Regulation of gene expression requires multiple levels of control

Since Mendel first discovered the inheritance of phenotypic traits, the definition of a 'gene' has been constantly evolving. A gene is now defined as a stretch of DNA encoding a protein or RNA which has a functional role in an organism (Brosius, 2009). The methods of regulating genes are numerous and complex, and can be roughly divided according to the spatial and temporal characteristics of the control; Pre-transcriptional, post-transcriptional and post-translational.

Physical changes to the DNA sequence can allow regulation; hypermutable regions of immunoglobulin genes are a well studied example. They rely on recombination to produce millions of sequence permutations allowing antibodies to be specific to every conceivable foreign antigen.

DNA and Chromatin methylation are both examples of epigenetic control of gene expression (Jaenisch and Bird, 2003). DNA methylation in vertebrates occurs by the transfer of a methyl moiety to either the 5^{th} carbon position of the cytosine pyramidine ring or the 6th nitrogen of the adenine purine ring. The methylation results in genomic imprinting; in early development gene expression is split between genes derived from maternal and paternal sources. The addition of the methyl group to cytosine effectively prevents translation of these sequences. It is often the case that transcription enhancers in the maternal genome are silenced by methylation to prevent malignant growth before fertilisation. Chromatin modification has been most widely studied in the Hox loci (Saha et al., 2006). Hox genes are so named after the 180 base pair homeobox, which encodes transcription factors responsible for body patterning in many species. The temporal and spatial co-linear characteristics of the expression are partly due to the function of Polycomb group (PcG) and Trithorax (Trx) proteins. PcGs causes a selective trimethylation of the lysine 27 residue of histone 3 which causes repression of the specific Hox gene (Kerppola, 2009). Trx cause the same modification to Lysine 4 of Histone 3 which leads to activation of the modified region (Kohler and Aichinger, 2010). It has now been shown that these modifications are regulated across the whole Hox gene cluster to provide a

'transcriptional window' which is related to relative position within the embryo (Soshnikova and Duboule, 2009, Soshnikova and Duboule, 2008). It is interesting to note that the high frequency of modifications within a small area allows them to be passed on during mitosis, ensuring that the tight regulation is maintained even during proliferation (Chang et al., 2002).

The most well studies, but perhaps most diverse, method of control of gene expression is via DNA binding proteins; transcription factors work in combination to selectively potentiate or depress gene expression in response to intracellular and extra cellular signals. Basal transcription factors are those which, in combination with RNA polymerase II, initiate gene transcription at a particular sequence (Sikorski and Buratowski, 2009). Core promoters are required for initiation of transcription; over 50% of these contain a TATA-box. The TATA box is a highly conserved sequence 20-25 base pairs upstream of the transcription start site that has a high Adenine and Thymine which, due to only having two hydrogen bonds, allows for easier unwinding when bound by the TATA-binding protein (TBP) (Hsu et al., 2008). The initiation of transcription by basal factors allows for transcription factor complexes to form. These complexes allow specificity over what genes become activated. The protein-protein interactions which occur are also highly regulated. The ASPP family of p53 regulators are rare in the fact they interact with the DNA binding regions of p53. Whereas other binding partners can influence the shape of p53 to cause differential binding to DNA, the ASPP family can directly allow or prevent interactions with promoters by blocking or facilitating a binding site (Trigiante and Lu, 2006).

Although in most cases, regulation of genes occurs in the immediate locale of expression, long range changes can also occur. Insulators are DNA sequences and the proteins that bind them that aid long-range control by either blocking the function of enhancers, or preventing the spread of repressive chromatin modifications (Carter et al., 2002). There are two models proposing how enhancer blockers function; the direct contact model and the tracking model. The direct contact model suggests the loops which are known to form within DNA could either separate a promoter from its

target region or conversely bring them into closer association. The tracker model suggests enhancer blockers actually prevent the tracking of a promoter along the DNA, thus preventing the activation of downstream sequences (Engel et al., 2008). Currently there is evidence which supports both models, and it is likely that both function simultaneous in cells (Carter et al., 2002).

