Hello, my name is Michael Barresi; I'm a developmental biologist and associate professor at Smith College. In this tutorial, I'll be discussing one of the most fundamental concepts of developmental biology and embryonic development. That of differential gene expression. It is estimated that an amazing 37 trillion cells make up the human body. Which is represented by about 200 different types of cells. How all of these different cell types come together to create the human body is absolutely amazing. But what I personally as a developmental biologist find so amazing and interesting is the fact that all of these different cells arose from a single cell, right at the very start of embryonic development. How is that possible?

How can one cell ultimately give rise to all of these different cell types whether it's photoreceptors from our eye that have a very characteristic shape in order to capture light or perhaps motor neurons going from our spine and forming functional connections with our muscle? Sometimes three feet long. Or perhaps it's the many different cell types of our blood, whether they be red blood cells or white blood cells. Or yet even greater diversity and importance from our gut epithelial cells that are structure boxes but have very unique structures to grab nutrients from the environment as we eat passing through the intestine. And the amazing skeletal muscle cells that have multiple nuclei and can be the entire length of your thigh? How is this diversity possible? Arising from one cell, to generate all of these cells. This cell originally came together from two haploid cells, the sperm and egg. When those parental genomes fuse, it created the single cell zygote. That cell then divides into two. And then that'll continue to divide into four, eight, 16 and so on and so forth, all seemingly looking identical. But eventually, over the course of development gives rise to this amazing diversity here. How is that possible?

Jacques Monod won the Nobel Prize 1965 for his contributions to the genetic control of enzyme and virus synthesis. He once stated, "It is the repeatedly stated" "and thus, unsolved problem of understanding how cells" "with identical genomes may become differentiated" "that of acquiring the property of manufacturing molecules" "with new or at least different specific patterns" "or configurations of understanding how cells" "with

identical genomes may become differentiated." This question speaks to the concept of genomic equivalence. Which states that every somatic cell in one's body contains the same chromosomes. And if they contain the same chromosomes then they may possibly contain the potential to in fact, give rise or become any cell in our body. So if you're looking at this photoreceptor or motor neuron, inside their nucleus resides the genomic information to create those cells. And the DNA making up that genomic information is exactly the same.

So why does the DNA provide the instructions only for a photoreceptor in this cell in the eye but yet in the gut, generates a completely different shape of a cell with a totally different function in the intestine? How is that possible? And in fact do each of these different cell types, the skeletal muscle, the red blood cell, the gut, the motor neuron, have the potential to in fact, make any cell? And if we go back in time all the way to that first embryo, where all that DNA originated from, could any of these cells in fact create a whole embryo?

It was not until Sir Ian Wilmut and his colleagues produced the first viable animal cloned from a somatic cell nucleus in 1997. They used somatic cell nucleo-transfer to incorporate the nucleus of a differentiated udder cell from a six year old adult sheep into an enucleated egg which is an oocyte lacking a nucleus. The resulting blastocyst was implanted into a surrogate sheep. Now while this approach was horribly inefficient and many epigenetic questions still persist, Dolly was born, survived, and reproduced naturally herself. This proves that a single somatic cell nucleus in the context of the egg cytoplasm does in fact possess all the necessary information to organize and generate all the cells of the embryo. This is amazing. How is that possible? And what does it say to the regulation of cell differentiation?

The answer to this question rests in what I believe to be one of the most important concepts in developmental biology, that of differential gene expression. While every cell has the same genes, only certain genes are being transcribed and made into functional proteins. At certain times and amounts while in other cells, those same genes may be

actively turned off, enabling differentiation along a completely different path. What you see here are three different cells deciding, trying to figure out what they're going to develop into. Represented here are five different genes, A through E. Each cell has the same exact DNA, so the same genes are represented here. However, over the course of development, only selected genes will actually be turned on or expressed. Meaning mRNA will be produced. When mRNA is produced, then in those selected cells, only proteins from those specific genes will be made. So in this first cell only A and B. In the second cell, B and C. And in this third cell, A and E. Different combinations of proteins begin to get expressed and turned on and these proteins will provide the structural information to help those cells differentiate and ultimately take on and develop into very different cell identities.

