Advanced Topics :

Viruses, Cancer, and the Immune System

Using this book: This book is designed to be used in both introductory and advanced cell biology courses. The primary text is generally on the left side of the vertical divider, and printed in black. Details that are usually left to an advanced course are printed in blue and found on the right side of the divider. Finally, additional biomedically relevant information can be found in red print on either side of the divider.

At this point, you should be fairly comfortable with the basic concepts of cell biology. The purpose of this chapter is to build on that basic knowledge and put it together into more complex systems. In addition, we will introduce some more advanced variations on some of the mechanisms and structures that were discussed in earlier chapters. The three topics, viruses, cancer, and immunity, are not only relevant as current news topics, but relate to one another through multiple pathways, which is why they are lumped together.

Viruses

Though a virus has both genetic material and protein components, it is not a living organism. It does not contain the capability of self-replication, and is completely reliant on the cellular biochemistry of whatever host cell it has infected. The minimal definition of a virus is a nucleic acid genome inside of a protein shell, or capsid. There are variations of this, such as virions (infectious viral unit) that have a membrane coat outside of the capsid, or some that have enzymes inside of the capsid alongside the genome. Again, none of these viral variations are able to fully replicate without cellular machinery. The cells that viruses can infect range across most living organisms. There are viruses specific for humans, some that only infect particular animals, some that infect plants, and even viruses that use bacteria as hosts. In current media reports of viral outbreaks in recent years, HIV, avian flu, swine flu, much is made of the origin of the virus with respect to its host/target organisms. However, most viruses are very specific about the cells that they infect. The narrow host range may be not only to particular species, but particular cell types within a particular species. Viruses with a broad host range are relatively rare. However, this does not preclude the possibility of new strains of virus evolving with different host ranges from their ancestral virii.

Viruses may have either RNA or DNA genomes, that may be linear or circular, and single or double -stranded. There are fewer variations of capsid structure. In general, capsids fall into two categories: helical and icosahedral. Helical capsids are actually made up of globular subunits that associate into a helical cylinder, with the genome lying inside

Viroids are also infectious non-living particles, but they are only genetic material, RNA, and have no protein capsid. To date, they are only known to infect plant hosts, and apparently spread by direct or extremely close contact with an infected plant. These infectious pathogenic RNA molecules are single-stranded and circular, and relatively small, roughly 200-400 nucleotides.

an interior groove of the helix (fig. 1c). The icosahedral capsids are also made up of many subunits that together form an approximately 20-sided polygon made from sides that are equilateral (or nearly so) triangles. If you play Dungeons and Dragons™ or know someone who does, then you have probably seen dice (a "d20") of this shape. Of course with capsids, but not dice, there can be some variation in the number of sides, the shape of the triangles, and the number of subunits that make up each face.

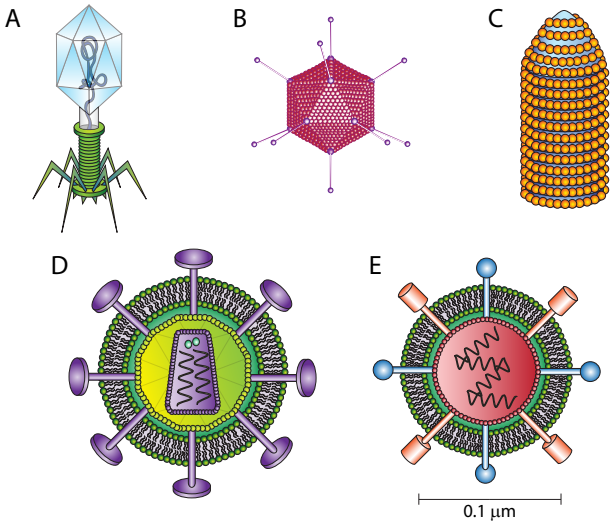

Figure 1. Viruses. (A) a T-4 bacteriophage, (B) adenovirus, (C) tobacco mosaic virus, (D) human immunodeficiency virus, (E) influenza virus.

External to the capsid, some viruses also have a membrane coat (viral *envelope*). As will be more clearly explained soon, this phospholipid bilayer comes from viruses that exit a host cell by exocytosis. Because it came from a host cell, the membrane can be used as a ruse by the virus to fool other potential host cells into misrecognizing the virus as a normal cell or cell debris, based on the receptors that recognize cellular proteins on the membrane. It may then be taken into the cell by receptor-mediated endocytosis in a mistaken attempt to recycle old cell debris, where it can proceed to infect the overgenerous host.

There are two classification systems for viruses, the International Committee on Taxonomy of Viruses (ICTV) has a Linnaean-like taxonomic system based on shared structural or biochemical properties (but not host specificity). Another system, also in use, is the Baltimore classification, in which viruses are classified into seven categories by the mechanism of mRNA production. That is, type I are the dsDNA viruses that make mRNA the "normal" way by direct transcription of the genome, type II are ssDNA viruses that

Viruses in all categories (ICTV class or Baltimore type) that can cause human disease. Some of the major ones are listed here.

Virus	Class	Disease
Adenovirus	Adeno	Febrile respiratory disease
		Pharyngoconjunctival fever
Epstein-Barr	Herpes	Infectious mononucleosis
Hepatitis A	Picorna	Acute hepatitis
Hepatitis B	Hepadna	Acute/chronic hepatitis
		Hepatic cirrhosis
		Hepatocellular carcinoma
Hepatitis C	Flavi	similar to Hepatitis B
Herpes Simplex		Cold sores, pharyngitis
Type 1	Herpes	Gingivostomatitis
Herpes Simplex		Aseptic meningitis
Type 2	Herpes	Genital herpes
HIV	Retro	AIDS
Influenza	Orthomyxo	Influenza
Measles	Paramyxo	Measles
Mumps	Paramyxo	Mumps
HPV	Papilloma	Cervical cancer
		Genital warts
Poliovirus	Picorna	Poliomyelitis
Rabies	Rhabdo	Rabies
Rubella	Toga	German measles

must first make a complementary strand to become dsDNA before transcription, type VI are ssRNA viruses that use reverse transcriptase (detailed later in this section) to first convert the RNA to a DNA intermediate before transcription, and so on.

Lytic "life" cycle of viruses

Viruses can interact with their hosts in two distinct ways: the lytic pathway and the lysogenic pathway. Some viruses are able to switch between the two pathways while others only use one. The distinguishing characteristic of the lytic life cycle is catastrophic death of the host cell by lysis and simultaneous release of viral particles. In figure 2, the stages of the lytic pathway are depicted. In this case, a T4 bacteriophage (the term "phage" is used for bacterial viruses) is used as an example. In step 1, the virus attaches to the cell wall. In step 2, the virus injects its genetic material (dsDNA) into the cytoplasm of the bacteria. In step 3, the viral DNA is being replicated and the genes on the viral DNA are being transcribed and translated into viral proteins. Expression from the host genomic DNA is arrested. In step 4, viruses are assembled from the proteins and DNA. And finally, once the viral factory has used up the cell's energy and material resources in making more viruses, it performs a final coup de grace, as cell is destroyed to free the viruses to exit and find more host cells. The T4 phage used in this example only undergoes this pathway and not the lysogenic pathway.

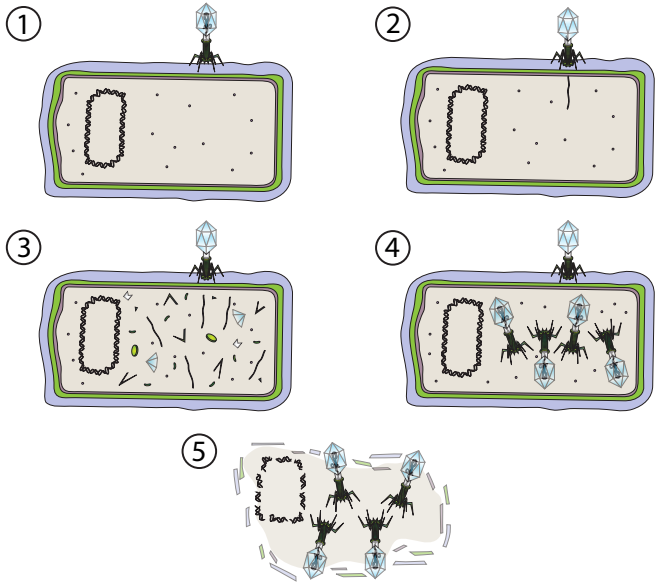

Figure 2. The Lytic Pathway

In eukaryotes, the mechanism is slightly more complicated by the nucleus. The DNA is transported into the nucleus, where the transcription and replication take place. Although the viral mRNA is transported out to the cytoplasm for translation as expected, the resulting capsid proteins are then imported back into the nucleus, where the virion particles are assembled. Lytic plant and animal viruses with RNA genomes can bypass the nucleus altogether, and the genome replication, protein synthesis, and particle assembly all occur in the cytoplasm.

