The Factor Fugue

Each Easter at Bardstown Road Presbyterian Church the congregation finishes off the service with rousing, if somewhat chaotic, rendition of Handel's Hallelujah Chorus. For historical reasons unbeknownst to me, the sheet music we are provided for this chorus is not only antiquated but also written in four-part harmony for professionals. I've never seen so many notes on a single page.

Undaunted, and wishing to participate because I am caught up in the moment, I decide on the bass voice since it seems to have fewer notes and words. As we are a small church and we don't have strong singers in each voice category, those of us choosing more under-represented voices are left on our own with no one's coattails to cling to. I start off reasonably well but things rapidly turn south when each category goes its own way and I am left to my own devises. Fortunately, every now and again, we are all brought back in register, "e.g., the kingdom of this world" and for the time being anyway, I am back on track.

I'm not entirely sure why, but my muddling through the Hallelujah chorus was the analogy that sprang to mind when I began considering a *Journal Club* having to do with human development. When sperm joins with egg, the process of human development begins; a process infinitely more complex than the Hallelujah chorus, but one that is played out at different levels simultaneously, with a myriad of individual notes, all of which have to come together in sync to produce a new human being.

So let's consider human development starting at the beginning, with the fertilized egg. The instructions for what will become a human being are embodied within the 3.5 billion bases pairs of DNA packaged on 46 chromosomes, half coming from mom and half coming from dad. The fertilized egg, a single cell, will ultimately give rise to the 50 or so trillion cells of the human body. Some of these cells become the rods and cones of the eye and give us vision, whereas others become precursors to red blood cells; the latter we all know, being charged with transporting oxygen through the blood. And these are just a small sampling of the potential fates of the trillions of cells arising from the original fertilized egg.

So what differentiates one cell's fate from another during human development? Does each cell type receive a different set of instructions? The answer to this, at least from my perspective, is yes and no. The written instructions laid down in the sequence of bases for in those 3.5 billion bases of DNA, with certain rare exceptions, are the same in all cells of the body regardless of whether they are a cell of the eye or the bone marrow. What differs from one cell to another are how these instructions are interpreted.

Major players tasked with interpreting the instructions laid down in the sequence of DNA are a tremendously important family of proteins referred to as **transcription factors**. Transcription is the process by which DNA is converted into RNA. RNA, in turn, is used to program ribosomes for the synthesis of proteins. It is the constellation of proteins synthesized within a cell that by in large give it its defining characteristics. So cells differentiate into a particular cell type by turning on genes to make one set of proteins while turning off genes coding for other sets of proteins. The decisions as to which set of genes to turn on and those to turn off are controlled in large part by transcription factors expressed in different cell types. As to what determines which transcription factors are expressed in a particular cell type, well, other transcription factors, a point I will return to later.

This discussion brings us to a very interesting review recently published in *Blood* by John Crispino and Mitch Weiss[1], two physician scientists that attended this year's DBA ICC Meeting. The title of their review was "Erythro/Megakaryocytic Transcription Factors Associated with Hereditary Anemia." This review, as the name implies, summarizes transcription factors that to date have been linked to hereditary anemia and suggests that additional transcription factors necessary for erythropoiesis may be encoded by some of the still unidentified genes responsible for congenital anemia, including DBA.

Notice that the title of the Crispino/Weiss article is not transcription factor associated with hereditary anemia, but transcription factor"**s**". Do we need more than one transcription factor to create a functional red blood cell? A priori, you might think that we might need at least two, one to turn on genes needed for red blood cell function, and perhaps another to turn off the genes for proteins that function in other specialized tissues. Alas, human development is vastly more complex than this simplistic notion would suggest.

Let's go back to my bass line in Handel's Hallelujah Chorus and envision my chosen bass line to be equivalent to the transcriptional regulation of the globin genes in developing red blood cells. Amongst other things this bass line needs to do is shut off γ globin synthesis at birth and replace it with β globin. But plenty of other things are going on at the same time in a developing red blood cell including the synthesis of heme, the synthesis of proteins that give red blood cells their unique shape, and the synthesis of factors that govern cell division to assure adequate number of red bloods cells. Each of these processes involved in creating a red blood cell is analogous to the different lines of voice that come together in harmony to produce the Hallelujah Chorus. One transcription factor, KLF1, appears to touch on each of these processes in red cell development.

According to Wikipedia, a transcription factor is a protein that binds to specific DNA sequences and regulates the flow of genetic information from DNA to mRNA. In this regard there is a region in front of the β globin gene where KLF1 binds. A mutation in this DNA sequence which interferes with KLF1 binding blocks β globin synthesis and leads to β thalassemia, a hereditary anemia. Other mutations in KLF1, in contrast, influence the production of proteins of the red blood cell membrane and give rise to hemolytic anemia. Thus, anemia can result from mutations in either the transcription factor or the site in DNA where the transcription factor binds and depending upon the different outcomes in terms of the number and types of genes affected, different forms of anemia can result.

