Drosophila: homeotic (HOM) genes. Antennapedia and Bithorax Complexes (WR lectures 3 and 4)

What is a regulatory mutation?

All cells in an organism contain the same genome - that is, the same genes arranged in the same order along the same chromosomes. This is a generalization - some genes are rearranged within certain cells, but this is the exception rather than the rule. For example, the genes coding for immunoglobulins (antibody molecules) are rearranged in B lymphocytes; this is how novel antibody structures are created in response to new invading organisms or viruses. Also, chromosomes are often broken and rejoined aberrantly in cancer cells. And, of course, germ cells (sperm and eggs) contain only half the number of chromosomes as normal somatic (body) cells.

Nevertheless, most cells contain identical DNA. What makes one type of cell different from another is not the genes they contain but which of these genes they express - i.e. which genes they transcribe into RNA, then translate into protein. A muscle cell differs from a neuron in a great many of its constituent RNAs and proteins. For example, the muscle cell transcribes the genes coding for muscle actin and myosin but not the gene encoding neurofilament; the reverse is true of the neuronal cell. What determines whether a particular gene is transcribed or not? This depends to a large extent on the other proteins that are present in the cell - in particular, proteins that are collectively known as "transcription factors".

Every gene consists of a protein coding sequence, which might be contiguous or broken up into a series of exons and introns, and which begins with a START codon (ATG) and concludes with a STOP codon (TAA, TAG or TGA). Apart from this, a gene must have regulatory sequences associated with it. These are stretches of DNA which do not themselves code for protein but which act as binding sites for RNA polymerase and its accessory molecules as well as a variety of transcription factors. Together, the regulatory sequences with their bound proteins act as molecular switches that determine the activity state of the gene - e.g. OFF or FULL-ON or, more often, something in between. The regulatory sequences include the promoter region together with enhancer elements.

Every gene has a promoter, which is the binding site for the basal transcriptional apparatus - RNA polymerase and its co-factors. This provides the minimum machinery necessary to allow transcription of the gene. The enhancer regions are found at a distance from the promoter, to either the 5' or 3' sides of the gene or within introns. They are typically short stretches of DNA (200bp, say), each made up of a cluster of even shorter sequences (25bp, say) that are the binding sites for a variety of cell- or region-specific transcription factors. Once bound, these transcription factor complexes interact with the basal transcriptional machinery at the promoter to enhance (or sometimes diminish) the transcription rate of the gene. Such interactions are possible because of the flexible nature of DNA which allows the enhancers to come close to the promoter by looping out the DNA in between (see diagram overleaf).
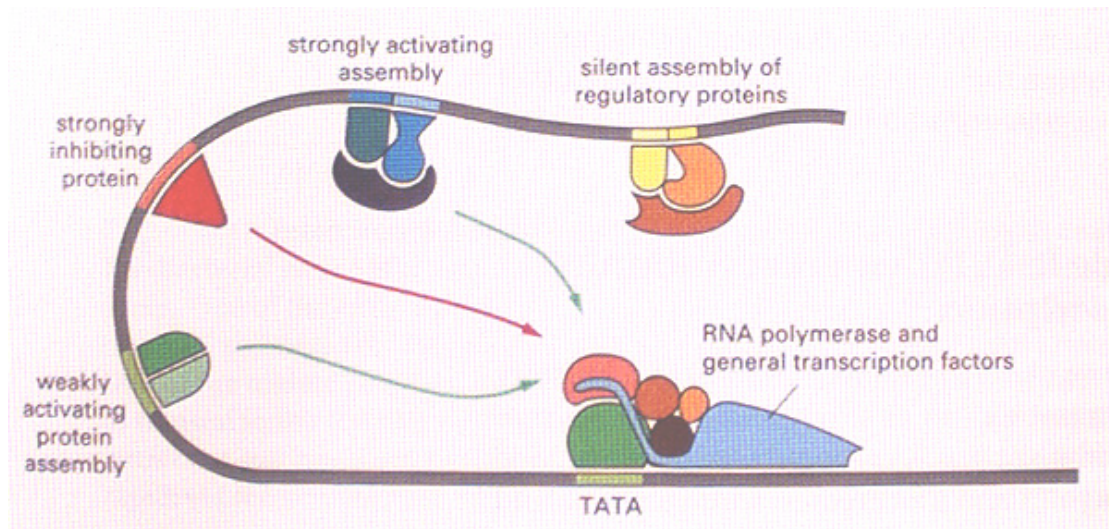
We can think of the activating function of enhancers as follows. Binding of RNA polymerase and the basal transcriptional machinery at the gene promoter is like switching on the engine and allowing it to idle in neutral. When the supplementary transcription factors bound to enhancer elements interact with the basal machinery, it is like putting the engine into gear and pulling away from the kerb. (Alternatively, for a repressor site it is like putting on the handbrake.)

