i,j} \). This regulatory influence can be either activating \((w_{i,j} > 0)\), or repressing \((w_{i,j} < 0)\). The values of \( w_{i,j} = 0 \) correspond to the lack of experimental evidence for the effect of gene \( j \) on gene \( i \). Put differently, only the nodes connected in Figure 7.2 have non-zero values of \( w \).

The total regulatory input to \( i \), \( r_i(t) \), is calculated by adding all the genes which interact with it:

\[
r_i(t) = \sum_j w_{i,j}u_j(t).
\]

The expression \( u_i(t+1) \) of a gene \( i \) to the regulatory input \( r_i(t) \) is “squashed” with a sigmoidal function,

\[
u_i(t+1) = \frac{m_i}{1 + e^{-(\alpha_i r(t) + \beta_i)}}.
\]

where \( m_i \) corresponds to the maximal observed expression level for the gene \( i \), and the parameters \( \alpha_i \), \( \beta_i \), and \( w_{i,j} \) correspond to the \( i \)-th gene’s intrinsic response to the regulatory inputs (or slope of the sigmoidal response function), the gene’s basal expression level, and the relative weights of the interactions respectively.

The expression levels of all \( N \) genes \( u_i(t) \) are updated simultaneously on each iteration. The simulations converge to a stable state of unchanging gene expression or cyclical set of gene expression states (see Figure 4.4). Such a steady state may depend on the initial conditions \( u(t=0) \), according to the gene expression landscape for the network studied.

For a given network the parameter space for \( \alpha, \beta, \) and \( W \) is evenly explored by a sufficiently large number of random trials. This procedure characterises the network and results in an average frequency of bistable outputs \( (P_B, \text{see section 7.2.3}) \) which should be free of the bias introduced by the choice of an arbitrary set of parameters.

The *TReMM* methodology has the advantage of representing gene regulation with a considerable degree of complexity while still being computationally tractable for the networks’ sizes studied. In this modelling methodology, the expression levels are updated simultaneously. The state of the cell at a certain time is represented by a vector \( u_i(t) \) which contains the absolute values of every gene \( i \) in the system.

The regulatory interactions between the genes are modelled with a weight. This regulatory input is then squashed with a sigmoidal function in order to model the response between maximal
repression / maximal expression,
\[ x_i(t+1) = \frac{1}{1 + e^{-(\alpha x_i(t) + \beta)}} , \]  
(4.6)

where \( \alpha_i \) and \( \beta_i \) are parameters to adjust the dose-response function. The positive parameter \( \alpha_i \) makes the response curve more steeply sloped as it approaches infinity, while \( \beta_i \) corresponds to the curve’s y-intercept, where the positive and negative regulatory inputs are equal (gene’s basal level of expression).

Multiplying the normalised response \( x_i \) by the maximum of the expression for a given gene, we can compute the absolute expression value

\[ u_i(t+1) = \frac{m_i}{1 + e^{-(\alpha \sum_{j} w_{ij} u_j(t) + \beta)}} , \]  
(4.7)

where \( m_i \) is the maximal expression of the gene \( i \), and corresponds to the maximal expression observed in any of the arrays for that given gene.

By computing the dynamics using any values to initialise \( u_i(t = 0) \), the system eventually reaches a steady state, namely, a state where the expressions of the genes do not change over time. The steady state can correspond to a fixed point or to limit cycle although in some cases, the dynamics could lead to chaotic behaviours.

While this approach has been used to predict the genetic pathways underlying observed expression data (Weaver et al., 1999; Morris et al., 2011; Saez-Rodriguez, Alexopoulos, Epperlein, Samaga, Lauffenburger, Klamt, and Sorger, 2009), it has not been proven to be valid to interpolate or extrapolate expression data points. In fact, despite the common assumption that the trajectories in silico would lead to a steady state with a corresponding real cellular steady state, it does not necessarily imply that the modelled trajectories exactly mimic the ones actually underlying the cellular processes.

### 4.2.4 Other continuous models

Other models that allow for a finer detail than linear models have been proposed. In the model of Nachman, Regev, and Friedman (2004), genes are determined by real-valued, non-linear regulation
functions that take the Michaelis-Menten form, together with mRNA decay rates. A number of similar approaches have been proposed (see for example Segal, Raveh-Sadka, Schroeder, Unnerstall, and Gaul, 2008).

Michaelis-Menten functions are used in another main class of continuous models which describe gene regulation based on ordinary differential equations. The changes in the values of the components of the network are described based on the levels of the other interacting units instantaneously. Analytical solutions other than Michaelis-Menten have been used, but larger networks usually require a numerical approach. An overview of this methodology is out of the scope of this thesis, a good review can be found in Klipp, Herwig, Kowald, Wierling, and Lehrach (2008).

Even more refined methods exist for the description of biological networks, which aim to integrate regulation and metabolism. These are generally grouped in the flux balance analysis (Palsson, 2006; Kauffman, Prakash, and Edwards, 2003).

Finally, another class of methods operate at the single-molecule level. These methods are useful when the number of molecules is small, and they incorporate the stochastic nature of regulation in the model. A review can be found in Gillespie (2007).

4.3 LTP in the context of functional genomics

A fundamental feature of most biological systems is that they exhibit non-linear behaviours and hence they are inherently difficult to analyse. However, an increasingly large amount of data available from genome sequencing projects and other -omic technologies allow for systematic analyses of gene expression. Both explanatory and predictive methods of gene regulatory networks have been proposed with success (Hasty, McMillen, Isaacs, Collins, et al., 2001). However, these efforts have generally aimed at describing simple systems which can be closely linked with experiments (for a brief review see Hasty et al., 2001).

Regulatory networks are quantified by the concentrations of the constitutive gene products. The system will be in a fixed point if small perturbations to this steady state are followed by an exponential return to the equilibrium. Ideally, a long-lasting response of a cell to a given input, requires the transcriptional machinery to transit from one steady state to another. The existence of more than one stable fixed point raises the possibility of a switch-like behaviour. In order to drive the cell from one state to another, a sufficiently large perturbation is needed.

LTP relies on gene transcription for its long-term persistence, and this stage is thought to be necessary for the acquisition of long-term memory (Nguyen et al., 1994; Abraham and Williams, 2003, 2008; Alberini, 2009). The gene regulatory network is driven from a pre-LTP state to a post-LTP state by signalling cascades following the calcium influx. This transition is believed to involve post-translational modifications, local protein synthesis, translational regulation, transcription and epigenetic changes. In addition, the LTP decay has been suggested to be a consequence of activity-dependent mechanisms (Villarreal, Do, Haddad, and Derrick, 2002), which supports a driven multistable system view. There are indeed a number of other studies that point towards the possibility that the plasticity in the hippocampus in living animals may not be very persistent (Abraham and Williams, 2003), and that long-term memories are gradually reorganised so that its storage becomes independent of the hippocampal region and distributed instead in the cortex (Squire and Alvarez, 1995; McClelland, McNaughton, and O’Reilly, 1995).

4.4 Justification for the choice of models

A number of studies suggest that the architecture of biological networks may be playing a fundamental role in the stability of cellular states (Albert, Jeong, and Barabási, 2000; Jeong, Mason,
Chapter 4. Models of gene regulatory networks

Barabási, and Oltvai, 2001; Milo, Shen-Orr, Itzkovitz, Kashtan, Chklovskii, and Alon, 2002) and hence be under evolutionary pressures (Li, Long, Lu, Ouyang, and Tang, 2004).

To characterize the topological properties of the networks involved in LTP (see Section 5.5) we have chosen logic-based models – dynamics with weight matrices (results in Chapter 7) and random Boolean networks (results outlined in Chapter 8). The choice of these models was guided by (1) the lack of any information on the rules for the interactions, which made unnecessary the use of complex continuous models, and (2) the relatively small sampling space of logic-based discrete models. Note that while discrete logic-based models for gene regulation may oversimplify the complexity underlying biological interactions, they do show remarkably rich dynamical behaviours and have a demonstrated biological significance (see e.g. Huang and Ingber, 2000; Huang, 2001; Albert and Othmer, 2003; Shmulevich et al., 2005; Serra, Villani, Graudenzi, and Kauffman, 2006). The attractors postulated to represent different cell states represent in this particular study the accessible pre- and post-LTP cell states in terms of gene expression.
5. Long-term potentiation microarray datasets

She said the root of man’s problems is memory. Without a past every day would be a new beginning

Ashes of time, Wong Kar-wai

5.1 Introduction

The microarray datasets used throughout this thesis were kindly provided as CEL files by the group of Joanna Williams at the Department of Anatomy of the University of Otago. The dataset consists of 30 Affymetrix GeneChip Rat 230.2 (TM) used and described in a series of two papers published in 2011 and 2012 (Ryan et al., 2011, 2012). We describe in this Chapter the replication of the results published in the aforementioned papers. Whereas the results presented in this chapter do not represent novel unpublished data, the aim is twofold – first, the datasets serve as a case study to describe the initial high-level analysis of microarray datasets and in addition, they represent the starting point to further explore the data using co-expression analysis (see Chapter 6) and compare the results. Additionally, the gene regulatory networks used in Chapters 7 and 8 are introduced here (see Section 5.5). Hence, the present chapter should be considered as a methodological introduction for the following chapters.

5.2 Microarray experiments

The experimental procedure to obtain the microarray image data is described in the original papers. For the scope of this thesis, the replication of the microarray analysis was only carried out using the CEL files as a starting point. The resulting normalised data was used in Chapter 6 to conduct the co-expression analysis.

We normalised the microarray data using the robust multi-array average (RMA) algorithm, implemented in the affy package of Bioconductor, a free, open source and open development software project for the analysis of genomic data (Irizarry et al., 2003). Bioconductor is based on the R programming language (R Core Team, 2013), also free and open source.
The RMA normalisation consists of three steps:

1. Background correction by convolution
   Separates probe-level signal from background signal (optical noise and non-specific binding).
   The probe-level model is fitted using PM intensity values only

2. Quantile normalisation
   Forces all arrays to have the same quantiles

3. Summarisation by median polish
   A robust measure of “centre”, using probe-level background-corrected and quantile-normalized intensities

Figure 5.1 shows the probe-level intensities before and after the normalisation by the RMA method. The output is a matrix of expression levels. Expression data from the arrays is represented as a data matrix with rows corresponding to probes and columns to arrays.

Figure 5.1: *Top-left panel shows the box plot of PM values of the raw data for each of the 30 arrays. In the top-right panel represents the density of the PM probe intensities. In the bottom panels, the same representations after RMA normalisations are represented.

5.3 Experimental design

The data consists of 30 arrays hybridised with the stimulated (S) and unstimulated (U) hemispheres of 15 rats. The animals were euthanised at different times following LTP induction (20 min, 5 h, and 24 h). These three different groups consisted of \( N = 4 \), \( N = 6 \), \( N = 5 \) respectively. For details see Ryan et al. (2011) and Ryan et al. (2012). The factorial design matrix is represented
Table 5.1: List of the time course microarrays.

<table>
<thead>
<tr>
<th>Array</th>
<th>Time</th>
<th>LTP</th>
<th>Array</th>
<th>Time</th>
<th>LTP</th>
<th>Array</th>
<th>Time</th>
<th>LTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20 min</td>
<td>S</td>
<td>9</td>
<td>5 h</td>
<td>S</td>
<td>21</td>
<td>24 h</td>
<td>S</td>
</tr>
<tr>
<td>2</td>
<td>20 min</td>
<td>U</td>
<td>10</td>
<td>5 h</td>
<td>U</td>
<td>22</td>
<td>24 h</td>
<td>U</td>
</tr>
<tr>
<td>3</td>
<td>20 min</td>
<td>S</td>
<td>11</td>
<td>5 h</td>
<td>U</td>
<td>23</td>
<td>24 h</td>
<td>U</td>
</tr>
<tr>
<td>4</td>
<td>20 min</td>
<td>U</td>
<td>12</td>
<td>5 h</td>
<td>S</td>
<td>24</td>
<td>24 h</td>
<td>U</td>
</tr>
<tr>
<td>5</td>
<td>20 min</td>
<td>S</td>
<td>13</td>
<td>5 h</td>
<td>S</td>
<td>25</td>
<td>24 h</td>
<td>U</td>
</tr>
<tr>
<td>6</td>
<td>20 min</td>
<td>U</td>
<td>14</td>
<td>5 h</td>
<td>U</td>
<td>26</td>
<td>24 h</td>
<td>U</td>
</tr>
<tr>
<td>7</td>
<td>20 min</td>
<td>S</td>
<td>15</td>
<td>5 h</td>
<td>S</td>
<td>27</td>
<td>24 h</td>
<td>S</td>
</tr>
<tr>
<td>8</td>
<td>20 min</td>
<td>U</td>
<td>16</td>
<td>5 h</td>
<td>U</td>
<td>28</td>
<td>24 h</td>
<td>U</td>
</tr>
<tr>
<td>17</td>
<td>5 h</td>
<td>S</td>
<td>29</td>
<td>24 h</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>5 h</td>
<td>U</td>
<td>30</td>
<td>24 h</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>5 h</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5 h</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The questions that can be addressed with a differential expression analysis are 1) which genes respond to LTP induction, and 2) which of those genes respond at the different times following LTP-induction. We fitted a linear model with a coefficient for each of the six factor combinations:

1. 20 min – S vs 20 min – U
2. 5 h – S vs 5 h – U
3. 24 h – S vs 24 h – U

For additional methodological details, see Methods section in Ryan et al. (2011).

5.4 Differential gene expression

A quick visual identification of genes that are significantly dysregulated can be accomplished by plotting the log-fold changes versus log-odds of differential expression (volcano plots). These plots combine a measure of statistical significance with the magnitude of the change. The volcano plots for the three time points are shown in Figure 5.2, where the outliers are also labelled.

**Figure 5.2:** The volcano plots in the figure represent the log-fold changes versus log-odds of differential expression for each of the time points
Table 5.2: Lists of differentially expressed genes with known transcripts for each of the times post-LTP induction. The genes with \(-0.15 > \ln(FC) > 0.15\) are ranked according to their p-value. Only the top 20 genes of each time point are shown. A more comprehensive list can be consulted in the Appendix A.1. Genes in red cells were significantly over expressed in the LTP-induced hemisphere, genes in green cells were significantly under expressed. FC: Fold Change

<table>
<thead>
<tr>
<th></th>
<th>20 min</th>
<th></th>
<th>5 h</th>
<th></th>
<th>24 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>ln(FC)</td>
<td>p-value</td>
<td>Gene</td>
<td>ln(FC)</td>
<td>p-value</td>
</tr>
<tr>
<td>Syt2</td>
<td>0.283</td>
<td>1.50E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptf1a</td>
<td>0.312</td>
<td>3.74E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myoc</td>
<td>0.277</td>
<td>4.30E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mmp3</td>
<td>0.290</td>
<td>4.67E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fzd5</td>
<td>0.236</td>
<td>6.41E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slc21a4</td>
<td>0.210</td>
<td>6.76E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erfa2</td>
<td>0.202</td>
<td>7.13E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pgap2</td>
<td>0.222</td>
<td>7.44E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prod2</td>
<td>0.343</td>
<td>8.05E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ith4</td>
<td>0.287</td>
<td>8.18E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trim14</td>
<td>0.287</td>
<td>8.23E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cmah</td>
<td>0.226</td>
<td>9.35E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memo1</td>
<td>0.212</td>
<td>9.82E-03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap1</td>
<td>0.249</td>
<td>1.12E-02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sem6aC</td>
<td>0.281</td>
<td>1.12E-02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oncuc1</td>
<td>0.226</td>
<td>1.13E-02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ard5A</td>
<td>0.233</td>
<td>1.16E-02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fat4</td>
<td>0.220</td>
<td>1.19E-02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp3a2</td>
<td>0.170</td>
<td>1.21E-02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tet3</td>
<td>0.359</td>
<td>1.25E-02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The list of the top 20 differentially expressed genes for each of the groups (20 min, 5 h, 24 h after induction of LTP) are shown in Table 5.2. The cut-off for fold change (FC) was set to \(-0.15 > \ln(FC) > 0.15\), which corresponds to a ratio of expression of around 1 : 1.16 between stimulated and unstimulated tissues. The genes are ranked according to their p-values. Note that only the genes with a known transcript are shown. As there are multiple probesets to a given transcript and some transcripts are only predicted, we include a more comprehensive table with all top 50 probesets per group in Appendix A.1. The number of differentially expressed genes at different times with respect to their unstimulated controls are represented in the Venn diagram of Figure 5.3.

The normalisation procedures carried out and outlined in the present chapter are necessary to the co-expression analysis performed in the following chapter. The results obtained in the differential expression analysis, on the other hand, are largely identical to the originally reported in Ryan et al. (2011, 2012). Minor differences are to be expected since the databases of transcript-gene mappings have been updated since the original work was published. Results are summarised in Figure 5.3 and Table 5.2.

5.5 Gene regulatory networks

The networks used in Chapter 7 and 8 were obtained from Ryan et al. (2011) and Ryan et al. (2012). They were originally identified by using the lists of differentially expressed genes as an input to Ingenuity pathway analysis (IPA; Ingenuity Systems, USA; www.ingenuity.com).

Based on the original input list of genes, the application generates a set of networks that can contain genes from the original list in addition to other genes which have a significant number of interacting partners from the list. Each network contains up to 35 genes and has an associated score which indicates the expected likelihood of not being generated by chance. The main advantage of IPA with respect to other tools of similar characteristics is the proprietary IPA knowledge base.
Figure 5.3: The Venn diagram in the centre of the figure shows the number of differentially expressed genes ($p < 0.05$) at the different times post-LTP. The fraction of the genes found to be differentially expressed for more than one time points are depicted in the heatmaps. Green represents under-expression and red represents over-expression. Only one probeset shows a differential expression on every time point (1376179_at). In order to visualise all the differential expression information overlap between time points, for this figure no cut-off on fold change was used.

(IPKB), queried for the network constructions. The IPKB is a manually curated interaction database of cellular proteins sourced from literature references.

A broader discussion of these networks can be found in the original publications (Ryan et al., 2011, 2012). The networks are depicted in Figure 5.4.
Figure 5.4: From the LTP networks previously identified by Ryan et al. (2011, 2012) at different times post-LTP induction, we represent in this figure the most significant for each time point. These networks were identified using IPA software.
6. Co-expression analysis of LTP microarray data

6.1 Introduction

The cellular and molecular mechanisms underlying LTP seem to be several and to operate at different scales. Currently, we know that LTP exhibits at least two mechanistically distinct phases of maintenance – a protein-synthesis independent early phase and a protein-synthesis dependent phase which lasts for longer periods.

A third biochemically distinct phase has been identified. This third phase is a translation- and transcription-dependent phase. The role of *de novo* gene transcription in the late phases of LTP has drawn considerable attention in the last decades. New gene expression seems to be of crucial importance in the downstream processes required to maintain late LTP. Within minutes of the high-frequency stimulus (HFS), the expression of immediate early early genes (IEGs) is induced by activated constitutive transcription factors. This wave of expression is followed by a second wave, the late response genes (LRGs; see Section 3.3.6).

The variety of functions held by the non-nuclear IEG products (or “effector IEGs”) offers a glimpse of the potential complexity underlying the genomic response following LTP induction. A number of studies have aimed at unravelling this tightly coordinated response by using a range of microarray technology platforms (Lee *et al.*, 2005; Park *et al.*, 2006; Wibrand, Messaoudi, Hävik, Steenslid, Lövlie, Steen, and Bramham, 2006; Hävik *et al.*, 2007a).

In the light of its apparent complexity it becomes necessary to investigate the effects that LTP induction elicits from an integrative perspective. In order to overcome the limitations imposed by the classical differential expression analysis and also to quantify temporal network reorganization, we describe in this Chapter the analysis of the microarray data from Ryan *et al.* (2011, 2012) described in Chapter 5 using weighted gene co-expression network analysis (WGCNA Zhang, Horvath, *et al.*, 2005). By clustering the genes according to the similarity in their co-expression...
patterns, we identified a set of modules of highly connected genes. This analysis allowed us to pinpoint genes of relevance based on their high intramodular connectivity. In addition, the functional enrichment in the different modules allows us to understand the cellular context in which these highly connected genes function – the degree of connectivity of a gene has been correlated with its essentiality. Essential genes are those genes that are critical for the survival of the organism or for the achievement of a particular cellular function, in this case, LTP (Giaever, Chu, Ni, Connelly, Riles, Veronneau, Dow, Lucau-Danila, Anderson, Andre, et al., 2002).

We inspected and quantified the reconfiguration of the molecular interaction structure within these gene modules along the different phases of LTP by using two complementary measures. First, a module-centric measure which indicates the average gain or loss of intramodular connectivity with time. Secondly, we considered a gene-centric measure to take into account the changes in expression similarities between pairs of genes in time.

6.2 Methods

6.2.1 Data sets and sample clustering

Raw gene expression data was obtained from Ryan et al. (2011) and Ryan et al. (2012) and corresponds to 30 brain tissue samples from 15 adult rats. More precisely, they induced LTP on freely moving animals unilaterally so that for each animal an unstimulated hemisphere could be used as control. At different times post-LTP induction, groups of animals were anaesthetized and decapitated: 20 min, 5 h, and 24 h ($N = 4, N = 5, N = 6$ respectively). These times are representative of the different phases of LTP.

The data had been previously normalised using the robust multichip average package (RMA; Irizarry, Bolstad, Collin, Cope, Hobbs, and Speed, 2003). In addition, the correlation between measurements on the same subject (within-animal correlation) was estimated and considered in the model fit. The t-statistic was generated using the LIMMA package, which uses a standard error moderated across all genes using a simple Bayesian model (Smyth, 2005).

The moderated paired t-test between the stimulated and unstimulated pairs of samples produced three lists, one for each temporal contrast, ranking the genes according to the significance of their differential expression (p-value). For computational reasons, we restricted the network analysis to the subsets of the top genes of each of the lists according to their p-value. Namely, we merged the top 1,700 genes of each of the temporal contrasts, so that the contribution to the final list of each of the different time groups was equivalent in terms of number of genes. This intersection resulted in a set of differentially expressed genes across early and late-LTP (L-LTP) of 4,804 genes with p-values ranging from $7.7 \times 10^{-6}$ to $6.7 \times 10^{-2}$.

The 30 brain samples were clustered according to the similarity in their gene expression profiles using a standard unweighted pair group method with arithmetic mean (UPGMA). The clustering dendrogram is shown in Figure 6.1. Samples tend to cluster together reflecting their belonging to the different time points. As there appear not to be outliers, every sample was kept for further analyses.

6.2.2 Construction of gene co-expression networks

The identification of modules of highly co-expressed genes for the three temporally distinct LTP-induced tissues and control was carried out using the WGCNA package (Langfelder and Horvath, 2008). The details are further reviewed in Zhang et al. (2005). For every pair of genes $i, j$ the Pearson correlation is calculated and transformed into an adjacency measure with a power function, $a_{ij}$. We adopted a scale-free topology criterion for choosing the parameter for the adjacency function. Figure 6.2 shows the scale-free fit index as a function of the parameter $p$ used for the power adjacency function. We chose $p = 5$ since at around that value the $R^2$ fit curve reaches
Figure 6.1: Clustering of the 30 brain samples according to the similarity in terms of gene expression profiles. Only the 4,804 most differentially expressed genes were used to perform the clustering.

saturation for the unstimulated and 24 h co-expression networks. Higher values of the exponent may lead to networks with very few connections (Zhang et al., 2005).

In order to identify modules of highly connected genes, the topological overlap (TO) (Ravasz, Somera, Mongru, Oltvai, and Barabási, 2002) is used as a similarity measure:

\[ w_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}} \]

where \( l_{ij} = \sum a_{iu} a_{uj} \) and \( k_i = \sum a_{iu} \) is the node connectivity and \( a_{ij} \) is the adjacency between the expression profiles of the genes \( i, j \). In this way we obtained four undirected weighted networks constituted each by the same 4,804 nodes but whose connections would differ depending on the values of TO. The first of the networks corresponds to the unstimulated tissues (unstimulated hemispheres) and the other three to the different times after LTP induction (20 min, 5 h, and 24 h). Figure 6.5 represents these weighted networks as heatmaps.

To identify the gene co-expression modules, the TO is used as a similarity measure to perform a hierarchical clustering. The default parameters for the cluster splitting are left unchanged (for more details, refer to Zhang et al., 2005). For each of the four groups we obtained specific gene co-expression modules based on their TO.
Figure 6.2: (a) Scale-free fit index as a function of the soft-thresholding power \( p \). The degree distributions corresponding to the red dashed lines are expanded in the panels of Figure 6.4. The \( R^2 \) fit to a scale-free distribution of the unstimulated control and the 24 h networks are both higher and saturate at lower values of \( p \) than the earlier networks (20 min and 5 h data). The latter reach a saturation only of around \( R^2 = 0.5 \).