I like to think of this process as if you were constructing a house. Every builder has a set of blueprints. And when they need to build certain portions of that house, they're only going to go to specific pages of that blueprint to gather the information they need. For instance, if they needed to work on the roof they're only going to go to that portion of the blueprints and figure out that they need shingles, rafters, perhaps a unique skylight for that particular house. However, if they're building the bathroom that day, they're not going to need those materials. So they'll only access other parts of the blueprint to figure out that they in fact need pipes, tiles, toilet, sink that are all necessary for the function of that bathroom.

Building a cell is exactly the same way. The genome represents the blueprints in this analogy. Only certain specific genes will be expressed, meaning accessed in those blueprints to build specific cell types while other information in the genome will simply be silenced and not looked at, not accessed, not transcribed or expressed. It is through this process of differential gene expression taking only those genes that are necessary to provide the structural and functional properties for a cell to differentiate, thus giving rise to the 200 or more different cell types that make us up. OK, let's take a step back. In order to understand the mechanisms of differential gene expression, I need to make sure we're on the same page with what a gene looks like, what it's anatomy is and really

what it means to be expressed. So the first core process involved in this is I need to make sure you understand a little bit about really the central dogma of biology. In which if this is a cell, outer membrane, and inside here is your double nuclear membrane, within the nucleus of course, is our DNA. DNA will provide the information to ultimately make the end product which are the proteins which provide the structural and functional properties of that given cell.

Now the DNA and genes on that DNA will ultimately be read and made into RNA, a single stranded molecule and that RNA can be processed, spliced into smaller pieces. And each one making its own protein and once processing is finished they will exit the nucleus and these messenger RNAs will interact with a ribosome and in that ribosome, undergo the process of translation. And in doing so, will make chains of amino acids, those chains of amino acids will begin to fold as that protein is modified and processed, different specific additions can be added to this protein to further increase it's level of specificity, and capabilities, it will then be transported to any part of the cell that is required. Perhaps this protein in fact needs to come back into the nucleus and play some important role in here to regulate gene expression.

For an example to understand gene anatomy we're going to explore the beta globin gene. Right here in this first segment of this schematic here is illustrated the beta globin gene. And this beta globin gene principally is made up by several different components. Two of the most important fall within this region which are the exons, one, two, and three illustrated here. And the introns. Only the exons will ultimately give rise to this protein. Provide the information to make and order the appropriate sequence of amino acids to give rise to this hemoglobin. However, on this DNA portion, transcription, in which there is a very specific start site, the transcription initiation, all of this up to the point of the transcription termination site will get transcribed and made into nuclear RNA. That nuclear RNA therefore will have the exon and introns within it. But remember, only those exons are going to be providing information for coding a protein. So it'll undergo a process of processing or splicing out those introns, putting together consecutive exons. Further modifications of a G cap and a poly(A) tail will be added to

the five prime and three prime ends of this mRNA that will then relieve the nucleus and undergo the process of translation in a ribosome to create that polypeptide of amino acids which will then fold and make into your protein of interest. However, what I want to bring your attention to is if you come back to the actual gene, there are some other very important information that gets to the heart of understanding transcriptional regulation and ultimately, differential gene expression. That is this promoter region. This promoter region, let's say, this aspect here, is a spot on the DNA that doesn't code for the protein. It is a location where other proteins and factors combine to and will ultimately trigger and enable transcription to happen. After all it is machinery that binds to that DNA and actually makes that RNA. Even further upstream, and in some cases downstream or even in introns will be sequences that provide information that will influence the ability of this promoter to initiate transcription. These sequences we call regulatory sequences, also known as enhancers or repressors. An enhancer domain can stimulate transcription while a repressor domain can inhibit transcription. But we'll come back to that.

This is all the anatomy that you really need to make sure you know to ultimately understand how genes are regulated. Okay, let's figure out how we can regulate gene expression differently in different cells. There are a lot of ways to regulate this. But for the purposes of this tutorial, I'm only going to really cover how gene expression is regulated at the gene level, transcriptional regulation. Simply put, in order for a gene to be expressed, the DNA actually has to be accessible for a variety of different proteins and factors to bind to it and regulate transcription. So regulation of this process is really about gaining access, opening up the DNA on a larger scale and at the individual gene level.