Frequently, a given gene is subject to complex regulation. That is, it might have to be transcribed at different times and in different places during development, or in response to different extracellular stimuli. In the present context (Drosophila embryogenesis) we have seen that the segmentation genes are expressed according to their position in the embryo. An example is the even-skipped (eve) gene, a pair-rule gene that is transcribed in alternate embryonic parasegments to generate a zebra pattern of seven stripes.

The transcriptional state of the eve gene - either ON or OFF according to which parasegment we are in - is under the control of a series of enhancer elements, one for each stripe. Each enhancer element contains binding sites for upstream segmentation gene products such as Bicoid and Kruppel (which, as you will recall, are themselves transcription factors). Thus the particular constellation of maternal effect genes, gap genes and other pair-rule genes that is expressed in a given parasegment determines whether or not one of the enhancer elements is fully occupied and consequently whether eve gene transcription is activated or not in that parasegment. The specificity of the enhancers can be demonstrated by removing just one of them (the element that specifies stripe 2, say) and inserting it upstream of a reporter gene like the bacterial beta-galactosidase (βgal). when this is introduced into the embryo, βgal is expressed in just one stripe - in the position of the eve stripe 2. Alternatively, particular enhancer elements can be deleted from the normal eve gene, resulting in deletion of the corresponding stripes of eve expression. Such a mutation - loss of an enhancer element - is called a regulatory mutation. It affects the spatial or temporal regulation of the gene without causing universal loss of the gene product.
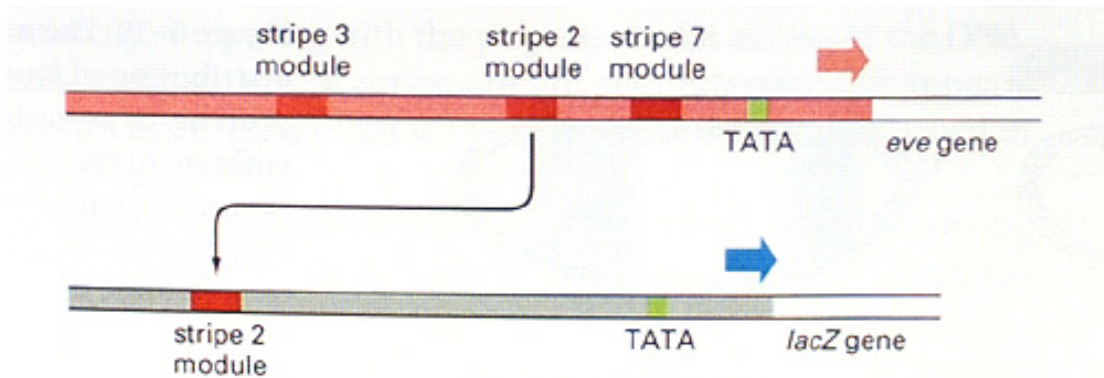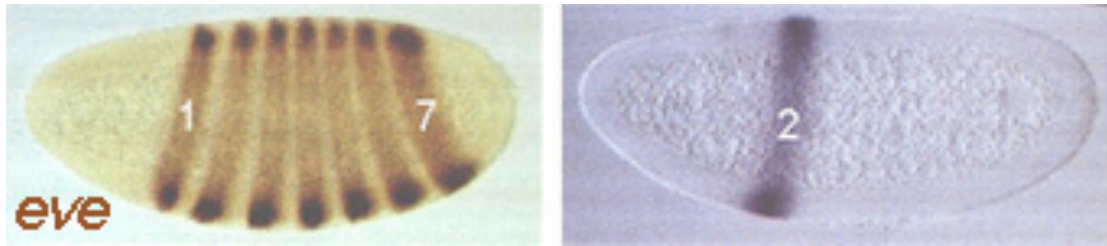
The lower part of this diagram shows part of the regulatory region of the *eve* gene - the part closest to the transcription start site (red arrow). In this region there are three enhancer elements that control *eve* expression in stripes 2, 3 and 7. Other elements, not shown in the diagram, lie further away from the transcription start site (to the left) and control eve expression in stripes 1,4,5 and 6. Taking just one of the elements, that for stripe 2, and linking it to a reporter gene gives the pattern shown above right - stripe 2 only. The wild-type *eve* expression pattern is shown above left. (From Gilbert, Developmental Biology 4th edition Chapter 15, p549 and Alberts et al. Molecular Biology of the Cell 3rd edition, Chapter 9 p428)