(b) Mean connectivity as a function of the soft-thresholding power \( p \).

### 6.2.3 Functional enrichment

The Gene Ontology (GO) project (Ashburner, Ball, Blake, Botstein, Butler, Cherry, Davis, Dolinski, Dwight, Eppig, et al., 2000) is a database that unifies descriptions of gene products from different sources. It consists of three structured vocabularies describing gene products in terms of their associated biological processes, cellular components and molecular functions.

Genes belonging to the same module are likely to be related functionally. In order to detect GO terms that are associated with a significant number of genes within a given module, we used the R package `topGO` (Alexa and Rahnenfuhrer, 2010). This tool finds significant shared GO terms or parents of those GO terms, associated with the set of genes. The statistical significance was assessed through Fisher’s exact test.

### 6.2.4 Measures of network reconfiguration

For each of the 58 modules obtained by WGCNA (9, 9, 24, and 16 modules for the control, 20 min post-LTP, 5 h post-LTP, and 24 h post-LTP respectively), we quantified the modular differential connectivity (MDC; Zhang, Gaiteri, Bodea, Wang, McElwee, Podtelezhnikov, Zhang, Xie, Tran, Dobrin, et al., 2013). The MDC is a module-centric measure and corresponds to the ratio of the average connectivity for any pair of module-sharing genes at time \( T_1 \) compared to that of the same genes at time \( T_2 \).

\[
\text{MDC}(T_1, T_2) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_{ij}^{T_1}}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_{ij}^{T_2}}
\]
where \( w_{ij} \) is the TO between two genes \( i \) and \( j \) in a given network (see Figure 6.3). The statistical significance was assessed by a false discovery rate (FDR) based on permutation of the gene labels. This corresponds to randomly permuting the rows and columns of the TO matrix of a given module. MDC > 1 indicates enhanced co-regulation between genes, whereas MDC < 1 indicates reduced co-regulation. The MDC was calculated for each of the 58 networks for each of the 3 temporal transitions (control \( \rightarrow \) 20 min, 20 min \( \rightarrow \) 5 h, 5 h \( \rightarrow \) 24 h).

In addition, we used a gene-centric measure to rank the pairs of genes according to the specificity of the TO value in the adjacent temporal points. We employed a measure based on Oldham, Horvath, and Geschwind (2006)

\[
S_{ij}(T_1, T_2) = \frac{w_{ij}^{T_1}}{\text{mean}(w_{ij}^{T_1})} + \frac{w_{ij}^{T_2}}{\text{mean}(w_{ij}^{T_2})}
\]

For a pair of genes \( i, j \) in a given module, a high \( S_{ij}(T_1, T_2) \) captures a high TO at time \( T_1 \), and a low TO at time \( T_2 \). A low \( S_{ij}(T_1, T_2) \) value captures the opposite relation, while a value of 0.5 means no change in the TO for those genes.

Each gene can also be ranked according to its degree, which identifies the main hubs within a module. As the networks are weighted, the degree is calculated simply by summing over each connection weight (TO) and normalising by the size of the module.

6.3 Results

6.3.1 Structural organisation of the temporally distinct networks

It is widely accepted that biological networks tend to have connectivity distributions which approximate scale-free distributions (Jeong, Tombor, Albert, Oltvai, and Barabási, 2000). In that aspect, the four distinct co-expression networks identified in this study exhibit different network properties in relation to the fit to a scale-free topology. Figure 6.2 depicts the fit to a scale-free topology model with different values for the soft thresholding power \( p \) used for the adjacency function. Note that for higher values of \( p \) the fit reaches saturation. From a qualitative point of view, the networks seem to reach the saturation at different values of \( p \), with the unstimulated control and the 24 h networks saturating earlier (\( p \approx 5 - 10 \)) than the 20 min and 5 h networks (\( p \approx 20 \)). In addition, the saturation value of the scale-free fit coefficient is higher for the control and 24 h than for the 20 min and 5 h networks.

A closer inspection to the scale-free model fit in Figure 6.4 reveals that in fact, at sufficiently high values of \( p \), both the unstimulated control and the 24 h networks fit to a scale free distribution. On the other hand, the earlier networks have few nodes with a low degree (truncated left side of the distribution). Considering that the scale-free architecture has been shown to display a high robustness against random perturbations (Albert et al., 2000), and that this property seems to have been selected for in biological networks (Jeong et al., 2000), it is plausible that transient topological rearrangements exhibit less stable architectures. We hypothesise that such could be the case of the earlier 20 min and 5 h networks – the homeostatic cell state acquired 24 h post-induction could have restored a scale-free architecture that is also observed in the unstimulated tissue. The intermediate networks at 20 min and 5 h post-LTP could be temporarily organised in topologies that resemble the characteristics of random networks.

As proposed by Zhang et al. (2005), we adopted the value of the fit as a criterion to select a power for the adjacency function. Given the above observations we opted to chose an exponent where the curve reaches saturation for the unstimulated control and the 24 h networks (\( p = 5 \)). For each temporal network generated after LTP-induction, we show in Table 6.1 the top hubs.
Table 6.1: Top hubs in the co-expression networks

### 20 min network

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>Gene Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1395900_at</td>
<td>Chtf8</td>
<td>CTF8, chromosome transmission fidelity factor 8 homolog (S. cerevisiae)</td>
</tr>
<tr>
<td>1385824_at</td>
<td>Cep350</td>
<td>centrosomal protein 350</td>
</tr>
<tr>
<td>1381003_at</td>
<td>Ikar2</td>
<td>IKAROS family zinc finger 2</td>
</tr>
<tr>
<td>1386234_at</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1391555_at</td>
<td>Ncoa3</td>
<td>nuclear receptor coactivator 3</td>
</tr>
<tr>
<td>1388079_at</td>
<td>Cacng8</td>
<td>Ca(^{2+}) channel, voltage-dependent, gamma subunit 8</td>
</tr>
<tr>
<td>1388684_at</td>
<td>Fnbp4</td>
<td>formin binding protein 4</td>
</tr>
<tr>
<td>1382979_at</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1387455_at</td>
<td>St8sia3</td>
<td>ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 3</td>
</tr>
<tr>
<td>1387795_at</td>
<td>Pola2</td>
<td>polymerase (DNA directed), alpha 2</td>
</tr>
</tbody>
</table>

### 5 h network

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>Gene Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1384230_at</td>
<td>Krtcap3</td>
<td>keratinocyte associated protein 3</td>
</tr>
<tr>
<td>1374827_at</td>
<td>Ndst2</td>
<td>N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 2</td>
</tr>
<tr>
<td>1383540_at</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1384860_at</td>
<td>Zfp84</td>
<td>zinc finger protein 84</td>
</tr>
<tr>
<td>1394492_at</td>
<td>RGD1563482</td>
<td>similar to hypothetical protein FLJ38663</td>
</tr>
<tr>
<td>1368005_at</td>
<td>Itrp3</td>
<td>inositol 1,4,5-triphosphate receptor, type 3</td>
</tr>
<tr>
<td>1371697_at</td>
<td>Pnpla2</td>
<td>patatin-like phospholipase domain containing 2</td>
</tr>
<tr>
<td>1368229_at</td>
<td>Sip1</td>
<td>survival of motor neuron protein interacting protein 1</td>
</tr>
<tr>
<td>1385928_at</td>
<td>Smad6</td>
<td>SMAD family member 6</td>
</tr>
<tr>
<td>1368321_at</td>
<td>Egr1</td>
<td>early growth response 1</td>
</tr>
</tbody>
</table>

### 24 h network

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>Gene Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1369067_at</td>
<td>Nf4a3</td>
<td>nuclear receptor subfamily 4, group A, member 3</td>
</tr>
<tr>
<td>1369398_at</td>
<td>Naalad1</td>
<td>N-acetylated alpha-linked acidic dipeptidase-like 1</td>
</tr>
<tr>
<td>1369558_at</td>
<td>I11r1</td>
<td>interleukin 1 receptor, type 1</td>
</tr>
<tr>
<td>1384999_at</td>
<td>Lcel1d</td>
<td>late cornified envelope 1D</td>
</tr>
<tr>
<td>1371003_at</td>
<td>Map1b</td>
<td>microtubule-associated protein 1B</td>
</tr>
<tr>
<td>1369237_at</td>
<td>Sלcfα7</td>
<td>solute carrier family 6 (neurotransmitter transporter, L-proline), member 7</td>
</tr>
<tr>
<td>1380864_at</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1397942_at</td>
<td>Cdc371l</td>
<td>cell division cycle 37 homolog (S. cerevisiae)-like 1</td>
</tr>
<tr>
<td>1370641_s_at</td>
<td>Cacna1i</td>
<td>Ca(^{2+}) channel, voltage-dependent, T type, alpha 1I subunit</td>
</tr>
<tr>
<td>1377276_at</td>
<td>Cdk5r2</td>
<td>cyclin-dependent kinase 5, regulatory subunit 2 (p39)</td>
</tr>
</tbody>
</table>
6.3 Results

Figure 6.3: The MDC measures the change in the average connectivity between the nodes within a module. In the example, the module magenta_20 is used as an example. It is constituted by a total of 64 nodes (probes in the array). The colour of the edges connecting the nodes in the networks correspond to their TO. The nodes are the same in the four networks, but their connectivity varies at different times. The y-axis represents the average TO of the module and the x-axis represents time. The asterisks (*) represent significant variation according to 10% FDR as described in the Methods. A simplified version of this plot is shown in the right-top corner inset for this particular module, and for each of the modules in Figure 6.7

6.3.2 Modular functional analysis

The identification of co-expression modules within the four temporal networks consists in the selection of a height for cutting the dendrograms. This is done automatically by the WGCNA package, and results in a number of densely connected subsets of networks corresponding to the branches of the dendrogram. This procedure was conducted independently for each of the four network datasets and resulted in control, 9 modules ranging from 69 to 2327 genes; 20 min, 9 modules ranging from 64 to 1184 genes; 5 h, 24 modules ranging from 31 to 535; and 24 h, 16 modules ranging from 42 to 1164 genes). The distribution is similar in the early phase of LTP when compared to the unstimulated tissue, whereas the networks generated for the later times (5 h and 24 h) appear to be more dissociated, splitting up into more modules for the same power threshold (Figure 6.5 and Table 6.2).

The intersections between the 58 modules identified are not always empty since each temporal expression dataset was analysed independently. As larger modules in the control and 20 min samples appear to segregate into different sets of genes, we chose to keep all the modules and not to merge overlapping modules. We show in Figure 6.6 the normalised overlap between modules.

The different modules were tested for Gene Ontology (GO) term enrichment. A representative top GO term for each module is shown in Table 6.2. For a more comprehensive list, see the
Figure 6.4: Degree distributions of the networks using different values of p. From left to right, unstimulated control, 20 min, 5 h, and 24 h networks. From top to bottom, p = 5, p = 10, and p = 20. The distributions are log-transformed both in the x and y axes. The black lines represent the linear model fit. For the 20 min and 24 h networks, an additional linear model fit is shown with a red line, using only the right decrementing side of the distribution. Values of $R^2$ are shown for the linear model fits.

Appendix A.2.

Table 6.2: List of modules identified by WCGNA with at least 50 genes

<table>
<thead>
<tr>
<th>Module</th>
<th>Size</th>
<th>Functional category</th>
<th>Top genes by $k_{GO}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cyan_24</td>
<td>97</td>
<td>endosome transport</td>
<td>Nog, Thbd, Zscan10</td>
</tr>
<tr>
<td>green_U</td>
<td>206</td>
<td>cation transmembrane transport</td>
<td>Camk4, Sf6gal1, Slc31a1</td>
</tr>
<tr>
<td>black_20</td>
<td>329</td>
<td>(+) reg. of endocytosis</td>
<td>Arhgap27, Kdm5b, Acrv1</td>
</tr>
<tr>
<td>brown_24</td>
<td>459</td>
<td>cofactor transporter activity</td>
<td>Mast2, Mst8, Fgd2</td>
</tr>
<tr>
<td>black_24</td>
<td>269</td>
<td>epithelial polarization</td>
<td>Cpn1, Thoc2, Pqcl3</td>
</tr>
<tr>
<td>brown_20</td>
<td>730</td>
<td>neuron part and cytoplasmic microtubule</td>
<td>Ddi2, Atg5i, Rab22a</td>
</tr>
<tr>
<td>red_20</td>
<td>455</td>
<td>axogenesis</td>
<td>Sic10a5, Tnfrs17, Slc4a11</td>
</tr>
<tr>
<td>yellow_U</td>
<td>222</td>
<td>BRCA1-A complex</td>
<td>Acap2, Dvl1, Alpk3</td>
</tr>
<tr>
<td>blue_20</td>
<td>779</td>
<td>leukocyte activation</td>
<td>Pias2, Atp6v1h2, Pldz3</td>
</tr>
<tr>
<td>green_20</td>
<td>461</td>
<td>response to axon injury</td>
<td>Igha, Tp53rp1, Tal1</td>
</tr>
<tr>
<td>pink_20</td>
<td>78</td>
<td>calmodulin-dependent kinase activity</td>
<td>Lmo2, Pacsin1, HmoX3</td>
</tr>
<tr>
<td>yellow_20</td>
<td>724</td>
<td>transcription from RNApolI promoter</td>
<td>Ndts2, Kcnj12, Pppn7</td>
</tr>
<tr>
<td>turquoise_20</td>
<td>1184</td>
<td>oxidoreductase activity</td>
<td>Brp11, Tsta3, Kdcl1</td>
</tr>
<tr>
<td>blue_24</td>
<td>753</td>
<td>reg. of endocrine process</td>
<td>Rtf1-Da, Mpl14, Ccnd1</td>
</tr>
<tr>
<td>turquoise_24</td>
<td>1164</td>
<td>neuron projection membrane</td>
<td>Akr3, Cacna1l, Bbm4</td>
</tr>
<tr>
<td>black_U</td>
<td>145</td>
<td>activation of prot kinase and membrane</td>
<td>Znf609, Cd24, Dab2</td>
</tr>
<tr>
<td>pink_24</td>
<td>238</td>
<td>fatty-acyl-CoA binding</td>
<td>Gtf3c6, Ak3l1, Ap2a2</td>
</tr>
<tr>
<td>yellow_24</td>
<td>355</td>
<td>proteasomal protein catabolism</td>
<td>Qtrt1, Sh3glb1, Hira</td>
</tr>
<tr>
<td>magenta_U</td>
<td>69</td>
<td>integrin binding</td>
<td>Uba6, Samd14, Atrx</td>
</tr>
<tr>
<td>pink_U</td>
<td>72</td>
<td>mitochondrial transport and apoptosis</td>
<td>nod3l, Rnasen, Glice</td>
</tr>
</tbody>
</table>

Continued on next page
6.3 Results

Table 6.2 – continued from previous page

<table>
<thead>
<tr>
<th>Module</th>
<th>Size</th>
<th>Functional category</th>
<th>Top genes by FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>green_24</td>
<td>281</td>
<td>histone demethylation</td>
<td>Gls, Junb, Fam153a</td>
</tr>
<tr>
<td>brown_U</td>
<td>691</td>
<td>synapse and reg. of secretion</td>
<td>Alox5, Kenj4, Dhrs9</td>
</tr>
<tr>
<td>greenyellow_24</td>
<td>143</td>
<td>tau-protein kinase activity</td>
<td>Hspb3, Hist2h2be, Hiat1</td>
</tr>
<tr>
<td>magenta_24</td>
<td>212</td>
<td>proteosomal protein catabolism</td>
<td>Hectd1, Nans, Sec1</td>
</tr>
<tr>
<td>midnightblue_24</td>
<td>86</td>
<td>clathrin-coated endocytic vesicle</td>
<td>Reg3a, Dmnt11, Cten</td>
</tr>
<tr>
<td>red_24</td>
<td>278</td>
<td>septin complex</td>
<td>Csep, Cc2d1a, Pdia6</td>
</tr>
<tr>
<td>purple_24</td>
<td>175</td>
<td>cAMP-mediated signaling</td>
<td>Epor, Xmppep3, Fam120b</td>
</tr>
<tr>
<td>blue_U</td>
<td>836</td>
<td>dephosphorylation and DNA binding</td>
<td>Gabra5, Cdkn2c, Kl</td>
</tr>
<tr>
<td>blue_5H</td>
<td>521</td>
<td>amino acid biosynthesis</td>
<td>Scn11a, Abi3, Clec10a</td>
</tr>
<tr>
<td>lightyellow_5H</td>
<td>88</td>
<td>CNS neuron axonogenesis</td>
<td>C1tnf3, Fbn1, St8sia3</td>
</tr>
<tr>
<td>black_5H</td>
<td>237</td>
<td>anion homeostasis and synapse assembly</td>
<td>Slic2a4, Pdzd4</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>142</td>
<td>reg. of DNA methylation</td>
<td>Fmod, Fam153a, Ankrd6</td>
</tr>
<tr>
<td>magenta_5H</td>
<td>208</td>
<td>progestosterone receptor signaling</td>
<td>Mpl35, Prkd3, Cull5</td>
</tr>
<tr>
<td>green_5H</td>
<td>336</td>
<td>GTP-Rho binding and mitochondrion</td>
<td>Prelid2, Ptrl2b1, Abca8</td>
</tr>
<tr>
<td>lightgreen_5H</td>
<td>95</td>
<td>T cell migration</td>
<td>H3f3b, Cnh2, Trps1</td>
</tr>
<tr>
<td>yellow_5H</td>
<td>358</td>
<td>chromatin DNA binding</td>
<td>Arglu1, Mcccl, Tmem206</td>
</tr>
<tr>
<td>turquoise_5H</td>
<td>535</td>
<td>DNA catabolism</td>
<td>Crrh1, Kdm6b, F8</td>
</tr>
<tr>
<td>purple_5H</td>
<td>165</td>
<td>histone H3-K27 methylation</td>
<td>Fyf21, Adey4, Kilh22</td>
</tr>
<tr>
<td>salmon_5H</td>
<td>151</td>
<td>oxidoreductase activity</td>
<td>Hs3r2, Hdac5, Ccch115</td>
</tr>
<tr>
<td>tan_5H</td>
<td>158</td>
<td>MAPK import into nucleus</td>
<td>Asb1, Tpr, Pex5l</td>
</tr>
<tr>
<td>brown_5H</td>
<td>458</td>
<td>K⁺ transport and Ras GTPase binding</td>
<td>Dhh, Cog7, Dgki</td>
</tr>
<tr>
<td>greenyellow_5H</td>
<td>162</td>
<td>response to Ca²⁺</td>
<td>Ftr3, Cngai1, Adra1d</td>
</tr>
<tr>
<td>red_U</td>
<td>178</td>
<td>reg. of GTPase activity</td>
<td>Nppa, Cyp8b1, Igifbp2</td>
</tr>
</tbody>
</table>

6.3.3 Network structure reconfiguration

The four distinct expression profiles studied represent different cellular states following LTP induction. We have studied the transitions from the pre-LTP state represented by the unstimulated hemispheres to the different gene expression snapshots post-LTP induction. For each temporal transition, control → 20 min, 20 min → 5 h, and 5 h → 24 h, we have examined the topological changes in the 58 modules identified using WGCNA.

Firstly, the fraction of modules showing a significant increase in MDC is similar along the time samples, whereas the fraction of modules with a significant loss of connectivity increases (stacked bars plot, Figure 6.7).

The response elicited rapidly after LTP induction at the gene expression level corresponds to the transition observed between the control and the 20 min dataset in terms of MDC. Out of the 58 modules, 57 show a gain of connectivity (MDC > 1, of which 20 are significant with FDR < 10% (a more strict FDR < 1% gives still a total of 10 significant modules). This marked increase in the average TO connectivity takes place in combination with the documented enrichment in upregulated genes within the set of differentially expressed genes (Ryan et al., 2011). The modules showing the highest significance (FDR < 1%) in the connectivity gain during this early LTP stage display an enrichment in functions such as membrane compartment, endosome transport and endocytosis. Other functions over represented in these networks are cation transmembrane transport, CaM-dependent kinase activity and transcription from RNApolI. Only the module with the function regulation of GTPase activity seems to undergo a loss of average TO in the early genetic response following LTP.

Ryan et al. (2012) reported that the general upregulation of LTP genes observed at 20 min lasts for at least 5 h using differential expression analysis. We note a global loss of connectivity between the 20 min and 5 h networks (0.187 to 0.089 average TO respectively). Yet, the proportion of modules exhibiting gain/loss/conservation of connectivity is very similar (17/17/24, respectively). More interestingly, the fraction of the modules with a significant loss of MDC corresponds to modules that showed an increase in MDC in the earliest transition. Amino acid biosynthesis, central neural system neuron axonogenesis, anion homeostasis and synapse assembly are some of the
functions assigned to the modules exhibiting a significant gain of connectivity from 20 min to 5 h. Other interesting functions over-represented are regulation of DNA methylation, GTP-Rho binding, and MAPK import into nucleus.

These modules subsequently undertake a loss of connectivity in the final transition to the 24 h phase, together with the modules enriched with the functions K\(^+\) transport, Ras GTPase binding and response to Ca\(^{2+}\). During this last transition, a large number of networks show an increase in average TO, with ascribed functions such as neuron projection membrane, activation of protein kinase, histone demethylation, synapse, regulation of secretion, tau-protein kinase activity, protein catabolism, clathrin-coated endocytic vesicle, and cAMP-mediated signalling.
Figure 6.6: Gene overlap between the modules generated from the different expression data corresponding to the different times post-LTP induction. The overlap was normalised by the size of the smallest module. The colour in the heatmap represents the FDR of the overlap. Red and blue represent a significant overlap and a significantly low number of genes in common respectively.

6.3.4 Overview of the time-points

20 min modules

The results of Ryan et al. (2012) evidenced a rapid LTP-induced increase in gene expression. We found that, in addition, the modules identified by WGCNA tend to increase their intramodular connectivity (average TO, see Section 6.2.2) – out of the 58 modules identified, only red_U shows a decrease in the average TO. Not surprisingly, this module corresponds to a set of genes co-expressed in the unstimulated hemispheres, with functions not specifically related with neural tissue. Out
Figure 6.7: Module reconfiguration following LTP. Each panel corresponds to one of the 58 modules identified. Time on the x-axis (control, 20 min, 5 h, and 24 h post-LTP induction) is plotted against the average TO of the module. For each transition between time points, green represents a significant gain of connectivity ($MDC > 1$; $FDR < 10\%$), red loss of connectivity ($MDC < 1$; $FDR < 10\%$), and grey conserved connectivity ($FDR > 10\%$). The number of modules showing each of these is represented in the bottom right panel.

Of the 57 modules that show an increase in intramodular TO ($MDC > 1$), 20 show statistical significance ($FDR < 10\%$, see Figure 6.7).

Among the modules that show a significant gain of connectivity, it is worth noting that some
display a significant loss of connectivity 5 h post-LTP. This abrupt peak in co-expression is exhibited by the 9 modules identified in the 20 min dataset (green_20, magenta_20, yellow_20, turquoise_20, blue_20, black_20, pink_20, brown_20, red_20, see Figure 6.8a). These networks do not show any overlap in genes. The rapid recruitment of some clusters of co-expressed genes during the first minutes in response to the experimental HFS is in agreement with the idea of the induction of IEGs discussed in previous chapters (see Section 3.3.6).

Figure 6.8: The three main trends in modular connectivity changes across time in the co-expression modules identified is shown. (a) Modules identified using the 20 min data all share a similar trend – they become tightly co-regulated at 20 min to lose the connectivity at 5 h. (b) Modules identified at 5 h exhibit a significant decrease in TO at 24 h. (c) Modules identified in the 24 h data together with black_U and brown_U modules (identified using only the control data) show a significant gain in intramodular connectivity at 24 h.

Green lines represent significant gain of intramodular connectivity (MDC > 1, FDR < 10%) and red lines represent significant loss of modular connectivity (MDC > 1, FDR < 10%). Grey lines represent no statistical significance for the MDC.

When the top 35 genes networks are subjected to an analysis of GO term enrichment, some interesting results arise. First, the red_20 cluster shows “long term synaptic depression” (GO:0060292)
as a significantly represented function \((p = 1.911e − 03, \text{ genes Pick1 and Plk2})\). The yellow_-20 network is densely connected with “regulation of ion transmembrane transporter activity” (GO:0032412) \((p = 3.897e − 04)\) and “synaptic transmission” (GO:0007268) \((p = 1.821e − 02)\) as salient overrepresented functions. The green_U network shows “convergent extension involved in axis elongation” (GO:0060028) \((p = 4.913e − 04)\) and “positive regulation of synaptic transmission, glutamatergic” (GO:0051968) \((p = 9.375e − 03)\) as overrepresented GO terms. Last, black_20 contains the only two genes annotated with the function “fasciculation of motor neuron axon” (GO:0097156) in the GO database, Epha3 and Epha4 \((p = 6.881e − 05)\).