Transcription factors, however, do not operate in isolation. The ultimate goal of a transcription factor is to influence transcription of mRNAs, which is catalyzed by RNA polymerase II and a host of what are termed basal transcription factors needed for transcription of most genes. So here I've introduced a new complexity in the transcription factor saga by differentiating between basal transcription factors, those that work with RNA polymerase II to transcribe most genes, and transcription factors like KLF1 which tend to be restricted to a particular tissue and influence transcriptional activity in a tissue- and gene-specific manner. Transcription factors like KLF1 however, perform their functions by interacting with components of the basal transcription machinery. For example, in the synthesis of β globin, KLF1 helps recruit the basal transcription machinery to the β gene, a process that wouldn't work very efficiently without the assistance afforded by KLF1. KLF1 works its magic by both binding directly to components of the basal transcription machinery and by altering the packaging of DNA to make sequences more accessible for basal transcription factors to bind.

Transcription factors, however, do not operate in isolation (II). KLF1 is just one of a number of tissue specific transcription factors that bind to the β globin gene. Another is GATA1, a transcription factor most of us in the DBA world are aware of, as it has recently been identified as a DBA gene[2]. GATA1 cooperates with KLF1 and together they provide an extra umph to β globin expression beyond that occurring with binding of either factor alone. Moreover, GATA1 binds to DNA sequences in the KLF1 gene and stimulates the expression of KLF1. So GATA1 can be thought of as regulating erythropoiesis both by regulating the expression of KLF1 and then cooperating with KLF1 in controlling the expression of erythroid specific genes.

But GATA1 has more on its plate than simply regulating erythropoiesis, as it is also involved in regulating megakaryopoiesis. GATA1 works at the point in hematopoiesis at the level of a common megakaryocyte/erythrocyte progenitor before these cells commit to either the red blood cell or megakaryocyte lineage. In going down the red cell lineage, GATA1 cooperates with KLF1 in regulating genes required for erythrocyte function, whereas in going down the megakaryocyte lineage it presumably cooperates with other transcription factors in specifying megakaryocytes and ultimately, platelets.

The location of GATA1 function in this common progenitor would seem at odds with it being a DBA gene (using the classical definition of DBA as a pure red cell aplasia where only the red cell lineage affected). One would predict that the loss of function of GATA1 function in specifying the megakaryocyte erythrocyte progenitor should affect both lineages; and in fact some mutations in GATA1 do[1]. It should be pointed out however, that only very specific types of mutations in GATA1 give rise to DBA and that these preferentially affect the role of GATA1 in erythropoiesis. Thus, not all mutations in GATA1 give the same clinical endpoint.

To get a handle on the genotype/phenotype relationships in GATA1 we have to go back to the concept that transcription factors do not work in isolation. The affects of GATA1 on RNA polymerase II activity are not direct but instead involve the ability GATA1 to bind other proteins, which serve as a bridge between GATA1 and the basal transcription machinery. As pointed out by Crispino and Weiss[1], the differential effects of mutations in GATA1 on erythropoiesis and megakaryopoiesis reflect outcomes on the interaction of GATA1 with these other bridging proteins that help mediate the ability of GATA1 to regulate genes involved in these different development pathways. Importantly, this view suggests that some of these other bridging proteins that interact with GATA1 and help specify the erythroid lineage may be candidate DBA genes. Thus, investigators doing whole exome sequencing as a part of the gene discovery process will most certainly be interested in following up on any polymorphisms they identify in genes encoding proteins that work with GATA1 in erythroid development.

Returning to the analogy that began this discourse, let's begin with my bass line that represents just one voice in Handel's Hallelujah Chorus. The Chorus, in turn, is but a part his opus, The Messiah, which is but a single component of total Handel's body of work. All of which began with Handel's fertile mind. Similarly, we have the erythroid lineage, which is but a part of hematopoiesis. Hematopoiesis, in turn, is derived from the mesodermal layer in primitive development. All of which is begun after the union of sperm and egg. At virtually every juncture in the mind-numbingly complex pathway by which a fertilized egg gives rise to red blood cells, transcription factors orchestrate the interplay of developmental, environmental and physiological cues that result in a new human being. In 2006 Swiers *et al.* attempted to describe the genetic regulatory networks that program hematopoietic stem cells and erythroid lineage specification[3]. Their circuitry diagrams were unbelievably complex involving 10 times the number of transcription factors discussed here and all this of course is likely just scratching the surface.

Much remains to be known regarding how mutations in GATA1 and perhaps other transcription factors cause DBA and whether these proteins will represent good targets for therapeutic intervention. The other pressing question is whether DBA arising from mutations in ribosomal proteins intersects with DBA caused by mutations in GATA1. Since the ribosome is needed to synthesize all proteins including GATA1, it is possible that mutations in ribosomal protein genes could influence erythropoiesis via affecting GATA1 expression thus providing a unified view of DBA pathophysiology.

1. Crispino, J.D., and Weiss, M.J. (2014). Erythro-megakaryocytic transcription factors associated with hereditary anemia. Blood 123, 3080-3088.
2. Sankaran, V.G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J.A., Beggs, A.H., Sieff, C.A., Orkin, S.H., Nathan, D.G., Lander, E.S., et al. (2012). Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. The Journal of clinical investigation 122, 2439-2443.
3. Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Developmental biology 294, 525-540.