Hopeful monsters….
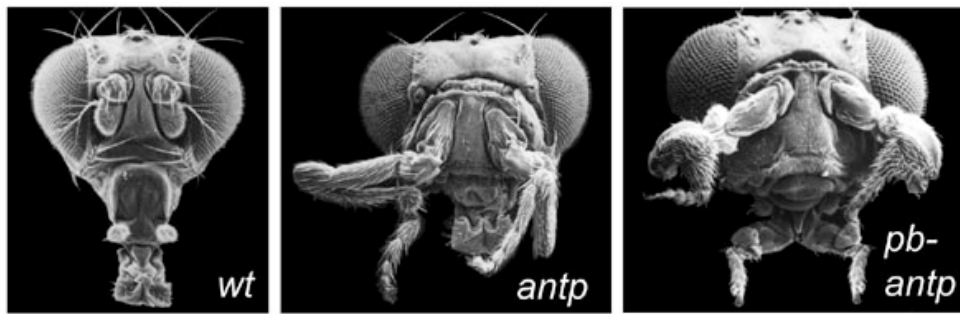
HOMEOTIC TRANSFORMATIONS involving the Antennapedia Complex (ANTC)

Adult fly heads: wild type, antennapedia (antennae transformed to 2nd thoracic legs) and a double mutant proboscipedia/ antennapedia (antennae to 2nd thoracic and proboscis to 1st thoracic legs). Note how "clean" the transformations are ie transformation of one head structure does not much affect adjacent structures. This illustrates how each structure is developmentally isolated from its neighbours - the fly is put together like a construction toy with interchangeable parts.

Adult fly heads: wild type, antennapedia (antennae transformed to 2nd thoracic legs) and a double mutant proboscipedia/ antennapedia (antennae to 2nd thoracic and proboscis to 1st thoracic legs). Note how "clean" the transformations are ie transformation of one head structure does not much affect adjacent structures. This illustrates how each structure is developmentally isolated from its neighbours - the fly is put together like a construction toy with interchangeable parts.

Note that the adult body parts develop from groups of precursor cells that are laid down during embryogenesis but which lie semi-dormant during larval life until their development is triggered again during pupation - when the fly undergoes metamorphosis. These groups of precursor cells are called imaginal discs (the imago is the final stage of an insect e.g. butterfly or adult fruit fly). There is an imaginal disc for each adult structure, e.g. leg disc, wing disc, genital disc, eye/antennal disc, etc. (see Wolpert, chapter 2, p 48 or Gilbert, chapter 18, p566).