5 h modules

We have described a set of networks that show a significant loss of intramodular connectivity at 5 h, which is then maintained in the later 24 h dataset. The general loss of intramodular connectivity in the weighted co-expression network is evident from Figure 6.5 and parallels the onset of a general downregulation of gene expression.

A different temporal trend, however, is exhibited by a number of modules identified at 5 h – a gain in connectivity between the unstimulated control and the 20 min is followed by a subsequent significant gain at 5 h to finally lose connectivity at 24 h. And yet, these set of modules show a rather mild average TO when compared to the networks discussed in the previous section (see Figure 6.8b). These results indicate that a strong and coordinated genomic response rapidly follows LTP induction, and that there exist different phases of LTP from the perspective of the genomic response. The reason why the high co-expression levels observed at 20 min are not transferred to other networks but rather lowered is unclear. However, the experimental evidence of a critical phase in the first minutes that follow the HFS in which gene expression is necessary for L-LTP may account for these observations.

The modules identified at 5 h which show a gain of intramodular TO at 20 min and at 5 h followed by a loss of TO at 24 h are summarised in Table 6.5. Perhaps the most representative module in the 5 h dataset is the purple_5H, which shows the lowest gene overlap with each of the networks identified at 20 min. Furthermore, from the top 30 genes ranked according to their average TO in the full weighted co-expression network at 5 h, 6 of them belong to the module. More than half of the genes in the module are related with metabolism and biosynthetic processes \((p = 1.302e − 07)\), in particular transcription. Regulation of gene expression (31 out of 87 known mapped genes) is also an enriched function \(p = 2.035e − 04\). Interestingly, when only the top 35 genes of the network are queried for functional enrichment, the most significant GO term is “synapse maturation” \((p = 1.140e − 03)\). Nrnx and Shank1 are among the top 35 genes. Nrnx1 is downregulated at 5 h and encodes for Neurexin 1, a membrane protein involved in synaptic signal transmission (De Wit, Sylwestrak, O’Sullivan, Otto, Tiglio, Savas, Yates III, Comoletti, Taylor, and Ghosh, 2009). Shank1 plays a role in the structural and functional organization of the dendritic spine by recruiting Homer to the membrane at postsynaptic sites, mediating the formation of multisynapse spines (Sala, Piëch, Wilson, Passafaro, Liu, and Sheng, 2001).

The main hubs are Ythdf1 (mRNA processing), Adcy4 (a membrane-bound, CaM-insensitive adenylcyclase), Gnal (a G protein involved in signal transduction), Hic1 (a transcriptional repressor), Zfp84 (a zinc finger protein), and Pnpla2 (which provides free fatty acids to tissues in situations of energy depletion).

Hic1 encodes for a transcriptional repressor of the Wnt pathway discussed in Section 6.3.5, since it was also identified in the magenta_20. The identification of this gene as a central hub in the 5 h TO matrix stresses its central role in LTP, possibly via Sirt1 inhibition, controlled by the formation of a Hic1/Sirt1 complex which controls the transcriptional expression of Sirt1 itself (see Section 6.3.5). G2e3 encodes for a negative regulator of apoptosis (Brooks, Helton, Banerjee, Venable, Johnson, Schoeb, Kesterson, and Crawford, 2008), while Trps1 encodes for a transcriptional repressor (Kaiser, Möröy, Chang, Horsthemke, and Lüdecke, 2003).
Another network identified at 5 h shows a low gene overlap with the networks discussed previously – darkgreen_5H. This small network also exhibits an enrichment in positive regulation of metabolic processes ($p = 2.137e-02$).

**Co-expression networks at 24 h**

The weighted co-expression network at 24 h shows a further decrease in TO. While the 20 min and 5 h networks are representative of the rapid response to the HFS, at 24 h LTP has consolidated and stimulated neurons have reached a new homeostatic level.

However, a representative trend exhibited by most of the co-expression networks identified at 24 h is a gain of intramodular connectivity (see Figure 6.8c).

![Figure 6.9: Expression of different genes across time in relation to the unstimulated hemispheres. Red indicates overexpression, green underexpression. (a) Egr family, (b) nBAF complex subunits, and (c) Interleukines ($p = 0.05, -0.15 > \ln(FC) > 0.15$)](image_url)

**6.3.5 Regulation of gene expression**

We have discussed in previous chapters the importance of gene expression for L-LTP. We pointed out that the late phase of LTP can be impaired if gene expression is inhibited during a short time window early after stimulation, but if the inhibition of RNA synthesis is applied later, L-LTP is not affected (Nguyen et al., 1994). Hence, the changes in gene expression strictly necessary for LTP must be triggered in the early stages post-LTP induction.

Some of the top genes in the co-expression networks play roles as gene expression regulators. In the red_20 module, the Wt1 gene encodes for a zinc finger that belongs to the Egr family of transcription factors. This family includes Egr1, Egr2, Egr3, Egr4, genes known to be involved in the early genomic response to LTP induction (see Section 3.3.6). While the Egr genes show consistent overexpression in the 20 min dataset (see Figure 6.9a), Wt1 RNA is underexpressed. This result is in agreement with its role in the brain, where it has been proposed to act as a growth suppressor by binding to and blocking/repressing the promoters of early growth-promoting genes normally activated by Egr genes (Haber, Sohn, Buckler, Pelletier, Call, and Housman, 1991).

In the magenta_20 module, a number of transcriptional modulators are also overexpressed – Erf, Hic1, Insm2, and Tnfaip2. Erf encodes for a protein activated upon phosphorylation by Erk1/2...
Chapter 6. Co-expression analysis of LTP microarray data

(Le Gallic, Sgouras, Beal, and Mavrothalassitis, 1999), a pathway involved in LTP (Section 3.3.4). According to the study of Le Gallic et al., Erf may be important in the control of cellular proliferation during the G0/G1 transition in the cell cycle. The other overexpressed genes co-expressed in the magenta_20 module, Hic1, and Ins2, are tumor suppressors, while Tnfaip2 seems to be involved in angiogenesis and development (Wales, Biel, El Deiry, Nelkin, Issa, Cavenee, Kuerbitz, and Baylin, 1995; Sarma, Wolf, Marks, Shows, and Dixit, 1992).

Homer1, an overexpressed gene which has been previously associated with LTP-related Ras signalling (Rosenblum, Futter, Voss, Erent, Skehel, French, Obosi, Jones, and Bliss, 2002) also appears in the magenta_20 module. Nevertheless, the candidate modulators for the Ras-Erk1/2 signalling cascade appear in different networks. In the pink_20 module, the Akirin2 gene may form a complex with Ywhab (which appears in the red_20 network) that acts to repress transcription of Dusp1, a dual-specificity phosphatase (Komiya, Kurabe, Katagiri, Ogawa, Sugiyama, Kawasaki, and Tashiro, 2008). DUSPs are able to dephosphorylate MAPKs, and Dusp1 in particular tends to inactivate p38 (MAPK11-14), JNK (MAPK8/10), and Erk1/2 (MAPK3/1) (Owens and Keyse, 2007). Interestingly, Dusp1, 2, 5, and 6 are all upregulated in the 20 min dataset. While the role of DUSPs in LTP as modulators of the Ras-Erk1/2 pathway has been suggested previously (Ryan et al., 2011), Akirin2 and Ywhab may also have a modulatory role in the pathway. In fact, while their expression is not significantly different from the controls for the 20 min dataset, they both show a significant underexpression in the 5 h dataset.

In the turquoise_20 co-expression network, other transcription factors are overexpressed – Atf6 and Gtf3c5. Atf6 encodes for a cAMP-dependent transcription factor that may be involved in the activation of transcription by the SRF (Zhu, Johansen, and Prywes, 1997).

Gene expression, however, is not only regulated via recruitment of transcriptional activators/repressors that bind promoters and other regulatory elements. The structure of the chromatin (DNA) influences the degree of access that the transcriptional machinery has to a given gene – the protein histones form octamers that pack the chromatin (Varga-Weisz and Becker, 1998). DNA has to be exposed for transcription to take place. The changes in chromatin structure that influence gene expression have been termed epigenetics. The two more studied mechanisms are DNA methylation/demethylation, which involves a chemical modification of the DNA itself, catalysed by DNA methyltransferases, and post-translational modifications of the histones (acetylation, phosphorylation and methylation).

In the 20 min dataset, the tumor suppressor Hic1 is overexpressed and acts as a Sirt1 repressor (Chen, Wang, Yen, Luo, Gu, and Baylin, 2005). However, the repression of Sirt1 requires the formation of a Hic1/Sirt1 complex (Chen et al., 2005). An additional gene in the same co-expression network, Bcl6, may have a role in the epigenetic control of gene expression. Bcl6 has been shown to be involved in the control of neurogenesis by epigenetic silencing (Tiberi, Van Den Ameele, Dimidschstein, Piccirilli, Gall, Herpoel, Bilheu, Bonnefont, Iacovino, Kyba, et al., 2012). This mechanism involves the recruitment of Sirt1.

Other co-expressed genes are known to regulate gene expression by modifications of chromatin structure. In the yellow_20 network, Phf21a, is a component of the BHC complex, a co-repressor complex that represses transcription of neuron-specific genes in non-neuronal cells by deacetylating and demethylating specific sites on histones (Shi, Matson, Lan, Iwase, Baba, and Shi, 2005). Accordingly, Phf21a is underexpressed in the 20 min dataset. In the green_20 module, Cxxc1 (underexpressed) binds to demethylated regions of DNA to exert its regulatory function.

In the turquoise_20 module, the Arid1a gene shows significant downregulation at 20 min post-LTP. It is involved in chromatin remodelling and belongs to the neuron-specific chromatin remodelling complex (nBAF complex) (Wang, Nagl, Wisker, Van Scoy, Pacchione, Yaciuk, Dallas, and Moran, 2004). The nBAF complex along with CREST plays a role regulating the activity of genes essential for dendrite growth and long-lasting forms of synaptic plasticity and
memory processes (Vogel-Ciernia and Wood, 2014). Also part of the same complex, Arid1b is an independent gene product that is overexpressed in one of the probes for 20 min. Different combinations of subunits in the nBAF complex are found in different tissues and cell-types. Another member of the nBAF complex, Smarcb1, appears in the brown_20 module, and shows a modest downregulation in the 24 h dataset. In fact, this diversity in the subunit composition can target different sets of genes (Ronan, Wu, and Crabtree, 2013). The combinatorial diversity of the nBAF complex may explain the complex pattern of expression exhibited by the different subunits in the three time datasets, as depicted in Figure 6.9b.

Also in the turquoise_20 network, the Ring1 gene encodes for a ubiquitin-protein ligase associated with the ubiquitination of the histone H2A, playing a central role in epigenetic gene regulation. It is overexpressed in the 20 min dataset (Cao, Tsukada, and Zhang, 2005). In the black_20 module, Setdb1 encodes for a histone methyltransferase.

Gene expression modulation is also central in the co-expression modules identified 5 h post-LTP induction. However, the top genes involved in regulation of gene expression seem to be related with either epigenetic regulation of chromatin structure or with transcriptional repression. As we discussed in the previous sections, the nBAF complex seems to have a central role in regulating gene expression necessary for LTP. In two of the networks identified at 5 h (midnightblue_5H and magenta_5H), components of the nBAF complex represent top hubs, Arid1a and Smarcb1. A number of other enzymes involved in the modification of histones appear as top hubs in other networks (Kdm6b, downregulated at 5 h and encoding a histone demethylase, and Hdac5, a histone deacetylase). In the lightgreen_5H module, H3f3b encodes for a variant histone H3 which replaces conventional H3, while in the darkred_5H module, Tada3l is a component of the PCAF complex, which functions as histone acetylase in a nucleosomal context (Ogryzko, Kotani, Zhang, Schiltz, Howard, Yang, Howard, Qin, and Nakatani, 1998). Finally, Zmynd11 encodes for a zinc finger transcriptional co-repressor which binds to histone H3. Differential expression of all these genes is found at 20 min or 5 h but never at 24 h.

Alongside the identified epigenetic regulators, a large number of transcriptional repressors are identified as hubs in the 5 h modules. In particular, the green_5H module contains Foxn3, Morc2, and Bhlhe41, genes encoding for transcriptional repressor (Scott and Plon, 2005; Shao, Li, Zhang, Liu, Liu, Zhao, Shen, and Li, 2010; Garriga-Canut, Roopta, and Buckley, 2001) Trps1 and Tbx3 are other transcriptional repressors identified as hubs in other 5 h networks.

Gene expression has still an important role at 24 h and many of the main hubs in the networks that exhibit a gain of intramodular connectivity are involved in gene expression regulation. In particular, epigenetic regulation appears as a mayor modulator of gene expression.

In the green_24 module, a main hub is constituted by the Junb gene, which is a well-characterised IEG in LTP (Chiu et al., 1988) and exhibits an upregulation in the 20 min dataset. Junb can dimerize with the Fos family oncogenes to form the DNA-binding AP-1 heterodimer. Other transcription factors are Hdgf and Cc2d1a (transcriptional repressors), and Pbx2.

6.3.6 RNA processing and ubiquitin-proteasome pathway

From the co-expression networks identified in the 20 min dataset, the role of RNA processing is apparent. One of the main hubs in the brown_20 network is the Tnrc4 gene (Celf3), which encodes for an RNA-binding protein involved in the regulation of mRNA alternative splicing. Perhaps important in the overlapping phases of new RNA synthesis and regulation. In particular, Celf3 activates the splicing of MAPT/Tau exon 10, whose missplicing has been shown to cause frontotemporal dementia (Wang, Gao, Wang, Lafyatis, Stamm, and Andreadis, 2004). In the same cluster, the protein encoded by Rnps3 is also involved in mRNA splicing (Benecke, Lührmann, and Will, 2005), while Pnisr encodes for an arginine/serine-rich protein that binds to poliA in RNA (Baltz, Munschauer, Schwanhäusser, Vasile, Murakawa, Schueler, Youngs, Penfold-Brown, Drew, ...
Dnajb5 encodes for an homolog of the Hsp40 chaperone, whose function consists in binding and stabilising proteins in a complex together with Hsp70, to subsequently transport them to the particular subcellular location where they are to exert their functions (Hartl, 1996). Chaperones have been shown to be localised in postsynaptic sites (Ohtsuka and Suzuki, 2000) and they have been proposed to be involved in the mechanisms that induce functionally active AMPA receptor during induction of LTP (Song, Kamboj, Xia, Dong, Liao, and Huganir, 1998). The overexpression is statistically significant in the three probes for Dnajb5 at 20 min ($p = 0.05$).

In the red_20 module, Ccdc86 and Lgals3 also encode for proteins involved in RNA processing. The former displays overexpression in the 20 min arrays and the latter encodes for Galectin-3, a pre-mRNA splicing factor (Haudek, Spronk, Voss, Patterson, Wang, and Arnoys, 2010).

Other differentially expressed genes involved in mRNA processing are Mmaa, Zranb2, and Eftud2 (all downregulated 20 min post-LTP), Syncrip, Fam98a, Fkbp11, Pabpn1. Perhaps the best example of the genes taking part in the regulation of mRNA post-translational processing is Btg2, which is highly overexpressed. Btg2, in the turquoise_20 network, is related to neuronal differentiation (el Ghissassi, Valsesia-Wittmann, Falette, Duriez, Walden, and Puisieux, 2002) and is considered to be necessary for neurite outgrowth (Miyata, Mori, and Tohyama, 2008).

In the top genes identified, at least 4 genes have a role in protein ubiquitination – Psmc4 and Siah2 (black_20, Asb14 (blue_20), Trim17 (turquoise_20), and Znrf2 (yellow_20). Siah2, which encodes for an E3 ubiquitin-protein ligase, shows a statistically significant pattern of overexpression in the 20 min arrays whereas Znrf2, another ubiquitin-protein ligase, shows downregulation. The latter is known to be present in presynaptic plasma membranes (Araki and Milbrandt, 2003).

An atypical E3 ubiquitin-protein ligase is encoded by the Zfp91 gene. It is indirectly involved in activating the NFκB pathway, a pathway shown to be involved in hippocampal synaptic plasticity (Albensi and Mattson, 2000; Ahn et al., 2008; Jin, Jin, Jung, Lee, Lee, and Lee, 2010) and long-term memory formation (Meffert, Chang, Wiltgen, Fanselow, and Baltimore, 2003; Kaltschmidt and Kaltschmidt, 2009). Furthermore, NFκB was identified as a central regulator in the most significant network identified at 20 min by IPA (Ryan et al., 2011). NFκB pathway may also be modulated by epigenetic mechanisms (Lubin, Roth, and Sweatt, 2008).

In the magenta_5H module, the Cul5 gene encodes for a fundamental component of a number E3 ubiquitin-protein ligase complexes, while Usp11 encodes for a protease able to remove ubiquitin from ubiquitinilated proteins. Both genes are downregulated at 5 h and the latter inhibits the degradation of target proteins and plays a role in the regulation of pathways leading to NFκB activation (Yamaguchi, Kimura, Miki, and Yoshida, 2007). Trim2 and Arih1 encode for E3 ubiquitin-protein ligases (the former involved in neuroprotective functions) and are both also differentially underexpressed.

Other genes involved in proteasomal degradation are Psmd11 and Psme, encoding a subunit of a proteasome regulator and a component of the 26S proteasome, respectively. However, these two genes do not show differential expression at the earlier times but only at 24 h.

In the 24 h modules, ubiquitination is still overrepresented, with Ube2g1 up, Hectd1, and Cul3 as main hubs. The former, overexpressed at 24 h while the other two are only downregulated at 5 h.

It is also worth mentioning that Rnasen, a hub in the pink_U module encodes for the ribonuclease 3, which is involved in the initial step of miRNA biogenesis and significantly underexpressed at 24 h. The importance of microRNAs in neural function has been studied (Mercer, Dinger, Mariani, Kosik, Mehler, and Mattick, 2008; Im and Kenny, 2012) and it has been suggested that they play a role in the late phases, in particular in growth and neurogenesis (Ryan et al., 2012).
6.3 Results

6.3.7 Intracellular signalling

The kinase-phosphatase system

The intracellular pathways activated following LTP can also be modulated by changes in gene expression.

Identified in the top 35 brown_20 module but not differentially expressed, Fgd2 product is capable of activating Cdc42, which in turn indirectly activates Mapk8, a c-Jun kinase (Dérijard, Hibi, Wu, Barrett, Su, Deng, Karin, Davis, et al., 1994). c-Jun is one of the most studied IEG involved in LTP, part of the AP-1 complex. Activated Mapk8 is also capable of phosphorylating Elk1 (Zhang, Yang, and Sharrocks, 2007), a transcription factor involved in the co-activation of SRE-dependent gene expression upon binding to SRF (see Figure 6.10 and Section 3.3.6). Dusp1 (dual specificity protein phosphatase 1), which shows a significant upregulation in the 20 min dataset, can dephosphorylate Mapk8.

Another overexpressed dual specificity protein phosphatase, Dusp13, appears in the same co-expression module. Although it shows no phosphatase activity on Mapk8, Erk2, p38 nor Map3k5, it does exhibit a phosphatase activity-independent regulatory role in Map3k5-mediated apoptosis, preventing Map3k5 inhibition by Akt (Park, Park, Kim, Song, Park, Lee, Kim, Choi, Kim, and Cho, 2010).

The CaMKII pathway is represented in the red_20 network by Camk2d, which encodes for the CaMKII delta subunit, the principally expressed form only in brain (Schworer, Rothblum, Thekkumkara, and Singer, 1993). The alpha subunit (Camk2a), central to the NMDA receptor-dependent LTP, is identified in the pink_20 network. In the same module, a CaMK-kinase (Camk1) is able to phosphorylate CaMKI, CaMKId, CaMKIg and CaMKIV (Edelman, Mitchelhill, Selbert, Anderson, Hook, Stapleton, Goldstein, Means, and Kemp, 1996). It also indirectly promotes cell survival by phosphorylating and activating Akt1, which in turn inhibits the pro-apoptotic Bcl2-antagonist of cell death (see also Section 6.3.7). In the brown_20 module by a potent and specific inhibitor of CaMKII (Camk2n1; Chang, Mukherji, and Soderling, 2001). While it shows no significant expression changes in the 20 min dataset, it is downregulated at 5 h.

In the brown_20 network, two phosphatases are identified among the 35 top genes – Ankrd52 (an overexpressed putative regulatory subunit of protein phosphatase 6, PP6), and Arpp19, a cAMP regulated phosphoprotein that specifically inhibits protein phosphatase 2A (PP2A) during mitosis (Irwin, Chao, Goritchenko, Horiuchi, Greengard, Nairn, and Benowitz, 2002). Arpp19 is only downregulated at 5 h post-LTP. The brain-specific isoform of the B regulatory subunit of PP2A (Ppp2r2c) is overexpressed at 20 min and part of the top 35 black_20. Also overexpressed is the regulatory subunit 3D of PP1 (Ppp1r3d), identified within the blue_20 network.

Three genes encoding for subunits of phosphatases, Ppp2r5a, Ppp3r2, and Ppp1r3b, are identified in the 24 h modules. Furthermore, Ppp2r5a and Ppp3r2 are both upregulated at 24 h.

Finally, it is worth mentioning the loss of intramodular connectivity exhibited by the brown_5H module. One of its main hubs is CaMKIV, which plays a pivotal role in memory consolidation by phosphorylation of CREB in the nucleus (Bito et al., 1996). However, the expression of the gene is not altered significantly, which suggests that its activation is mainly driven by post-translational changes (CaM-Ca^{2+} binding and phosphorylation) rather than changes in expression. It does show, however, downregulation earlier in the 5 h dataset.

The PI3K-Akt-mTOR pathway

Akt1 (also known as protein kinase B, or PKB) appears as a central gene in the pink_20 co-expression module, although it only shows a slight overexpression at 5 h (1.6-fold change). PDGF-BB stimulates the PI3K-Akt pathway by binding to Pdgfrb (Hanai, Tokuda, Ohta, Matsushima-Nishiwaki, Takai, and Kozawa, 2006), a receptor which we identified in our brown_20 module.

Identified in the same module as Akt1, the overexpressed Cds2 gene encodes for an enzyme that provides CDP-DAG, precursor of PI.
In the *pink_20* module we find Dgkg, a gene that encodes for a DAG kinase (DGK) gamma, upregulated at 20 min. This enzyme catalyses the conversion of DAG phosphatidic acid (PA). Both molecules constitute messengers and hence the switch among the two regulates different cellular responses (Merida, Avila-Flores, and Merino, 2008). In fact, PA can bind and stabilise the mTOR complex (Fang, Vilella-Bach, Bachmann, Flanigan, and Chen, 2001). While the gamma isoform is highly expressed in retina and brain (Sakane, Yamada, Kanoh, Yokoyama, and Tanabe, 1990), the epsilon isoform has been demonstrated to modulate LTP (de Turco, Tang, Topham, Sakane, Marcheselli, Chen, Taketomi, Prescott, and Bazan, 2001).

Finally, it is worth mentioning the possibility of formation of 2-arachidonoylglycerol (2-AG) from DAG via DAG lipase (DGL). 2-AG is an endocannabinoid present at relatively high levels in the central nervous system and interacts with cannabinoid receptors. We highlighted in Section 6.3.9 the overexpression of the endocannabinoid receptor Cnr1.

The PI3K-Akt-mTOR pathway is represented in some of the 5 h modules. Dab2, downregulated at 5 h and hub in the *lightcyan_5H*, which modulates the balance between PI3K-Akt-mediated cell survival and MAP3K5-JNK signalling pathways (Xie, Gore, Zhou, Pong, Zhang, Yu, Vessella, Min, and Hsieh, 2009). Plekha1 in the *lightyellow_5H* module binds specifically to PIP$_2$ (Kimber, Deak, Prescott, and Alessi, 2003).

Finally, in the *darkred_5H* network, Ip6k1 encodes for an inositol hexakisphosphate kinase overexpressed at 5 h.

**NFκB and other signalling cascades**

The modulation of the Ras-Erk cascade is under-represented within the top genes, but some candidates, such as Plk2 (Polo-like kinase 2, *red_20*) and Rasip1 (Ras interacting protein 1, *blue_*20) are overexpressed. The former encodes for a Ser/Thr-protein kinase known to be involved in synaptic plasticity by regulating Ras protein signalling (Lee, Lee, Rozeboom, Lee, Udagawa, Hoe, and Pak, 2011). It has also a kinase-independent activity that occludes long-term depression (LTD) via disrupting the interaction between AMPA receptor and NSF (Evers, Matta, Hoe, Zarkowsky, Lee, Isaac, and Pak, 2010).

In addition to the Ras-Erk cascade, several modulators of the transcriptional activator NFκB appear in the early co-expression networks.