The response to extracellular signals occurs via transcription factor cascades; a signal will be received at the membrane and lead to sequential activation of effectors. A well studied example is during Long term potentiation (LTP) of nervous signals (Ooi and Wood, 2008). The simultaneous presence of glutamate and depolarisation of the membrane will lead to activation of NDMA glutamate receptors; this causes activation of the membrane associated enzyme adenyl cyclase which in turn increases the production of cyclic AMP (cAMP). cAMP bind to the cAMP response element (CRE) of target DNA with the help of the CRE binding protein (CREB) (Kandel, 2001). This cascade of signals allows gene expression to be adapted to the specific local environment of an individual cell.

Regulation doesn't stop after transcription. The mRNA produced by DNA polymerases can undergo a number of changes, either by alternative splicing, or by RNA interference. Alternative splicing is where the composition of exons is varied resulting in different proteins being produced (Keren et al., 2010). In Drosophila Ultrabithorax, a Homeobox gene, undergoes alternative splicing to produce six isoforms which act as transcription factors. The isoforms have different binding affinities to Decapentaplegic, which is needed for patterning of abdominal muscles. Changing the expression of the isoforms along the Anterior-posterior axis will result in a loss of muscle patterning (Reed et al., 2010). RNA interference takes two forms; small interfering RNA's (siRNA)(Elbashir et al., 2001) and micro RNA's (mirRNA)(Lee et al., 1993). Both groups are similar in that they are genomically encoded sequences which are initially transcribed as longer, double stranded, hairpin RNAs before being cleaved by in the DICER enzyme and incorporated into the RNA-induced silencing complex (RISC)

(Jinek and Doudna, 2009). Only one strand, termed the guide strand, is incorporated into the RISC, and when mature are roughly 20-22 nucleotides in length (Thakur, 2004). The difference between siRNA and mirRNA is in target specificity. siRNAs bind with precision to one target only, and thus cause an equally specific down regulation in the corresponding protein. MirRNAs on the other hand tend to have imperfect base pairing, and the result is a less selective down regulation of a number of genes (Pillai et al., 2007). The field of small non coding RNAs is expanding exponentially, and it is likely that their role in development in particular will become more apparent as more research is dedicated to them.

Once mRNA has been translated, the immature proteins often undergo modifications to specify them to a role, or to direct them to a particular cellular compartment. Phosphorylation is common; it is estimated up to 33% of proteins undergo this modification (Jensen, 2006). This can occur directly after translation or later in the cell to change the activity of a protein. This is also key process during LTP; adenyl cyclase activates protein kinases including protein kinase A (PKA) and Calcium and calmodulin dependant kinase II (CaMK II). These proteins phosphorylate AMPA glutamate receptor subunits which are held in intracellular pools, leading to translocation of the subunits to the membrane allowing a greater response to glutamate (Derkach et al., 2007). De-phosphorylation is also essential. The phosphatase Calcineurin is activated in response to GABA signalling (Groth et al., 2003). GABA is the main inhibitory neurotransmitter in the brain, and calcineurin mediates it effect by removing the activating phosphate groups from the AMPA subunits.

Proteins are regularly cycled within cells; thus alongside synthesis, protein breakdown must occur. Ubiquitin is a small protein monomer expressed in every eukaryotic cell; its covalent attachment to another protein is used as a marker for proteasomal degredation in the lysosome (Peng et al., 2003). This breakdown is the final level of regulation of gene expression; controlling the life span of a gene product ensures that gene expression can be temporally modified.

In summary, our understanding of gene regulation has improved greatly in the last decade. Long established mechanisms such as alternate splicing and transcription factor binding have now been joined by newer methods such as RNA interference and Chromatin remodelling.

Words: 1427

Bibliography:

BROSIUS, J. 2009. The fragmented gene. *Ann N Y Acad Sci,* 1178**,** 186-93.

CARTER, D., CHAKALOVA, L., OSBORNE, C. S., DAI, Y. F. & FRASER, P. 2002. Long-range chromatin regulatory interactions in vivo. *Nature Genetics,* 32**,** 623-626.

CHANG, H. Y., CHI, J. T., DUDOIT, S., BONDRE, C., VAN DE RIJN, M., BOTSTEIN, D. & BROWN, P. O. 2002. Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 12877-12882.