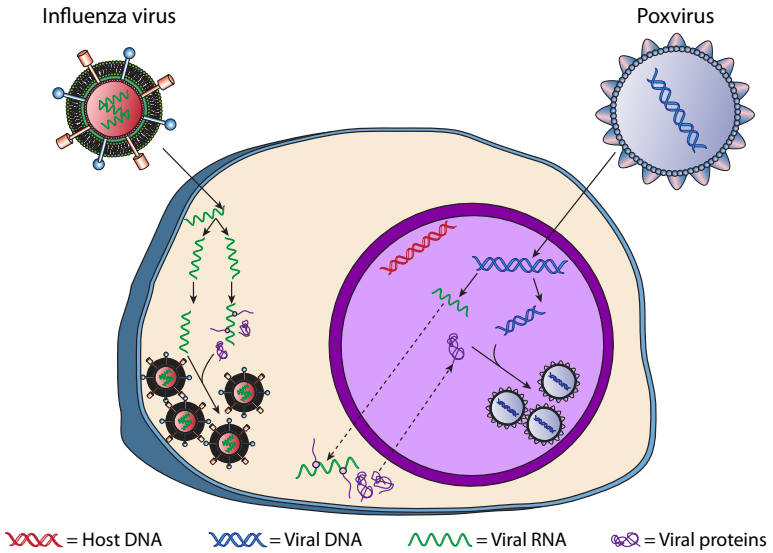

Figure 3. Lytic viruses in eukaryotes: an RNA virus (influenza) replicates completely in the cytoplasm, while the DNA virus (poxvirus) uses both cytoplasm and nucleus.

The lytic pathway can produce a huge number of viral particles between infection and lysis, as many as several tens of thousands, for example from a rabies-infected cell. Therefore, this pathway is well suited for conditions in which potential host cells are plentiful. On the other hand, this is a waste of resources if there are relatively few potential hosts. Imagine a few bacteria that have floated off from the colony: if a phage infected a bacteria in the main colony, commandeering the bacteria to create thousands of viral particles, most of those particles would infect new hosts and make many thousands more soldiers in this viral army. But if the virus infected one of the breakaway bacteria, then once it killed its host by lysis, the viral particles would have few, if any, other potential hosts, and eventually all the viral particles just break down from various environmental conditions. What would a better survival strategy for the virus in such a situation?

The Lysogenic Pathway

A better option for some bacterial viruses is called the lysogenic pathway. The bacteriophage that have this option, as well as a lytic pathway, are known as temperate phage. In this pathway, the virus goes into dormancy by integrating into the host genome, and remaining transcriptionally quiescent until environmental conditions change and reflect a likelihood of more host cells to infect (fig. 4). Lambda (λ) is an example of a

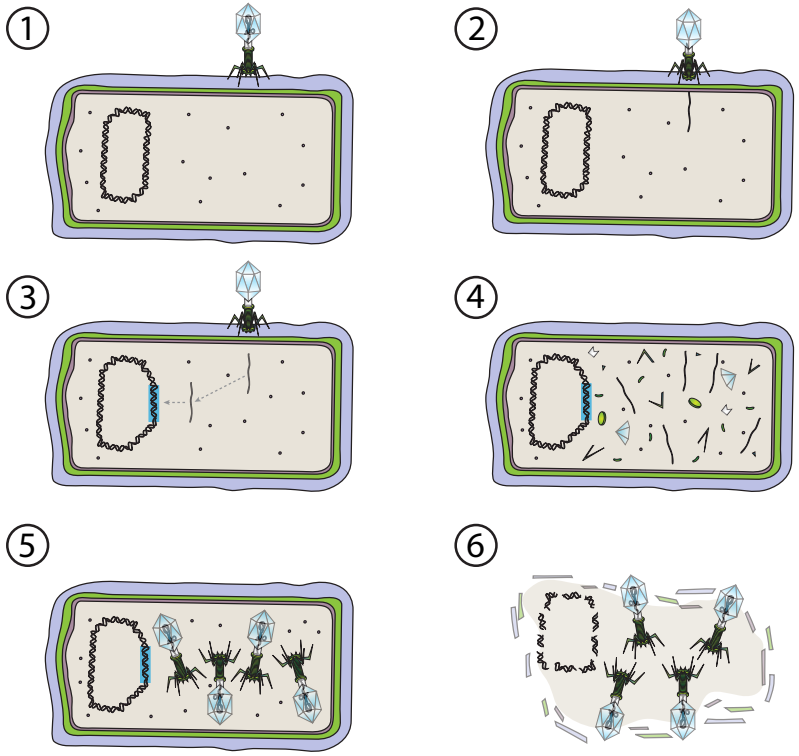

Figure 4. The lysogenic pathway.

temperate bacteriophage. The initial stages of infection and genome injection are the same as the lytic cycle, but under conditions that encourage lysogeny, the viral genome is integrated into the host genome in step 3. In λ integration into E. coli, this occurs by reciprocal recombination at a 15-base pair sequence known as the att λ site and is facilitated by the Int gene product. As long as the environmental conditions are not conducive to bacterial reproduction (and thus limited number of possible host cells), the viral genome remains mostly hidden and inactive. The only significant exception is a gene encoding a λ repressor that prevents the next step and keeps the virus dormant.

That next step is the excision of the l phage DNA from the host chromosome, and subsequent replication and transcription of the viral DNA (fig. 4, Step 4). Then, like before, the final steps are assembly and accumulation of virions, and eventual breakdown of cellular structure and release of the viral particles.

Although it is not referred to as lysogeny, some animal viruses can behave similarly. The most prominent example is the Baltimore Class VI viruses - commonly known as retroviruses, one of which is HIV. The path of a retrovirus through a eukaryotic host cell is depicted below (fig. 5). HIV has an envelope, which is studded with transmembrane proteins that are recognized by the host cell, binding the virus to the cell surface and initiating receptor-mediated endocytosis (1). After the endocytosis, the membrane envelope of the virion and the vesicular membrane fuse to release the capsid and its contents (2). After the capsid dissociates in the cytoplasm, the two strands of viral RNA are released along with a special polymerase: reverse transcriptase, which reads an RNA template and synthesizes DNA. Reverse transcriptase also uses that new DNA to synthesize a complementary DNA strand so that it eventually produces a double-stranded DNA version of the viral genome (3). This viral dsDNA is transported into the nucleus where it integrates into the host genome using another viral protein, integrase

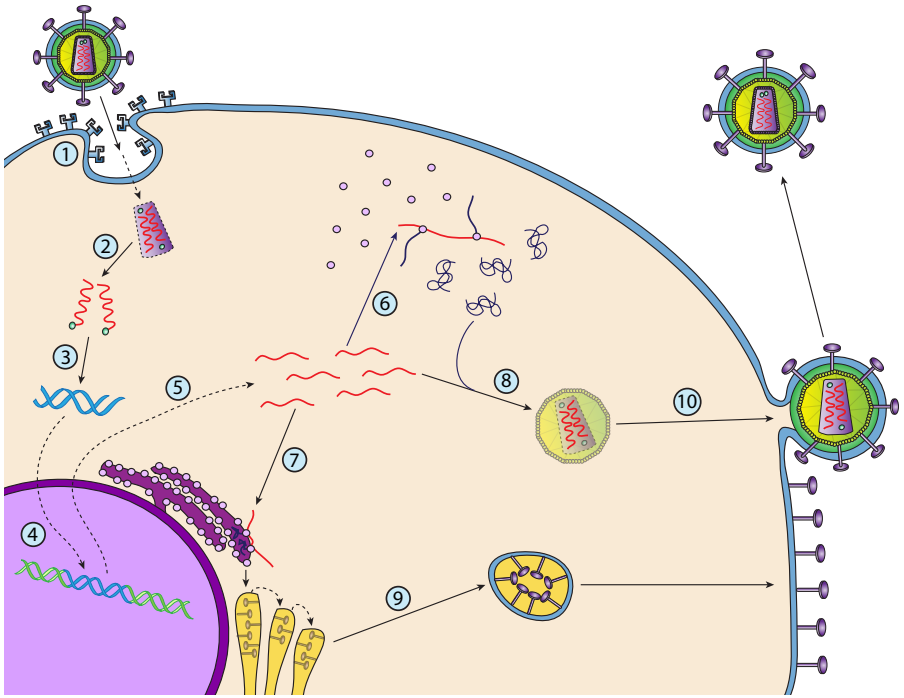

Figure 5. Infection and reproduction pathway of retrovirus in a eukaryotic host cell.

In order for packaging into the tight space constraints afforded by capsids, viral genomes must be highly economical. For example, the HIV genome (fig. 6) has several genes that overlap.

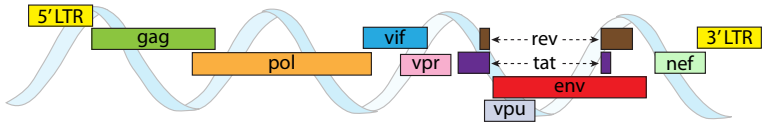

Figure 6. The HIV genome.

Or, in the case of curtoviruses, ssDNA plant viruses (e.g. beet curly top virus), the genome not only has overlapping genes, it is even bi-directional (fig. 7) encoding gene products in both strands of DNA after the ssDNA has been converted to dsDNA.

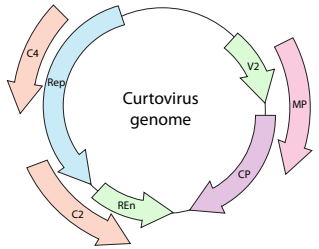

Figure 7. A curtovirus genome.

