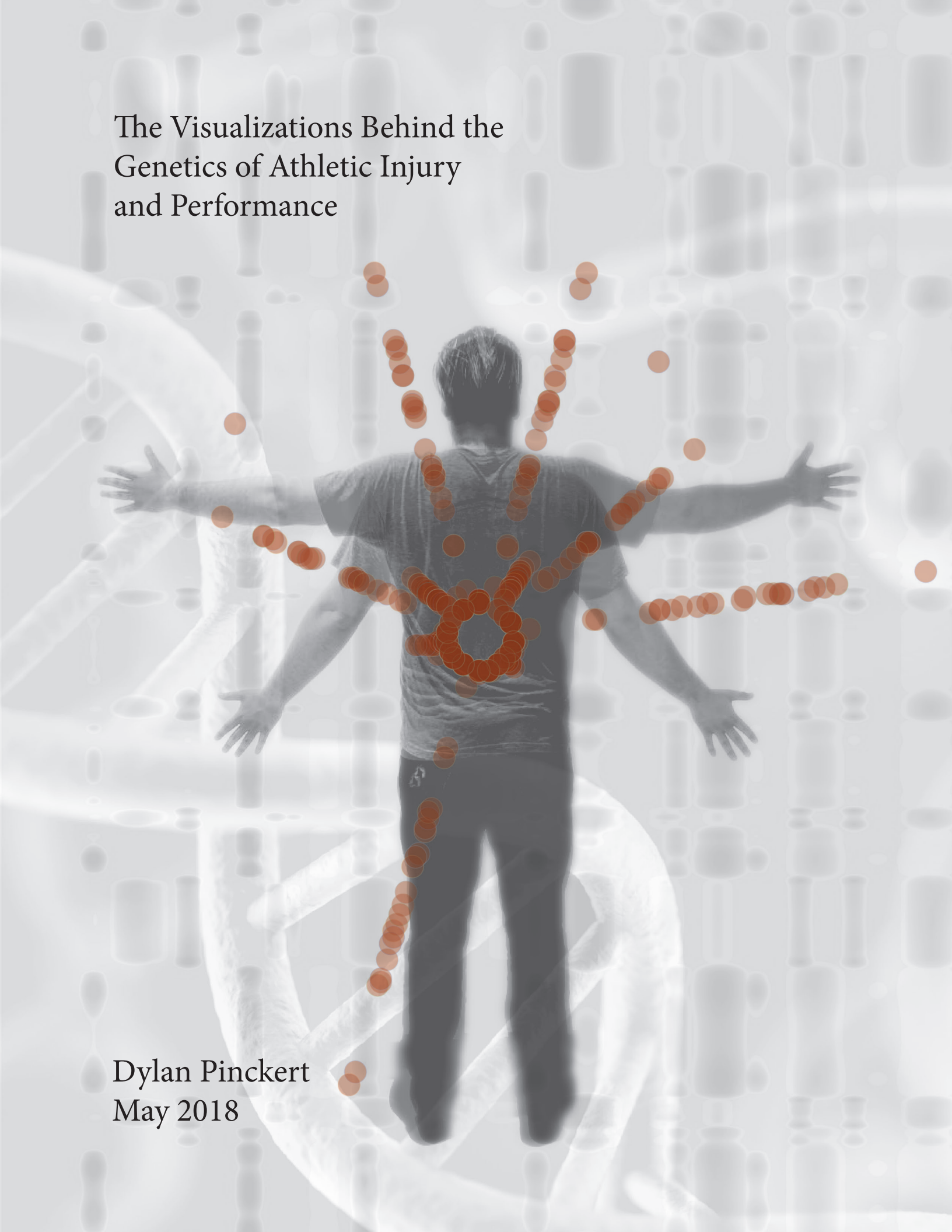
2018

The Visualizations Behind the Genetics of Athletic Injury and Performance

Dylan Pinckert
Connecticut College

The Visualizations Behind the Genetics of Athletic Injury and Performance



Dylan Pinckert
May 2018

Connecticut College
Art Department
Honors Thesis 2018

Faculty Advisers: Andrea Wollensak (Studio Art)
Kristine Hardeman (Biology)

Acknowledgments:

To my advisers for meeting and emailing with me constantly for encouragement and guidance

To my family for helping me every step of the way

Table Of Contents

Introduction

It all starts with numbers… Numbers have the power to create almost any response in a viewer in science. When numbers represent something, they take on a new name called Data. Data has many uses including, but not limited to, statistical analysis, political gains/losses, education, and discovery. Scientists use data to study and confirm their hypotheses in the name of discoveries and learning. They also use them to advance new concepts and ideas that will forge new research to help the greater good. To present new data, you must think about what you want to communicate.

A table of numbers may not tell the common viewer much of anything without an extensive background in the subject. Sources of the data and the context of its generation may not be apparent in a list or table. This is where data visualization comes into play. Data Visualization involves the production and application of the visual portrayal of data. The operation and practice of data visualization is very unique process and it can be entirely dependent on many different variables.

The difference from looking at a set of numbers versus a visualization is staggering. From a graph, one can see proportions, amounts, time, and space which lets the viewer better analyze and understand what they are viewing. The field of data visualization has been growing recently as the need to understand vast amounts of information has increased. Edward Tufte, considered the master of data visualization, has written and lectured extensively on what it means to visualize data and the need to represent data in accurate ways. The inspiration for this thesis came from his description of the 1986 Space Shuttle Challenger disaster. Tufte illustrates how an unfortunate approach to communicating large amounts of complex data obscured a simple aspect that could have saved the lives aboard

the shuttle craft (Tufte, 2005). This specific event and Tufte's analysis and my couriosity of combining both my biology and art majors brought on the intradisciplinary question for this thesis.

How can the complex information used in genetic analysis be visually represented to enhance a viewer's interpretation?

This thesis takes the concepts and theories put forth by many artists/designers and applies them to a very specific field, the application of genetics to sports performance and injuries. Like data visualization, the field of human genetics is rapidly expanding field due to the growing ability to identify very small parts of the human genetic structure. This creates an opportunity to apply principles of data visualization to communicating research in this field of science. The application of visualization principles to data regarding genetics in connection with athletic injury and performance is the foundation of the Art Installation Project *Genetic Visualizations of Athletic Performance and Injury*.

Interpretation, Analysis, and Complex Information

For this project, the guidelines put forth by Edward Tufte of how a graphical display should be designed were closely considered.

“Show the data (Tufte, 2007)”

This guidline means to say present the information without tainting it with other data that does not belong or anything from an outside source to curve interpretation. Upon obtaining the data for this project, there were data sets that came in addition to the intended information. To have the intended data in its final form, I had to pick out individual data points because it did not pertain to the project or the information I was targeting. Data visualization is the act of showing the data.

“Encourage the viewer to think about the substance rather than about the methodology, graphic design, or the technology of graphic production (Tufte, 2007)”

The viewer needs to think critically about what the information means. Does the data say that I will lose money? Does the graphic say that the population is growing or decreasing? The viewer should be asking these questions, rather than “look at that beautiful spiral with the different colors, I wonder what the lines mean…” This is a question that this project tries to avoid eliciting from a viewer. The actual data, which represents information that can be useful, is the medium of abstracted expression. Data can be manipulated to form different abstract forms and still be an accurate representation. The role of design is to be as objective as possible because there is a message that data gives, the role of the artist is to communicate that message.

“Avoid distorting what the data states (Tufte, 2007)”

This is a classic statement that can be easily applied to politics. Political science tends to run on data generated by the population at large. Great examples of misconstrued data have been presented by a a statistician named Stephanie Deviant. A couple of her examples included graphs from Fox News. One specific graph during the Obama era that incorrectly represented the country’s unemployment rate was called into question by many media sources. Media analysts then go into a deeper discussion of fake news that we will stay away from (Deviant, 2014). This specific graph had a data point incorrectly placed that would have showed the unemployment rate from the President Obama era actually decreasing significantly rather than staying stagnant (Deviant, 2014). This graph had a powerful pull on viewers because it showed that President Obama’s unemployment rate was stagnant which caused many to sway away from supporting him versus the actuality of a decreasing percentage which would have caused more of the public to support him. A cardinal rule for my work is that it must be accurate. If it can’t be accurate, then it is not worth presenting because it would be falsifying what the data is telling.