In the *green_20* module, Ubfd1 is hypothesised to act as NFκB regulator (Fenner, Scannell, and Prehn, 2009), while Ncoa3 (*turquoise_20*) is involved in the co-activation of the NFκB pathway via its direct interaction with the Nfkb1 subunit. It also acts as a histone acetyltransferase (Chen, Lin, Schiltz, Chakravarti, Nash, Nagy, Privalsky, Nakatani, and Evans, 1997).

Map3k3 constitutes a hub in the *midnightblue_5H* module and shows a downregulation at 5 h. It encodes for a component of a protein kinase signal transduction cascade which mediates activation of the NFκB, AP1 and DDIT3 transcriptional regulators (Ellinger-Ziegelbauer, Brown, Kelly, and Siebenlist, 1997), the former two known to be key regulators in LTP gene expression. In the *lightyellow_5H* module, the Snip gene also regulates the NFκB signalling pathway (Kim, Flanders, Reffey, Anderson, Duckett, Perkins, and Roberts, 2001), but is downregulated at 24 h.

**6.3.8 Extracellular signalling and cell adhesion**

In the *red_20* module, Hmox2 (heme oxygenase 2) is implicated in the production of carbon monoxide in brain, a molecule that acts as a neurotransmitter (Rotenberg and Maines, 1990). Svp2b encodes for a synaptic vesicle glycoprotein. It plays a role in the control of regulated secretion in neural cells (Janz and Südhof, 1999) and exhibits upregulation at 20 min. The phosphorylation of Svp2 modulates binding to synaptotagmin, a membrane-trafficking protein (Pyle, Schivell, Hidaka, and Bajjalieh, 2000). In the *turquoise_20* module, Gjb5, encodes for an overexpressed gap junction protein.
Two receptor tyrosine kinases of the ephrin-receptor family (EphA3 and EphA4) are identified in the black network. These receptors are involved in bidirectional signalling into neighbouring cells upon membrane-bound ephrin binding.

Figure 6.10: Activation of MAPK pathways through the GTPase intermediates, leading to the expression of delayed-response genes. Figure adapted from Liu et al. (2007)

Extracellular signalling is associated with cell-cell and cell-matrix adhesion. In the magenta network, the Cadm3 gene is involved in cell-cell adhesion. In the green network, Adam23, Cldn2, and Tnr are involved in cell-adhesion and interactions with different cells and matrix components. Adam23 is overexpressed, and Tnr, which encodes for Tenascin-R, has been shown to be central in LTP in mice (Saghatelyan, Dityatev, Schmidt, Schuster, Bartsch, and Schachner, 2001). In the blue network, Itgb4 and Itgb7 encode for different integrin beta subunits, also involved in
cell adhesion.

Cadherin-3 (Cdh3 gene) and a cadherin-related protein (Mucdhl gene) are the main cell adhesion proteins found in the 24 h co-expression networks. The Ina gene, which encodes for a neuronal intermediate filament is also a major hub in the green_24 module, involved in the morphogenesis of neurons (Fliegner, Ching, and Liem, 1990). Whereas both Cdh3 and Ina genes are overexpressed only at 20 min, their high TO within the networks suggest that the regulation of cell adhesion still has a central role at later stages of LTP and may be of importance for LTP maintenance.

6.3.9 Changes in membrane composition, endocytosis and exocytosis

Gene expression changes can have an effect on membrane composition via synthesis of proteins that regulate membrane channels trafficking, or by the expression of genes that encode for channel subunits.

Synthesis of new channel subunits

In the brown_20 co-expression module Gabbr1 encodes for a subunit of the GABA receptor (see Section 3.1.3). It is worth noting that Gabbr1 expression changes significantly compared to the unstimulated control only in the 5 h dataset, in which the gene is downregulated around 15-fold.

Other genes that constitute channel subunits are identified in other co-expression modules. Cnr1, for example, encodes for the cannabinoid receptor 1, a receptor highly expressed in the hippocampus. In region CA1, the activation of these receptor leads to impairment of LTP and LTD by reducing presynaptic neurotransmitter release (Misner and Sullivan, 1999). Gpr26 (pink_20 module) is overexpressed and encodes for an orphan receptor that stimulates adenylate cyclase and hence elevates intracellular cAMP.

In the blue_20 module, Grm2 encodes for a G-protein coupled receptor for glutamate. Another glutamate G-protein coupled receptor appears in the yellow_20 module, Grm6. The latter is overexpressed at 20 min. These type of receptors trigger a G-proteins signalling cascade upon glutamate binding, modulating down-stream effectors involved in synaptic plasticity (Devi, Markandeya, Maddodi, Dhinra, Vardi, Balijepalli, and Setaluri, 2013). Slc6a11 encodes for the sodium- and chloride-dependent GABA transporter 3, which terminates the action of GABA by its high affinity sodium-dependent re-uptake into presynaptic terminals (Borden, Dhar, Smith, Branche, Gluchowski, and Weinschank, 1993).

Also in the yellow_20 module, Htr3b encodes for one of the several known serotonin receptors. Serotonin serves as a neurotransmitter, hormone and mitogen. Binding of serotonin to the receptor causes a fast depolarising response.

The top genes in the 5 h networks are particularly enriched in genes functionally related with changes in post-synaptic membrane composition. Of note, Adrb1 and Adrb2 are both beta-adrenergic receptor subunits. These receptors mediate the catecholamine-induced activation of adenylate cyclase through the action of G proteins, and the former is upregulated at 20 min. Other receptors are Scn11a (sodium channel), Kcnk3 (voltage-insensitive, background potassium channel, which acts as an outward rectifier), Kcna2 (voltage-sensitive potassium channel), and Gabre and Gabrr1 (GABA receptor subunit epsilon and rho-1, respectively).

In the modules identified 24 post-LTP that show an gain of intramodular TO, a number of genes encoding for membrane channels and receptors appear as top hubs. Kcnj1 and Kcnj4 both encode for inward rectifier potassium channels, involved in the potassium homeostasis. Other hub genes that encode for membrane channels are shown in Table 6.3.

Expression of channel trafficking modulators

EAAC1 is the primary neuronal glutamate transporter from the extracellular environment in brain (Lin, Orlov, Ruggiero, Dykes-Hoberg, Lee, Jackson, and Rothstein, 2001), and it has been suggested to regulate the exposure of postsynaptic ionotropic GluRs to glutamate (He, Janssen, Rothstein,
Table 6.3: Top hubs in the co-expression networks that function as membrane channels

<table>
<thead>
<tr>
<th>Module</th>
<th>Gene</th>
<th>Channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>blue_24</td>
<td>Kcnj1</td>
<td>Inward rectifier K⁺ channel</td>
</tr>
<tr>
<td>turquoise_24</td>
<td>Kcnj4</td>
<td>Inward rectifier K⁺ channel</td>
</tr>
<tr>
<td>midnightblue_24</td>
<td>Cacna1i</td>
<td>Voltage-dependent L-type Ca²⁺ channel</td>
</tr>
<tr>
<td>magenta_U</td>
<td>Grm6</td>
<td>Metabotropic glutamate receptor</td>
</tr>
<tr>
<td>pink_U</td>
<td>Neto2</td>
<td>Accessory subunit of neuronal kainate-sensitive glutamate receptors</td>
</tr>
<tr>
<td>lightcyan_24</td>
<td>Abcc9</td>
<td>ATP-sensitive K⁺ channel</td>
</tr>
</tbody>
</table>

and Morrison, 2000). While EAAC1 was not identified by the co-expression analysis, Arl6ip5 (GTRAP3-18), a negative modulator of EAAC1, appears in the brown_20 network.

As we mentioned earlier, trafficking of AMPA receptors is a central mechanism in LTP (see Section 3.3). A number of genes identified in the analysis play such a role. Pick1 (red_20) is an adapter protein that may be involved in clustering of various receptors and is involved in the regulation of AMPA receptor trafficking. Via interaction with Arp2/3 complex, it has a role in cytoskeletal remodelling shown to regulate synaptic plasticity in excitatory synapses (Hanley and Henley, 2005; Rocca, Martin, Jenkins, and Hanley, 2008). Interestingly, the Arp2/3 complex appears in the pink_20 module (see Section 6.3.10).

In the black_20 module, Hip1 is involved in the regulation of AMPA receptor trafficking in the in an NMDA-dependent manner (Metzler, Li, Gan, Georgiou, Gutekunst, Wang, Torre, Devon, Oh, Legendre-Guillemin, et al., 2003). Rab12 and Rab8b encode for small Rab GTPases, key regulators of intracellular membrane trafficking.

Cnih2 (in the yellow_20 module) gene encodes for a cornichon, a protein involved in the modulation of AMPA receptor trafficking and gating properties. In the same network and in the turquoise_20 network, two transmembrane AMPAR regulatory proteins (TARPs) are identified in the top 35 genes, Cacng4 and Cacng8, respectively. They encode for the voltage-dependent calcium channel gamma-4 and gamma-8 subunits respectively (Milstein, Zhou, Karimzadegan, Bredt, and Nicoll, 2007).

Other central genes in the co-expression networks modulate the pathways leading to changes in the protein characteristics following LTP at 5 h. Clathrin-mediated exocytosis has been demonstrated to be required for some forms of LTP (Wang and Linden, 2000), and Ap2b1 in the black_5H network encodes for a component of the adaptor protein complex 2 (AP-2), which regulates protein transport via clathrin-transport vesicles (Owen, Collins, and Evans, 2004). In the lightgreen_5H module, Cnih2 encodes for the protein cornichon homolog 2, identified also in the yellow_20 module as a 20 min hub and involved in the regulation of trafficking and gating properties of AMPA receptors but only differentially upregulated at 5 h. Another gene that appears as a hub in a 20 min module and a 5 h module (salmon_5H) is Arl6ip5, which regulates intracellular concentrations of taurine and glutamate and downregulated at 5 h (see Section 6.3.9).

Finally, in the lightcyan_5H, Homer2 appears as a top hub although it is only overexpressed in the 20 min dataset. Homer2 protein binds Shank1, a hub in the purple_5H network and previously described (see Section 6.3.4). The different Homer proteins (Homer1 and Homer3), despite not representing important hubs in the co-expression networks, do show a significant overexpression at different times (Homer1 at 20 min and 24 h, Homer3 at 5 h). Consistently, the different Homer proteins are known to be differentially regulated and may play important roles in the maintenance of the synaptic plasticity. In the hippocampus, for example, Homer1 and Homer2 are mainly localized in the CA1 region and CA1-CA2 region, respectively, whereas Homer3 is concentrated in the CA2-CA3 regions (Shiraiishi-Yamaguchi and Furuichi, 2007).
Putting modulators of membrane composition with a role at 24 h may involve Dlg4, which interacts with the cytoplasmic tail of NMDA receptor subunits, required for synaptic plasticity associated with NMDA receptor signalling. Furthermore, overexpression or depletion of the Dlg4 gene changes the ratio of excitatory to inhibitory synapses in hippocampal neurons (Prange, Wong, Gerrow, Wang, and El-Husseini, 2004). Ap2a2 in the pink_24 module encodes for the alpha-2 subunit of the AP-2 complex, involved in protein transport. The beta subunit, Ap2b1 was found as hub in the black_5H network.

6.3.10 Neuronal structural changes via cytoskeletal remodelling

A number of genes identified in the co-expression networks have roles in the structural reorganisation of the cytoskeleton. Table 6.4 summarises these findings.

<table>
<thead>
<tr>
<th>Module</th>
<th>Gene</th>
<th>Encoded protein</th>
<th>Function</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>brown_20</td>
<td>Diaph3</td>
<td>Protein diaphanous homolog 3</td>
<td>Recruits profilin to the membrane and promotes actin polymerisation</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>Sema6a</td>
<td>Semaphorin-6-A</td>
<td>Promotes reorganization of the actin cytoskeleton. Plays a role in axon guidance in the developing CNS</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>Tnn</td>
<td>Tenascin-N</td>
<td>Involved in neurite outgrowth and cell migration in hippocampal explants in mouse. OVEREXPRESSED</td>
<td>(3)</td>
</tr>
<tr>
<td>red_20</td>
<td>Fmn1</td>
<td>Formin-like protein 1</td>
<td>Involved in the regulation of cell morphology and cytoskeletal organization. Required in the cortical actin filament dynamics and cell shape</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>Parvb</td>
<td>Beta-parvin</td>
<td>Participates in the reorganization of the actin cytoskeleton</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>Snap</td>
<td>SRC kinase signalling inhibitor 1</td>
<td>Regulates dendritic spine morphology. Involved in calcium-dependent exocytosis</td>
<td>(6)</td>
</tr>
<tr>
<td>magenta_20</td>
<td>Actr10</td>
<td>Actin-related protein 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yellow_20</td>
<td>Wasl</td>
<td>Neural Wiskott-Aldrich syndrome protein</td>
<td>Regulates actin polymerization by stimulating the actin-nucleating activity of the Arp2/3 complex. UNDEREX-PRESSED</td>
<td>(7)</td>
</tr>
<tr>
<td>green_20</td>
<td>Add3</td>
<td>Gamma-adducin</td>
<td>Membrane-cytoskeleton-associated protein that promotes assembly of spectrin-actin network</td>
<td>(8)</td>
</tr>
<tr>
<td></td>
<td>Cof6A1</td>
<td>Collagen</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sfrp5</td>
<td>Secreted frizzled-related protein 5</td>
<td>Involved in axis elongation in the retina</td>
<td>(9)</td>
</tr>
<tr>
<td>punk_20</td>
<td>Arpc3</td>
<td>Actin-related protein 2/3 complex subunit 3</td>
<td>Component of the Arp2/3 complex, which is involved in regulation of actin polymerization. Mediates the formation of branched actin networks. (See also Pick1 in 6.3.9)</td>
<td>(10)</td>
</tr>
<tr>
<td></td>
<td>Krt18</td>
<td>Keratin type I</td>
<td>Plays a role in filament reorganisation</td>
<td>(11)</td>
</tr>
<tr>
<td></td>
<td>Plekho1</td>
<td>Pleckstrin homology domain-containing family O member 1</td>
<td>Involved in actin regulation. May inhibit Akt1 upon binding.</td>
<td>(12)</td>
</tr>
<tr>
<td></td>
<td>Pacsin1</td>
<td>Synaptapin-1</td>
<td>Plays a role in the reorganization of the actin cytoskeleton and in neuron morphogenesis, also involved in neurite formation, neurite branching and the regulation of neurite length</td>
<td>(13)</td>
</tr>
<tr>
<td>black_20</td>
<td>Sgcg</td>
<td>Gamma-sarcoglycan</td>
<td>Cytoskeleton component that forms a link between the F-actin cytoskeleton and the extracellular matrix. Muscle specific. OVEREXPRESSED</td>
<td>(14)</td>
</tr>
</tbody>
</table>

Actin polymerisation represents the major cytoskeletal remodelling cellular system in the 24 h modules, with cardiac muscle actin standing as a hub in the magenta_U module. Hspb3 and Enah are both involved in actin polymerisation, the former being an inhibitor, both only overexpressed at...
20 min. Other cytoskeletal modulators are Map1b, involved in the tyrosination of alpha-tubulin in neuronal microtubules, and Sept3, a GTPase filament-forming specific of neural tissue.

### 6.3.11 Neurogenesis and cell death

Although a number of genes with putative roles in cell-cycle appear in the early co-expression networks, perhaps the most representative is Pola2, the gene encoding for the DNA polymerase alpha subunit B. It is overexpressed and central to the *turquoise_20* module.

In the *brown_20* module, Eps15 encodes for an epidermal growth factor receptor substrate involved in cell growth regulation. In the *red_20* network, Ddr2 is underexpressed at 20 min and encodes for a cell surface receptor for fibrillar collagen and regulates cell differentiation, remodelling of the extracellular matrix, cell migration and cell proliferation. In the *pink_20* module, Bcl2l14 encodes for a Bcl2-like protein involved in apoptosis. Bcl2 is inhibited by Akt1, also present in the module and discussed previously (Section 6.3.7).

The long form of the GTPase PIKE-L, which activates nuclear PI3K, interacts with Homer1c and Homer2a together with mGluR1a/5 to form an mGluR1a/5-Homer-PIKE-L complex which leads the activation of PI3K and the prevention of neuronal apoptosis (Rong, Ahn, Huang, Nagata, Kalman, Kapp, Tu, Worley, Snyder, and Ye, 2003). In the *black_5H* and *darkred_5H* modules respectively, the Aven and Dapk2 genes encode for regulators of apoptosis and are both overexpressed at 5 h. Aven protects against apoptosis and Dapk2 is involved in cell survival, apoptosis, and autophagy (Kawai, Nomura, Hoshino, Copeland, Gilbert, Jenkins, and Akira, 1999; Chau, Cheng, Kerr, and Hardwick, 2000).

The Cdk5r2 gene encodes for a neural Cdc2-like kinase involved in cell cycle and another hub, Rif1, functions as a cell cycle checkpoint in humans (Silverman, Takai, Buonomo, Eisenhaber, and de Lange, 2004). On the other hand, the Plekhg2 gene and Arhgef7 are involved in the regulation of apoptosis. The former encodes for a pleckstrin homology domain-containing protein – other proteins with the same domain were identified in other networks and discussed previously (see Sections 6.3.10 and 6.3.7). In the *brown_U* module, Robo1 is involved in axon guidance, while Ptk2, in the *greenyellow_24*, regulates axon growth and neuronal cell migration, axon branching and synapse formation and is involved in a plethora of cellular functions related with LTP (Burgaya and Girault, 1996).

### 6.3.12 Immunity-related genes

Genes involved in the modulation of T-cell-mediated immunity are Cd8b, a T-cell surface glycoprotein involved in the regulation of immune response (*pink_24* module), Ctl4, an inhibitory receptor acting as a major negative regulator of T-cell responses (Prange et al., 2004) (*midnightblue_24*) and is overexpressed at 20 min. Ptpn22, which encodes for a negative regulator of T-cell receptor signalling (*purple_24*) is overexpressed at 5 h.

Other immunity-related genes are Cst7 (Cystatin-F) (20 min), Itgb7 (Integrin beta-7) (20 min) and Ilf5 (Interleukin-5) (20 min). Interestingly, these genes are differentially expressed at 20 min.

Furthermore, a number of Interleukins have been related with LTP (IL-1b, Cunningham, Murray, O’Neill, Lynch, and O’Connor (1996); IL-2, Tancredi, Zona, Velotti, Eusebi, and Santoni (1990); IL-6, Bellinger, Madamba, Campbell, and Siggins (1995); IL-10,Kelly, Lynch, Vereker, Nolan, Queenan, Whittaker, O’Neill, and Lynch (2001); IL-18, Curran and O’Connor (2001)) and they show a significant upregulation especially in the earliest time-point (20 min, see Figure 6.9b).

### 6.4 Discussion

We reanalysed the microarray data in the differential expression analysis presented in Chapter 5, which served as a starting point to our study. Our results are largely identical to the originally
Chapter 6. Co-expression analysis of LTP microarray data

reported in Ryan et al. (2011, 2012), with expected minor differences caused by the updated databases of transcript-gene mappings, together with the cut-offs used in the lists of overexpressed genes.

Compared to traditional differential expression analysis, co-expression is less error-prone for genes with a large expression variation. Perhaps more importantly, genes which are not detected by differential expression can (potentially) be detected by this type of analysis if they activate other genes which change enough to be detected. Arguably, a pair of genes with a high value of co-expression are likely to be forming complexes, pathways, or participate in the same cellular circuits (Eisen, Spellman, Brown, and Botstein, 1998).

We found that the rapid increase in gene expression observed by differential expression analysis is complemented by the increase in intramodular connectivity (average TO, see Section 6.3). Note that out of the 58 modules identified, only one exhibits a significant decrease in average TO from the control to the 20 min time point, while a total of 20 show a significant increase. This suggests that the rapid genomic response that follows LTP induction does not only involve a marked upregulation of gene expression, but also a tight coordination of the components that ultimately allow the transition to a new homeostatic cellular state.

The transitional 5 h dataset shows a loss of intramodular connectivity that parallels the onset of a general downregulation of gene expression, similar to the 24 h co-expression network. We believe that the early phase following stimulation is critical in the onset of the genomic changes that are known to be essential for L-LTP. A fundamental fraction of the genes that are transcribed rapidly after LTP induction may be of crucial importance at later times (Nguyen et al., 1994).

6.4.1 LTP-induced transcription

A large number of transcriptional modulators appear as hubs in the 20 min co-expression modules, described in more detail in Section 6.3.5. Our results stress the importance of the Egr family, been previously reported to be expressed following LTP induction (see Section 3.3.6 and Figure 6.9). In addition to these expected results, we find that the Wt1 gene, a member of the same family, appears as a key regulator in the co-expression network at 20 min. This observation, together with the its role as a repressor of the other Egr family members (Haber et al., 1991) suggests that it may play an important role in regulating Egr gene expression after LTP.

The Homer family is implicated in LTP though its association with the Ras-Erk1/2 pathway (see Section 3.3.4). Homer1 is overexpressed at 20 min and 24 h and has been extensively studied as an IEG. Likewise, Homer2 is overexpressed at 20 min whereas Homer3 is overexpressed at 5 h. It is known that these genes are differentially concentrated in different hippocampal regions (Shiraishi-Yamaguchi and Furuichi, 2007). Hence, it is conceivable that they may play important roles at different times after LTP induction. Note, however, that only Homer2 appears as a hub in a co-expression module. Another important hub is Shank1, whose protein recruits Homer to the membrane at the postsynaptic sites (Sala et al., 2001). These findings are in agreement with the role of Homer1 as an IEG and further suggest that other members of the Homer family may be modulating gene expression at other times after induction, possibly even participating in LTP consolidation and maintenance.

The co-expression analysis also highlight the importance of the constitutive transcription factor NFκB, whose activity can be modulated in a transcription-independent manner. Its central role in LTP has been described elsewhere and discussed in previous sections (see Freudenthal et al. (2004); Meberg et al. (1996), and Sections 3.3.6 and 6.3.7). In particular, our results suggest that ubiquitin signalling may be of particular importance in the NFκB pathway during LTP. Within the top genes co-expressed at 20 min, the Zfp91 and the Siah2 genes encode for E3 ubiquitin-protein ligases indirectly involved in activating the NFκB pathway. In a later co-expression network identified at 5 h, the genes Cul5 and Usp11, which encode for components of the ubiquitin/protease system are
6.4 Discussion

Both underexpressed. The latter regulates pathways leading to the activation of NFκB.

Other highly co-expressed genes may be regulating the pathway via direct or indirect interaction with NFκB or its inhibitor IκB at different times post-induction, although mainly in the first time point – e.g. Ubfd1, Ncoa3, Map3k3. The Snip gene, appears as a hub at 24 h, which suggests that the modulation of the NFκB pathway may not be restricted to a short time frame following HFS, consistent with learning-related NFκB activity in the nucleus, which shows a peak immediately after training and a second peak between 6 and 12 h after training (Freudenthal and Romano, 2000), both necessary for memory consolidation (Merlo, Freudenthal, and Romano, 2002).

6.4.2 Epigenetic control of gene expression

The modification of chromatin structure arises as a fundamental mean of gene expression regulation – we find that a substantial number of highly co-expressed enzymes is involved in epigenetic control, DNA methylation/demethylation and post-translational modifications of histones.

The inhibition of DNA methyltransferases, has been shown to prevent LTP induction in vitro and to affect the methylation levels in the promoters of Reelin and Bdnf in the hippocampus (Levenson, Roth, Lubin, Miller, Huang, Desai, Malone, and Sweatt, 2006). In vivo, it has been shown that hippocampus-dependent learning can trigger changes in the methylation levels at the Reelin, PP1, and Bdnf gene promoters. These changes were transient and normal levels were observed at 24 h after training (Miller and Sweatt, 2007; Lubin et al., 2008). Furthermore, mice with a conditional knockdown of Dnmt1 and Dnmt3a, two DNA methyltransferases, show impairment of LTP and memory formation (Feng and Fan, 2009).

Although histone acetylation seems to be less determinant for memory formation, and inhibiting the action of histone acetylases can lead to improve memory formation (Levenson et al., 2006; Guan, Haggarty, Giacometti, Dannenberg, Joseph, Gao, Nieland, Zhou, Wang, Mazitschek, et al., 2009). An exception may be Sirt1. In a work carried out by Michán, Li, Chou, Parrella, Ge, Long, Allard, Lewis, Miller, Xu, et al. (2010), the authors evidenced the crucial role of Sirt1, a histone deacetylase, in the induction of LTP. Sirt1 was able to inhibit the transcription of miR-134, a small non-coding RNA molecule (microRNA or miRNA), able to bind to CREB mRNA and inhibit its transcription. Furthermore, LTP induction in hippocampal slice cultures from Sirt1 loss-of-function mice was impaired. Similarly, overexpression of miR-134 resulted in LTP impairment. In the case of the loss-of-function mutation, LTP was restored by the knockdown of miR-134. Furthermore, a recent large-scale study by Joilin, Guévremont, Ryan, Claudianos, Cristino, Abraham, and Williams (2014) documented a generalised and rapid down-regulation of a large number of miRNAs following LTP induction.