Given the need for economy, what genes are found in viruses? One of the most studied viral genomes, bacteriophage λ, contains genes encoding five transcriptional control proteins (which ones are expressed depends on whether the phage is in a lysogenic or lytic mode), a binding protein that controls degradation of a transcriptional activator, 17 capsid proteins, an excisionase that controls excision and insertion of the phage genome in the host genome, an integration protein that inserts the phage genome into the host's, and 3 genes participating in lysis of the host cell.

The HIV genome depicted above is much smaller than λ, at around 9 kilobases compared to 48 kb, but again, the theme is to use cellular proteins when possible, and encode viral genes if necessary. So, *gag* encodes capsid proteins, *pol* encodes reverse transcriptase, integrase, and HIV protease (which cleaves the gag and pol gene products into their functional proteins), *vif* acts against a common host cell antiviral enzyme, *vpr* regulates nuclear import, *tat* strongly increases transcription of HIV genes, *rev* exports viral RNA from the nucleus, *vpu* is needed for budding of particles from the host, *env* encodes viral envelope glycoproteins, and *nef* promotes survival of infected cells. The LTR regions are very strong promoters to drive high expression of these genes.

(4). The integrated viral DNA is called a provirus. The provirus can lay dormant, but if it is activated, then it is transcribed and the resulting viral RNA is transported out of the nucleus (5). Some of the viral RNA encodes enzymes like reverse transcriptase and integrase, or capsid proteins, all of which are made in the cytoplasm (6), but some encode membrane bound glycoproteins, which are translocated into the ER (7) and eventually processed through the Golgi and incorporated into the plasma membrane (9). Once the virion has been assembled (8), it binds to the viral transmembrane proteins, nucleating an exocytic "vesicle" (10) which is the virion complete with viral envelope.

In considering viruses with respect to the rest of this integrative chapter, there are two overriding ideas to keep in mind. First, viral survival is based on numbers: it needs to make huge numbers of its components to cast as wide a net for new host cells as possible. To do this, viral promoters are usually much stronger than host cell promoters, simultaneously driving more viral gene expression while preventing host gene expression (by dedicating cellular resources to virus production). Second, because of fast generation times, the rate of viral mutation and evolution is far faster than normal eukaryotic genomes. In addition, if the virus uses its own polymerase (such as reverse transcriptase or RNA replicase), the mutation rate rises even more because there is no error-checking by viral polymerases.

Cancer

Cancer encompasses a set of genetic diseases that lead to uncontrolled cell proliferation in multicellular organisms. The discussion of cancer also happens to be useful in a cell biology course, because it ties together many of the concepts that you just spent most of the semester learning. Although it can be caused in part by an outside agent, the development of cancer is essentially a series of uncorrected mistakes by a cell's regular processes. It can strike plants as well as animals, and because of intense re-

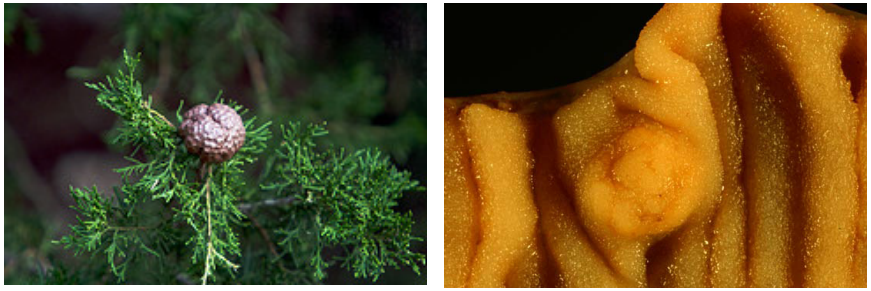

Figure 9. (left) A tumor on a cypress branch. (right) A tumor of the small intestine. Cypress tumor photo by W. Calder, cc licensed 2009. Small bowel tumor by E. Uthman, public domain 1999.

Recent structural examination of the HIV genome suggests that the structure of HIV RNA itself may play a significant role in its propagation inside of host cells. Figure 8, from Watts et al, *Nature* **460**:711-716, 2009, shows a predicted secondary structure of the genome. The authors suggest that the RNA structure actually may interact with ribosomal elongation to control the folding of the viral proteins. They also postulate the extension of this argument to include important genetic information encoded not just in the nucleotide sequence, but the secondary structure and tertiary structure of any RNA virus.

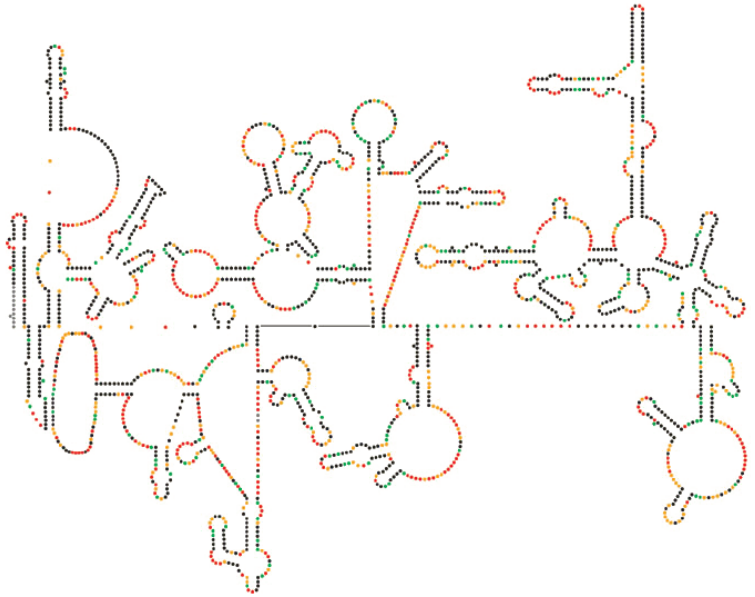

Figure 8. Secondary structure of HIV-1. (reprinted by permission from Macmillan Publishers Ltd: Nature 460:697, 2009)

search and subsequent deeper understanding of the cellular events that lead to cancer, it can now be treated in humans with some degree of success, depending on the type, location, and progression of the tumor.

Abnormal replication of a cell generally leads to the formation of a tumor, which is simply a solid mass of abnormally growing cells, usually clonal colonies of one or a few original tumorigenic cells. However, a tumor is not necessarily cancerous. A benign tumor is one that is ensconced within an extracellular matrix sheath, does not spread beyond that sheath, and whose growth is slow or limited. In contrast, a cancerous or malignant tumor grows quickly due to uncontrolled proliferation, expands significantly beyond its original boundaries, invading new tissue, and can metastasize, spreading through the circulatory system. Once this happens, not only is it no longer possible to remove all of the cancerous cells by surgical excision of the primary tumor, it is also nearly impossible to know how many secondary tumors have formed or where they formed, since the metastatic cancer cells in the bloodstream may theoretically exit almost anywhere. However, in reality, certain tumors metastasize preferentially to particular target tissues/organs, presumably based on molecular markers on the surface of the cells or in the extracellular matrix. Metastasis is considered the greatest medical problem with respect to cancer treatment. If cancer is detected after metastasis has occurred, the chances of survival drop dramatically.

At the cellular level, cancerous cells differ from normal cells in a number of important ways. Normal cells are regulated by the cells around them, and by adulthood, most cells are inhibited from proliferation by contact with their neighboring cells. In vitro, this can be demonstrated by the observation that non-cancerous proliferative cells such as epithelial cells can proliferate until the culture dish bottom is completely covered

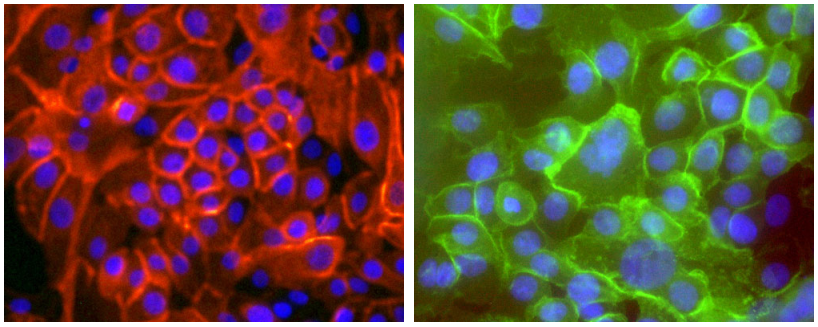

Figure 10. Normal human breast cells in culture at left. At right, similar cultured cells that have been transformed (i.e. they are now cancerous). Note the irregularity of both cell and nuclear morphology. Membranes are arbitrarily stained in different colors; chromosomes are stained blue in both panels. Photos from Ince et al, Cancer Cell 12:160-170, 2007.

(confluence), but once that happens, proliferation stops. This phenomenon is known as contact inhibition. If cancerous cells are allowed to proliferate in culture, they do not stop after the surface is covered, and instead can mound up on one another. The cell surface and internal cellular organization of cancer cells is often disorganized in comparison to normal cells. Finally, cancer cells usually appear de-differentiated in comparison to their original cell type. If the original cell type was a flat cell, the cancerous cell would be more rounded and three-dimensional. This is an expected consequence of becoming a cancerous cell. Not only is proliferation deregulated, cell surface protein expression is altered to promote metastasis.