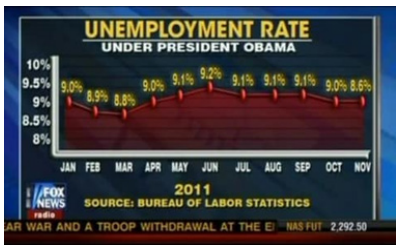


Figure 1: A graph misrepresenting data presented by *Fox News* (Deviant, 2014)

“Present many numbers in a small space (Tufte, 2007)”

There should be sufficient numbers to state the message. This can be a balancing act to keep from having an overwhelming visual, such as Figure 2. In the genetics of humans, there can be upwards of three billion data points. In this project, there were tens of thousands of data points brought down to a visualization size of inches. To successfully do this, I layered the data and separated it into different categories to show the vastness of the small information of genetic data.

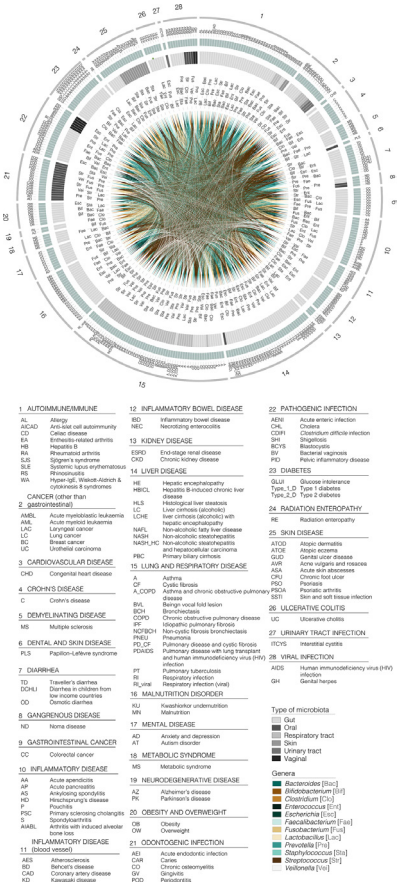


Figure 2: A graph put representing a summary of diseases and dissorders that have to do with environmental changes to make up the community of the microorganisms presented. There are numerous layers of information, breaking it down into smaller chucnks to attempt to make it legible (Rojo et al, 2017 page 45).

"Make large data sets coherent (Tufte, 2007)"

When one has a large excel file with tens of thousands of data points, it is a challenge to try and keep it together and make it all look like one set. It is imperative if the viewer is to understand the data that it makes sense as a whole. It is very easy to make the numbers look chaotic and incoherent with much information. An example of this is Figure 2. Rojo et al. presented an article that explored the entire human microbiome, or simply, the bacterial organisms that live on and inside humans. Their data was extremely extensive because their figure includes too many layers of information for people to understand. The figure includes too much data to be legible and needs to brought down to a smaller size to be legible. This was one of the challenges for my project, making the data as coherent as possible.

"Encourage the eye to compare different pieces of data (Tufte, 2007)"

Data is meant to be compared and analyzed so the viewer can make their own conclusions. In communicating quantitative data, correlations and context are important because viewers want to find that correlation and not look at chaos. It is human nature to find a pattern, and so I tried to look for patterns in the data to see if there was a large group or category that looked like there was a trend. which there was.

"Reveal the data at several levels of detail, from a broad overview to the fine structure (Tufte, 2007)"

This specific guideline can be interpreted in different ways because of the amounts of details data can contain. Details can be hard to express with extreme amounts of data, but they can also really bring out the beauty in the information. This can be done by categorizing the data and showing hierarchies of information to orient the viewer on what is big and what is small. This is what I did when I was categorizing my data.

I essentially cut all of the data into many pieces and made it possible to view in both as individual pieces and as a whole.

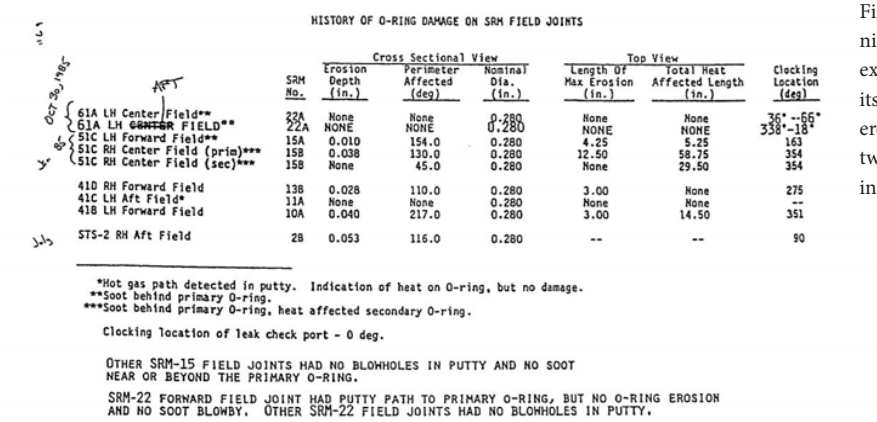
“Serve a reasonably clear purpose: description, exploration, tabulation, or decoration (Tufte, 2007)”

Each graphical presentation has a purpose, whether that be aesthetic, informational, or poetic. The intent must be evident. During the course of this project, many objectives were explored, such as education towards the common viewer, abstract design aesthetic, and poetic beauty. This specific guideline was hard to follow since the data obtained was so vast and could say more than just a few things.

“Be closely integrated with the statistical and verbal descriptions of a data set (Tufte, 2007)”

To make a graphic out of the data presented, it is imperative that the creator understands the jargon associated with the data in the graphic, if applicable, and in the data description. In many ways the designer of the graphic visualization becomes an interpreter between a technical field and a non-technical audience. This is the role of the artist/designer.

As mentioned, most of the inspiration for this project came from Tufte’s analysis of the 1986 Space Shuttle Challenger disaster. Tufte (2005) concluded that in the data NASA engineers had to review prior to making the decision on whether to launch the shuttle in freezing conditions, that the answer was obscured by the amount and form of the data. In his analysis, he included the pictures of the charts faxed to NASA before the launch (Figures 3 and 4).When looking at the numbers, it is hard for someone to understand what they mean. The representations of the data did not allow the NASA controllers to make a fully informed decision. If the information had been visually communicated in a clearer and less complex way, the decision to launch may have been



HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
22A	None	None	0.280	None	None	36° -66°
22A	NONE	NONE	0.280	NONE	NONE	338°-18°
15A	0.010	154.0	0.280	4.25	5.25	163
15B	0.038	130.0	0.280	12.50	58.75	354
15B	None	45.0	0.280	None	29.50	354
13B	0.028	110.0	0.280	3.00	None	275
11A	None	None	0.280	None	None	--
10A	0.040	217.0	0.280	3.00	14.50	351
2B	0.053	116.0	0.280	--	--	90

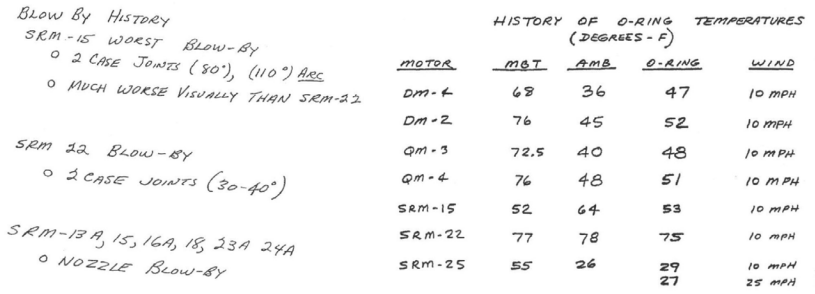
*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

Figure 3: One of the charts sent to NASA the night before Challenger was launched. Tufte explains how this specific table does not do its job in reporting the incident of field joint erosion on STS 61-C, which was launched two weeks before challenger's launch and not included in the chart (Tufte, 2005).