In the light of these results, the overexpression observed in the 20 min dataset of the tumor suppressor Hic1 is difficult to interpret, since it is known to transcriptionally repress Sirt1 (Chen et al., 2005). The functions of Hic1, however, are not restricted to the inhibition of Sirt1. In addition, the repression of Sirt1 requires the formation of a Hic1/Sirt1 complex, which acts on its promoter region (Chen et al., 2005). Hence, downregulation of Sirt1 is directly dependent on its own protein levels, and this pathway may only represent a homeostatic mechanism to prevent overexpression of the deacetylase, which can increase cell proliferation cancer risk (Liu, Liu, and Marshall, 2009). The Bcl6 gene, an epigenetic regulator involved in the control of neurogenesis also requires the recruitment of Sirt1 (Tiberi et al., 2012).

The large number of transcriptional regulators that directly affect chromatin structure that appear as hubs in the co-expression networks identified 20 min post-LTP – e.g. Phf21a, Cxxc1, Ring1, Setdb1, and members of the neuron-specific chromatin remodelling complex (nBAF) – suggest that the epigenetic regulation is central to the early stages after the induction of LTP. However, in the 5 and 24 h datasets, other members of the nBAF complex appear as highly co-expressed, together with a significant number of epigenetic regulators, suggesting that the control of gene expression
Chapter 6. Co-expression analysis of LTP microarray data

6.4.3 Regulation of signalling pathways via changes in gene expression

Regulation of the kinase-phosphatase system

The modulation of the Ras-Erk1/2 pathway appears to happen mainly in the 20 min dataset. In addition to the discussed Homer1, some candidate modulators such as the complex formed by Ywhab and Akirin2 show a high co-expression. This complex represses the transcription of Dusp1, which is responsible for the inactivation of Erk1/2 (see Section 6.3.5). While the Dusp family has been implicated in LTP previously, our results suggest that Akirin2 and Ywhab may also have an important role despite not showing a significant overexpression in the 20 min dataset. Downstream of Erk1/2, Efr encodes for a transcription factor involved in the control of cell cycle.

Perhaps one of the most extensively studied molecular systems involved in LTP is the kinase-phosphatase system. Changes in gene expression also modulate this structure – a large number of kinases and phosphatases were identified across the different co-expression modules (see Section 6.3.7).

It is well known that PKA boosts CaMKII in the CaMK switch in the early stages of LTP (see Section 3.3.1) by phosphorylation of 11, which in turn inhibits PP1. While the CaMKII-PP1 system may not be necessary for an early-LTP molecular switch as proposed by Lisman in 1985, the crosstalk between CaM-kinases and phosphatases has a fundamental role in LTP (Allen, Hvalby, Jensen, Errington, Ramsay, Chaudhry, Bliss, Storm-Mathisen, Morris, Andersen, et al., 2000; Greengard, Allen, and Nairn, 1999). However, the high co-expression and significant upregulation at 24 h of a number of genes encoding for phosphate subunits suggests that kinase-phosphatase regulation may also be relevant in the persistence of LTP.

Regulation of the Akt pathway and lipid signalling

The involvement of Akt (or PKB) in LTP in the hippocampus has been suggested to be mediated by phosphatidylinositol 3-kinase (PI3K). PI3K may be required for the expression and maintenance of L-LTP in the CA1 region of the hippocampus (Sanna, Cammalleri, Berton, Simpson, Lutjens, Bloom, and Francescon, 2002; Karpova, Sanna, and Behnisch, 2006) through extracellular signals in a kinase-independent manner (Opazo, Watabe, Grant, and O’Dell, 2003). PI3K phosphorylates the D-3 position of the inositol ring of phosphoinositides in the plasma membrane, which converge with phosphoinositide-dependent kinase-1 (Pdk1) to activate Akt. This convergence is also of a spatial nature, since the activated phosphoinositides are restricted to the membrane. Full activation of Akt is the result of phosphorylation by the mTOR protein kinase. PI3K activity increases in response to $Ca^{2+}/CaM$, binding to AMPA receptors and driving their insertion during LTP (Man, Wang, Lu, Ju, Ahmadian, Liu, D’Souza, Wong, Taghibiglou, Lu, et al., 2003).

Not surprisingly, Akt1 is central to a module in the 20 min dataset, even though its differential expression is not significant (only moderately overexpressed at 5 h). Interestingly, the analyses performed by Ryan et al. (2011) on the same expression data did not identify Akt as an important gene in their Network 2, which contains PI3K. That same network had as a central hub the platelet-derived growth factor-binding protein homodimer (PDGF-BB), known to stimulate the pathway by binding the receptor Pdgfrb (Hanai et al., 2006), which we also identified in a 20 min module. Upon ligand binding, Pdgfrb activates the Akt signalling pathway via diacylglycerol (DAG)-inositol triphosphate (IP$_3$) (Kashishian, Kazlauskas, and Cooper, 1992). It can also stimulate the Ras-Erk1/2 pathway (Yokote, Mori, Hansen, McGlade, Pawson, Heldin, and Claesson-Welsh, 1994) and it is involved in the activation of c-fos transcription (Kruijer, Cooper, Hunter, and Verma, 1983).

The identification in the same module of other members of the DAG-IP$_3$ pathway as Akt1, such as Cds2, suggests that these pathways may be acting in conjunction. The Cds2 gene, also
overexpressed, encodes for an enzyme that provides CDP-DAG, precursor of PI (PI $\xrightarrow{\text{PI3K}}$ PIP$_2$).

The enzymatic reactions represented in Figure 6.11 can be summarised as

$$\text{PIP}_2 \xrightarrow{\text{PI3K}} \text{PIP}_3$$

$$\text{PIP}_2 \xrightarrow{\text{PLC}} \text{DAG} + \text{IP}_3$$

In the first case, PIP$_3$ then activates Akt, which subsequently activates mTOR. On the other hand, the DAG-IP$_3$ pathway requires the action of phospholipase C (PLC) upon PIP$_2$. The cleavage of the transmembrane molecule frees IP$_3$, that diffuses in the cytoplasm acting as a second messenger. Upon binding to IP$_3$ receptors in the endoplasmic reticulum, more Ca$^{2+}$ is released into the cytoplasm. There is evidence that IP$_3$ plays an important role in signalling in some types of synaptic plasticity (Sarkisov and Wang, 2008). In turn, DAG acts in conjunction with the rise in Ca$^{2+}$ to activate PKC and facilitate its translocation to the membrane (Nishizuka, 1995).

Together with the identification of other components of the lipid signalling pathway at 20 min and presented in more detail in Section 6.3.7, these findings indicate that lipid signalling actively participates in the induction of LTP, further supporting the notion that PI3K gene-expression pathway may be of importance in the early stages of LTP (Sanna et al., 2002).

The regulation of the PI3K-Akt-mTOR pathway by changes in gene expression may also occur at 5 h, as evidenced by the identification of the genes Dab2 and Plekha1. These components of the pathway – Dab2 is significantly downregulated at 5 h – are highly co-expressed at 5 h, suggesting that the regulation of this pathway could be coordinating LTP-related molecular events at different time scales.

### Other overrepresented functions

The results reported in Chapter 6 highlight a number of other overrepresented functions. In the first place, it is worth commenting the importance of the regulation of the protein product for L-LTP. In particular, we find a large number of genes involved in the ubiquitin-proteasome pathway. In this respect, a study by Dong, Upadhya, Ding, Smith, and Hegde (2008) already suggested that proteasome activity obstructs the maintenance of L-LTP by interfering with transcription and translation.

The 5 h and 24 h networks are to be enriched in a higher number of genes involved in the ubiquitin-proteasome pathway compared to the earlier networks, suggesting that protein degradation via ubiquitination may be amplified following the early phase of LTP.

Several genes co-expressed at 20 min have functions related with extracellular signalling and anchoring. The ephrin-receptors identified stress the importance of bidirectional signalling.

The involvement of this trans-synaptic signalling in hippocampal synaptic plasticity has been documented in several studies. Using hippocampal slices from gene-targeted mice, the requirement of EphA4 for the early stages of LTP at the CA3-CA1 synapses has been documented. Similarly, EphB receptors are thought to be necessary in the postsynaptic membrane for the induction of mossy fiber-CA3 synapses (Contractor, Rogers, Maron, Henkemeyer, Swanson, and Heinemann, 2002). EphA5, in turn, is known to be involved in guidance of hippocampal axons during development (Klein, 2008). Genes involved in cell adhesion were also identified at different times in the co-expression networks.

Gene expression changes also influence membrane composition. The synthesis of both transmembrane channel subunits and proteins that modulate their trafficking to the membrane is central to synaptic plasticity. Not surprisingly, a large number of genes with membrane functions were
Figure 6.11: Depicted in the figure are the different compounds involved in lipid signalling. Black arrows indicate enzymatic reaction, red arrows represent activation and blue dashed lines represent translocation.

identified in the co-expression networks. Indeed, across the three temporal datasets, a substantial number of channels and trafficking modulators appear as hubs.

For example, we found that Gabbr1, a hub in a 20 min module, was downregulated – around 15-fold compared to the control – only in the 5 h dataset. This downregulation may be directly implicated in facilitating LTP induction, as demonstrated by an experimental study carried out in hippocampal slices (Wigström and Gustafsson, 1983).

Other receptors known to be involved in LTP and identified as hubs in the co-expression analysis are cannabinoid receptors – Cnr1, a subunit highly expressed in the hippocampus whose activation leads to impairment of LTP and LTD (Misner and Sullivan, 1999) –, and receptors for glutamate coupled to G-proteins.

As we mentioned in previous sections, the trafficking of AMPA receptors is a central mechanism in LTP (see Section 3.3). Not surprisingly, we found a considerable number of gene playing this role. We want to stress the importance of cornichons, proteins involved in the modulation of AMPA receptor trafficking and gating properties. These proteins target the AMPA receptor to the synaptic membrane and regulate their rates of activation, deactivation, and desensitisation (Schwenk, Harmel, Zolles, Bildl, Kulik, Heimrich, Chisaka, Jonas, Schulte, Fakler, et al., 2009). Together with transmembrane (TARPs) – also identified as highly co-expressed in the 20 min dataset – cornichons seem to have a central role in the control of AMPA receptor gating in the hippocampus (Kato, Gill, Ho, Yu, Tu, Siuda, Wang, Qian, Nisenbaum, Tomita, et al., 2010).

An interesting hub identified in the analysis is the gene encoding for Reelin, which has a variety of functions related to microtubule regulation in neurons, neuronal migration, and brain
development (Rice and Curran, 2001). It also participates in the composition, recruitment and traffic of NMDA receptor subunits in the hippocampus (Qiu, Zhao, Korwek, and Weeber, 2006), in the generation of dendrites, and in the formation of dendritic spines (Niu, Yabut, and D’Arcangelo, 2008). A study carried out by Weeber, Beffert, Jones, Christian, Förster, Sweatt, and Herz (2002) showed that LTP can be augmented in mice hippocampal slices by perfusing Reelin, whereas this effect is abolished in very low density lipoprotein receptor (VLDLR) and apolipoprotein E receptor (apoER) knockout mice. Both are receptors for apolipoprotein E. The modulatory effect of Reelin on LTP is hence likely to be dependent on both these receptors. This regulation may be related with the recruitment of NMDA receptor subunits. Interestingly there is no significant overexpression of the gene in any of the datasets and this gene was not identified with the classical differential expression analysis with the exception of a 1.75 : 1 ratio of underexpression in the 5 h dataset.

Neuronal structural changes are known to follow LTP induction. For example, the dendritic spines show morphological changes following the stimulation patterns that elicit LTP. The changes in morphology are associated with the polymerization of actin (Fischer, Kaech, Knutti, and Matus, 1998; Muller, Toni, and Buchs, 2000). The enlargement of spine heads follows closely the pharmacology, amplitude, time course, and spatial localisation with synaptic potentiation, and this effect is observed soon after potentiation (Matsuzaki et al., 2004). The formation of new spines and the budding of dendritic filopodia are also modulated by synaptic activity (Ziv and Smith, 1996). However, new filopodia or spines require longer times, at least 20 min after the induction of LTP Maletic-Savatic, Malinow, and Svoboda (1999); Engert and Bonhoeffer (1999).

The actin cytoskeleton is particularly rich in dendritic spines (Matus, 2000). The role of actin in LTP seems to be central, as demonstrated by the impairment of LTP that actin inhibitors cause (Kim and Lisman, 1999; Krucker et al., 2000). The inhibition of actin depolymerisation may be the process underlying the changes in the actin cytoskeleton caused by LTP in vivo (Fukazawa, Saitoh, Ozawa, Ohta, Mizuno, and Inokuchi, 2003). Changes in actin polymerisation stand as the main cytoskeletal remodelling mechanism later in the 24 h modules.

We have pointed out that changes in cell morphology play a fundamental role in LTP in vivo. However, morphological changes in the existing synapses are not the only macroscopic changes that occur following LTP induction. The creation of new synapses and neurogenesis are processes linked to LTP and long-term memory (Snyder, Hong, McDonald, and Wojtowicz, 2005; Bruel-Jungerman, Davis, Rampon, and Laroche, 2006). In fact, the hippocampal region in particular is known to produce new neurons in adults, a process driven by learning processes (see for example Dayer, Ford, Cleaver, Yassaee, and Cameron, 2003; Gould, Beylin, Tanapat, Reeves, and Shors, 1999; Shors, Miesegaes, Beylin, Zhao, Rydel, and Gould, 2001; Raber, Rola, LeFavour, Morhardt, Curley, Mizumatsu, VandenBerg, and Fike, 2004).

We briefly discussed the role of the protein Reelin, involved in AMPA receptor trafficking and spine formation 6.3.9. Interestingly, Reelin seems to be involved in the regulation of postnatal neurogenesis (Pujadas, Gruart, Bosch, Delgado, Teixeira, Rossi, de Lecea, Martínez, Delgado-García, and Soriano, 2010). Perhaps the most representative gene involved in cell-cycle, however, encodes for a DNA polymerase subunit. It is overexpressed and central to a 20 min co-expression module. The number of genes involved in regulation of neurogenesis suggests that neuronal proliferation has an important role in hippocampal LTP as suggested by other previous studies (Ryan et al., 2011; Bruel-Jungerman et al., 2006).

Finally, our findings also support the observation that immune response genes are implicated in LTP (Håvik et al., 2007a). In particular, we identified immunity-related genes as hubs of some of the modules that exhibit a gain in intramodular connectivity at 24 h.

In conclusion, we suggest that the stabilisation of LTP requires the control of multiple cellular systems which act in coordination. Although these are not restricted to gene expression changes, we argue that the genomic response plays a fundamental role in L-LTP. Indeed, we speculate that
the partial overlap in time and space exhibited by these mechanisms confers an increased stability by redundancy to the observed cellular changes following LTP, which is a central discussion in the present thesis. This interpretation justifies the different experimental results on necessity and sufficiency of the key molecular species involved in LTP found across the literature (see Section 3.3). The stability of the molecular networks will be further discussed in the following sections.

6.5 Concluding remarks

The analyses outlined in the present chapter suggest that the rapid genomic response that follows LTP does not only involve a marked upregulation of gene expression, but also a tight coordination of its components. The early phase following stimulation must be critical in the onset of LTP, consistent with the notion that a fraction of the genes transcribed rapidly after LTP induction are of crucial importance at later times (Nguyen et al., 1994).

In addition, it is apparent that the stabilisation of LTP requires the control of multiple cellular systems. These include the regulation of gene expression (in particular epigenetic regulation), changes in the LTP-related signalling pathways, changes in membrane composition, and structural changes at the neuronal level via cytoskeleton remodelling. While some of the findings reported in the present chapter support previous reports in the literature, it is worth stressing that a number of genes and pathways have not been previously reported to be central to LTP. In particular,

- Wt1, a repressor of the Egr gene family
- Different members of the Homer gene family may be playing roles at later stages of LTP
- nBAF complex regulation
- Akirin2 and Ywhab, modulating the Ras-Erk1/2 pathway
- Protein degradation by ubiquitination, in particular in the NFκB pathway
- Activatino of NFκB immediately following LTP induction and then again 24 h post-LTP
- Lipid signalling
- Cadherin and its role in morphogenesis at 24 h post-LTP

<table>
<thead>
<tr>
<th>Module</th>
<th>Top 5 genes</th>
<th>Protein</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>blue_5H</td>
<td>Scn11a</td>
<td>Sodium channel protein type 11 subunit alpha</td>
<td>Mediates voltage-dependent sodium ion permeability of excitable membranes</td>
</tr>
<tr>
<td></td>
<td>Ab3</td>
<td>ABI gene family member 3</td>
<td>In vitro, reduces cell motility</td>
</tr>
<tr>
<td></td>
<td>Clec10a</td>
<td>C-type lectin domain family 10 member A</td>
<td>Probable role in regulating adaptive and innate immune responses</td>
</tr>
<tr>
<td></td>
<td>Alp2a1</td>
<td>Sarcoplasmic/endoplasmic reticulum calcium ATPase</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Cdkn2apn1</td>
<td>CDKN2AP N-terminal-like protein</td>
<td>Catalyses the hydrolysis of ATP coupled with the translocation of calcium from the cytosol to the sarcoplasmic reticulum</td>
</tr>
</tbody>
</table>

Table 6.5: Gene modules identified at 5 h with gain of TO at 20 min and 5 h and a subsequent loss at 24 h
<table>
<thead>
<tr>
<th>Module</th>
<th>Gene</th>
<th>Protein</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>lightyellow 5H (CNS neuron axogenesis)</td>
<td>C1qtnf3</td>
<td>C1qtnf3 protein</td>
<td>Positive regulation of Erk1/2 and Akt cascades</td>
</tr>
<tr>
<td></td>
<td>Fbln1</td>
<td>Fibulin-1</td>
<td>Role in cell adhesion and migration. Could play a significant role in modulating the neurotrophic activities of APP, particularly soluble APP</td>
</tr>
<tr>
<td></td>
<td>St8sia3</td>
<td>Alpha-2,8-sialyltransferase</td>
<td>Protein glycosylation</td>
</tr>
<tr>
<td></td>
<td>Haver1</td>
<td>Hepatitis A virus cellular receptor 1 homolog</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Anks1b</td>
<td>Ankyrin repeat and sterile alpha motif domain-containing protein 1B</td>
<td>May participate in the regulation of nucleoplasmic coilin protein interactions in neuronal and transformed cells. Overexpression can down-regulate APP processing</td>
</tr>
<tr>
<td>black 5H (anion homeostasis, synapse assembly)</td>
<td>Ap2b1</td>
<td>AP-2 complex subunit beta</td>
<td>Involved in protein transport via transport vesicles</td>
</tr>
<tr>
<td></td>
<td>Slc2a4</td>
<td>Glucose transporter type 4</td>
<td>Insulin-regulated facilitative glucose transporter</td>
</tr>
<tr>
<td></td>
<td>Pdcd4</td>
<td>Protein Pdcd4</td>
<td>PDZ domain containing protein</td>
</tr>
<tr>
<td></td>
<td>Aven</td>
<td>Cell death regulator Aven homolog</td>
<td>May protect against apoptosis</td>
</tr>
<tr>
<td></td>
<td>Agtbp1</td>
<td>Agtbp1 protein</td>
<td>In mice deglutamylates target proteins. Accumulation of tubulin polyglycylamidation causes neurodegeneration</td>
</tr>
<tr>
<td>darkred 5H (chondroitin sulfate metabolic process)</td>
<td>Dcaf4</td>
<td>Protein Dcaf4</td>
<td>May function as a substrate receptor for CUL4-DDB1 E3 ubiquitin-protein ligase complex</td>
</tr>
<tr>
<td></td>
<td>Ppp</td>
<td>Protein Ppp</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Kras</td>
<td>GTPase Kras</td>
<td>Oncogene Ras protein with GTPase activity</td>
</tr>
<tr>
<td></td>
<td>Dapk2</td>
<td>Death-associated kinase 2</td>
<td>Ca²⁺/CaM-dependent kinase. Involved in cell survival and apoptosis</td>
</tr>
<tr>
<td></td>
<td>Ip6k1</td>
<td>Inositol hexakisphosphate kinase 1</td>
<td>Enzyme involved in conversion of inositol phosphates</td>
</tr>
<tr>
<td>cyan 5H (regulation of DNA methylation)</td>
<td>Fmod</td>
<td>Fibromodulin</td>
<td>May have a primary role in collagen fibrillogenesis</td>
</tr>
<tr>
<td></td>
<td>Fam135a</td>
<td>Protein Fam135a</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Ankrd6</td>
<td>Similar to Diversin protein</td>
<td>Positive regulator of JNK cascade</td>
</tr>
<tr>
<td></td>
<td>Fndc3c1</td>
<td>Protein Fndc3c1</td>
<td>Unknown</td>
</tr>
<tr>
<td>midnightblue 5H (chaperone mediated protein folding)</td>
<td>Myo2</td>
<td>Myozienin-2</td>
<td>Plays an important role in the modulation of calcineurin signalling</td>
</tr>
<tr>
<td></td>
<td>Itsn2</td>
<td>Intersectin-2</td>
<td>Regulates endocytic membrane traffic</td>
</tr>
<tr>
<td></td>
<td>Zltp324</td>
<td>Protein Zltp324</td>
<td>Zinc finger containing protein</td>
</tr>
<tr>
<td></td>
<td>Cenpq</td>
<td>Centromere protein Q</td>
<td>Involved in mitotic progression and chromosome segregation</td>
</tr>
<tr>
<td></td>
<td>Arti1</td>
<td>E3 ubiquitin-protein ligase ARIH1</td>
<td>E3 ubiquitin-protein ligase, which catalyses polyubiquitination of target proteins</td>
</tr>
<tr>
<td>magenta 5H (progesterone receptor signalling pathway)</td>
<td>Mrpl35</td>
<td>Mitochondrial ribosomal protein L35</td>
<td>Ribosomal protein</td>
</tr>
<tr>
<td></td>
<td>Prkd3</td>
<td>Serine/threonine-protein kinase D3</td>
<td>Converts transient DAG signals into prolonged physiological effects, downstream of PKC</td>
</tr>
<tr>
<td></td>
<td>Cul5</td>
<td>Cullin-5</td>
<td>Core component of multiple E3 ubiquitin-protein ligase complexes</td>
</tr>
<tr>
<td></td>
<td>Pqcp</td>
<td>Interferon regulatory factor 2-binding protein-like</td>
<td>Zinc finger containing protein</td>
</tr>
<tr>
<td></td>
<td>Smarb1</td>
<td>SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1</td>
<td>Core component of the BAF complex</td>
</tr>
<tr>
<td>green 5H (regulation of amino acid transport)</td>
<td>Prelid2</td>
<td>PRELI domain-containing protein 2</td>
<td>Phospholipid transport</td>
</tr>
<tr>
<td></td>
<td>Prl2b1</td>
<td>Prolactin-2B1</td>
<td>May act as a hormone</td>
</tr>
<tr>
<td></td>
<td>Atfca8</td>
<td>ATR-binding cassette sub-family A member 8</td>
<td>Lipophilic drug transporter</td>
</tr>
</tbody>
</table>