Cancer is considered a *genetic* disease because it is caused by alterations to the DNA. However, it is rarely an *inherited* disease. An inherited disease would mean a disease that can be passed from one generation to the next, implying that the disease-causing DNA mutation is found in the gametes (sperm or egg) of the stricken adult. Most cancers are due to spontaneously arising mutation in the DNA of one or a few somatic cells, and not a systemic aberration. Spontaneous mutation in the germ cells are possible, but most potentially cancer-causing ones lead to non-viable offspring. So, although it is exceedingly rare for cancer to be inherited, however, it is much more common to inherit a predisposition or increased chance of developing a cancer.

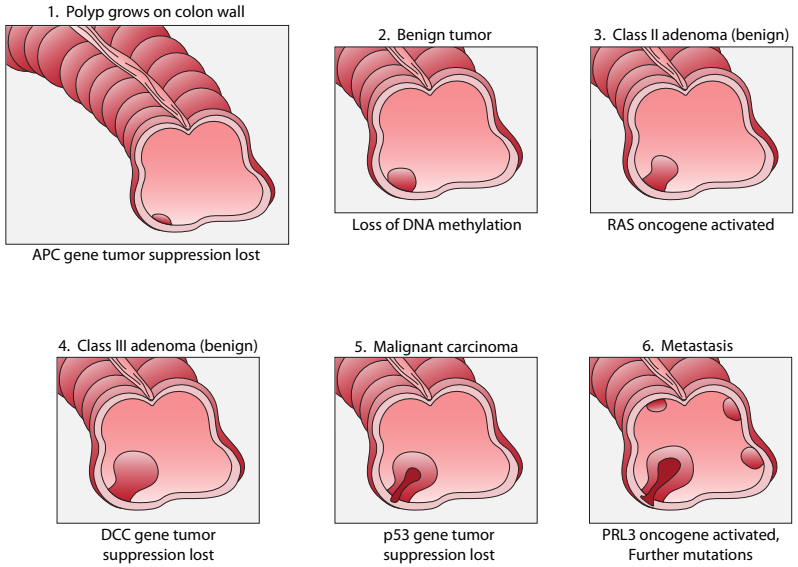

Figure 11. Development of colon cancer takes time and multiple mutations.

Differentiation is a key part of normal metazoan development. All cells come from the fertilized egg, and even after several divisions, the cells are very similar. Eventually, though, they begin to specialize for their particular physiological functions, whether as lung cells, brain cells, or bone cells, and that process of specialization is differentiation. In cancer cells, this process is partially reversed, as the cell reverts to a less specialized, more primitive state.

An individual cancer-causing mutation generally creates a problem that can be corrected by some other cellular mechanism. Therefore, development of cancer comes about through the accumulation of multiple mutations and not the acquisition of just one. The best studied example of this gradual development of cancer is colon cancer (fig.11). There is a fairly characteristic progression of mutations in the genes APC, RAS, DCC, TP53, and PRL3. Note that the progression depicted here is not inevitable: the presence of polyps does not lead invariably to colon cancer. Furthermore, intervention can be highly successful if it occurs early in the progression, so oncologists need to consider a range of risk factors in weighing the cost and benefits of medical intervention. RAS and PRL3 are oncogenes, while APC, TP53, and DCC are tumor suppressor genes.

Oncogenes

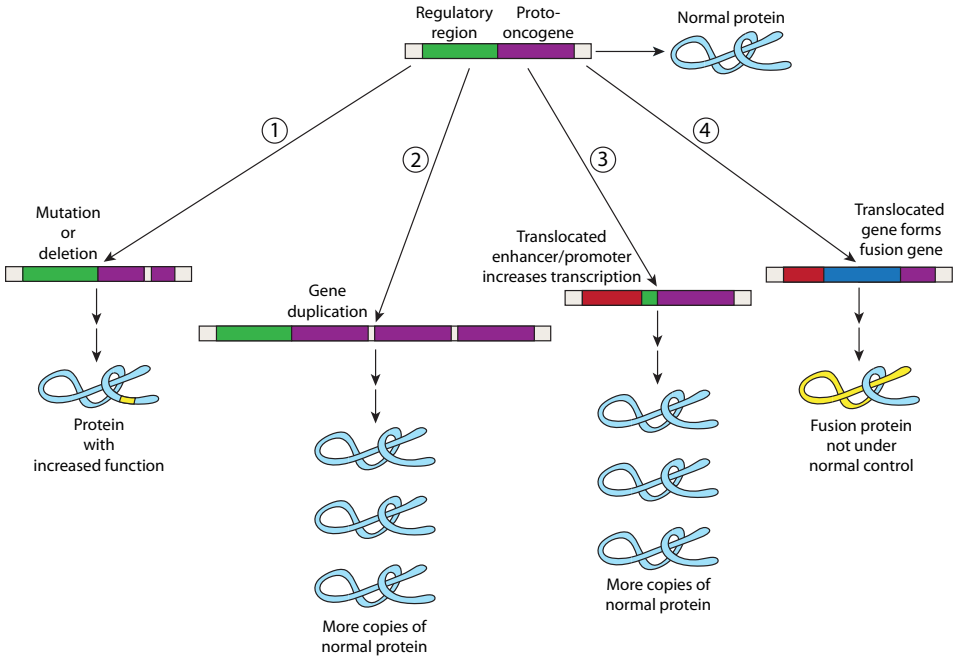

Figure 12. Conversion of protooncogenes to oncogenes. (1) Due to mutation of the coding region, the protein has higher physiological activity. (2) Gene duplication leads to multiple copies of the gene each expressed normally, making many more copies of the protein than normal. (3) Mutation of the regulatory region or translocation of a stronger enhancer or promoter to the protooncogene can lead to enhance transcription, and therefore more protein. (4) Translocation of another gene inline with part of the coding region, can put the activity of the protooncogene under the control of modifiers of the translocated gene, and thus lead to overactivity.

Oncogenes are generally dominant gain-of-function mutations of normal cellular genes called protooncogenes. These protooncogenes are themselves positive regulators of the cell cycle, but they are regulated by other factors, either extracellular signals or intracellular mechanisms. Mutations that turn them into oncogenes specifically remove all or some of this regulation. They thus become overactive, and try to push the cell cycle forward leading to increased proliferation. These mutations can also be classified into a few general mechanistic categories. These (fig. 12) are mutations to the coding region that increase physiological activity, gene duplications resulting in more copies of the gene at the DNA level and thus more at the protein level, mutations to the regulatory region of the gene or that alter regulation of gene expression, thus increasing copy number of the protein, and finally, translocations that replace part of the coding region, resulting in a chimeric protein whose activity may be under a different control scheme than normal.

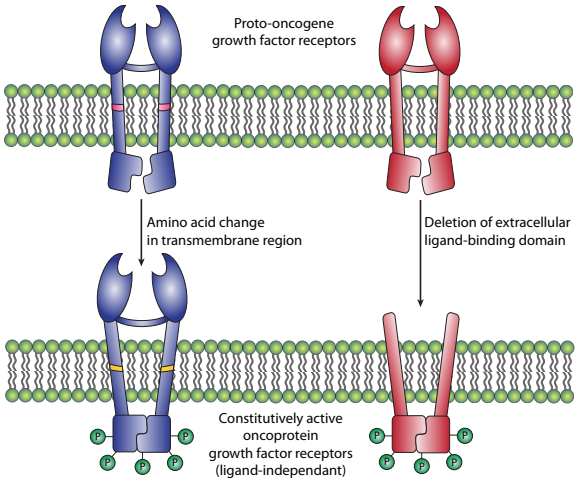

Figure 13. Conversion of a mitogen receptor protooncogene into an oncogene by point mutation leading to amino acid change in the transmembrane region (left) or by truncation of ligand-binding domain.

Examples of two types of mutations are illustrated to the right (fig. 13) with a mitogen receptor as the protooncogene. In the first case, the transmembrane portion of the receptor has been mutated, causing an amino acid change that alters the conformation not just of the transmembrane region, but of the cytoplasmic kinase domain, which becomes constitutively active, regardless of whether a ligand has bound outside or not. In the second case, the entire extracellular domain has been removed due to a mutation of an amino acid codon into a stop codon or translocation, and the resulting receptor is always active, also independent of ligand binding.

Some kinds of retroviral infection can accomplish the conversion of a protooncogene to an oncogene by inserting viral DNA near the promoter region of the protooncogene. Because the viral promoters tend to be very strong, they can induce overexpression of the protooncogene product. In avian species, avian leukosis virus is known to cause tumors by insertion near the c-myc oncogene, while in humans, another retrovirus, HTLV (human T-lymphotropic virus), can cause acute disease (tropical spastic paraparesis), but may also cause T-cell leukemia and lymphoma.

What functions are characteristic of protooncogenes? Mitogen receptors, as already described above, and exemplified by the receptor tyrosine kinases EGFR (epidermal growth factor receptor), VEGFR (vascular endothelial growth factor receptor), RON (recepteur d'origine nantais, a macrophage stimulating protein receptor), and ErbB2 (also HER2/neu, another human EGF receptor). Growth factors themselves may also be protooncogenes, such as FGF-5, one of several oncogenes in the fibroblast growth factor family, or c-sis, an oncogenic form of PDGF (platelet-derived growth factor). Signal cascade proteins, often either tyrosine or serine/threonine kinases or other regulatory enzymes, are a large group of protooncogenes (e.g. Src family tyrosine kinases, BTK family tyrosine kinases, cyclin-dependent Ser/Thr kinases, Ras-family small GTPases). Finally, various transcription factors (e.g. Ets, Myc, E2F families), can effectively be mutated into oncogenes.