Blow By History
SRM-15 WORST Blow-By
- 2 Case Joints (80°), (110°) Arc
- Much worse visually than SRM-22

SRM 22 Blow-By
- 2 case joints (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A
- Nozzle Blow-By

HISTORY OF O-RING TEMPERATURES (DEGREES-F)

MOTOR	MBT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

Figure 4: Another chart sent to NASA hand written the night before the launch. The forecasted O-ring temperature for the challenger launch is at the bottom section under O-ring temperatures at 29° and 27° Fahrenheit where the "nozzle blow-by" occurred for those SRM rockets. The "blow by" is soot found on the rocket, which meant that pressurized gases flamed over the O-ring and reached one of the tanks causing damage to the rocket (Tufte, 2005).

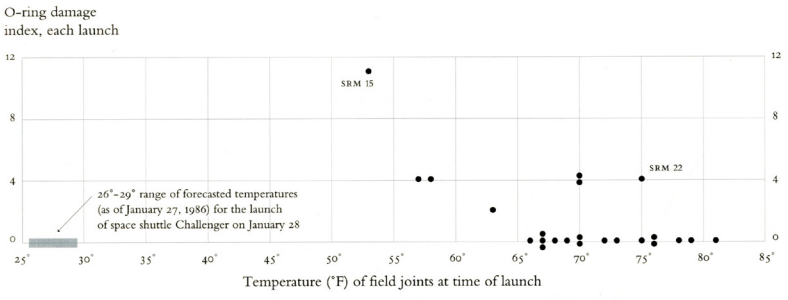


Figure 5: This is a scatterplot of all of the data of O-ring damage that was available, even with data that was not included on the table pictured in Figure 3. This graph clearly indicates a trend saying that the damage index of the O-ring rises as the temperature lowers. The forecasted temperature is indicated and it is obvious that the index, if measured in a test, would be exponentially higher (Tufte, 2005).

avoided (Tufte, 2005). Figure 5 illustrates how the data should have been presented and how it would have told NASA that the launch needed to be cancelled. A question that was constantly brought up in the process of this project: Can the biologic data be visualized abd represented in a way that will be understood successfully to a broad audience?

These data visualization principles and the drive to produce legible work has kindled my desire to produce graphical work that has to do with genetics. Genetics combined with athletic injury and performance research really enhances the unique potential of the data. Concepts of genetics applied to sports are hard to convey to the general public since there is so much understanding needed to interpret the data.

Because DNA is on the micro level and so loaded with information, the best method of showing genetic data is with DNA mapping. “DNA mapping portrays vast amounts of data for the code of life is very very long. To describe and understand DNA is an inherently high-resolution problem requiring high resolution displays (Tufte, 2010).” Recent DNA mapping has shown connections between the locations and occurrences of certain gene sequences and the prevalence of athletic injury. Using the same methodology, there have been connections made to relative strength of athletic performance. By applying Tufte’s principals to relationships revealed through DNA mapping, conclusions regarding aspects of athleticism is more clearly revealed to a non-technical audience.

There have been few artists who have attempted to bring complicated biology to the understanding of a common viewer, however, there are still the few. Natalie Miebach is an artist who specializes in 3-D and scientific visualizations. She uses weaving as her main medium and states in her artist statement that weaving is “a simple, yet highly effective grid.”

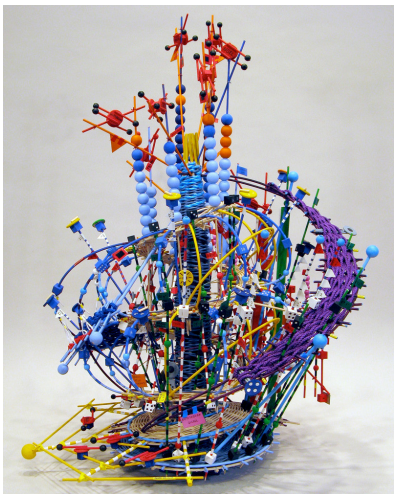


Figure 6: Natalie Miebach’s *Hurricane Noel*. The specific pieces of this sculpture map the different factors and elements that came from Hurricane Noel. Meibach also assigned each element with musical notes so the sculpture can actually be played (Sierzputowski, 2016).

She makes a very small divide between functionality and sculpture. Miebach specifically looks at the role visual aesthetics play in translation and understanding weather information. As demonstrated in Figure 6, Miebach's work looks to create 3-dimentional data that reaches out to create an abstract shape. My thesis and art work looks at similar ideas, however, my projects translate a more specific kind of data that is difficult to translate into a 3-dimensional form, while Meibach is more abstract and less true to scale .

Another example of a designer attempting to make visual representations of large amounts of data is Jake Barton. Barton is an artist who took a very difficult subject and brought it down to a level to where kids could understand. He took large amounts of data and information from synthetic biology and morphed it into hands-on activities understandable to the general public (Stinson, 2016). The interactive presentations are based on trial and error activities, "see what happens…Learn by doing which disguises that fact that visitors are working on some complicated stuff (Stinson, 2016)." This relates to my thesis project because both Barton's and my work result from sophisticated and complicated information that is aimed at becoming simpler. However, while Barton's project is highly interactive, my exhibit uses a set of connected displays that allow the audience to discover relationships within the genetic process.

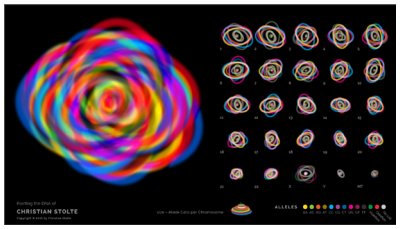


Figure 7: Christian Stolte's *Paint My DNA.* An abstract take on genetic information.

The artist with the most similar process to my thesis project is Christian Stolte. The methods Stolte uses are more abstract and colorful, but the data being worked with is similar (Stolte, 2016). Instead of using colors and abstract visualizations, this project changed aspects of graphs and added transparency to create a combined figure rather than creating blurred colored circles of different sizes. Stolte's work is also just 2-dimentional while this project utilizes both 2 and 3 dimensions.

Data: The Approach

The information used in this project comes from the growing field of applying genetics to athletic ability and injury and my own independent study. This research application is providing data regarding injury risk for professional athletes (Synovitz and Eshanova, 2014) (Williamson, 2014) and even everyday people or non athletes. The intention of this research is to provide athletes and the general public with the knowledge of their genetics to help prevent injury. However, this research has the potential to put many professional athletes out of a job depending on their specific genetics and the probability of injury. This can even go to the point of preventing athletes from achieving professional status because of their risk. There is still much research to be done considering the general population and not just elite athletes. (Please refer to Appendix 1 for further readings and information on the specific genes from my research addressed in this project.)

The biggest challenge in my project was bringing the data down to a manageable size. As noted earlier, genetic data can reach to 3 billion data points. Various data types were experimented with including codon data, allelic data, and chromosomal data. They all had massive data set sizes in common. After experimenting with the data as a whole, it was very clear that the data needed to be simplified after some viewers could not understand what was going on. Each data point gave the information of genotype for each gene. Genotype is the make up of an individual, which determines a specific characteristic, or phenotype, of that organism or individual (Griffiths, 2015). Initially, there were over 668,000 data points on a single Excel spread sheet with 3 categories for each one of the points. The data was divided into their respective chromosomes, with the X

and Y chromosomes being combined to be number 24. The chromosomes were then broken up into the genotype letter combinations with the combinations without a value removed. There were a total of 22 possible combinations. After breaking down this massive data into 24 categories and then 22 subcategories, the information was much easier to interpret and it was simpler to create graphs.