So the two types of modes that I'm going to talk about is one at the chromatin level, meaning first really getting access to regions of the genome, regions of chromosomes, and then more at the gene level, transcriptional regulation. You may already know that DNA is not freely floating around a eukaryotic nucleus, but is rather tightly bound to and wrapped around a series of proteins called histones, forming repeated structures called

nucleosomes. Specifically, 147 base pairs of DNA is complexed with eight histones forming what is called a nucleosome. And these nucleosome structures are regularly repeated, about every 70 base pairs throughout the genome. Envision a pearl necklace. When DNA is coiled around these histones and tightly packaged up it forms what you call heterochromatin. During which access to genes is closed. However, when the coiling is loosened, it takes on a form or state called euchromatin, and access to genes for transcription is now open. So ultimately regulating this chromatin from heterochromatic and euchromatic states will influence how and where within the genome certain genes can become expressed. How specifically are these states of chromatin regulated between condensed and loosened configurations?

Gaining access to DNA for transcription requires an open promoter region. Which means the nucleosome assembly of histones has to be loosened or unwound. This is accomplished by adding and eliminating small organic groups to the nucleosome, specifically methyl and acetyl residues are transferred to and from the tail regions of individual histone proteins. So right here, in a condensed nucleosome state, you see these little lines coming off of these histone nucleosome structures. On those tails are these small molecules that are showing represented methylated additions. This establishes this condensed configuration. If these methyl groups are removed and replaced with acetyl groups on those tails, then it'll establish an uncondensed nucleosome structure.

A general accepted rule is that acetylations often promote access and active transcription in that particular region of the genome while most methylations will serve to condense chromatin and repress transcription. Now we know we can modify histones by adding methyl groups or acetyl groups to open it up or condense it. In the end, giving access or restricting access to that promoter region for other proteins to bind to and regulate transcription. So the other part of gene regulation that I want to discuss is actually regulation at that promoter region. Those other proteins that may come and bind and turn genes on or turn genes off. Appropriately, these other proteins are called transcription factors.

If you've ever pondered how a certain developmental process happens, how does the limb form, or perhaps, what went wrong? Well, most cases you can blame it on a transcription factor. In fact, if you change just one transcription factor in some cases it can have profound, dramatic effects on how that embryo forms. As an example, loss of the gene ultrabithorax, which is a homeobox containing transcription factor causes a complete change in the segmentation pattern in the fly. Such that if forms two sets of perfectly developed wings. It's amazing. How could perfectly developed wings form in the wrong location just from altering one gene? In this case, one transcription factor?

This is a dramatic example of how transcription factors have the power to really control gene expression and ultimately, development and even disease. All transcription factors by definition have the ability to bind DNA and influence gene transcription in doing so. There are a variety of different DNA binding domains affiliated with different families of transcription factors. But in either case, they're binding to very specific sequences on the DNA that will then influence transcription. So if this is our gene and our gene has a transcriptional start site, and there is a promoter region, and somewhere say, in this example, are there upstream, is a specialized sequence that we're going to call an enhancer. This enhancer is on the same gene or on the same chromosome, rather, to influence gene x. Transcription factors, many different shapes, many different kinds, have those abilities to bind DNA with great specificity. So they'll come down and recognize this enhancer and bind to it. And form a transcription factor complex.

When this happens, it actually contorts and changes the shape of the DNA.
Such that your DNA regions of enhancer and promoter literally change shape and come very close to each other. Enabling a very important enzyme to bind to that promoter, which is RNA polymerase. In this way, these transcription factors are able to influence the binding of RNA polymerase to that promoter and initiate RNA synthesis of that gene. Regulatory sequences such as enhancers and repressors are usually found on the same chromosome as the gene they're influencing. And thus are called cis-regulatory elements. These cis-regulatory elements are not just on off switches, but actually function more like dimmable switches, depending on how the complex of transcription

factors influence the tertiary structure of DNA and regulate the rates of transcription from that promoter. Remember, every cell has the same DNA which means they all have the same enhancer sequences, too. What differs is the specific combination of transcription factors that that particular enhancer experiences to regulate transcription.

What you see in this image is an example of a particular gene we're calling Gene A. And the regulation that it can be expressed in different cells or different tissues of this mouse embryo. What you see are two specific enhancers. One is a brain enhancer and the other one is a limb enhancer. This RNA can be expressed and turned on in those regions when those enhancers have bound transcription factors. So in the two examples below, Gene A is expressed in the brain in this first one in which transcription factors, combination of specific transcription factors that may only be present in the brain bind to that enhancer and regulate that particular Gene A RNA in that tissue. Whereas the cells of the limb have a different repertoire of transcription factors and they are capable of binding to that other enhancer which still regulates Gene A but now that's happening in the limb because of that repertoire of transcription factors. In this next example, this is a real gene called PAK 6. Which itself is a very important regulating transcription factor for many different developmental processes.