Tumor Suppressor Genes

Tumor suppressor genes normally do what would be expected from their name. Whereas the oncogenes mostly drive the cell cycle forward, the tumor suppressor genes' primary functions are to temporarily stall the cell cycle so that DNA repair mechanisms can have time to work. However, if repair is unsuccessful after a few attempts, the tumor suppressor gene product may then trigger apoptosis rather than allow a damaged cell to replicate and potentially create another genetically damaged cell. Thus, the presence of an oncogene in a cell will not necessarily lead to development of cancer because a functioning tumor suppressor gene might prevent the cell from replicating. Equally, if a tumor suppressor gene is knocked out but there is no oncogene present, then the cell is unlikely to be immediately cancerous because although a cellular "emergency brake" is nonfunctional, if there is nothing to drive the cell through its cycle any faster or more frequently than usual, then the "brake" is never needed anyway.

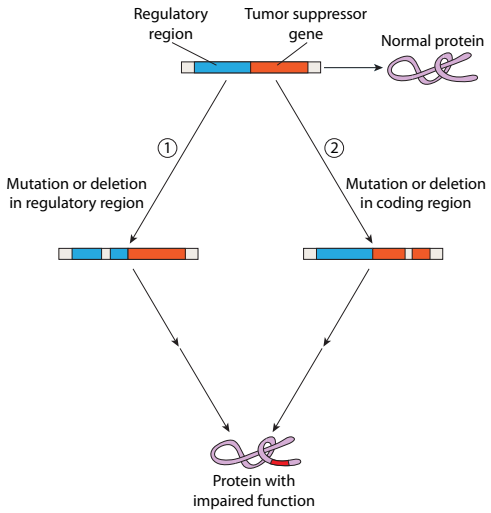

Figure 14. Tumor suppressor gene mutations can lead to cancer.

Like oncogenes, tumor suppressor genes can work (or not work, as would be the case in cancer) in several ways. Here is an example with the breast cancer-associated genes, BRCA1 and BRCA2. These gene products are involved in DNA repair (chapter 7). When BRCA1 or BRCA2 is knocked out, the cell loses its ability to use that DNA repair pathway. There are other repair pathways, and even if there weren't there may not be any serious lesions to the DNA, so the cell could behave normally for the time being. What is important from a cancer standpoint, is that each safety/repair mechanism that is lost increases the likelihood that an additional mutation may cause the cell to become cancerous.

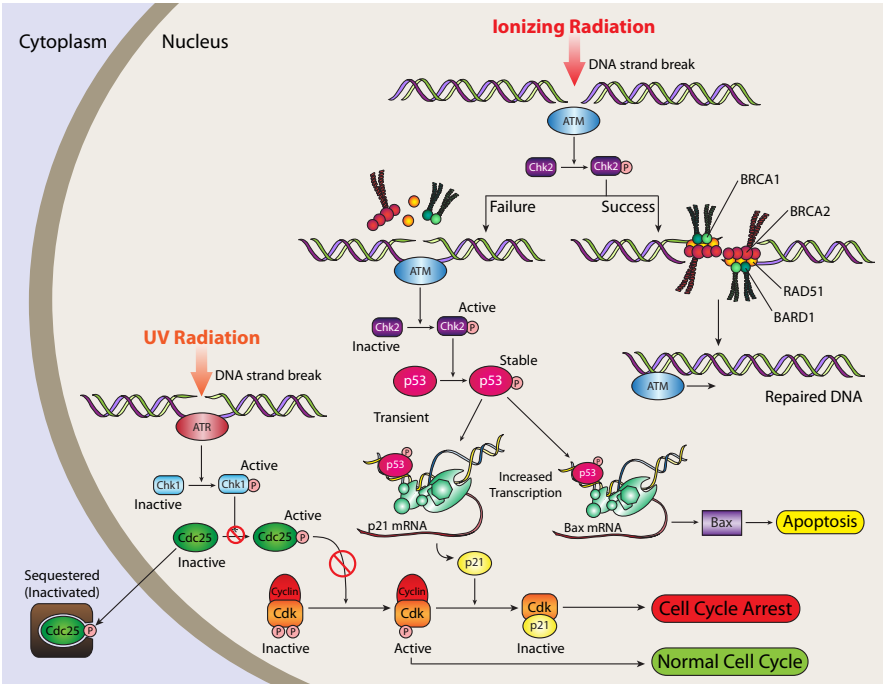

Figure 15. Cell cycle arrest due to DNA damage. ATM detects the double strand break, and activates Chk2 and BRCA1. Chk2 also activates BRCA1, which with BRCA2 forms a repair complex. However, if BRCA1 is not immediately available, the cell needs to go into a holding pattern until one becomes available. Therefore, Chk2 activates p53, which induces transcription of p21, which binds to cdk, preventing association with cyclin, and thus preventing cell cycle progress. If this continues for long, some of the p53 activates transcription of Bax, which will induce apoptosis to kill off a cell with damaged DNA. When p53 is hit with a loss of function mutation, the cell does not die, and it attempts to replicate even with damaged DNA, which may lead to more mutations in the subsequent generation, if it is successful in reproduction. Without p53, the accumulation of errors in successive generations increases. The mechanism of buying time for the cell to make repairs is not limited to the ATM-BRCA situation. The left side of the figure shows another response to DNA damage that leads to cell cycle arrest.

It should be clear now how recessive loss-of-function mutations in a tumor suppressor gene can lead to an inherited predisposition to cancer. As diploid organisms, we have two copies of each gene in our cells, so losing one to mutation does not wipe out the protective function. Thus, if nothing happens to the other one, then the cell is fine. It is just a question of probability. Losing the function of one is a very low probability event, but the probability of losing both copies is extremely small. Thus, even though it is "only 1 step" on the way to losing the protection of this particular tumor suppressing function, it is a very large difference in probabilities. Of course, keep in mind that even complete loss of a single tumor suppressor gene is usually not enough to lead immediately to cancer, and still other mutations must occur to take advantage of the weakened cell defenses and push it towards a cancerous state.

Human Cancers

Although some oncogenes and tumor suppressor genes have a restricted distribution that hints at likely tumor locations, many of the genes are widespread and even ubiquitous. It is presently unclear, therefore, why certain types of cancer are linked to particular mutated genes, but there are a number of strongly correlated cases.

Retinoblastoma is a cancer of the eye that usually strikes at a relatively young age. It has been linked to the RB gene, which encodes a repressor of E2F, a transcription factor that would normally turn on genes needed for S phase progression. Only 10% of individuals who inherit the RB loss-of-function mutation escape the development of the cancer. It also turns out that people with the RB mutation have a higher incidence of developing other tumors as well, although generally later in life. Perhaps the higher rate of damage to retinal cells (due to light exposure) leads to greater susceptibility.

Breast cancer is another disease that has strong links to mutations in certain genes. Loss of function mutations to the BRCA1 gene encoding a DNA repair protein lead to a five-fold higher risk of developing breast cancer in a woman's lifetime. Although mutations to other tumor suppressors (including p53, PTEN, CHEK2, ATM) most hereditary breast cancers have a link to BRCA1 or BRCA2 mutation. On the oncogene side, breast cancer tumors consistently show expression of CYCD1 (a cyclin) mutations, and depending on the type of tumor, HER2/Neu may be linked as well.

Lung cancers are among the most common - the second highest in men (prostate is higher) and women (breast is higher) alike, and make up approximately 1 in 3 cancer deaths annually. Several oncogenes of the myc family: N-myc, L-myc, and c-myc, as well as H-ras have been linked to various lung cancers. Loss of p53 and RB are also associ-

Cancers are classified by the tissue type in which the tumors first arise. Thus, *carcinomas,* which are the most common type (~85% of human cancers), come from epithelial cells arising from either the embryonic ectoderm (skin and nerve cells) or endoderm (gut lining). *Leukemias* (~4%) arise from white blood cells. *Lymphomas* (~5%) reflect aberrant growth of lymphocytes in spleen or lymph nodes. *Sarcomas* (~2%) arise from connective tissue of mesodermal origin, such as bone cancers.

ated with the development of lung cancers, and perhaps not coincidentally, tobacco smoking is associated with p53 mutations. Interestingly, despite being so common, so far, there have been no particular oncogenes associated with prostate cancer, nor any hint of prostate-specific tumor suppressor susceptibilities.

Metastasis

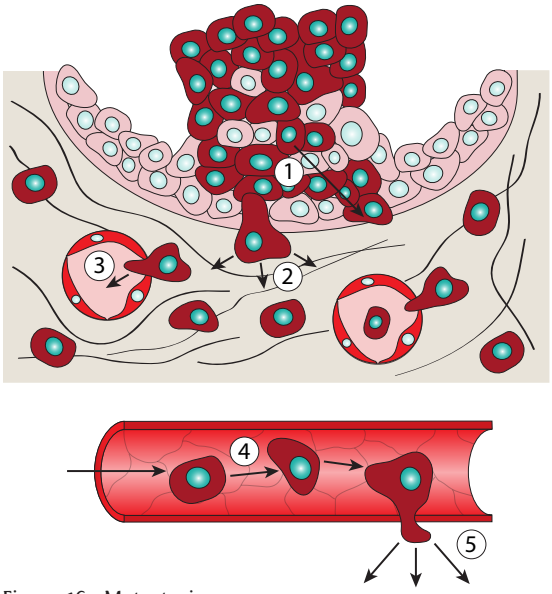

Figure 16. Metastasis.