Many graphs were created and experimented with including radial charts, bar charts, scatter plots, maps and line graphs. Everything still looked technical and was not contributing to communicating the emotional aspects associated with revealing an individual human's basic genetic makeup. This was the point where abstraction was experimented with.

Different software was used to investigate with different variables, colors, transparencies, ratios, and viewpoints. I was particularly interested in transparencies because of the depth it gives to 2-dimensional images Three interrelated design concepts and projects came from this experimentation. The aspiration was to provide for the viewer a path to understanding the complexity of the genetic structure and a path to empathic connection with the artist that serves as the base data in this self-portrait.

The Projects:
Gene Maps of Athlete-Linked Genes

Figure 8: *Gene Maps of Athlete-Linked Genes*

Figure 8.1: Chromosome 9 in *Gene Maps of Athlete-Linked Genes.* These specific genes shown, called TNC and Col5A1 effect tendon and ligament health.

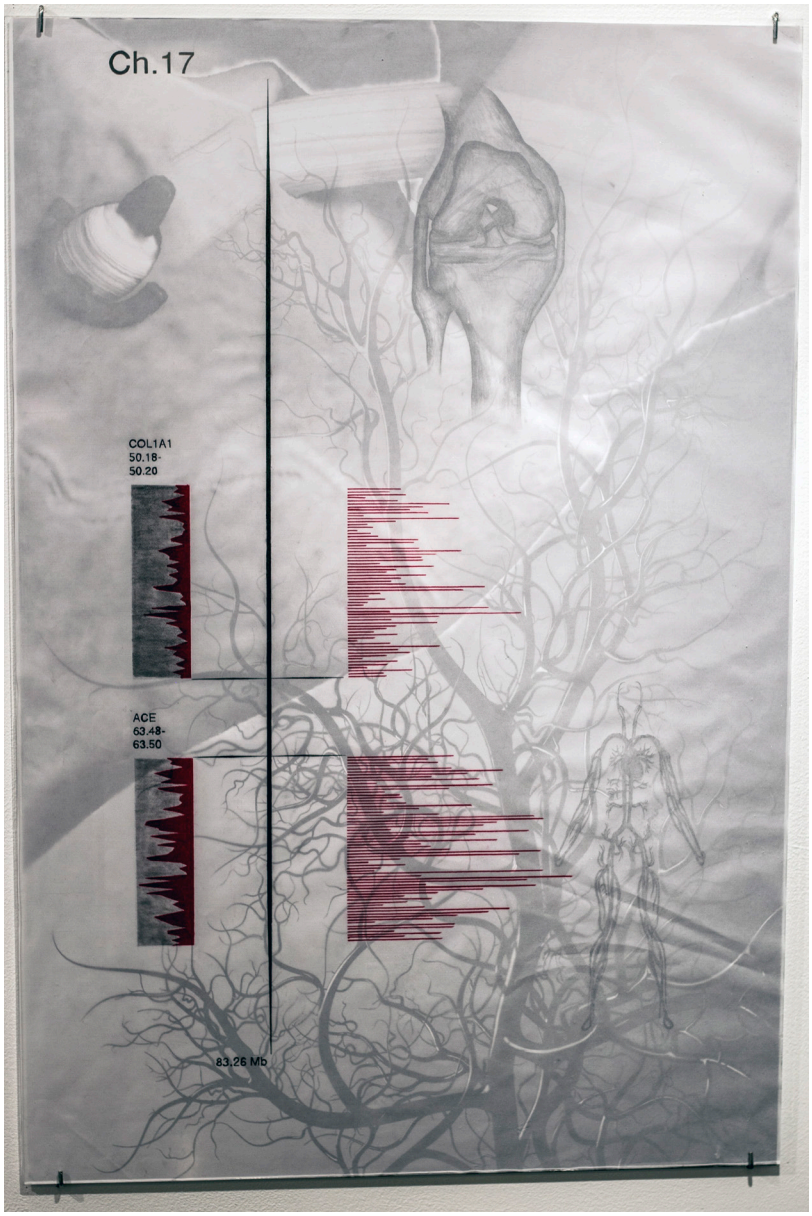


Figure 8.2: Chromosome 17 in *Gene Maps of Athlete-Linked Genes.* The Col1A1 gene shown on top effects the ACL, while the ACE gene effects cardio ability.

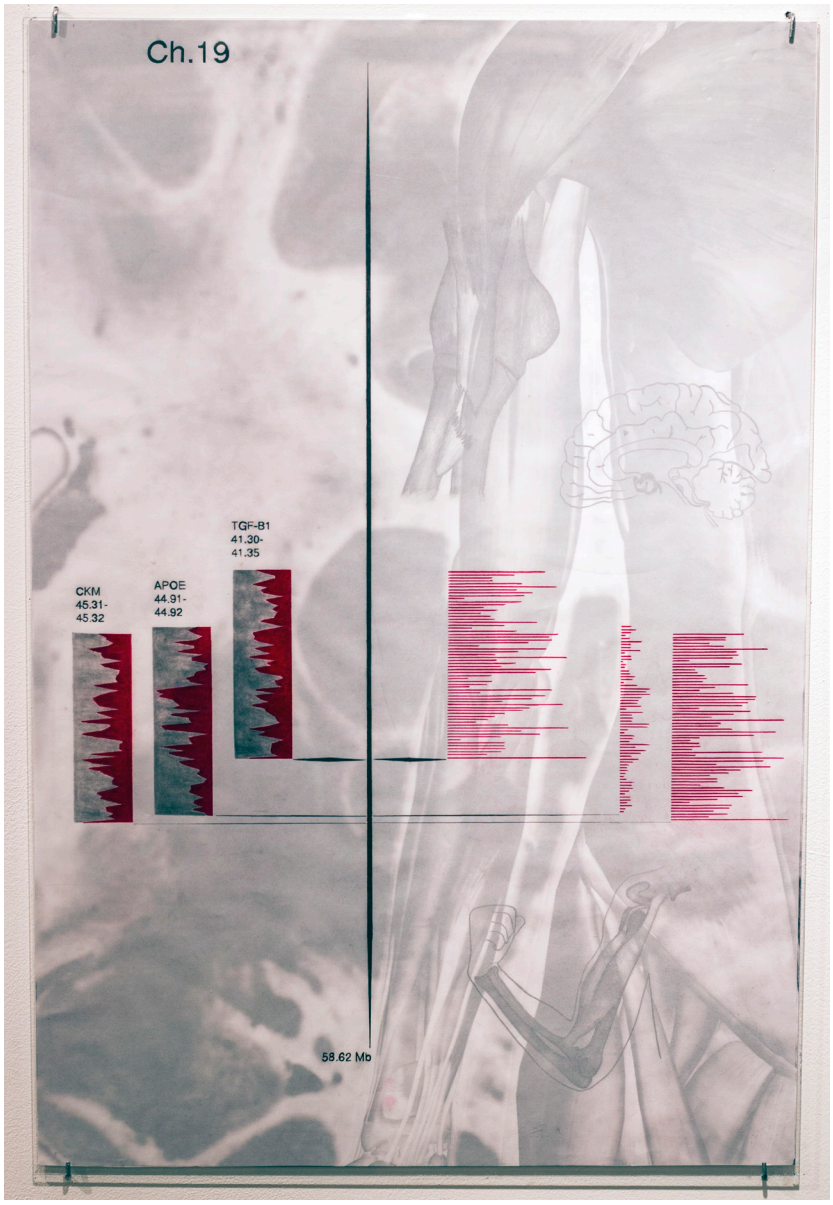


Figure 8.3: Chromosome 19 in *Gene Maps of Athlete-Linked Genes.* The three identified genes on this chromosome have very different functions. TGF-B1 effects tendon repair in the body. The APOE gene has been associated with concussions and the CKM gene produces an enzyme that effect power and stamina in the muscles.