And you can see that there are numerous enhancer regions here. One is a pancreas enhancer another one is a lens enhancer, and another region, this regulatory region happens to be in the neural tube. All of these three enhancers are upstream of the gene but in fact there's a little retinol enhancer over here nestled within the gene. So remember, these regulatory sequences can be anywhere, upstream, downstream, nestled within an intron, located throughout the gene. What scientists have been able to do, is actually take advantage of these regulatory sequences so that we can actually study development.

For instance, there have been transgenic mice made using this pancreas enhancer and this lens enhancer of PAK 6 gene, in which they took those sequences and used them to direct and drive the expression of a specific reporter. In this case called lac Z or beta-

galactosidase which is an enzyme that can be used to catalyze a color precipitate in the embryo. So when this image here, you see blue colored cells where the lens would be developing as well as blue colored cells where the pancreas is developing. You don't see blue anywhere else in this embryo but only in those two locations. Proving that these two enhancer regions of the sequence actually do drive expression in those specific cells of those specific tissues. As this example shows differential gene expression is of course best understood simply by seeing.

Here is an in situ hybridization for a gene called Crestin. Found to be differentially expressed in neural crest cells of the zebra fish embryo. From its head going all the way down to its tail you see purple expression. Those are all individual neural crest cells expressing this particular gene called Crestin. Neural crest cells are one of my favorite types of embryonic cell types, because they are actually migratory stem cells. They leave the nervous system, venture out throughout the embryo to populate a variety of tissues and differentiate into a variety of different cell types, whether it's from pigment cells or cartilage cells, or sensory neurons, they are just an amazing plastic cell type that is really a good example of differentiation. And what you see here is one particular gene called Crestin that's helping us to visualize these.

If we look a little bit closer, you can actually see individual populations of cells beginning to be positioned in the location where they're migrating towards the craniofacial structures. So these neural crest cells, as stem cells, will be migrating away and going and populating regions of the head and eventually differentiating to cartilage and bone, just like using the lac Z reporter. Scientists have used cis-regulatory elements to drive other reporters in cell type specific manners. Such as the regulatory sequences for the flea one gene.

We’re used to drive expression of the green fluorescent protein seen here, in this transgenic zebra fish embryo. So this embryo is in fact showing you those pouches of pharyngeal arches that are populated by these neural crest cells migrating down to create those craniofacial structures. My laboratory is interested in understanding how

environmental toxins influence neural crest migration and ultimately, craniofacial development. So this is a great example of how understanding the principles of differential gene expression have allowed us to literally watch development happen over time. So let's have a quick recap.

Chromatin modification and transcription factor bind into cis-regulatory elements are two of the more significant ways in which gene expression can be regulated so precisely in timing, in space, and in amount. During early development important initiating transcription factors will begin to turn on or repress the expression of a multitude of regulatory genes. One of the greatest realizations about gene regulation has been that mutations in the coding regions of genes has not really been the driving force of evolution but rather duplication and mutation in the cis-regulatory elements of a gene has provided a mechanism to subtly change gene expression patterns in a developing embryo. Without having to alter the function of that protein. In this way, the co-option of existing proteins could be used in new tissues at new times ultimately to create new forms. Some of which would be advantageous for survival, while perhaps others not.

As I end this tutorial on differential gene expression, please know that there are many ways to regulate both the gene and the protein which will influence how cells differentiate. These might include methylation of the DNA itself. Also alternative splicing of the RNA into a variety of different proteins is an important way to change and increase the multitude of the types of structural and functional properties a cell can have. Also proteins themselves can be modified, whether it's by adding various residues to them to alter their ability to be secreted, or perhaps where they're located in a given cell, which stricken them to the nucleus-users, sending them to the cell membrane. Perhaps they're even degraded directly. Either way, there is a variety of different ways genes and proteins can be regulated, many of which I think we still have a lot to learn and discover these many diverse mechanisms of gene and protein regulation really make possible the amazing and incredible diversity of cell types that make us up.