The onset of metastasis signals a drastic change in the prognosis of a cancer patient. While pre-metastatic tumors can certainly be dangerous or painful, treatment can be fairly localized, e.g. surgical excision and directed radiation therapy. Once the tumor metastasizes it must be treated systemically due to the potential for secondary tumors literally anywhere in the body. This presents a problem because the tumor cells are derived from the body's cells and are mostly indistinguishable by the body's immune system. The primary mechanism for anti-cancer drugs is to target fast proliferation, since most cancer cells proliferate much faster than most normal cells, but this still kills off some of the body's naturally fast-proliferating cells, such as the epithelial cells lining the gut. More recently, other approaches to anti-cancer drug treatments have been developed; most notably, anti-angiogenesis drugs to starve tumors by preventing them from developing or recruiting new blood vessels. As tumors grow, the ability to absorb nutrients from the environment decreases for the innermost cells of a solid tumor.

Metastasis (fig. 16) starts with downregulation of cell-cell adhesions (1). This may include inside-out signaling to integrin receptors, or downregulation of cadherin expression, and other methods for allowing the cell to separate from the rest of the tumor. Non-metastatic tumors are surrounded by a capsule of extracellular matrix that contains the tumor in its location. To escape this capsule, the metastasizing cell must secrete proteases (usually metalloproteases) that can break down the ECM proteins (2). Once out into the looser connective mesenchymal tissue, the metastatic cell increases

its locomotive activity and heads for a blood vessel. Intravasation (3) into a small, low-flow blood vessel allows the cell to be carried to nearly any destination in the body by the circulatory system (4). At some point, the metastatic cell will attach to the interior wall of a blood vessel, and exit the circulation (5). The molecules and situations that determine the point of exit are not clear yet, although there are clearly preferred sites of metastasis for some types of tumors. Presumably, there is specific recognition and adhesion occurring based on cell adhesion molecule expression on tumor cell and target tissue surfaces.

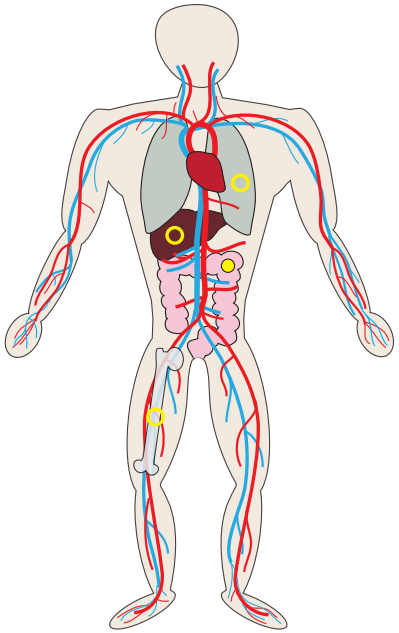

Figure 17. Common sites (open yellow circles) for the metastasis of colon cancers (filled yellow circle).

The Immune System

Immunology is a full semester course at most universities, so this section will only touch on a few basic concepts that should be easily accessible to the student who has nearly completed the cell biology course. At its core, immunology is about adaptation. That is, since an animal has no preconception of the various potential infections it may be subject to, it must have a system in place that is flexible enough to deal with almost anything that comes along. Obviously, the systems are not perfect, but considering the wide range of pathogens, immune systems are remarkably efficient. In humans, there are two types of immune response to infection: the innate response, which is relatively nonspecific, and the adaptive, or acquired, response, which has more specificity.

The innate immune responses are common to all animals, and act on large classes of pathogens. For example, Toll-like receptors on phagocytes recognize a variety of bacterial surface molecules such as the flagellin specific to bacterial flagella, or the peptidoglycan components of bacterial cell walls. When these receptors are activated, the phagocyte goes into action, enveloping the offending bacteria or virus, and breaking

it down. This depends on recognizing the external surface, so bacteria or viruses that do not have a recognizable molecule on their surface are able to escape this particular line of defense.

Defensins, which are found on a variety of surfaces (skin, cornea, gut) as well as in circulation, are small (18-45 amino acids) cysteine-rich cationic proteins that bind to a variety of pathogenic viruses, bacteria, and fungi. It is unclear how they may work against viruses, other than perhaps attacking infected host cells, but against bacteria and fungi the mode of operation is generally to bind to the cell membrane and form a pore that allows ions and other small molecules to flow out killing the pathogen. Complement, a group of proteins (~20) circulating in the blood, can act similarly against pathogenic cells.

Finally, natural killer (NK) cells, lymphocytes that target any cell that does not carry cell surface proteins that are normally found on cells from the animal, can kill not only attacking cells, but virally infected cells that have stopped producing their normal proteins (including the recognition protein) because they are busy producing viral proteins. NK cells can even be effective against some cancer cells if they have downregulated cell surface protein expression as part of their de-differentiation and deadhesion.

The adaptive immune system, which is only found in vertebrates, is what most people think of when the human immune system is mentioned. We and other vertebrates also have an innate immune system, but all the molecules and cells that normally come to mind — antibodies, T-cells, B-cells — are part of the adaptive immune response. There are two components to the adaptive response, a humoral response, in which proteins (antibodies) floating in the blood bind to the infectious agent and prevent if from binding to cells or targeting it for the cellular response, which is mediated by T cells that can specifically recognize and kill the targeted pathogen.

Antibodies

Front and center in the adaptive immune response are antibodies. These proteins may be either secreted by or attached to the surface of B cells, the lymphocytes that differentiate either in bone marrow (adult) or liver (fetus), as opposed to those called T cells, which differentiate in the thymus gland. Incidentally, if you see sweetbreads on a menu, that would be thymus. Yum. You can take that seriously or sarcastically depending on how you think my tastes run.

Back to the antibodies. The different types of antibodies, IgA, IgD, IgE, IgG, and IgM, are all based on the IgG structure (fig. 18), which is roughly Y-shaped, and composed of two heavy chains and two light chains. These chains have disulfide bond-stabilized loops (recall the Ig-like loops in the cell adhesion molecules a few chapters back?), and the combination of the distal light chain loop and distal heavy chain loop make the antigen binding site. The antigen is defined as the molecule, or more specifically the part of a molecule that is recognized by the particular antibody. Since the antibody is meant to mediate highly specific recognition of a wide variety of invading pathogens, there must be a way to create at least as many different antibodies. This is made possible by the process of *DNA rearrangement*. This mechanism is also used to generate diversity in T-cell receptors, which are quite different structurally, but also need to be available with an extremely wide variety of specific binding sites.

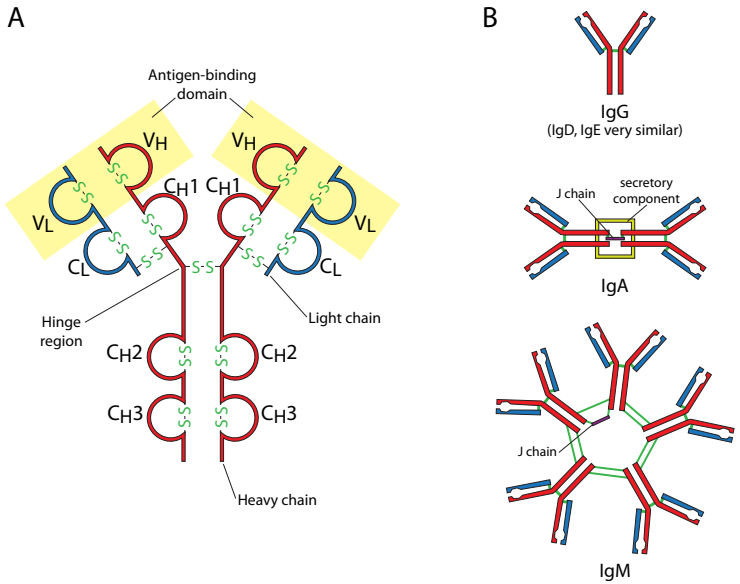

Figure 18. Antibodies. (A) an antibody is composed of two heavy chains (red) and two light chains (blue). Each has a variable region, and constant region(s). (B) IgA and IgM are built upon multiple IgG-like structures.

DNA Rearrangement

One of the central assumptions throughout our study of the cell has been that although the RNA and proteins in any cell may differ, any cell of a given organism other than the gametes should have the same DNA. This is not the case with B cells or T cells. In these cells, part of the maturation process is to create a unique arrangement of different domains to form a specific antibody (or T-cell receptor). The germline DNA, or the DNA that is found in all other somatic cells of the organism, contains many different such segments, but only a few are put together to make the antibody/TCR. This is a stochastic process, and with this kind of rearrangement happening on both heavy chain genes and two different light chain genes (designated κ and λ), there are well over 10 trillion (10^{13}) different combinations for generating immunoglobulins in humans, and even more combinations for T-cell receptors! How is this accomplished?