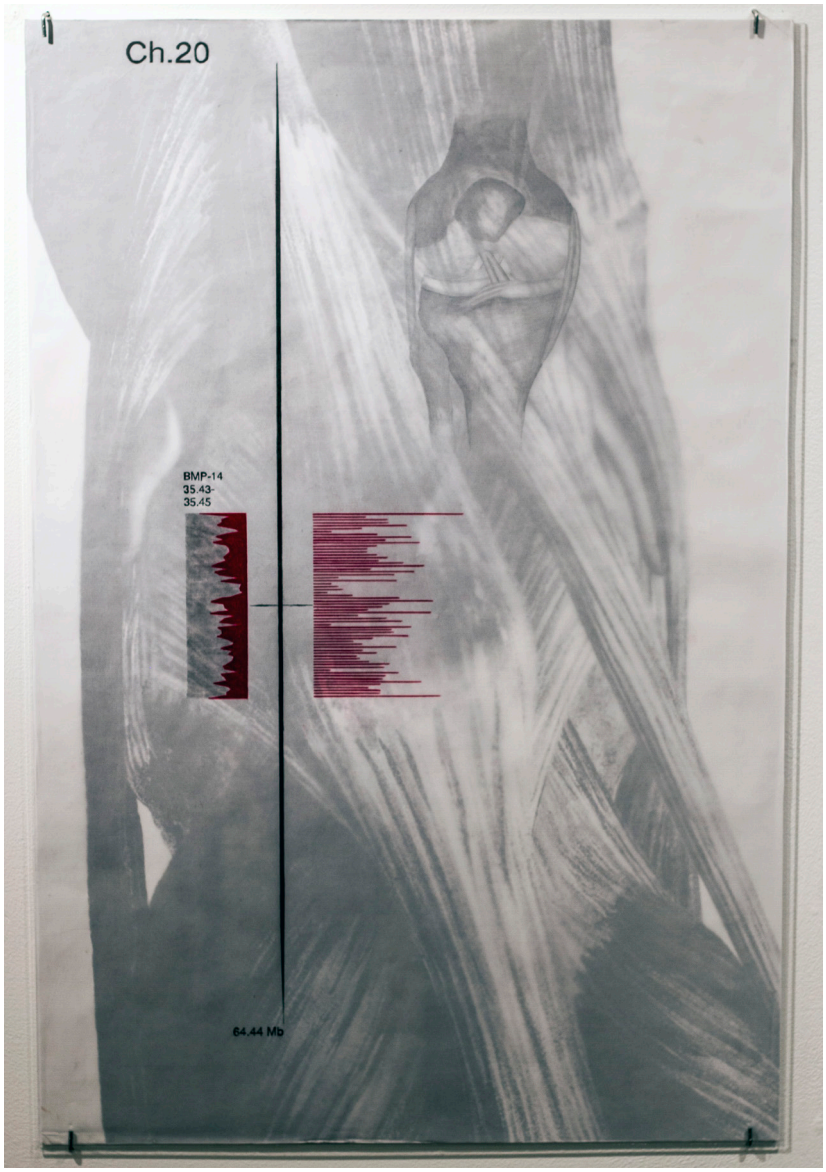


Figure 8.4: Chromosome 20 in *Gene Maps of Athlete-Linked Genes.* This specific gene, shown as BMP-14, has been shown to effect the strength and repair of tendons and lligaments.

Figure 8.5: Chromosome 22 in *Gene Maps of Athlete-Linked Genes.* This gene, identified as PPAR-a has a direct effect to heart heatlh, specifically the gowth ability of the left ventricle.

This series of drawings contain three components: a gene map with base data, illustrations pertaining to the map, and a background detail of a corresponding human organ. There are also three functions in this visualization: the series dispays proportions, distribution, and location.The map shows where certain genes from my research, shown in Appendix 1, are located on their respective chromosomes with their specific location under their name. The human organ illustrations show the locations where the gene affects the human body. In this map, there are two graphs. The line graph shows the codon counts of the gene, or the raw contents of the gene. A codon is a sequence of three nucleotides (single pieces of DNA) that together form a unit of genetic code in a DNA or RNA molecule (Griffiths, 2015).

Shown in detail in Figure 8.6, on the left of the map is an area graph showing each codon's percentage area in the gene. The shaded box containing the graph only represents 5% of the total area of the gene which really shows how massive the amount of information carried on a single gene can be. The height of the diamond shaped markers are indications of the gene's size in relation to the corresponding chromosome as well as where on the chromosome the genes are located. The color of the graphs is a dark red color because it resembles the color of blood which gives the drawings a life like characteristic and makes the graphs pop.

The chromosome map shows the complexity of the genetic data. In addition there is a pattern to be seen in the values to the other charts. The question here is: Does this help people interpret the data? Yes, because patterns inherently give a viewer a path to interpretation. The two most distinct patterns in the drawings are the comparable shapes between the area graph and the line graphs. The shapes are similar because the area graphs are percentages instead of absolute values and so

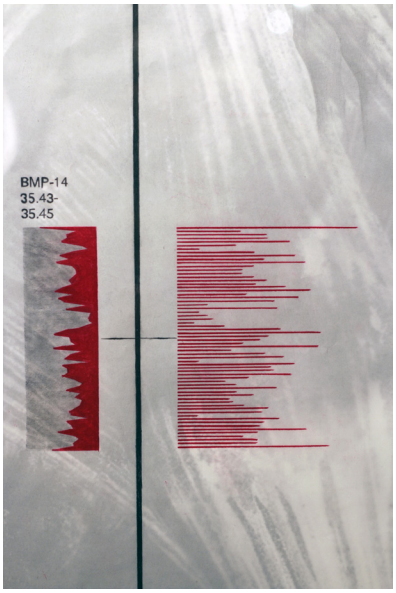


Figure 8.6: A detail of Chromosome 20 in *Gene Maps of Athlete-Linked Genes.* Under the name are numbers indicating position on the chromosome relative to the total number listed at the bottom of the chromosome (not shown). The units used are Mega Base Pairs (Mb). A Mega Base Pair is equal to one million base pairs. A base pair is the smallest component of DNA.

depict the solid shape of the line graphs. The other pattern is among the drawing as a whole. If the viewer looks carefully at the10th line though out the series, they will see that there is a similarly low value on all of the drawings. This shows that almost every gene in the cell carries some similar characteristics.

At the infancy of this part of the project, I explored what I could do by hand. This was because I was very much insistent on using my computer to generate graphs and visualizations for me to make everything look “perfect.” I came to the realization, however, that hand work and physical craft of the project was missing. It was about the human genome, but it was generated by a computer. I believe that adding the human touch to the project would show the human imperfections in this series of visualizations. Initially, I started out with just the drawings on a flat white background. It was very simple and not very pleasing to look at.To improve upon it, I started to think about what these genes actually effect and not what they were themselves. I wanted to keep the gene map thecentral focus of the pieces, so I drew the affected body parts behind the vellum to use the transparency of the material to create a heirarchy of data. They looked better, but it still looked like a textbook. I then took detail pictures of body parts, organs, and surgical procedures and edited them heavily to make them show through the transparency of the vellum. This is what gives the drawings depth.