Continued on next page
Table 6.5 – continued from previous page

<table>
<thead>
<tr>
<th>Module</th>
<th>Gene</th>
<th>Protein</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>darkturquoise_5H (negative regulation of</td>
<td>Foxn3</td>
<td>Checkpoint suppressor 1</td>
<td>Transcriptional repressor. May be involved in DNA damage-inducible cell cycle arrests</td>
</tr>
<tr>
<td>protein autophosphorylation)</td>
<td>Morc2</td>
<td>MORC family CW-type zinc finger protein 2</td>
<td>Involved in lipogenesis</td>
</tr>
<tr>
<td></td>
<td>Atad2</td>
<td>ATPase family AAA domain-containing protein 2</td>
<td>May be required for histone hyperacetylation</td>
</tr>
<tr>
<td></td>
<td>Jun</td>
<td>Transcription factor AP-1</td>
<td>Increases gene expression upon cAMP signaling pathway stimulation</td>
</tr>
<tr>
<td></td>
<td>Ing3</td>
<td>Inhibitor of growth protein 3</td>
<td>Component of the NuA4 histone acetyltransferase complex</td>
</tr>
<tr>
<td></td>
<td>Rtn6a</td>
<td>RNA-binding protein 8A</td>
<td>Core component of the splicing-dependent multiprotein exon junction complex</td>
</tr>
<tr>
<td></td>
<td>Fkhb15</td>
<td>FK506-binding protein 15</td>
<td>May be involved in the cytoskeletal organization of neuronal growth cones</td>
</tr>
<tr>
<td>grey60_5H (regulation of branching involved</td>
<td>Egln1</td>
<td>Egl nine homolog 1</td>
<td>Cellular oxygen sensor</td>
</tr>
<tr>
<td>in salivary gland morphogenesis)</td>
<td>Gabre</td>
<td>GABA(A) receptor subunit epsilon</td>
<td>Major inhibitory neurotransmitter in the vertebrate brain</td>
</tr>
<tr>
<td></td>
<td>Btb16</td>
<td>BTB/POZ domain-containing protein 16</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Ift30</td>
<td>Gamma-interferon-inducible lysosomal thiol reductase</td>
<td>Lysosomal thiol reductase</td>
</tr>
<tr>
<td></td>
<td>Tsen54</td>
<td>tRNA-splicing endonuclease subunit Sen54</td>
<td>Subunit of the tRNA-splicing endonuclease complex</td>
</tr>
<tr>
<td>lightcyan_5H (pyrimidine nucleotide</td>
<td>Adrb1</td>
<td>Beta-1 adrenergic receptor</td>
<td>Receptor that binds epinephrine and norepinephrine</td>
</tr>
<tr>
<td>metabolic process)</td>
<td>Dab2</td>
<td>Disabled homolog 2-interacting protein</td>
<td>Involved in PI3K-Akt and Map3K5-JNK signaling pathways</td>
</tr>
<tr>
<td></td>
<td>Ribb12</td>
<td>Ribonucleo-related protein 3</td>
<td>May be involved in regulated intramembrane proteolysis and the subsequent release of functional polypeptides</td>
</tr>
<tr>
<td></td>
<td>Cyp11b1</td>
<td>Cytochrome P450 11b1, mitochondrial</td>
<td>Forms corticosterone</td>
</tr>
<tr>
<td></td>
<td>Rprmt</td>
<td>Reprimo-like protein</td>
<td>Unknown</td>
</tr>
<tr>
<td>lightgreen_5H (regulation of T-cell</td>
<td>H3f3b</td>
<td>Histone H3.3</td>
<td>Variant histone H3</td>
</tr>
<tr>
<td>migration)</td>
<td>Cmhl2</td>
<td>Protein cornichon homolog 2</td>
<td>Regulates the trafficking and gating properties of AMPA receptors</td>
</tr>
<tr>
<td></td>
<td>Trps1</td>
<td>Zinc finger transcription factor Trps1</td>
<td>Transcriptional repressor</td>
</tr>
<tr>
<td></td>
<td>Tnfaip2</td>
<td>Tumor necrosis factor alpha-induced protein 2</td>
<td>May play a role as a mediator of inflammation and angiogenesis</td>
</tr>
<tr>
<td></td>
<td>Fam46a</td>
<td>Protein Fam46a</td>
<td>RNA binding</td>
</tr>
<tr>
<td>turquoise_5H (DNA catabolic process)</td>
<td>Cfrh1</td>
<td>Corticotropin-releasing factor receptor 1</td>
<td>G-protein coupled receptor. Promotes the activation of adenylate cyclase, leading to increased intracellular cAMP levels</td>
</tr>
<tr>
<td></td>
<td>Kdm6b</td>
<td>Lysine-specific demethylase 6B</td>
<td>Histone demethylase</td>
</tr>
<tr>
<td></td>
<td>F8</td>
<td>Coagulation factor VIII</td>
<td>Blood coagulation</td>
</tr>
<tr>
<td></td>
<td>Rad52</td>
<td>DNA repair protein RAD52 homolog</td>
<td>DNA repair</td>
</tr>
<tr>
<td></td>
<td>Exoc8</td>
<td>Exocyst complex component 8</td>
<td>Involved in the docking of exocytic vesicles with fusion sites on the plasma membrane</td>
</tr>
<tr>
<td>yellow_5H (positive regulation of response</td>
<td>Arglu1</td>
<td>Arginine and glutamate-rich protein 1</td>
<td>Unknown</td>
</tr>
<tr>
<td>to interferon-gamma)</td>
<td>Mcccl</td>
<td>Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial</td>
<td>Amino acid degradation</td>
</tr>
</tbody>
</table>

Continued on next page
### 6.5 Concluding remarks

<table>
<thead>
<tr>
<th>Module</th>
<th>Gene</th>
<th>Protein</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>purple_5H (histone H3-K27 methylation)</td>
<td>Tmem206</td>
<td>Transmembrane protein 206</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Ypel4</td>
<td>Protein yippee-like 4</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Tiparp</td>
<td>TCDD-inducible polyA polymerase</td>
<td>Formation of polyA chains</td>
</tr>
<tr>
<td>salmon_5H (positive regulation of vasoconstriction)</td>
<td>Fgf21</td>
<td>Fibroblast growth factor 21</td>
<td>Stimulates glucose uptake in differentiated adipocytes</td>
</tr>
<tr>
<td></td>
<td>Adcy3</td>
<td>Adenylate cyclase type 4</td>
<td>Membrane-bound, CaM-insensitive adenylyl cyclase</td>
</tr>
<tr>
<td></td>
<td>Klh22</td>
<td>Kelch-like protein 22</td>
<td>Substrate-specific adapter of a BCR (BTB-CUL3-RBX1) E3 ubiquitin ligase complex</td>
</tr>
<tr>
<td></td>
<td>Gnal</td>
<td>Guanine nucleotide-binding protein G(o1f) subunit alpha</td>
<td>Involved in various transmembrane signaling systems</td>
</tr>
<tr>
<td></td>
<td>Kcnj12</td>
<td>ATP-sensitive inward rectifier potassium channel 12</td>
<td>Inward rectifying potassium channel activated by PI2P and that probably participates in controlling the resting membrane potential</td>
</tr>
<tr>
<td>tan_5H (cellular response to heat)</td>
<td>H3st2</td>
<td>Heparan sulfate glucosamine 3-O-sulfotransferase 2</td>
<td>Sulfotransferase</td>
</tr>
<tr>
<td></td>
<td>Hdac5</td>
<td>Histone deacetylase 5</td>
<td>Deacetylation of lysine residues on core histones</td>
</tr>
<tr>
<td></td>
<td>Ccdc115</td>
<td>Coiled-coil domain-containing protein 115</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Gabbr1</td>
<td>GABA(A) receptor subunit rho-1</td>
<td>Major inhibitory neurotransmitter in the vertebrate brain</td>
</tr>
<tr>
<td></td>
<td>Atf6p5</td>
<td>PRAT family protein 5</td>
<td>Regulates intracellular concentrations of glutamate. Negatively modulates SLC1A1/EAAC1 glutamate transport activity</td>
</tr>
<tr>
<td></td>
<td>Asb1</td>
<td>Ankyrin repeat and SOCS box protein 1</td>
<td>Involved in protein ubiquitination</td>
</tr>
<tr>
<td></td>
<td>Tpr</td>
<td>Nucleoprotein TPR</td>
<td>Component of the nuclear pore complex</td>
</tr>
<tr>
<td></td>
<td>Pcx5l</td>
<td>PEX5-related protein</td>
<td>Accessory subunit of hyperpolarization-activated cyclic nucleotide-gated channels</td>
</tr>
<tr>
<td></td>
<td>Pank4</td>
<td>Pantothenate kinase 4</td>
<td>Plays a role in the physiological regulation of the intracellular coenzyme A concentration</td>
</tr>
<tr>
<td></td>
<td>Znf609</td>
<td>Zinc finger protein 609</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
7. Stability of an early LTP network using TReMM

7.1 Introduction

Memory processes in the brain are dependent on the capacity of neurons to undergo long-lasting enhancement of synaptic effectiveness. LTP is one of such mechanisms, providing a cellular model of associative, specific, and long-lasting storage.

Although LTP is generally divided into three phases over time only the intermediate- and late-phases (often referred as LTP2 and LTP3 respectively), seem to be dependent on protein synthesis when induced in vivo at perforant path synapses. As well as long-term memory, late-LTP (L-LTP) is dependent on both changes in gene expression and protein synthesis (Nguyen et al., 1994; Abraham and Williams, 2003, 2008; Alberini, 2009).

The new transcripts and proteins that characterise the L-LTP could be targeted to the specific synapses from the nucleus by means of a synaptic “tag” generated at the potentiated synapses (Frey and Morris, 1997). A first wave of transcription brought about by the activated constitutive transcription factors (TFs) would consist of effector genes and inducible TFs which, in turn, would be responsible for a second wave of transcription (Ryan et al., 2011; Abraham and Williams, 2003).

Although several genes have been identified to belong to some of those categories (see Abraham and Williams, 2008, 2003; Lynch, 2004) it is unclear how do they coordinate in order to lead to the persistence of LTP. Several efforts have been taken towards defining a wider picture of the biochemical networks underlying the maintenance of LTP (Lee et al., 2005; Park et al., 2006; Håvik et al., 2007a; Valor and Barco, 2010; Ryan et al., 2011). One of the pathways leading to de novo gene expression is triggered by the activation by phosphorylation of cAMP response element-binding protein (CREB) by the phosphorylated calcium- and cAMP-dependent kinase.
CREB functions as a hub in the gene regulation. Its phosphorylation can be driven by a plethora of different kinases (Schulz, Siemer, Krug, and Höllt, 1999) and in turn, its activation leads to the expression of transcriptionally linked genes.

Once phosphorylated, CREB acts as a TF stimulating the expression of other TFs and other effector genes by binding to the promoter CRE region. The transcriptional response will constitute a first wave of gene expression. Other TFs which are constitutively expressed may also become active following LTP induction.

The first wave of de novo transcription brought about by the LTP-activated TFs consists of activity-induced TFs, structural proteins, signal transduction proteins, growth factors, and enzymes. These set of genes have been referred to as immediate-early genes (IEGs). These TFs are believed to drive the expression of delayed effector genes (see Section 3.3.6 and Abraham and Williams (2003, 2008); Miyashita, Kubik, Lewandowski, and Guzowski (2008); Tischmeyer and Grimm (1999) for comprehensive reviews).

Several studies have more recently covered the gene expression activation with studies of coregulation based on microarrays (see for example Park et al., 2006; Ryan et al., 2011), identifying sets of differentially expressed genes after the induction of LTP. It appears that the expression state of the LTP-related transcriptional regulatory network (LTP-GRN) shifts from a resting state (pre-LTP state) to a post-LTP state, where LTP has been induced via experimental high frequency stimulus. These expression states have been measured experimentally using high density arrays by Ryan et al. (2011). They identified a set of differentially expressed genes whose interactions served to identify biologically relevant networks and pathways related with LTP. Their network analyses suggest that some genes may play central roles during LTP induction without showing a significant alteration in their expression levels (such is the case of NF-κB). Other genes identified by the IPA software as major hubs in LTP-regulated gene expression networks were SRF, Egr1 and CREB. Erk also appears as a central hub in the network and its activity seems regulated by the phosphatase activity of DUSPs, which in turn appear to have a coordinated upregulation. In addition, DUSP and Egr families regulate the Ras-Erk1 signalling pathway, which is known to have an essential role in LTP.

In the present chapter, we analyse the most significant network published in Ryan et al. (2011) to investigate its topological properties from a dynamical perspective. We model the transcription regulation through weight matrices, and we show that the particular topology of the LTP-GRN is more likely to hold bistability when compared to random networks. On a similar perspective, we classify the LTP-related genes according to their relevance to the bistable output of the network.

7.2 Methods

7.2.1 Weight matrices dynamics

Continuous models based on differential equations entail a main disadvantage compared to discrete-time models. Despite experimental evidence can be gathered regarding the networks’ topology (the who interacts with who), there is a lack of quantitative information relative to the biochemical parameters underlying gene interactions (the how the interacting partners interact). In other words, the strength of the interactions between genes and proteins are difficult to measure. On the other hand, the widely-used Boolean models for gene regulation might oversimplify the complexity underlying biological interactions.

For this study, we have adopted the TReMM algorithm (Transcription Regulation Modelled with Matrices) described by Weaver et al. (Weaver et al., 1999). The TReMM methodology characterises the gene expression values from a state of maximal expression to a state of minimal repression. These intervals are gene-specific, and during a simulation a gene can take any continuous value in the range defined by its maximal and minimal expression. The details of the methodology are
7.2 Methods

described in detail in Section 4.2.3.

In particular, we are interested in the existence of such a landscape for the LTP-GRN, where these attractors would represent the pre- and post-LTP cell states, potentially accessible from different initial conditions. Those initial conditions could be specified by the cell’s environment.

7.2.2 LTP-GRN and gene expression data

For this study we have chosen the highest scoring LTP-GRN published in Ryan et al. (2011), constituted by a total of 35 genes (see Figure 7.2), which was constructed using the IPA software and is based on biological information sourced from literature references (see Section 5.5).

The initial conditions for the time course experiments, \( u(t = 0) \), as well as the maximal expression levels \( m \) were obtained from real expression data. In the original experiments the animals were stimulated unilaterally so that the unstimulated hemisphere could serve as a within-animal control Ryan et al. (2011, 2012). We used the values of the microarrays hybridized with the RNA from the stimulated hemisphere as post-LTP state, and the unstimulated hemisphere values as the pre-LTP state.

7.2.3 Frequency of bistable outputs \( P_B \)

We define as bistable a network dynamic simulation which converges onto two different stable terminal states when initialised with different expression values. In this study, we have set those starting expression values according to the LTP-induced microarray data and the control microarray data, namely, running for each network and each set of \( \alpha \), \( \beta \), and \( W \) parameters two parallel dynamics representing pre- and post-LTP conditions respectively. The presence of two different attractors in the gene expression state space each relatively closer to either the pre- or the post-LTP initial conditions, will lead to two different terminal states. These cases are referred to as bistable outputs. We have chosen the experimental expression values obtained by Ryan et al. since they represent real estimates of the gene expression levels.

By sampling the space of parameters and testing whether the output states of the pre- and post-LTP parallel dynamics reach different points in the N-dimensional space, we calculate the frequency of bistable outputs or \( P_B \), which characterises the network’s expression state space around the pre- and post-LTP initial conditions. In this work, we have randomly sampled the space of parameters by 10,000 trials with \( w_{ij} \in [-1, +1], \beta_i \in [-1, +1], \) and \( \alpha \in (0, +3] \).

The value of \( P_B \) is free of the bias which arbitrary parameters would introduce, and serves as a descriptor of how a particular topology is likely to show a sensitivity to the realistic initial conditions observed in the in vivo experiments. Put differently, the attractors in the gene expression state space represent stable phenotypes, while \( W \) represents the neuron’s genotype.

7.2.4 Random and knock-out networks

In order to characterise the LTP-GRN, we have constructed a total of 100 random networks by preserving the same number of nodes and edges as in the LTP-GRN and without assuming any fixed degree distribution. These networks have been characterised in terms of their \( P_B \), and the values are represented in Figure 7.1.

Similarly, we have calculated the \( P_B \) for the 35 networks resulting from the knock-out of each gene in the LTP-GRN network. Some of these knock-out networks show an inferior propensity to bistability (a lower \( P_B \)), meaning that the removed gene has a key role in the capacity of the LTP-GRN to reach bistable outputs. The \( P_B \) values for the removal of each gene are represented in Figure 7.2, where a darker shade of red means higher influence in the network’s bistability propensity.
Chapter 7. Stability of an early LTP network using TReMM

7.3 Results

7.3.1 LTP-GRN bistable frequency

The frequency of bistable outputs for the LTP-GRN ($P_B = 0.086$) is in sharp contrast with the frequencies obtained for the random networks with the same number of nodes and edges (Figure 7.1, avg. $P_B = 0.031$). Furthermore, the value of the $P_B$ for the LTP-GRN is higher than any of the random networks. Assuming normality for the distribution of $P_B$ in the random networks’ population, ($p = 0.019$, Shapiro-Wilk test), the probability of a random network with a higher $P_B$ than the LTP-GRN is extremely low ($P < 3.65 \times 10^{-07}$).

![Figure 7.1: Distribution of 100 random networks according to their bistable output frequencies $P_B$ (10,000 simulations for each network in the parameter space). The $P_B$ distribution for the random networks has $\mu = 0.031$ and $\sigma = 0.011$. The LTP-related network shows a much higher $P_B = 0.086$, which lies about 5 standard deviations away from the mean assuming normality for the random network distribution.]

7.3.2 Knock-out effects

For each $i$-th gene of the $N = 35$ genes in the LTP-GRN, an independent set of knock-out runs were performed. In these runs, the $i$th gene of the network was removed and not considered in the simulations, setting the $i$th row and column of $W$ to zero. The resulting 35 networks with $N = 34$ were characterized in terms of their frequency of bistable outputs, $P_B$.

Figure 7.2 summarises these results by showing in a red scale the effect of removing that particular gene from the network. NF-κB has a low influence on $P_B$, while the Egr members appear to be of a fundamental importance, which supports previous findings (Abraham et al., 1993). Interestingly, the level of expression of NF-κB does not change after LTP induction (Ryan et al., 2011), suggesting a regulation by post-translational modification.

7.4 Discussion

While evolution has equipped cells with a robust and plastic analog system which confers to living cells the ability to respond in a graded manner to environmental inputs, some aspects of the cellular
behaviour are of a Boolean nature – this is the case of cell cycle checkpoints, cell-fate transitions, and apoptosis to name a few. Likewise, LTP as a long-lasting phenotypic phenomenon represents one such Boolean characteristic, and its maintenance relies on gene expression.

Back in 1969, Stuart Kauffman proposed that cell types are attractors in the gene expression state space (Kauffman, 1969a). Similarly, our hypothesis extends this idea to the phenotypic pre- and post-LTP neuronal states. High-frequency stimulus would drive the state of a neuron from one gene expression attractor to another in vivo.

From such a perspective, we have classified in the study described in the present chapter the LTP-related genes according to their relevance to the bistable dynamic output of the network. It is noteworthy how the Egr family stands out in Figure 7.2 and in the results of the co-expression analysis. Their role in LTP has been already documented (Abraham et al., 1993; Williams, Dragunow, Lawlor, Mason, Abraham, Leah, Bravo, Demmer, and Tate, 1995; Jones et al., 2001), and the results presented here further stress their importance in the dynamic context of the other genes. In particular, Egr2 appears to be the most determinant gene for the network to show
bistability in the weight matrices dynamics (Section 7.3.1). Note however that this ranking only represents the relative importance of the genes for bistable behaviour. For example, NFκB shows a relatively low influence on the network’s $P_B$, while it is believed to be involved actively in LTP and synaptic plasticity (Freudenthal et al., 2004; O’Mahony, Raber, Montano, Foehr, Han, Lu, Kwon, LeFevour, Chakraborty-Sett, and Greene, 2006). The low $P_B$ shown by NFκB is not trivial, since it also constitutes a hub in the network, and a high dependence on it for the bistable frequency might be expected. Interestingly, NFκB does not show an altered expression level after LTP induction (Ryan et al., 2011), which is consistent with its constitutive expression observed in a number of studies and the notion of its activation by phosphorylation of a bound inhibitor (see Section 3.3.6). The results of the co-expression discussed in the previous section confirm the central role of NFκB signalling in LTP.

The capacity of showing different stable points in the gene expression state space must be of a crucial importance for the gene regulatory networks involved in such step-like responses. We show that this is the case for an LTP-related gene regulatory network by using realistic expression values from microarray data.

Whether this property of the LTP-GRN extends to most biological networks or only to the networks involved in Boolean decisions remains unclear and deserves further insight. Indeed, a long-lasting cellular process such as LTP is expected to function under a changing environment. Neurons are subject to both internal and environmental noise, which compromises the stability of the gene expression states. This compromise between plasticity (two-attractors hypothesis) and robustness to noise has not been addressed in this study. The role of gene expression attractors in robustness and evolvability of biological networks, from a more general perspective, emerges as an interesting generalisation of this analysis but beyond the scope of this study.

These results were published in the proceedings of the 2012 IEEE International Joint Conference on Neural Networks (Nido, Williams, and Benuskova, 2012).
8. Stability of LTP networks using Boolean models

What matters in life is not what happens to you but what you remember and how you remember it

Gabriel García Márquez

8.1 Introduction

It has been suggested that natural genetic regulatory networks do not possess ordered dynamics, but rather lie in the so-called critical regime, the boundary between order and chaos (Balleza et al., 2008). The resulting compromise between stability to small perturbations and sensitivity to specific perturbations may confer an evolutionary advantage as biological networks have to be able to discriminate background noise (robustness against unspecific changes in the environmental variables) from the meaningful signals that deserve the triggering of a specific response. Likewise, LTP as molecular mechanism that functions as a mnemonic device, must bear these central properties.

In fact, a number of studies suggest that network architecture plays an important role in the stability of cellular states (Albert et al., 2000; Jeong et al., 2001; Milo et al., 2002) and hence be under evolutionary pressures (Li et al., 2004). It can be expected that topological properties of networks differing in their functions or acting at different time scales will show different properties, tuned according to a particular compromise between robustness and plasticity. In this aspect, networks involved in early stages of a persistent mechanism are likely to be endowed with plastic properties, while networks regulated later on and responsible for the maintenance of the new homeostasis, will show a higher dynamic stability. A lack of quantitative data to date has made dynamic modelling a rather rare approach to understanding the genetic mechanisms controlling LTP maintenance.

The remarkable persistence of LTP and its highly specific nature, that is only inputs that are active during high-frequency stimulation are potentiated (Bonhoeffer, Staiger, and Aertsen, 1989), highlight LTP as a good model mechanism to study how order and chaos in gene network dynamics are tuned by living cells. By studying LTP-related networks identified at different times after LTP induction (20 min, 5 h and 24 h post-LTP) from a dynamic systems perspective, we aim to gain a
Chapter 8. Stability of LTP networks using Boolean models

deep understanding of the relationship between function and network structure in gene regulatory networks. In addition, the present study emphasises the key role played by new gene expression in LTP consolidation an maintenance, which is still debated.

In this chapter, we further extend the study presented in Chapter 7 by using a random Boolean network model and taking advantage of the additional gene expression data available for different times after LTP induction (5 h and 24 h, Ryan et al., 2012). The goal of this work is to characterise the dynamic differences between temporally distant networks, which are regulated by the same environmental trigger. We compare the post-LTP regulated networks with random models and we use the yeast transcriptional network as a benchmark. We show that the dynamics exhibited by the latest (24 h) network are less sensitive to perturbations than the earlier networks – an effect consistent with a role in the consolidation of synaptic plasticity.

8.2 Methods

8.2.1 Random Boolean Networks

Random Boolean networks (RBN) represent one of the simplest models for gene regulatory networks. In Boolean networks, there are \( N \) nodes representing the genes which can only take on or off values at a given time, \( x_i(t) \in \{0, 1\} \). The values of the vector \( x(t) \) are updated according to the Boolean rules \( B = (f_1, ..., f_N) \) and the values of the vector \( x(t-1) \) in the previous iteration. For details on the methodology, see Section 4.2.1.

In particular case of RBNs, both the set of Boolean rules \( B \) and the values of the state vector \( x(t = 0) \) are initialised randomly. Subsequently, \( x(t) \) is updated synchronously based on the vector at \( t-1 \) and according to the Boolean rules. Given an initialised vector at \( t = 0 \) and the set \( B \) of truth tables, this type of dynamics are deterministic. The set of rules \( B \) will stay the same over time during the simulation, whereas the state vector \( x(t) \) will evolve until reaching an attractor. For a given network topology, a large number of simulations has to be run to properly sample random initial conditions \( x(t = 0) \) and \( B \).

As described in Section 4.2.1, the dynamics of RBNs can fall in the ordered or chaotic regime, depending on their dynamic stability (Fox and Hill, 2001). For randomly connected networks with a Poisson and power-law degree distribution, it has been shown that the border between the ordered and chaotic regimes is given by \( 2p(1-p)K = 1 \), where \( p \) denotes the bias for the fraction of output values that are 1 in the Boolean rules, \( p = P(B_i = 1) \), and \( K \) the constant in-degree (Aldana-Gonzalez et al., 2002; Aldana and Cluzel, 2003).

There are different strategies to study the degree of dynamical order in a Boolean network. A popular method consists in measuring the effects of small perturbations on the outcome of the simulations – two initial states that differ in a number of positions, \( h(0) \) (known as Hamming distance), are evolved in parallel for \( \tau \) time steps. The new Hamming distance, \( h(\tau) \) is calculated. The Hamming distance \( h \) can be averaged over a large number of initial conditions and Boolean rules (denoted with \( H(\tau) \)). This methodology is very useful when combined with the Derrida plots (see Section 4.2.1). The Hamming distances in ordered networks tend to decrease on average after evolving the network in time, while the chaotic or disordered networks show the opposite behaviour, \( H(\tau) > H(0) \).

8.2.2 Network topologies

The networks used in this study were sourced from the literature or generated randomly according to a set of criteria.