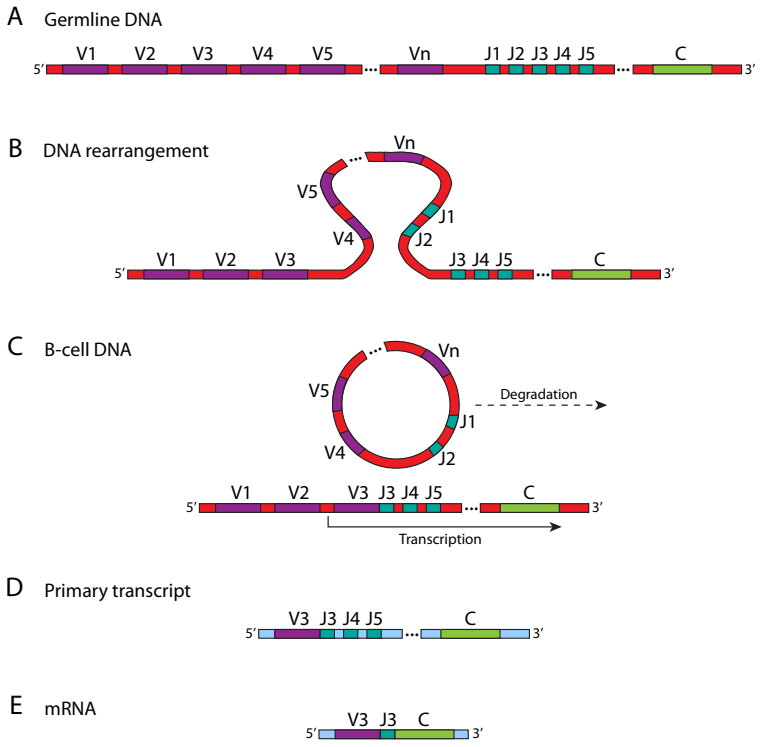

Figure 19. DNA rearrangement of a κ light chain gene. (A) The germline DNA has approximately 40 V region genes, 5 J segments, and one C region gene. (B) By action of the V(D)J recombinase, a random V region is brought close to a random J region, and the intervening sequence is cut out (C). (D) The gene may still contain multiple J segments, but RNA splicing removes all but one, leaving the final mRNA (E) with one V, one J, and one C region.

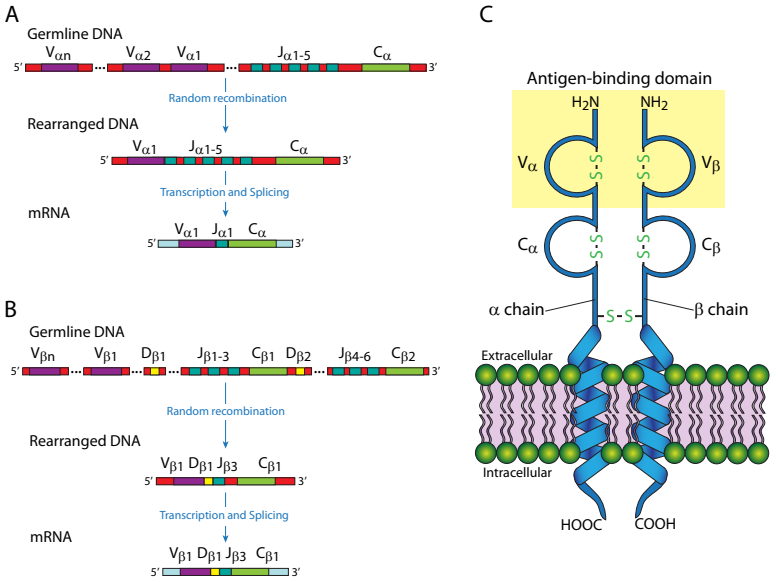

Figure 20. T-cell receptor genes also under DNA rearrangement to generate diversity like antibodies. In fact, there is actually greater diversity in TCR than in immunoglobulins. (A) rearrangement of α chain and (B) β chain. (C) protein structure of TCR.

Figure 19 shows the DNA rearrangements that take place in generating κ chain diversity. The λ chain locus has a slightly different arrangement, and has only 30 V genes, with 4 J segments and 4 C genes. The heavy chain has an extra domain: there are 40 V genes, which are linked to one of 25 D segments, then 6 potential J segments, and one C gene. These rearrangements, although they look something like the RNA splicing that we saw earlier in this course, are happening at the DNA level. Once it has happened, that cell and its progeny can no longer make the other combinations because those parts of the genome have been cut out and destroyed. This is distinctly different from alternative splicing of RNA, in which the genetic information is still there, and under different conditions could still generate other variations of the gene product.

The enzyme that produces this diversity is a complex called the V(D)J recombinase. The recombination occurs in two parts: first double-stranded breaks are made at recombination signal sequence (RSS) sites, then the breaks are repaired by the general double-stranded break repair mechanism. Depending on which J segment is chosen, there may be more than one left in the gene after the rearrangement. However, only the one closest to the V segment is used, and the others are spliced out of the primary transcript (by normal RNA splicing) in the process of connecting the C segment to the V and J for the final mRNA. Although this process generates great diversity, there is another mechanism that can generate further diversity under certain circumstances.

Somatic hypermutation causes rearranged V segments to mutate at 10^5 times the rate of other DNA! This mechanism is carried out by Activation-Induced Cytidine Deaminase (AID), which converts cytidines to uracils, generating a G:U mismatch that is "corrected" by repair polymerases without strong error-correction. This hypermutation is initiated by the activation of a B cell by recognizing and binding to a ligand. As we will see in the next paragraphs, when that happens, the B cell initiates rapid proliferation and this is when the somatic hypermutation takes effect, so that many of the B cells will carry highly similar but subtly different antibodies than the initial B cell that recognized and was activated by the antigen. The idea is some of these subtle mutation may lead to antibodies with higher affinity for the antigen and therefore faster response the next time this particular pathogen tries to infect the organism.

The reason that this kind of DNA rearrangement is necessary is that antibody "design" is not a reactive system, but a proactive system. A common misconception is that the immune system encounters a pathogen and creates antibodies that fit it. Unfortunately, there is no known mechanism by which a cell can "feel" the shape of something and create a protein that matches it. Instead, the immune system pre-emptively makes as many different antibodies (and TCRs) as possible, so that initially, most of the B and T cells in the body are actually genetically different. If an infection occurs, most of

RAG1 and RAG2, lymphocyte-specific recombination activating genes, recognize the RSS sites and make the double-stranded cuts. Once the cuts are made, the excised portion joins its ends together to form a circular signal joint (SJ) which is then degraded. The coding portions have asymmetric cut ends that fold into hairpin formations and prevent their fusion. These hairpins are broken by Artemis, a nuclease recruited by DNA-dependent protein kinase (DNA-PK), which also brings together XRCC4, XLF, DNA ligase IV, and a DNA polymerase. The XRCC and XLF alight the DNA ends, recruits a terminal transferase that randomly adds nucleotides to the ends, then DNA polymerase λ or μ fills in the overhangs, and the ligase completes the join. Interestingly, this process adds even more variability to the immunoglobulin, since Artemis cuts the hairpin at random, and the terminal transferase also adds nucleotides at random.

the B and T cells will bump into the pathogen and bounce right off, not recognizing it, but some of them will have the right antibody combination to bind to a part of the pathogenic invader. Although B cells can recognize surface antigens on their own, in most cases, a helper T cell is needed to activate the B cell (fig. 21). This process is initi-

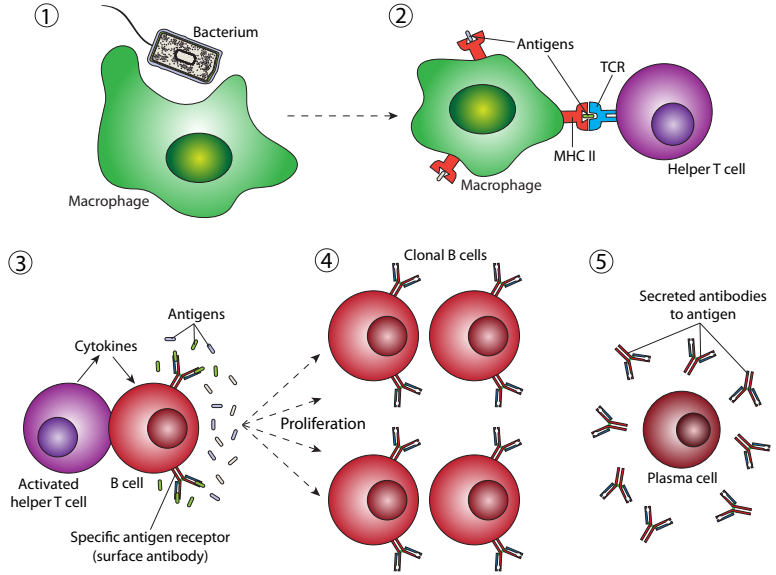

Figure 21. Activation of B cells by helper T-cells.

ated by a macrophage non-specifically ingesting a pathogen (1), breaking it apart, and presenting bits of it on its cell surface in partnership with MHC (major histocompatibility complex). A helper T-cell with a TCR that can recognize the antigen presented by the macrophage binds to it (2) and that leads to activation of the T cell. The activated helper T-cell binds to and activates a B cell (3) that has also bound to the antigen of interest, leading to massive B-cell proliferation (4), thus providing the body with many more copies of cells that have the right antibody to locate and fight the infection. This does two things: it provides lymphocyte reinforcements to specifically deal with a particular pathogen (but not other B cells, fig. 22), and once the pathogen has been eliminated, there is a larger circulating pool of these cells to respond more quickly to any subsequent infection by this particular type of pathogen. Finally, some of the B cells will differentiate into plasma cells (fig. 21-5), that secrete antibodies into the bloodstream to provide a humoral response. Others become memory cells, which are like B cells in that the antibody is on its cell surface and not secreted, but they can be thought of as "pre-activated" and can respond more quickly than naive B cells to re-infection.