The Projects:
The Central Dogma of Biology

Figure 9: *The Central Dogma of Biology*

Figure 9.1: *The Central Dogma of Biology: Transcription*

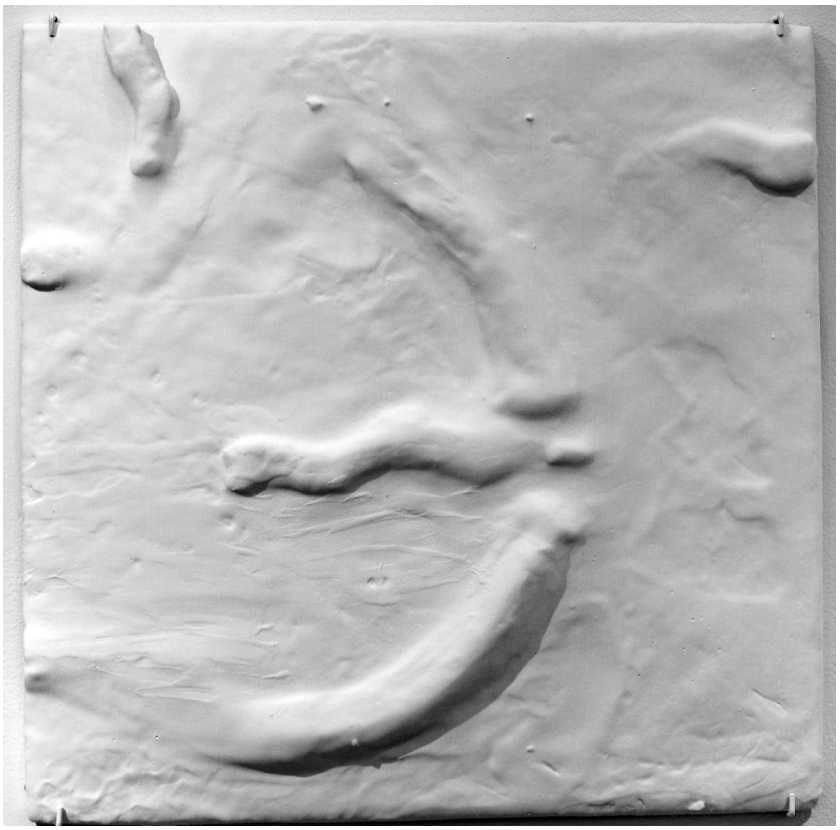


Figure 9.2: *The Central Dogma of Biology: RNA Leaving the Nucleus*

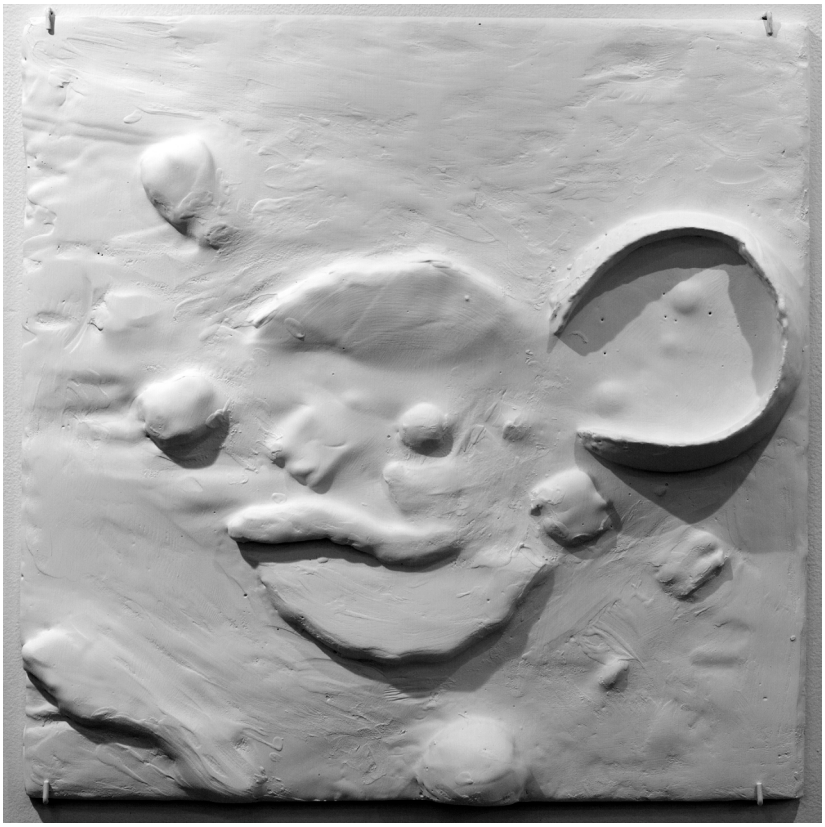


Figure 9.3: *The Central Dogma of Biology: Translation*

Figure 9.4: *The Central Dogma of Biology: Golgi Apparatus*

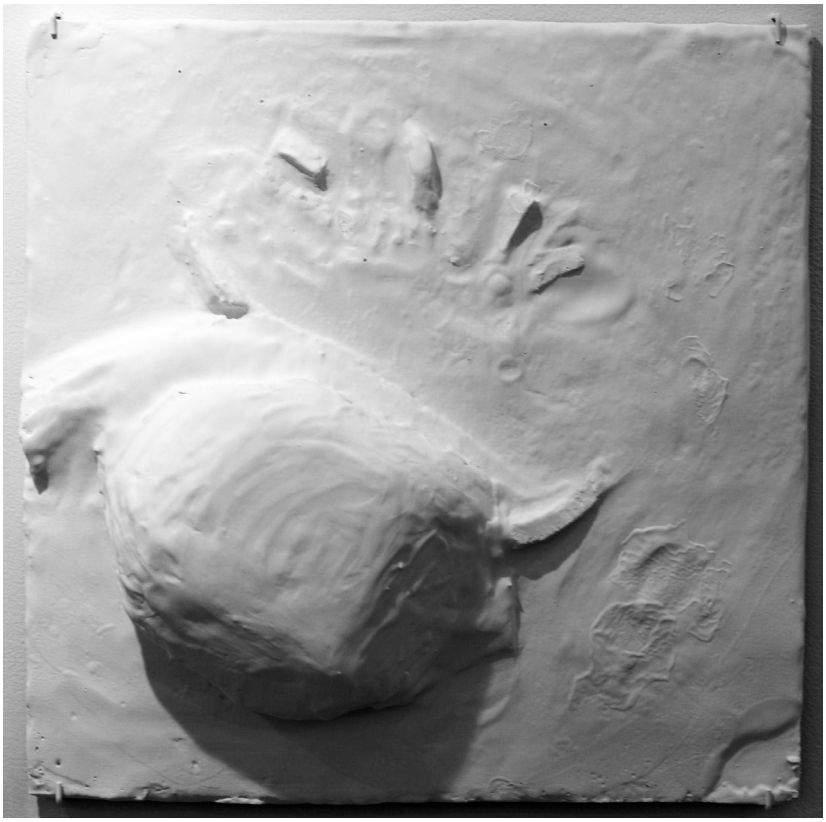


Figure 9.5: *The Central Dogma of Biology: Vesicle Release*

The central dogma of biology is simply the process of how cells use DNA to create functional proteins that the cell uses to survive. This process applies to every living organism known to humans. Some of the differences between organisms are, in fact, the mechanisms of this operation, the form the DNA is in, and the types of enzymes and proteins produced. This series of plaster panels illustrates that process in 5 steps in a moving flow manner. Styrofoam coated in plaster and complex compound sculpt the shapes in this process. This was done in plaster to illustrate the viscosity of the environment since it takes place inside a cell. It was important to show the qualities of the fluid environment in this process because everything happens in a very chaotic manner. The reliefs illustrate one concept with many processes with the added factor of movement and flow.