Ryan et al. (2011, 2012) carried out a series of microarray experiments at different times after LTP induction. Specifically, their analyses aimed to identify genes regulated 20 min, 5 h, and 24 h after LTP induction. For each of these groups, they provided the three highest
scoring networks according to the analysis provided by the IPA software (Ingenuity Systems, USA; www.ingenuity.com), making a total of 9 networks. In the present study we will analyse the highest scoring network for each time point. The 20 min network is represented in Figure 8.1a, whereas all the networks used are depicted in Figure 5.4 in Section 5.5. In addition, we have used the yeast transcriptional network as a benchmark for RBN modelling, so that our results can be readily compared to the analyses available in the literature – more specifically to the papers by Kauffman, Peterson, Samuelsson, and Troein (2003) and Karlsson and Hörnquist (2007). The results presented in those studies used the same network published in (Lee, Rinaldi, Robert, Odom, Bar-Joseph, Gerber, Hannett, Harbison, Thompson, Simon, et al., 2002). We will refer to this network simply as the yeast network.

![Figure 8.1: Types of network topologies used in the study. (a) First of the three 20 min post LTP networks (Ryan et al., 2011). (b) Example of a rewired network, where the edges have been swapped in order to preserve the same node degree. (c) Random network with the same number of nodes and edges. The in- out-degree distribution is not preserved with respect to (a). The networks are represented in a circular layout with the nodes in ascending order of degree, counter-clockwise starting from the bottom](image)

Although there are a number of different algorithms to compute random networks, in order to compare the dynamic properties of the LTP networks we have used two different sets of constraints to generate them. First, an ensemble of 100 random networks is generated for each of the 4 biological networks studied (20 min, 5 h, 24 h and yeast) by preserving the same number of nodes and edges. The 4 ensembles will be referred to simply as Random networks, see Figure 8.1c.

To discriminate the effects of the network’s local structure from the effects of the general topology, a second type of random architecture consists of ensembles of 100 random networks where each node has the same number of in- and outgoing edges as the original biological network. To preserve these local topological properties, a randomly chosen pair of edges is swapped such that if \( A \rightarrow B \) and \( C \rightarrow D \), the resulting network has \( A \rightarrow D \) and \( C \rightarrow B \). Swapping is prevented if either \( A \rightarrow D \) or \( C \rightarrow B \) exist already (Kannan, Tetali, and Vempala, 1997). Within each of these 4 ensembles (referred to as Rewired networks, see Figure 8.1b for an example), each network has the same in- and out-degree distribution as the original LTP network from which they originated.

Finally, we aimed to study the effect of noise at the genetic level by generating all the 35 possible knock-out networks for each of the original LTP networks.
8.3 Results

8.3.1 LTP-related networks are dynamically similar to the yeast network

To compare the effect that different topologies have on the dynamic behaviour of networks we used an RBN model. Specifically, we compared the three most significant networks derived using IPA at three time points (20 min, 5 h and 24 h) following the induction of perforant path LTP (Ryan et al., 2011, 2012), with the yeast network obtained from (Lee et al., 2002).

Figure 8.2: Histograms of the degree for the LTP networks and the yeast transcriptional network
Kauffman et al. (2003) studied the yeast network by using a similar RBN model and showed that the yeast network was more stable around fixed points than the rewired networks. In our approach we do not assume any structure in the Boolean rules (neither canalizing nor nested functions Kauffman (see 1993); Kauffman et al. (see 2003) and this difference is difficult to grasp (Figure 8.3). Interestingly, when we applied the same analysis to the LTP-related networks compared to their rewired counterparts, we found this tendency towards a more ordered regime to be even more noticeable (blue lines in Figure 8.4). The dynamics fall in the chaotic regime (note that for perturbations in which a fraction of \( < 0.5 \) of the genes differ, these networks are above the diagonal, Figure 8.4) but this behaviour can be suppressed by using forcing structures in the Boolean functions (Kauffman, Peterson, Samuelsson, and Troein, 2004). For the purpose of comparing between the different time-scales networks, we opted not to assume any structure and we used a flat distribution.

![Figure 8.3: Derrida plots for the yeast network, with \( H(0) \) vs \( H(5) \). Each point in the plot is an average over 1,000 random rule assignments for 100 random initial conditions. Hamming distances are normalised by the number of nodes.](image)

### 8.3.2 Knock-out analyses

Figure 8.5 summarises the results obtained by simulating the knock-out networks. A total of 35 knock-out networks for each LTP network are represented in the figures. To highlight the effect of the degree on the dynamics shown by the original network, the knock-out networks corresponding to the removal of the top 2 genes in terms of degree are represented with different colours (red, blue) – Yeast: Yap6, Hms1; 20 min: Egr1, NFkB; 5 h: Hdac1, Smarca4; 24 h: Dlg4, Phlda3. Figure 8.6 stresses these differences by plotting the Hamming distances after 5 iterations.

### 8.4 Concluding remarks

The comparison of the dynamic stability between the networks obtained at different times after LTP-induction show that the regime becomes less chaotic with time after LTP induction (see Figure 8.4). In particular, the 24 h network appears to be considerably more ordered than the yeast, and ordered dynamics prevail under perturbations higher than a fraction of 0.38 genes. In contrast, the dynamic stability exhibited by the 20 min and 5 h networks is comparable to the one observed for the yeast. This observation is readily accounted for in Figure 8.7, where two states differing in...
Figure 8.4: Derrida plots for the LTP networks (from left to right: 20 min, 5 h, 24 h). Random networks for each of them are represented in red lines, Rewired networks in blue. The top plots show the $H(0)$ distance plotted against $H(1)$, while the bottom plots represent $H(0)$ vs $H(5)$ for the same simulations. Despite showing slightly different behaviours, the curves corresponding to the real networks are above the diagonal for the same small perturbations. Contrarily to these dynamics, random Poissonian and scale-free networks dynamics have been shown to be ordered, and lie below the diagonal (Aldana, 2003; Aldana and Cluzel, 2003). Each point in the plots is the average over 1,000 random rule assignments for 100 random initial conditions (increasing these numbers has no effect on the results). The dashed diagonal, $H(0) = H(1)$, represents the edge between order and chaos. Hamming distances are normalised by the number of nodes.
8.5 Discussion

The results reported in the present chapter include dynamical characteristics of networks identified at different times post-LTP. These results reinforce the view by which the architecture of the networks is under a selective pressure. Yet, it remains unclear the contribution of the structural properties to the overall robustness of biological circuits. This tendency towards the stable regime represents only one mechanism yielding robust behaviour and does not rule out other genetic mechanisms (Wagner, 2005).

While it is clear that late-LTP is dependant on new protein synthesis, it is likely that late-LTP in vivo relies both on local translation of mRNA located in the synapses and on de novo transcription. It has been suggested that the sub-processes requiring nuclear transcription in LTP could be involved in neural maintenance, while translation would confer synaptic specificity (Vickers, Dickson, and Wyllie, 2005). Although a more detailed picture of the mechanisms that serve as a trigger for the rapid transcription of the immediate early genes (IEGs) is needed, the key regulators have been pinpointed (see Section 3.3.6 for a detailed review). Contrarily, little is known regarding the regulation of gene expression in the maintenance phase of LTP.
Chapter 8. Stability of LTP networks using Boolean models

Figure 8.6: Derrida plots for the knock-out experiments. The plots show the initial Hamming distances vs the Hamming distances after 5 iterations. Red and blue lines correspond to the knock-out of the two genes with highest degree.

It seems that the protein synthesis-dependent first hours of LTP are maintained by pre-existing mRNA. However, it has been noted that there is a critical temporal window after LTP induction in which a rapid nuclear response takes place (Nguyen et al., 1994). This early phase is in the order of minutes, and is characterised by a rapid upregulation of gene expression which persists for at least 5 h. The network identified 20 min after induction represents these early response genes. The set of genes identified 5 h post-LTP induction are closely related to the 20 min early responding genes, as demonstrated by the expression profiles (Ryan et al., 2012). The nature of this rapid transcriptional response following LTP induction suggests that the underlying mechanisms are facilitating a switch-like response. In this line, Saha, Wissink, Bailey, Zhao, Fargo, Hwang, Daigle, Fenn, Adelman, and Dudek (2011) documented recently the presence of stalled RNA polymerase II in LTP IEGs, which they interpreted as a mechanism for the rapid neuronal induction observed. However, other mechanisms may be acting jointly at different levels to complement the gene expression trigger.

The 24 h post-LTP induction represents a temporal and functionally different data set, as indicated both by the overlap in gene expression (Ryan et al., 2012) as well as by the fact that mRNA-synthesis inhibitors are only effective in blocking LTP starting after about 4-6 h (Vickers...
Figure 8.7: Effect of a small perturbation, $H(0) = 1$, over 5 time steps for the LTP and yeast networks. Initial states differing only on one position are sampled during 5 subsequent time steps for each LTP network. As shown in Figure 8.1, this small difference tends to be amplified in these biological networks. The latest network appearing after LTP induction (24 h, blue), shows a less pronounced tendency to extend the perturbation. This result is consistent throughout all the initial perturbations in the chaotic regime shown in Figure 8.1

...et al., 2005). We hypothesise that if the 24 h network is representative of a new homeostatic state brought about by LTP induction, its architecture should confer an enhanced stability.

Interestingly, the temporal effect on the vulnerability of the networks discussed is mirrored by what is known about the vulnerability of LTP and memory itself. Previous studies have shown that LTP can be reversed within hours of induction, but then becomes resistant to reversal. It is of particular relevance to our studies that this resistance to reversal is dependent on new protein synthesis. Thus our new data support the conclusion that the LTP-related gene networks contribute to the stabilisation of LTP.

In light of this it is interesting to consider how on-going potential for plasticity in neural networks is maintained over time, if LTP or memories become stable. Such on-going plasticity is crucial within the central nervous system. And it is apparent that memories are at least in part labile. Interestingly, LTP does appear to remain vulnerable to reversal in some instances, for example, seemingly stable LTP in the dentate gyrus can be reversed by exposure to enriched environments up to two weeks post-induction (Abraham et al., 2002). The molecules within the most significant network formed at 24 h post-LTP, and that identified as the most stable network here, relate to growth and/or neurogenesis, functions relevant to consolidation of plasticity. In contrast, the next most significant gene network contains molecules relating to histone modification and epigenetic control of gene expression, and is less stable in nature. Potentially, the vulnerability of this network to change may contribute to maintenance of on-going plasticity in the hippocampus. In summary, as persistent LTP is associated with extensive alteration in gene networks our observations suggest that, as LTP remains plastic so to will aspects of the associated genomic response.

While RBN modelling for gene networks is fraught with the disadvantages of the simplifying assumptions of the model, a number of studies have shown the validity of these simple models to study some biological properties of living systems (see for example Kauffman, 1993; Li et al., 2004; Davidich and Bornholdt, 2008) as well as in other fields such as neural networks, social
networks and protein interaction networks (Aldana, 2003). In the present work we show that while LTP-related networks possess a dynamic behaviour expected from biological networks, some characteristics are specific to their sub-functions. In fact, while the networks studied are all regulated by the same cellular mechanism, the architectures seem to confer a higher stability to the latest network (24 h).

These results are in line with a model in which the rapidly induced networks following LTP induction exhibit a higher sensitivity to perturbations so that a switch-like response can be triggered in response to the signals that induce LTP in the neurons. In turn, networks associated with the late phase of LTP have to possess a more stable architecture, contributing to the homeostatic response underpinned by gene expression. Processes acting longer time scales related with the 24 h network identified by co-expression analysis and differential expression analysis are associated with growth and/or neurogenesis via microRNA epigenetic regulation.

Although analysis of the IPA generated networks has provided validated and biologically meaningful data (Ryan et al., 2011, 2012; Joilin et al., 2014), some limitations are inherent to the methodology. First, potential key interactions may be excluded as the interactions of only 35 genes per network have been considered. Secondly, the architecture of each network is directly dependent upon the information contained within the IPA Knowledge base, a manually curated database, which makes these networks susceptible to false negatives. Thirdly, genes that are modestly but consistently regulated at each time point will be excluded if they do not reach the inclusion criteria at any time. Finally, the analysis is incompatible with the identification of a control network, allowing the characterization of a pre-LTP homeostatic state, as networks are based on differentially expressed genes. The connectivity distributions observed in the co-expressin networks further support the hypothesis that architectural properties on the LTP networks become more stable with time.

Taken together, our results support the idea that networks regulated at different levels possess different dynamic characteristics adjusted to their respective time frames.

The results described in this chapter have been published in the journal Frontiers in Molecular Neuroscience (Nido, Ryan, Benuskova, and Williams, 2015).
As we have discussed in previous chapters, at least three distinct phases have been identified following high-frequency stimulus (HFS) required for LTP. It is unclear how these phases are interrelated, and some evidence seems to suggest that while the underlying mechanisms are distinct, they overlap in time (Park et al., 2014). This should come as no surprise if we recognize that neurons capable of modifying synaptic efficacy for long periods of time are to overcome the limitations imposed by protein and mRNA half-life (in the scale of minutes or hours, Price, Guan, Burlingame, Prusiner, and Ghaemmaghami, 2010; Schwanhäusser, Busse, Li, Dittmar, Schuchhardt, Wolf, Chen, and Selbach, 2011). Protein and mRNA turnover has to be surmounted by a machinery capable of transmitting information across different levels of hierarchical associations which operate at different time scales.

The interplay between mechanisms acting at different levels of organisation and time scales ensures that neurons undergoing LTP can robustly perform their functions and maintain homeostasis. Gene expression changes, in particular, are known to be necessary for the maintenance of late-LTP. As gene expression is regulated at different times following the stimulation necessary to trigger LTP, we have aimed to study the characteristics of these genomic networks taking advantage of the availability of microarray data. The use of computational models has allowed us to further characterise the biological networks involved in LTP and delve deeper into the complexity of this form of synaptic plasticity.

The first integrative checkpoint of the signal elicited by HFS is represented by the putative role of CaMKII as a molecular switch. The capacity of autophosphorylation demonstrated for CaMKII may be sufficient to turn the transient change in intracellular Ca$^{2+}$ caused by the neurotransmitter release from the presynaptic cell into a longer-lasting signal (see Section 3.3.1). This switch sets off the molecular changes needed to increase the post-synaptic sensitivity to subsequent stimuli. The local protein pools suffice to carry out these early changes. However, as we have discussed previously, protein synthesis becomes necessary to enter the longer-lasting late-LTP phase. Protein
turnover is overcome by replenishing the protein pools with newly-synthesised molecules.

In parallel, the signal is transmitted to the nucleus through molecular signalling cascades involving other proteins (PKA, PKC, CaMKIV). These signals seem to converge in the activation of inducible transcription factors that act on SRE- and CRE-containing genes and the subsequent induction of gene expression. Thus the original signal crosses over to an additional level. Furthermore, the signal driving the synaptic consolidation of LTP spreads to influence the morphology of the cell, inducing morphological changes in the stimulated synapses. These changes can be seen under the electron microscope (Yuste and Bonhoeffer, 2001).

This picture gives an idea of the different levels and complexity of the hierarchical associations involved in LTP at the cellular level. Furthermore, LTP only constitutes a model for the cellular mechanism that may underlie long-term memories, and it is becoming clear that LTP comprises rather a family of different processes by which neurons integrate and process the information to change their synaptic weights. For example, it has been argued recently that in the Schaffer collateral-commissural pathway at least three mechanistically different forms of synaptic plasticity co-exist, all NMDAR-dependent (Park et al., 2014). These forms can overlap partially in time, and the combination of these processes can increase their functional utility.

The genomic component of LTP operates within a single neuron. Yet, it only represents one tier of data processing. This tier is “nested” on the another tier of computation constituted by the neural networks, ultimately responsible for the consolidation of memories and learning. This “nested” network architecture is, in reality, a highly modular network recollective of the fractal structure found in many naturally occurring structures (Song, Havlin, and Makse, 2005). Subsequent hierarchically higher levels have been described in cortical connectivity, in which the local neural circuits and columns are connected via nerve fiber projections between brain areas (Scannell, Burns, Hilgetag, O’Neil, and Young, 1999).

From this perspective, LTP represents the experimental representation of a cellular mechanism which acts at the interneuronal level by changing the strengths of the interactions (black arrows in Figure 9.1). The decision, however, is the result of the integration of the incoming information by the cellular machinery at the molecular level. Interestingly, the brain’s genomic response to learning and other neuronal stimulation is not a new evolutionary trait to mammals, but rather seems to be an ancient process, as evidenced by the fact that gene expression changes have been documented, for a number of experimental organisms and tissues (Clayton, 2000).

In this context, in the present thesis we analysed general topological properties of the co-expression networks corresponding to different time points (control, 20 min, 5 h, and 24 h). Notably, these networks exhibit a different connectivity distribution (see Figures 6.2 and 6.4). Interestingly, the unstimulated control and the 24 h networks fit to a scale free distribution fairly well while on the contrary, the co-expression networks corresponding to the 20 min and 5 h datasets have a truncated distribution. As higher robustness is expected of scale free distributions (Albert et al., 2000), this observation is consistent with the notion that cellular networks involved in the homeostatic states (pre- and post-LTP induction) are more stable, while transient topological rearrangements are characteristic of intermediate networks (20 min and 5 h).

Moreover, using the RBN paradigm we have shown that the network derived at 24 h exhibited an enhanced stability when compared to those derived at earlier times post-LTP. This temporal effect on the vulnerability of the networks is mirrored by what is known about the vulnerability of LTP and memory.

In summary, we have presented a view of LTP as a biological process in which a transient signal sets a new homeostatic state that is “remembered” by the cellular systems. Central to this process is the regulation by gene expression, in which the central role played by the Egr transcription factors after LTP induction was highlighted by differential expression and co-expression analyses. In addition, we found a rapid enrichment in connectivity at 20 min followed by a systematic decrease.
This observation provides a potential explanation for the down-regulation of gene expression at 24 h documented by previous studies. From a systems perspective, we have provided evidence that these networks will show less stable architecture, while networks recruited later will exhibit increased stability, consistent with the fact they are more directly related to LTP consolidation. The architecture exhibited by a control and the 24 h LTP co-expression networks fit well to a scale-free distribution, known to be robust against perturbations, whereas the earlier 20 min and 5 h networks showed truncated distributions. Moreover, using random Boolean network simulations we have shown that the network derived at 24 h exhibited an enhanced stability when compared to those derived at earlier times post-LTP. This temporal effect on the vulnerability of the networks is mirrored by what is known about the vulnerability of LTP and memory. Taken together, these results suggest that a new homeostatic state is achieved 24 h post-LTP, and defines an integrated view of the genomic response following LTP induction by which the stability of the networks regulated at different times parallel the properties observed at the synapse.

Figure 9.1: Network representation of a two-level nested network. The molecular species (proteins, RNA, genes and other metabolites) are represented as small nodes. They are organised within a neuron (dashed circles) and they serve as a processing unit. Some nodes will act as sensors or inputs (blue nodes) of information from other neurons. The information is then processed and an output sent to other neurons (red nodes). The nested networks inside the neurons can be different according to the neuron types and regions of the brain, represented as light red and light green in the figure. In this model, LTP would act changing the weight of the connections between neurons (black arrows).


References


References


References


References


McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3), 419.


References


### A.1 List of differentially expressed genes

#### A.1.1 20 min post-LTP

Top 50 differentially expressed genes, $p$-value $< 0.05$, $-0.15 < \ln(\text{Fold Change}) < 0.15$

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>p-value</th>
<th>ln(Fold Change)</th>
<th>Gene Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1390166_at</td>
<td>8.80E-05</td>
<td>3.80E-01</td>
<td>Syt2</td>
</tr>
<tr>
<td>1369053_at</td>
<td>1.50E-03</td>
<td>2.83E-01</td>
<td></td>
</tr>
<tr>
<td>1382848_at</td>
<td>3.43E-03</td>
<td>-2.49E-01</td>
<td></td>
</tr>
<tr>
<td>1369803_at</td>
<td>3.74E-03</td>
<td>3.12E-01</td>
<td>Ptf1a</td>
</tr>
<tr>
<td>1387313_at</td>
<td>4.30E-03</td>
<td>2.77E-01</td>
<td>Myoc</td>
</tr>
<tr>
<td>1368657_at</td>
<td>4.67E-03</td>
<td>2.90E-01</td>
<td>Mmp3</td>
</tr>
<tr>
<td>1381302_at</td>
<td>5.22E-03</td>
<td>2.21E-01</td>
<td></td>
</tr>
<tr>
<td>1384726_at</td>
<td>5.28E-03</td>
<td>-2.07E-01</td>
<td></td>
</tr>
<tr>
<td>1398033_at</td>
<td>5.44E-03</td>
<td>3.27E-01</td>
<td></td>
</tr>
<tr>
<td>1375816_at</td>
<td>5.46E-03</td>
<td>-3.10E-01</td>
<td></td>
</tr>
<tr>
<td>1382084_at</td>
<td>5.75E-03</td>
<td>3.40E-01</td>
<td></td>
</tr>
<tr>
<td>1391508_at</td>
<td>5.96E-03</td>
<td>2.56E-01</td>
<td></td>
</tr>
<tr>
<td>1396041_at</td>
<td>6.14E-03</td>
<td>3.15E-01</td>
<td></td>
</tr>
<tr>
<td>1382360_at</td>
<td>6.41E-03</td>
<td>2.36E-01</td>
<td>Fzd5</td>
</tr>
<tr>
<td>1398162_at</td>
<td>6.55E-03</td>
<td>2.31E-01</td>
<td></td>
</tr>
<tr>
<td>1387974_a_at</td>
<td>6.76E-03</td>
<td>2.10E-01</td>
<td>Slc21a4</td>
</tr>
<tr>
<td>1380224_at</td>
<td>7.13E-03</td>
<td>2.02E-01</td>
<td>Eftud2</td>
</tr>
<tr>
<td>1370203_at</td>
<td>7.44E-03</td>
<td>2.22E-01</td>
<td>Pgap2</td>
</tr>
<tr>
<td>1381214_at</td>
<td>7.45E-03</td>
<td>2.22E-01</td>
<td></td>
</tr>
<tr>
<td>1378870_at</td>
<td>7.95E-03</td>
<td>2.24E-01</td>
<td></td>
</tr>
<tr>
<td>1393260_at</td>
<td>7.99E-03</td>
<td>2.19E-01</td>
<td></td>
</tr>
<tr>
<td>1389645_at</td>
<td>8.05E-03</td>
<td>3.43E-01</td>
<td>Prodh2</td>
</tr>
<tr>
<td>1368707_at</td>
<td>8.18E-03</td>
<td>-2.87E-01</td>
<td>Iih4</td>
</tr>
<tr>
<td>1377292_at</td>
<td>8.21E-03</td>
<td>2.59E-01</td>
<td></td>
</tr>
<tr>
<td>1386753_at</td>
<td>8.23E-03</td>
<td>2.87E-01</td>
<td>Trim14</td>
</tr>
<tr>
<td>1397923_at</td>
<td>8.38E-03</td>
<td>2.01E-01</td>
<td></td>
</tr>
<tr>
<td>1395033_at</td>
<td>9.18E-03</td>
<td>2.28E-01</td>
<td></td>
</tr>
<tr>
<td>1398504_at</td>
<td>9.34E-03</td>
<td>2.41E-01</td>
<td></td>
</tr>
<tr>
<td>1380432_at</td>
<td>9.35E-03</td>
<td>2.26E-01</td>
<td>Cmah</td>
</tr>
<tr>
<td>1394127_at</td>
<td>9.53E-03</td>
<td>2.22E-01</td>
<td></td>
</tr>
<tr>
<td>1397256_at</td>
<td>9.55E-03</td>
<td>2.37E-01</td>
<td></td>
</tr>
<tr>
<td>1376083_at</td>
<td>9.82E-03</td>
<td>2.12E-01</td>
<td>Memo1</td>
</tr>
<tr>
<td>1378799_at</td>
<td>1.00E-02</td>
<td>-3.22E-01</td>
<td></td>
</tr>
<tr>
<td>1373259_at</td>
<td>1.03E-02</td>
<td>2.32E-01</td>
<td></td>
</tr>
</tbody>
</table>
### A.1.2 5 h post-LTP

Top 50 differentially expressed genes, $p$ - value $< 0.05$, $-0.15 < \ln(\text{Fold Change}) < 0.15$