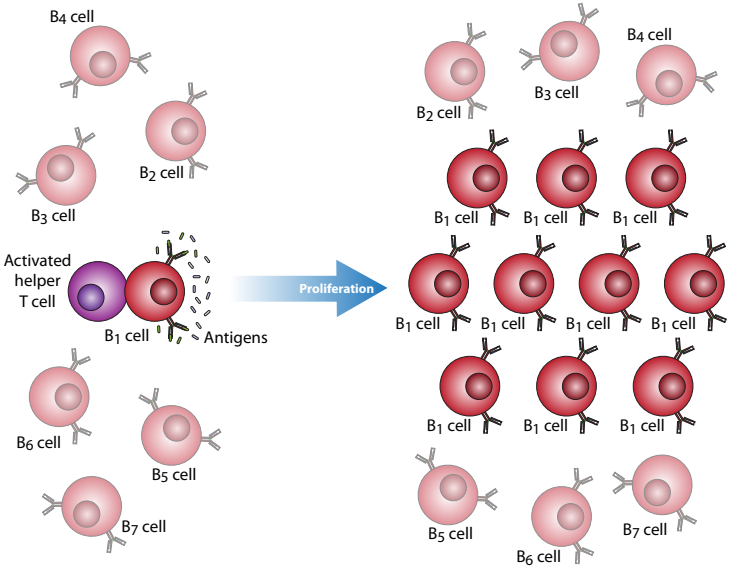

Figure 22. Amplification of only those B cells possessing antibodies that can recognize the infectious particle(s).

The big question that should have been lurking in the back of your mind through all this is, how do the antibodies and T-cell receptors tell what's foreign and what's part of one's own body? We'll get to that shortly. First, recall the activation of the helper T-cell. It occurs when the T-cell receptor recognizes an antigen from an ingested patho-

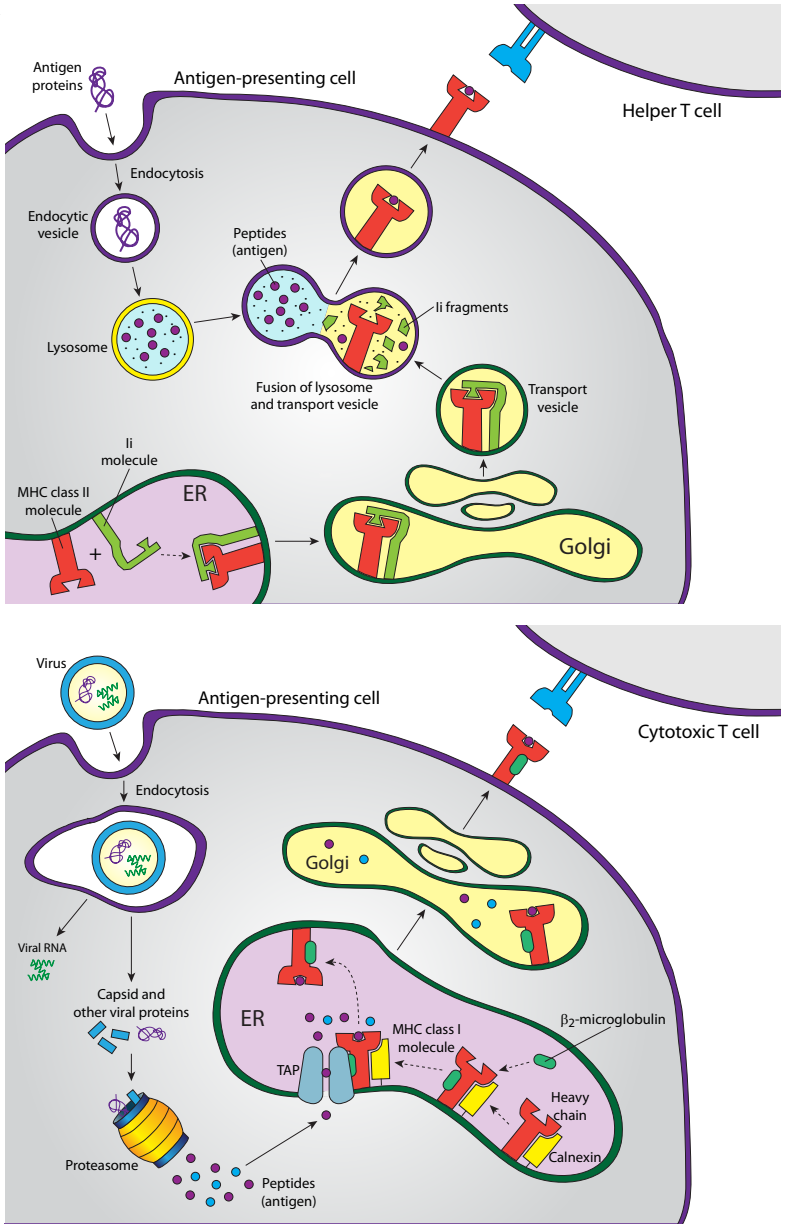

Figure 23. Antigen presentation by MHC class II (top) and class I (bottom) to T cells.

gen being presented by the MHC class II molecule on an antigen presenting cell (e.g. a macrophage). Figure 23 shows the pathway from ingestion of the pathogen to the presentation of its molecular parts on an MHC molecule. A similar pathway also applies to presentation of antigens on MHC class I molecules in conjunction with cytotoxic (killer) T cells. The cytotoxic T cells work against compromised cells, whether they are infected by a virus or another pathogen (fig. 24). There are two major pathways to

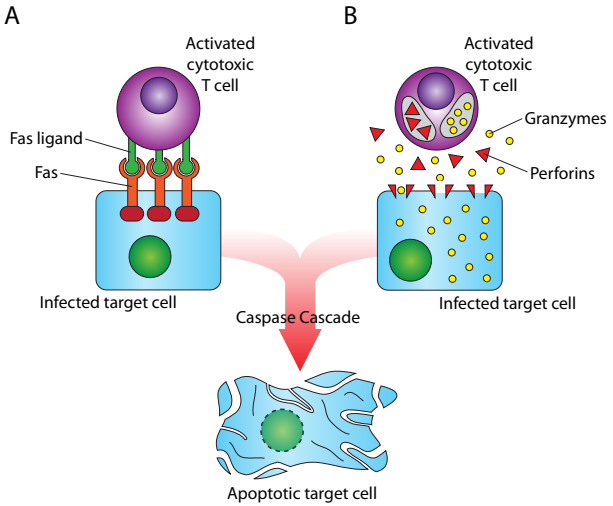

Figure 24. Cytotoxic T cells first recognize an infected cell by the T cell receptor, thich then leads to (A) binding of T-cell Fas ligand to Fas on the target cell, or alternatively, (B) secretion of granzymes and perforins. Both lead to activation of a caspase cascade and subsequent apoptosis of the infected cell.

killing the infected cell. One is the activation of a "death receptor", Fas, which induces a signal transduction cascade to activate caspases and apoptosis. The other pathway is the release of granzymes and perforins. The perforins drill into the membrane of the target cell and become pores that allow, among other things, granzymes to enter the cell, where they activate caspases by proteolysis, and again induce apoptosis. An important part of this is that the T cell receptor recognizes the antigen in combination with the MHC molecule that is presenting it. Furthermore, there are many variations of MHC molecules due there being 6 loci each with many known alleles.

So, what does any of this have to do with self vs non-self recognition? Early in the development of the immune system, the MHC does not present bits of digested pathogens, it presents bits of the organism's own cells that have gone through a proteasome or lysosome. At this early time, T cells behave somewhat differently, and if the T cell receptor binds strongly, the T cell commits apoptosis. This gets rid of TCR genes that strongly react to the organism's own cells. If the T cells do not react at all, apoptosis is also invoked, because the TCR genes that cannot recognize the MHC will not be useful in an immune response. Only those T cells that very weakly bind to the self-presenting MHC survive (fig. 25).

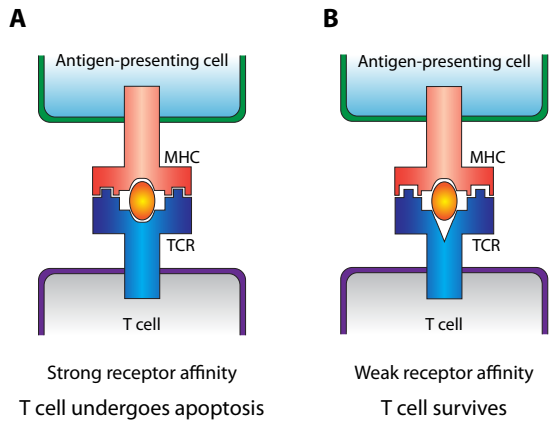

Figure 25. Self-recognition by T-cell receptor binding of MHC proteins presenting antigens derived from the organism's own cells. If there is strong recognition (A), the T cell dies to prevent cytotoxic attacks on its own cells. If it is weak and the TCR recognizes the MHC but not the antigen, the T cell survives (B).