DERKACH, V. A., OH, M. C., GUIRE, E. S. & SODERLING, T. R. 2007. Regulatory mechanisms of AMPA receptors in synaptic plasticity. *Nature Reviews Neuroscience,* 8**,** 101-113.

ELBASHIR, S. M., HARBORTH, J., LENDECKEL, W., YALCIN, A., WEBER, K. & TUSCHL, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature,* 411**,** 494-498.

ENGEL, N., RAVAL, A. K., THORVALDSEN, J. L. & BARTOLOMEI, S. M. 2008. Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking. *Human Molecular Genetics,* 17**,** 3021-3029.

GROTH, R. D., DUNBAR, R. L. & MERMELSTEIN, P. G. 2003. Calcineurin regulation of neuronal plasticity. *Biochemical and Biophysical Research Communications,* 311**,** 1159-1171.

HSU, J. Y., JUVEN-GERSHON, T., MARR, M. T., WRIGHT, K. J., TJIAN, R. & KADONAGA, J. T. 2008. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes & Development,* 22**,** 2353-2358.

JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics,* 33**,** 245-254.

JENSEN, O. N. 2006. Interpreting the protein language using proteomics. *Nature Reviews Molecular Cell Biology,* 7**,** 391-403.

JINEK, M. & DOUDNA, J. A. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature,* 457**,** 405-412.

KANDEL, E. R. 2001. Neuroscience - The molecular biology of memory storage: A dialogue between genes and synapses. *Science,* 294**,** 1030-1038.

KEREN, H., LEV-MAOR, G. & AST, G. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics,* 11**,** 345-355.

KERPPOLA, T. K. 2009. Polycomb group complexes - many combinations, many functions. *Trends in Cell Biology,* 19**,** 692-704.

KOHLER, C. & AICHINGER, E. 2010. Antagonizing Polycomb group-mediated gene repression by chromatin remodelers. *Epigenetics,* 5**,** 20-23.

LEE, R. C., FEINBAUM, R. L. & AMBROS, V. 1993. THE C-ELEGANS HETEROCHRONIC GENE LIN-4 ENCODES SMALL RNAS WITH ANTISENSE COMPLEMENTARITY TO LIN-14. *Cell,* 75**,** 843-854.

OOI, L. & WOOD, I. C. 2008. Regulation of gene expression in the nervous system. *Biochemical Journal,* 414**,** 327-341.

PENG, J. M., SCHWARTZ, D., ELIAS, J. E., THOREEN, C. C., CHENG, D. M., MARSISCHKY, G., ROELOFS, J., FINLEY, D. & GYGI, S. P. 2003. A proteomics approach to understanding protein ubiquitination. *Nature Biotechnology,* 21**,** 921-926.

PILLAI, R. S., BHATTACHARYYA, S. N. & FILIPOWICZ, W. 2007. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends in Cell Biology,* 17**,** 118-126.

REED, H. C., HOARE, T., THOMSEN, S., WEAVER, T. A., WHITE, R. A. H., AKAM, M. & ALONSO, C. R. 2010. Alternative splicing modulates Ubx protein function in Drosophila melanogaster. *Genetics,* 184**,** 745-58.

SAHA, A., WITTMEYER, J. & CAIRNS, B. R. 2006. Chromatin remodelling: the industrial revolution of DNA around histones. *Nature Reviews Molecular Cell Biology,* 7**,** 437-447.

SIKORSKI, T. W. & BURATOWSKI, S. 2009. The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology,* 21**,** 344-351.

SOSHNIKOVA, N. & DUBOULE, D. 2008. Epigenetic regulation of Hox gene activation: the waltz of methyls. *Bioessays,* 30**,** 199-202.

SOSHNIKOVA, N. & DUBOULE, D. 2009. Epigenetic Temporal Control of Mouse Hox Genes in Vivo. *Science,* 324**,** 1320-1323.

THAKUR, A. 2004. RNA interference revolution. *Electronic Journal of Biotechnology,* 7**,** 39-49.

TRIGIANTE, G. & LU, X. 2006. ASPPs and cancer. *Nature Reviews Cancer,* 6**,** 217-226.