Mutant flies like those above were noticed as long ago as the 1890s. Most of the early mutations were dominant - ie only one copy of the mutant gene is required to affect a transformation (e.g. Antennapedia: antenna-to-leg). Later, recessive *Antennapedia* mutations were discovered, which cause the reverse transformation (leg-to-antenna). Many more analogous mutants were isolated that cause transformations of abdominal as well as thoracic and head segments. The mutations were genetically mapped and shown to fall in two gene complexes, the Antennapedia Complex (ANTC) and the Bithorax Complex (BX-C) on separate chromosomes. ANT-C controls the identites of segments anterior to T3, while BX-C controls T3-A8. In the 1970s Ed Lewis undertook a detailed genetic analysis of the BX-C for which he gained a Nobel Prize (with Christiane Nusslein-Volhard and Eric Wieschaus), nearly twenty years later in 1995.

Lewis found that, in general, dominant mutations in the BX-C changed segment identity to that of the next more posterior segment while recessive mutations caused transformations in the opposite direction. For example, the recessive mutation *bithoraxoid (bxd)* causes the first abdominal segment (A1) to take on the character of the third thoracic segment (T3). Since T3 normally carries a pair of legs as well as halteres (balancers), *bxd* mutants have four pair of legs instead of the normal three, as well as two pairs of halteres. Other recessive BX-C genes are *bithorax (bx),* mutations of which convert the anterior half of T3 to the anterior half of T2 (i.e. T3a -> T2a), *postbithorax (pbx)* which converts the posterior half of T3 to the posterior half of T2 (T3p ->T2p). Thus, the double recessive mutant *(bx, pbx)* carries two pairs of wings.


bx, pbx
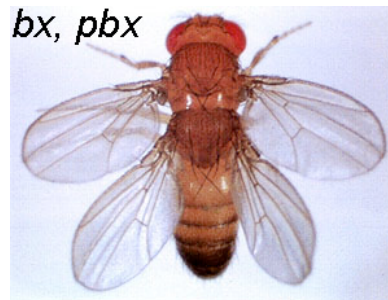
In addition, Lewis identified recessive mutations *infra-abdominal-2 (iab2)*, that changed A2->A1, *iab3* (A3->A2), *iab4* (A4->A3) and so on up to *iab8* (A8->A7). Dominant mutations in the BX-C include *Contrabithorax* (*Cbx*: T2->T3), *Hyper-abdominal (Hab*: A1->A2), *Ultra-abdominal (Uab*: A2->A3) and *Miscadestral pigmentation (Mcp*: A4->A5).

Lewis also noticed that large deletions that took out the entire BX-C converted all the thoracic and abdominal segments to T2, suggesting that T2 represents the "ground state" that is generated when no BX-C genes are active, and that fates more posterior to that require the activation of one or more BX-C genes. Lewis proposed a model whereby an additional BX-C gene is activated in each more posterior segment; no genes are active in T2, *bx+pbx* are active in T3, *bx+pbx+bxd* are active in A1, *bx+pbx+bxd+iab2* are active in A2 and so on until all of the BX-C genes are active in A8. In order to achieve this stepwise activation of BX-C genes, Lewis postulated that there was a gradient of a repressor molecule, highest in T2 (where no BX-C genes are expressed) and lowest in A8 (where all are expressed). Further, he postulated that the promoter region of each gene in the BX-C had a different affinity for the repressor, with iab8 having the highest affinity (and thus remaining repressed at lower concentrations of the putative repressor than any other BX-C gene). This model was consistent with the finding that mutations in *extra sex combs (esc)*, a regulatory gene that lies outside the BX-C locus, can apparently activate all of the BX-C genes in all segments T2-A8, generating a mutant fly in which all segments resemble A8.

This model, based purely on genetics, was a remarkable achievement and contains the essence of the truth. However, when the BX-C was cloned and analyzed it was found to be wrong in one important respect: there are only three transcription units (protein coding units) not nine or ten as Lewis had proposed. These units are named *Ubx, Abd-A* and *Abd-B*, in that order along the chromosome. They are expressed in broad overlapping domains in the embryo, *Ubx* from T3 to A8, *Abd-A* from A2 to A8 and *Abd-B* from A5 to A8. Note that they are expressed in progressively more posterior