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>$p$ - value</th>
<th>$\log_2(\text{Fold Change})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lrba</td>
<td>5.67E-04</td>
<td>2.56E-01</td>
</tr>
<tr>
<td>Plaur</td>
<td>4.0E-03</td>
<td>2.41E-01</td>
</tr>
<tr>
<td>LOC679811</td>
<td>1.07E-03</td>
<td>2.68E-01</td>
</tr>
<tr>
<td>Homer1</td>
<td>1.73E-03</td>
<td>2.45E-01</td>
</tr>
<tr>
<td>Psmd11</td>
<td>1.76E-03</td>
<td>2.58E-01</td>
</tr>
<tr>
<td>Rgds11</td>
<td>2.07E-03</td>
<td>3.56E-01</td>
</tr>
<tr>
<td>Fbxl6</td>
<td>2.14E-03</td>
<td>2.91E-01</td>
</tr>
<tr>
<td>Lbd3</td>
<td>2.20E-03</td>
<td>3.69E-01</td>
</tr>
<tr>
<td>Rert1</td>
<td>2.22E-03</td>
<td>3.04E-01</td>
</tr>
<tr>
<td>RGD1559904</td>
<td>2.36E-03</td>
<td>4.69E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.44E-03</td>
<td>3.54E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.45E-03</td>
<td>3.14E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.45E-03</td>
<td>1.86E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.46E-03</td>
<td>4.69E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.47E-03</td>
<td>3.54E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.47E-03</td>
<td>3.14E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.48E-03</td>
<td>1.86E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.49E-03</td>
<td>4.69E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.50E-03</td>
<td>3.54E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.51E-03</td>
<td>3.14E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.52E-03</td>
<td>1.86E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.53E-03</td>
<td>4.69E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.54E-03</td>
<td>3.54E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.55E-03</td>
<td>3.14E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.56E-03</td>
<td>1.86E-01</td>
</tr>
<tr>
<td>Rgds1</td>
<td>2.57E-03</td>
<td>4.69E-01</td>
</tr>
</tbody>
</table>
A.2 Top Gene Ontology terms for co-expression networks

Table A.4: Top Gene Ontology terms for each of the modules ($p < 0.01$)

<table>
<thead>
<tr>
<th>Module</th>
<th>GO Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>black_20</td>
<td>positive regulation of receptor-mediated endocytosis</td>
</tr>
<tr>
<td></td>
<td>regulation of reactive oxygen species metabolic process</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Module</th>
<th>GO Term</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>regulation of hydrolase activity</td>
</tr>
<tr>
<td></td>
<td>negative regulation of transcription by competitive promoter binding</td>
</tr>
<tr>
<td></td>
<td>energy derivation by oxidation of organic compounds</td>
</tr>
<tr>
<td></td>
<td>ribonucleoside metabolic process</td>
</tr>
<tr>
<td>black_24</td>
<td>morphogenesis of a polarized epithelium</td>
</tr>
<tr>
<td></td>
<td>establishment of planar polarity</td>
</tr>
<tr>
<td></td>
<td>establishment of tissue polarity</td>
</tr>
<tr>
<td></td>
<td>post-anal tail morphogenesis</td>
</tr>
<tr>
<td></td>
<td>acetyltransferase activity</td>
</tr>
<tr>
<td></td>
<td>canonical Wnt receptor signaling pathway</td>
</tr>
<tr>
<td>black_5H</td>
<td>coated pit</td>
</tr>
<tr>
<td></td>
<td>anion homeostasis</td>
</tr>
<tr>
<td></td>
<td>neuromuscular junction</td>
</tr>
<tr>
<td></td>
<td>sarcolemma</td>
</tr>
<tr>
<td></td>
<td>synapse assembly</td>
</tr>
<tr>
<td></td>
<td>cholesterol metabolic process</td>
</tr>
<tr>
<td>black_U</td>
<td>negative regulation of tissue remodeling</td>
</tr>
<tr>
<td></td>
<td>membrane budding</td>
</tr>
<tr>
<td></td>
<td>interleukin-1 secretion</td>
</tr>
<tr>
<td></td>
<td>regulation of interleukin-1 secretion</td>
</tr>
<tr>
<td></td>
<td>protein transport</td>
</tr>
<tr>
<td></td>
<td>response to nitric oxide</td>
</tr>
<tr>
<td>blue_20</td>
<td>vitamin metabolic process</td>
</tr>
<tr>
<td></td>
<td>positive regulation of leukocyte activation</td>
</tr>
<tr>
<td></td>
<td>positive regulation of cell activation</td>
</tr>
<tr>
<td></td>
<td>lipid transporter activity</td>
</tr>
<tr>
<td></td>
<td>water-soluble vitamin metabolic process</td>
</tr>
<tr>
<td></td>
<td>secretion of lysosomal enzymes</td>
</tr>
<tr>
<td>blue_24</td>
<td>regulation of endocrine process</td>
</tr>
<tr>
<td></td>
<td>endocrine hormone secretion</td>
</tr>
<tr>
<td></td>
<td>retin development in camera-type eye</td>
</tr>
<tr>
<td></td>
<td>cellular response to corticosteroid stimulus</td>
</tr>
<tr>
<td></td>
<td>response to estrogen stimulus</td>
</tr>
<tr>
<td></td>
<td>steroid catabolic process</td>
</tr>
<tr>
<td>blue_5H</td>
<td>cellular amino acid biosynthetic process</td>
</tr>
<tr>
<td></td>
<td>vacuolar membrane</td>
</tr>
<tr>
<td></td>
<td>aspartate family amino acid catabolic process</td>
</tr>
<tr>
<td></td>
<td>aspartate family amino acid metabolic process</td>
</tr>
<tr>
<td></td>
<td>positive regulation of oxidoreductase activity</td>
</tr>
<tr>
<td></td>
<td>vacuole</td>
</tr>
<tr>
<td>blue_U</td>
<td>nucleus</td>
</tr>
<tr>
<td></td>
<td>protein phosphatase type 2A complex</td>
</tr>
<tr>
<td></td>
<td>phosphoric diester hydrolase activity</td>
</tr>
<tr>
<td></td>
<td>DNA binding</td>
</tr>
<tr>
<td></td>
<td>sequence-specific DNA binding RNApolII TF activity</td>
</tr>
<tr>
<td></td>
<td>dephosphorylation</td>
</tr>
<tr>
<td>brown_20</td>
<td>cytoplasmic microtubule</td>
</tr>
<tr>
<td></td>
<td>organic hydroxy compound transport</td>
</tr>
<tr>
<td></td>
<td>Rho guanyl-nucleotide exchange factor activity</td>
</tr>
<tr>
<td></td>
<td>monoamine transport</td>
</tr>
<tr>
<td></td>
<td>microtubule binding</td>
</tr>
<tr>
<td>brown_24</td>
<td>cofactor transporter activity</td>
</tr>
<tr>
<td></td>
<td>heterocyclic compound binding</td>
</tr>
<tr>
<td></td>
<td>modified amino acid binding</td>
</tr>
<tr>
<td></td>
<td>organic cyclic compound binding</td>
</tr>
<tr>
<td></td>
<td>female meiosis</td>
</tr>
<tr>
<td></td>
<td>cofactor transport</td>
</tr>
<tr>
<td>brown_5H</td>
<td>potassium ion transport</td>
</tr>
<tr>
<td></td>
<td>insulin-like growth factor receptor binding</td>
</tr>
<tr>
<td></td>
<td>cell body membrane</td>
</tr>
<tr>
<td></td>
<td>Ras GTPase binding</td>
</tr>
<tr>
<td></td>
<td>response to nutrient</td>
</tr>
<tr>
<td></td>
<td>dendrite cytoplasm</td>
</tr>
<tr>
<td>brown_U</td>
<td>regulation of secretion</td>
</tr>
<tr>
<td></td>
<td>plasma membrane</td>
</tr>
<tr>
<td></td>
<td>regulation of multicellular organismal process</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Module</th>
<th>GO Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>cyan_24</td>
<td>cell periphery</td>
</tr>
<tr>
<td>cyan_24</td>
<td>regulation of system process</td>
</tr>
<tr>
<td>cyan_24</td>
<td>negative regulation of secretion</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>gastric acid secretion</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>early endosome to late endosome transport</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>single-stranded DNA binding</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>negative regulation of pathway-restricted SMAD protein phosphorylation</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>acid secretion</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>DNA repair</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>regulation of DNA methylation</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>negative regulation of IkB kinase/NFkB cascade</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>DNA alkylation</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>DNA methylation</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>negative regulation of viral transcription</td>
</tr>
<tr>
<td>cyan_5H</td>
<td>osteoblast proliferation</td>
</tr>
<tr>
<td>darkgreen_5H</td>
<td>6-phosphofructo-2-kinase activity</td>
</tr>
<tr>
<td>darkgreen_5H</td>
<td>sugar:hydrogen symporter activity</td>
</tr>
<tr>
<td>darkgreen_5H</td>
<td>cation:sugar symporter activity</td>
</tr>
<tr>
<td>darkgreen_5H</td>
<td>phosphofructokinase activity</td>
</tr>
<tr>
<td>darkgreen_5H</td>
<td>cytoplasmic ubiquitin ligase complex</td>
</tr>
<tr>
<td>darkgreen_5H</td>
<td>DNA replication factor A complex</td>
</tr>
<tr>
<td>darkgrey_5H</td>
<td>M/G1 transition of mitotic cell cycle</td>
</tr>
<tr>
<td>darkgrey_5H</td>
<td>negative regulation of exit from mitosis</td>
</tr>
<tr>
<td>darkgrey_5H</td>
<td>regulation of extracellular matrix constituent secretion</td>
</tr>
<tr>
<td>darkgrey_5H</td>
<td>positive regulation of extracellular matrix constituent secretion</td>
</tr>
<tr>
<td>darkgrey_5H</td>
<td>glycine metabolic process</td>
</tr>
<tr>
<td>darkgrey_5H</td>
<td>glycine biosynthetic process</td>
</tr>
<tr>
<td>darkred_5H</td>
<td>chondroitin sulfate metabolic process</td>
</tr>
<tr>
<td>darkred_5H</td>
<td>chondroitin sulfate proteoglycan metabolic process</td>
</tr>
<tr>
<td>darkred_5H</td>
<td>lipid phosphorylation</td>
</tr>
<tr>
<td>darkred_5H</td>
<td>phosphatidylinositol phosphorylation</td>
</tr>
<tr>
<td>darkred_5H</td>
<td>aminoglycan metabolic process</td>
</tr>
<tr>
<td>darkred_5H</td>
<td>glycosaminoglycan metabolic process</td>
</tr>
<tr>
<td>darkturquoise_5H</td>
<td>negative regulation of protein autophosphorylation</td>
</tr>
<tr>
<td>darkturquoise_5H</td>
<td>response to oxidative stress</td>
</tr>
<tr>
<td>darkturquoise_5H</td>
<td>response to hydrogen peroxide</td>
</tr>
<tr>
<td>darkturquoise_5H</td>
<td>regulation of protein autophosphorylation</td>
</tr>
<tr>
<td>darkturquoise_5H</td>
<td>response to reactive oxygen species</td>
</tr>
<tr>
<td>darkturquoise_5H</td>
<td>oxygen transport</td>
</tr>
<tr>
<td>green_20</td>
<td>response to axon injury</td>
</tr>
<tr>
<td>green_20</td>
<td>regulation of interphase of mitotic cell cycle</td>
</tr>
<tr>
<td>green_20</td>
<td>response to calcium ion</td>
</tr>
<tr>
<td>green_20</td>
<td>quaternary ammonium group transport</td>
</tr>
<tr>
<td>green_20</td>
<td>reflex</td>
</tr>
<tr>
<td>green_20</td>
<td>alcohol binding</td>
</tr>
<tr>
<td>green_24</td>
<td>interaction with host</td>
</tr>
<tr>
<td>green_24</td>
<td>regulation of gliogenesis</td>
</tr>
<tr>
<td>green_24</td>
<td>regulation of organ formation</td>
</tr>
<tr>
<td>green_24</td>
<td>adenohypophysis development</td>
</tr>
<tr>
<td>green_24</td>
<td>embryonic skeletal system development</td>
</tr>
<tr>
<td>green_24</td>
<td>regulation of astrocyte differentiation</td>
</tr>
<tr>
<td>green_5H</td>
<td>regulation of amino acid transport</td>
</tr>
<tr>
<td>green_5H</td>
<td>steroid binding</td>
</tr>
<tr>
<td>green_5H</td>
<td>xenobiotic metabolic process</td>
</tr>
<tr>
<td>green_5H</td>
<td>response to xenobiotic stimulus</td>
</tr>
<tr>
<td>green_5H</td>
<td>cellular response to xenobiotic stimulus</td>
</tr>
<tr>
<td>green_5H</td>
<td>gamma-aminobutyric acid secretion</td>
</tr>
<tr>
<td>green_U</td>
<td>outflow tract septum morphogenesis</td>
</tr>
<tr>
<td>green_U</td>
<td>labyrinthine layer morphogenesis</td>
</tr>
<tr>
<td>green_U</td>
<td>embryonic placenta morphogenesis</td>
</tr>
<tr>
<td>green_U</td>
<td>cranial suture morphogenesis</td>
</tr>
<tr>
<td>green_U</td>
<td>craniofacial suture morphogenesis</td>
</tr>
<tr>
<td>green_U</td>
<td>lateral sprouting from an epithelium</td>
</tr>
<tr>
<td>greenyellow_24</td>
<td>intrinsic to mitochondrial outer membrane</td>
</tr>
<tr>
<td>greenyellow_24</td>
<td>integral to mitochondrial outer membrane</td>
</tr>
<tr>
<td>greenyellow_24</td>
<td>superior temporal gyrus development</td>
</tr>
<tr>
<td>greenyellow_24</td>
<td>tau-protein kinase activity</td>
</tr>
</tbody>
</table>

*Continued on next page*
<table>
<thead>
<tr>
<th>Module</th>
<th>GO Term</th>
</tr>
</thead>
</table>
| greenyellow_5H | regulation of myeloid leukocyte differentiation  
|              | glycoprotein binding  
|              | macrophage differentiation  
|              | regulation of macrophage differentiation  
|              | response to calcium ion  
|              | protein tyrosine/threonine phosphatase activity                           |
| grey60_5H   | regulation of branching involved in salivary gland morphogenesis  
|              | cochlea development                                                      |
| lightcyan_24 | regulation of cellular response to stress  
|              | optic nerve development  
|              | optic nerve morphogenesis  
|              | optic nerve structural organization  
|              | regulation of eosinophil differentiation  
|              | interleukin-5 receptor binding                                            |
| lightcyan_5H | pyrimidine nucleotide metabolic process  
|              | pyrimidine nucleoside triphosphate metabolic process  
|              | pyrimidine ribonucleoside triphosphate metabolic process  
|              | pyrimidine ribonucleotide metabolic process  
|              | pyrimidine nucleoside metabolic process  
|              | pyrimidine ribonucleoside metabolic process                             |
| lightgreen_5H | regulation of T cell migration  
|              | positive regulation of T cell migration  
|              | tumor necrosis factor superfamily cytokine production  
|              | T cell migration  
|              | regulation of lymphocyte migration  
|              | positive regulation of lymphocyte migration                             |
| lightyellow_5H | central nervous system neuron axonogenesis  
|              | establishment of nucleus localization  
|              | nucleus localization  
|              | dendritic spine morphogenesis  
|              | dendritic spine organization  
|              | dendritic spine development                                              |
| magenta_20  | pattern binding  
|              | polysaccharide binding  
|              | protein activation cascade                                               |
| magenta_24  | proteasomal protein catabolic process  
|              | cullin-RING ubiquitin ligase complex  
|              | response to virus  
|              | ubiquitin ligase complex  
|              | defense response to virus  
|              | proteasomal ubiquitin-dependent protein catabolic process                |
| magenta_5H  | progesterone receptor signaling pathway  
|              | response to gonadotropin stimulus  
|              | microvillus  
|              | blastocyst development  
|              | epithelial cell differentiation involved in kidney development  
|              | mesenchymal to epithelial transition                                     |
| magenta_U   | integrin binding  
|              | inflammatory response                                                    |
| midnightblue_24 | clathrin-coated endocytic vesicle  
|              | chaperone-mediated protein folding  
|              | mitogen-activated protein kinase binding  
|              | negative regulation of B cell activation  
|              | regulation of lymphocyte differentiation                                |
| midnightblue_5H | negative regulation of cytokine biosynthetic process  
|              | neutral amino acid transmembrane transporter activity  
|              | actinin binding  
|              | alpha-actinin binding  
|              | neutral amino acid transport  
|              | mesodermal cell differentiation                                           |
| pink_20     | calmodulin-dependent protein kinase activity  
|              | protein serine/threonine kinase activity  
|              | neural crest cell migration  
|              | positive regulation of proteasomal ubiquitin-dependent protein catabolic process  
|              | positive regulation of proteasomal protein catabolic process             |

Continued on next page
<table>
<thead>
<tr>
<th>Module</th>
<th>GO Term</th>
</tr>
</thead>
</table>
| pink_24 | Golgi vesicle transport  
|        | fatty-acyl-CoA binding transporter activity  
|        | sensory perception of chemical stimulus  
|        | mannose metabolic process  
|        | cholesterol storage  
|        | plasma lipoprotein particle assembly  
| pink_5H | RNApolII core promoter proximal region sequence-specific DNA binding TF activity involved in negative regulation of transcription  
|        | RNApolII transcription regulatory region sequence-specific DNA binding TF activity involved in negative regulation of transcription  
|        | sexual reproduction  
|        | taurine binding  
|        | immunoglobulin production  
|        | activation-induced cell death of T cells  
| pink_U | mitochondrial transport  
|        | regulation of intrinsic apoptotic signaling pathway  
|        | regulation of mitochondrial membrane permeability  
|        | regulation of release of cytochrome c from mitochondria  
|        | positive regulation of intrinsic apoptotic signaling pathway  
|        | intrinsic apoptotic signaling pathway  
| purple_24 | negative regulation of vascular permeability  
|        | glycogen granule  
|        | midbody  
|        | chemical homeostasis  
|        | regulation of glycogen catabolic process  
|        | cAMP-mediated signaling  
| purple_5H | histone H3-K27 methylation  
|        | regulation of neutrophil chemotaxis  
|        | endothelial cell apoptotic process  
|        | regulation of endothelial cell apoptotic process  
|        | positive regulation of lymphocyte differentiation  
|        | regulation of primary metabolic process  
| red_20 | perinuclear region of cytoplasm  
|        | regulation of axonogenesis  
|        | dendrite development  
|        | regulation of neuron projection development  
|        | regulation of cell projection organization  
|        | cellular component morphogenesis  
| red_24 | septin complex  
|        | septin cytoskeleton  
|        | myeloid leukocyte activation  
|        | interleukin-23 production  
|        | regulation of interleukin-23 production  
|        | negative regulation of interleukin-17 production  
| red_5H | maternal placenta development  
|        | regulation of JNK cascade  
|        | multicellular organismal movement  
|        | musculoskeletal movement  
|        | positive regulation of natural killer cell mediated immune response to tumor cell  
|        | positive regulation of natural killer cell mediated cytotoxicity directed against tumor cell target  
| red_U | organonitrogen compound metabolic process  
|        | regulation of GTP catabolic process  
|        | regulation of GTPase activity  
|        | positive regulation of GTPase activity  
|        | cellular amide metabolic process  
|        | regulation of nucleoside metabolic process  
| royalblue_5H | chromatin remodeling  
|        | carboxypeptidase activity  
|        | cytosolic part  
|        | exopeptidase activity  
| salmon_24 | striated muscle adaptation  
|        | positive regulation of steroid biosynthetic process  
|        | cGMP biosynthetic process  
|        | rRNA processing  
|        | rRNA metabolic process  

Continued on next page
<table>
<thead>
<tr>
<th>Module</th>
<th>GO Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>salmon_5H</td>
<td>positive regulation of steroid metabolic process</td>
</tr>
<tr>
<td></td>
<td>positive regulation of vasoconstriction</td>
</tr>
<tr>
<td></td>
<td>ethanolamine-containing compound metabolic process</td>
</tr>
<tr>
<td></td>
<td>protein targeting to mitochondrion</td>
</tr>
<tr>
<td>tan_24</td>
<td>Golgi stack</td>
</tr>
<tr>
<td></td>
<td>chromosome organization</td>
</tr>
<tr>
<td></td>
<td>macromolecule methylation</td>
</tr>
<tr>
<td></td>
<td>mitotic sister chromatid segregation</td>
</tr>
<tr>
<td></td>
<td>methylation</td>
</tr>
<tr>
<td></td>
<td>chromatin modification</td>
</tr>
<tr>
<td>tan_5H</td>
<td>cellular response to heat</td>
</tr>
<tr>
<td></td>
<td>neuromuscular synaptic transmission</td>
</tr>
<tr>
<td></td>
<td>lipoprotein biosynthetic process</td>
</tr>
<tr>
<td></td>
<td>lipoprotein metabolic process</td>
</tr>
<tr>
<td></td>
<td>MAPK import into nucleus</td>
</tr>
<tr>
<td></td>
<td>RNA import into nucleus</td>
</tr>
<tr>
<td>turquoise_20</td>
<td>oxidoactivity and reductant activity</td>
</tr>
<tr>
<td></td>
<td>negative regulation of defense response</td>
</tr>
<tr>
<td></td>
<td>flavin adenine dinucleotide binding</td>
</tr>
<tr>
<td></td>
<td>regulation of natural killer cell mediated immunity</td>
</tr>
<tr>
<td></td>
<td>positive regulation of natural killer cell mediated immunity</td>
</tr>
<tr>
<td></td>
<td>regulation of histone deacetylation</td>
</tr>
<tr>
<td>turquoise_24</td>
<td>leading edge membrane</td>
</tr>
<tr>
<td></td>
<td>negative regulation of protein catabolic process</td>
</tr>
<tr>
<td></td>
<td>response to external stimulus</td>
</tr>
<tr>
<td></td>
<td>positive chemotaxis</td>
</tr>
<tr>
<td></td>
<td>cellular component movement</td>
</tr>
<tr>
<td></td>
<td>neuron projection membrane</td>
</tr>
<tr>
<td>turquoise_5H</td>
<td>DNA catabolic process</td>
</tr>
<tr>
<td></td>
<td>axoneme</td>
</tr>
<tr>
<td></td>
<td>SCF ubiquitin ligase complex</td>
</tr>
<tr>
<td></td>
<td>TBP-class protein binding</td>
</tr>
<tr>
<td></td>
<td>regulated secretory pathway</td>
</tr>
<tr>
<td></td>
<td>transcription initiation from RNApolII promoter</td>
</tr>
<tr>
<td>turquoise_U</td>
<td>polysaccharide catabolic process</td>
</tr>
<tr>
<td></td>
<td>glycogen catabolic process</td>
</tr>
<tr>
<td></td>
<td>glucan catabolic process</td>
</tr>
<tr>
<td></td>
<td>cellular polysaccharide catabolic process</td>
</tr>
<tr>
<td></td>
<td>polysaccharide metabolic process</td>
</tr>
<tr>
<td></td>
<td>methylated histone residue binding</td>
</tr>
<tr>
<td>yellow_20</td>
<td>transcription from RNApolII promoter</td>
</tr>
<tr>
<td></td>
<td>hormone biosynthetic process</td>
</tr>
<tr>
<td></td>
<td>vascular endothelial growth factor production</td>
</tr>
<tr>
<td></td>
<td>regulation of vascular endothelial growth factor</td>
</tr>
<tr>
<td></td>
<td>regulation of transcription from RNApolII promoter</td>
</tr>
<tr>
<td></td>
<td>positive regulation of protein import into nucleus (translocation)</td>
</tr>
<tr>
<td>yellow_24</td>
<td>proteasomal protein catabolic process</td>
</tr>
<tr>
<td></td>
<td>proteasomal ubiquitin-dependent protein catabolic process</td>
</tr>
<tr>
<td></td>
<td>cell-cell adhesion involved in gastrulation</td>
</tr>
<tr>
<td></td>
<td>negative regulation of gene expression</td>
</tr>
<tr>
<td></td>
<td>transcription cofactor activity</td>
</tr>
<tr>
<td></td>
<td>negative regulation of macromolecule biosynthetic process</td>
</tr>
<tr>
<td>yellow_5H</td>
<td>chromatin DNA binding</td>
</tr>
<tr>
<td></td>
<td>thyroid hormone receptor activator activity</td>
</tr>
<tr>
<td></td>
<td>positive regulation of response to interferon-gamma</td>
</tr>
<tr>
<td></td>
<td>positive regulation of interferon-gamma-mediated signaling pathway</td>
</tr>
<tr>
<td></td>
<td>electron carrier activity</td>
</tr>
<tr>
<td></td>
<td>androgen metabolic process</td>
</tr>
<tr>
<td>yellow_U</td>
<td>BRCA1-A complex</td>
</tr>
<tr>
<td></td>
<td>photoreceptor cell maintenance</td>
</tr>
<tr>
<td></td>
<td>positive regulation of response to DNA damage stimulus</td>
</tr>
<tr>
<td></td>
<td>visual perception</td>
</tr>
<tr>
<td></td>
<td>ion channel binding</td>
</tr>
<tr>
<td></td>
<td>calcium-dependent phospholipid binding</td>
</tr>
</tbody>
</table>