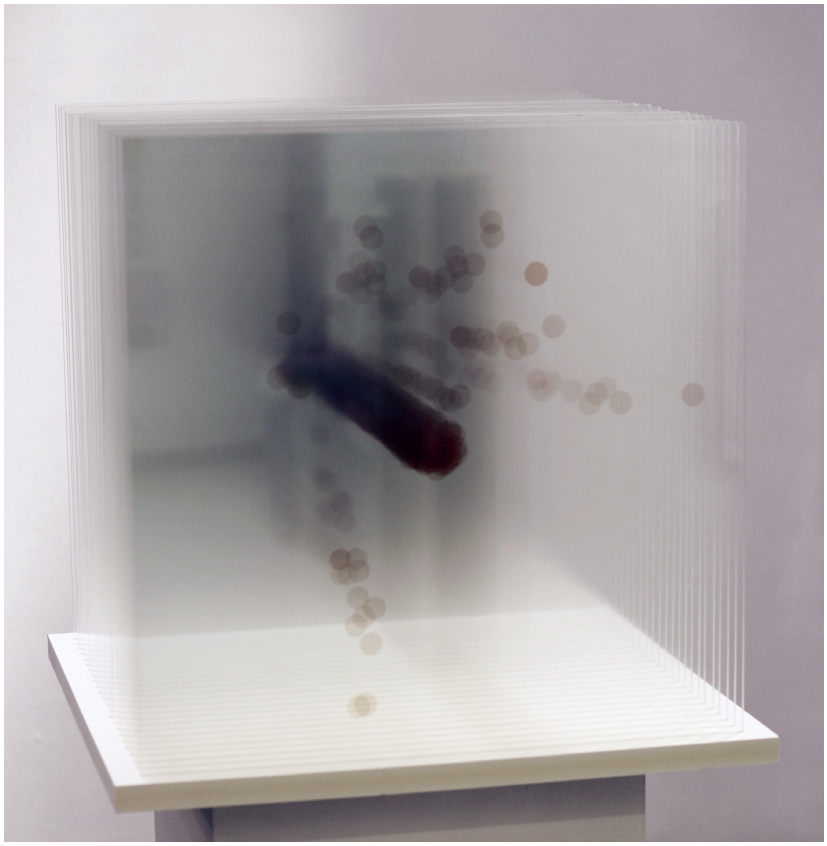
The process starts with the DNA transcription (Fig. 9.1). DNA transcription is when the raw DNA material is transcribed into a molecule called RNA which is essentially half of a DNA molecule with some molecules replaced for others (Griffiths, 2015). The second panel (Fig. 9.2) is RNA leaving the nucleus through a nuclear pore into the outer cell. In panel three (Fig. 9.3), translation takes place. Translation is when the RNA is translated into protein molecules (Griffiths, 2015). The fourth panel (9.4) illustrates the process the protein going through the Golgi apparatus, a series of warpped walls that adds to the proteins, to become a usable protein to the body. Lastly, the protein makes its way to the cell membrane via a protein transport, called a vesicle, and out into the body (Fig 9.5).

Plaster was chosen as a medium because it represents a 3-dimensional world on a flat wall in relief. There are organic shapes and flowing liquids that show the different substances and dimensions inside a cell. The shapes were created by cutting and sculpting pieces of styrofoam and pasting the

shapes on a piece of flat drywall. The drywall was then covered with plaster, not once, but multiple times to get the consistency of liquid. I then painted the plater panels the same white as the wall it is mounted on. It would not be correct to represent the work as a completely flat plane because the cells in the human body function with cytoplasm inside them. This specific project presents an encapsulating concept that applies to the other projects presented. This work represents the central dogma of biology and supports the other works.

The Projects:
Plexigram: A Self Portrait

Figure 10: *Plexigram: A Self Portrait*



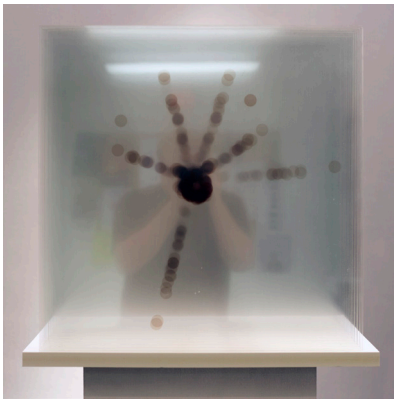


Figure 10.1: A frontal view of *Plexigram: A Self Portrait*

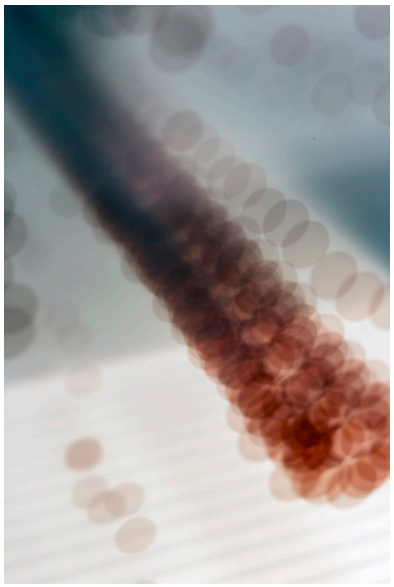


Figure 10.2: An angled view of *Plexigram: A Self Portrait*

The data used in this endeavor is my own DNA genotype data taken from a DNA sequencing service. During my experimentation and explorations of graphs on computer software, I focused on sparkline graphs, or microcharts. These graphs are small graphics designed to give a quick representation of numerical values, which represents the data more directly since the graph shows just values. Shown in Figures 10 through 10.2, the graphs on these plexiglass slides are sparkline graphs that were created to show how many genotypes appear on each of the 23 pairs (24 is the data from X and Y combined) of chromosomes in my genome. Stacking these plexiglass slides illustrate a data visualization concept called small multiples (Tufte, 2013). Tufte defines the concept of small multiples as a series of similar graphs or charts using the same scale and axes, allowing them to be easily compared (Tufte, 2013). Individually, the slides provide some general information and specific information for the corresponding chromosome. When each slide is combined with each other, the information can be viewed as a whole and as a complete work or by extension, the many sections together describe a complete form of me. Also when combined, there is an image of a hand that can be distinguished, which was completely unintentional. The color and transparency keep the occurring theme from the drawings and also give a sense that the circles on the graphs are blood samples. The transparency works well with this project because of the small multiples creating a distinguished image with all of the combined slides.

This specific project was inspired initially by John Cage and his piece “Not Wanting to Say Anything About Marcel.” His artwork was created by layering pieces of transparent plexiglass squares with information on them, specifically text. The text and graphics give a sense of linear time as the images become more obscured by the addition of more squares in the array.

Conclusion:

While the work of Edward Tufte on data visualization was an inspiration for this project, another part or layer of synthesis and fusion was required to create a piece that could communicate and illustrate the structure of the chosen scientific data accurately as well as provide visual richness and emotional engagement enough for a viewer to be involved with the piece. There were several techniques adapted from the Tufte guidelines that presented interesting, readable, and structured displays of the genetic data from my genotype, but it took examining a number of alternatives before the pieces seemed to transcend numeric data and move into a satisfying self-portrait. Emotional engagement doesn't necessarily flow directly from a set of numbers.

The sheer amount of numbers became difficult to deal with without significant computer assistance. The categorization and manipulation of the data into a form that could be worked with visually was very time consuming. Data sets on this scale require sophisticated computer processing tools or an extensive pre-process timeframe. It became very satisfying when eventually useful patterns within the data structure became apparent in some specific charting. When these patterns came into focus I found some freedom to integrate another level of engagement in the pieces by manipulating scale, medium, and technique. I developed the hand drawings to bring in a human touch to the numerical data while maintaining accurate depictions. I used the hand molded sculptures to portray the organic nature of the content.

The most difficult aspect was in finding the third piece of the triad to go back to a holistic image of my self-portrait. Discovering patterns led me to the notion of adding a third

dimension to the visualization with transparency. A sequence of charts on the wall seemed to emphasize the bits of information and not to speak to the whole person metaphor I wanted. I experimented with transparencies of bar graphs that produced interesting overlapping city-scapes but they were not evocative of the biologic data I was dealing with. By adapting a radial set of coordinates for the charted data a more organic image began to take place. It was very satisfying when another anthropomorphic image of a human hand began to emerge from the data. The hand of my Genotype.