domains that correspond to their relative positions along the chromosome. If the *Ubx* protein coding sequence is deleted, segments A1 and T3 develop as T2 (i.e. the sum of Lewis's mutations bx, pbx and bxd). If *Abd-A* is deleted, then T2-A1 develop more-or less normally (judging by the bristles on the larval cuticle) but A2-A4 develop as A1 (the sum of iab2, 3 and 4). If *Abd-B* is deleted, T2-A4 develop normally but A5-A8 develop as A4 (the sum of *iab5, 6,7* and *8*) (see below).
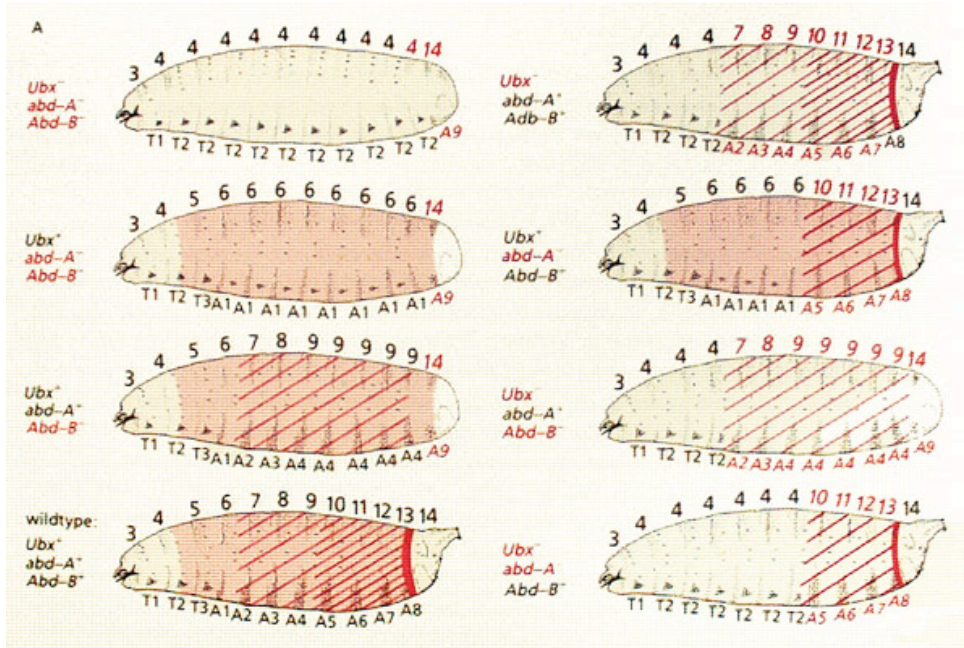
[red text/digits means deleted or abnormal, black means wild-type. The segment designations beneath the larvae refer to the bristle pattern on the cuticle. From "The making of a fly" by Peter A Lawrence (Blackwell Scientific Publications, 1992) Fig 5.3 p117.]

How do *Ubx*, *Abd-A* and *Abd-B* relate to Lewis's mutations? It turns out that most of the mutations Lewis identified do not fall within protein coding sequences, but in regulatory elements that control the expression of Ubx, abd-A and Abd-B. For example, mutation of the cis-acting regulatory control element bxd prevents expression of *Ubx* in segment A1, *iab2* prevents expression of *abd-A* in A2, and mutation of *iab6* abolishes expression of *Abd-B* in A6.
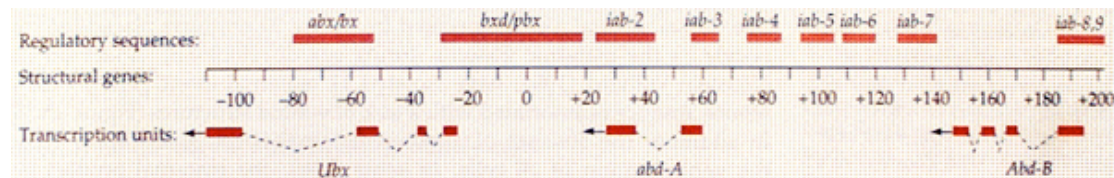
Reading:

1. Wolpert Chapter 5 pp 182-185 (not a very full treatment)
2. Gilbert Chapter 9 pp 283 - 286