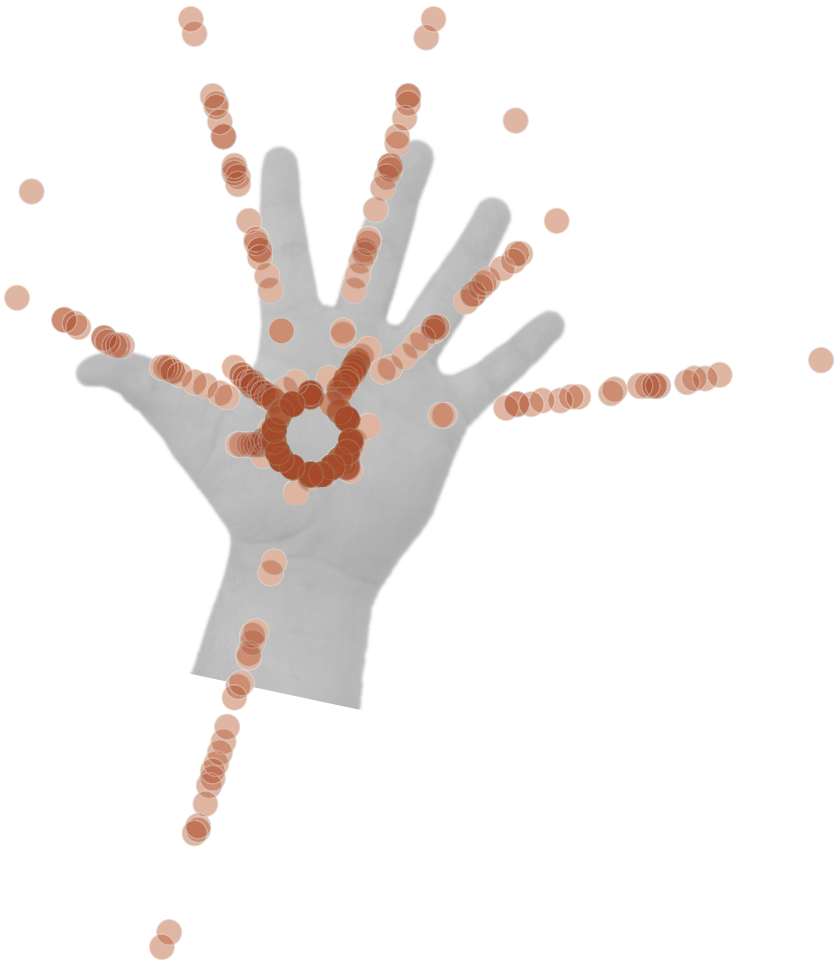


Figure 11: The combined slides from *Plexigram: A Self Portrait* with my hand.

Appendix 1: Research From Pinckert, 2017

The following genes have been identified as having connections to specific athletic injury:

ACE

a. Polymorphism in intron 16 gene location 17q23.3 (I/D) [Ahmetov, 2015]

b. Angeotesin-1 converting enzyme, part of the renin-angiotesin system responsible for controlling blood pressure [Guth & Roth, 2013]

. c. Was discovered in 1998 to be associated with human performance. 287 pair insertion is associated with lower serum and tissue ACE activity [Guth & Roth, 2013]

e. I/I associated with endurance and D/D associated with strength [Guth & Roth, 2013]

i. In a study of swimmers, II, ID, and DD were all significantly different [Tsianos et al, 2004]

ii. (II/ID/DD) Short distances = 6%/47%/47%
Long distances= 18.8%/75%/6.2%

f. Knowledge of ACE may guide an athlete as to which practice or sport suits them [maffulli et al, 2013]

ACTN3

a. Polymorphism that leads to a premature stop codon (X) rather than arginine (R) at position 577 [Guth & Roth, 2013] Location 11q13.1 [Ahmetov, 2015]

b. Codes for the protein α-actin-3, structural sacromeric protein found only in fast type II muscle fibers [Guth & Roth, 2013]

c. Protein helps to form a lattice structure that anchors together actin-containing thin filaments and stabilizes the muscle contractile apparatus [Ahmetov, 2015]

i. Mouse model provides mechanistic insights to humans, loss of αactin-3 in mouse results in a shift in a variety of properties of fast fibers towards those characteristic of slower fibers [Yang et al, 2009]

Collagen related genes

a. COL5A1 location 9q34.2-q34.3 encodes the pro-α1 chain of type V collagen and is the rate limiting component of the trimer assembly. (rs 12722 C/T polymorphism) [Ahmetov, 2015]

i. In a study of Exercise associated muscle cramping, EAMC, COL5 A1 CC genotype was significantly overrepresented (P=0.031) among the control group when compared to the group with self-reported EAMC history, which shows that genotype TC + TT are a positive contributor to risk of history of EAMC [O'Connell et al, 2013]

ii. CC genotype underrepresented in in female participants with anterior cruciate ligament ruptures, which means females with CC genotype had a decreased risk of ACL ruptures [Posthumus et al, 2009]

b. COL1A1 gene contains a functional Sp1 binding site polymorphism, rs1800012 location 17q.21.33, which is a G to T substitution. [Collins et al., 2010]

i. Associated with cruciate ruptures, shoulder dislocations, ACL ruptures, and Achilles tendon ruptures [Collins et al, 2010]

TNC

a. The tenascin-C gene location: 9q32-q34 encodes for a structural component of tendons [Mokone, 2006]

b. Related to ABO blood group, in close proximity to tenascin-C area, also has an impact on tendon injury and ideal candidate genetic marker of tendon injury for the following reasons [Mokone, 2006]:

i. Several investigators have suggested that either the ABO gene or a closely linked gene(s) on the tip of the long arm of chromosome 9 may be associated with Achilles tendon injuries

ii. The TNC gene has been mapped to chromosome 9q32-q34, which is in close proximity to the ABO gene

iii. The gene encodes for tenascin-C, which is an important structural component of tendons

iv. Tenascin-C expression is regulated by me chanical stimuli and altered during tendon injury. The TNC gene contains a guanine thymine (GT) dinucleotide repeat polymorphism (a tandem repeat consisting of a repeated 2–base pair sequence of varying lengths in different people) within intron 17 (an intervening DNA sequence within a gene which does not encode for a protein).

TGF-B1

a. Shown to be a stimulant for collagen production. Location 19q41.3-41.35 [Hou et al., 2009]

b. “Procollagen” shown, when locally injected, to promote repair with procollagen types 1 and 3 [Hou et al, 2009]

c. Accelerates collagen protein synthesis, crosslink information, and matrix remodeling in tendon healing. [Hou et al, 2009]

MMP3

a. Matrix metallopeptidase 3 gene, have regulatory roles that in maintaining extracellular matrix homeostasis. [Maffulli et al, 2013]

b. Associated with achilies tendinopathy and collagen genes and TNC [Raleigh et. al, 2009]

i. Along with COL5A CC genotype, MMP3 A allele had the lowest risk of AT [Raleigh et al, 2009].

BMP-14 and GDF5

a. Bone morphogenic protein 14 [Bolt et al, 2007] and Growth and differentiation factor 5. BMP-14 and GDF5 are closely related. [Chapman et al, 2008]

b. It was shown that BMP-14 increases tendon tensile strength in rats of the Achilles tendon using florescent proteins to identify, improves tendon healing at the site of repair [Bolt et al, 2007]

c.GDF5 location 20q 11.22, has been shown to play a role in joint formation and articular cartilage homeo stasis. Rs143383 T allele is associated with increased risk of Osteoarthritis [Chapman et al, 2008]

APOE

a. APOE, apoliprotein E, gene location 19q13.2 has 3 alleles ε2, ε3, and ε4. Most common is ε3 found in more than 50% of world population [Terrell et al, 2008]

i. Found that there was a nearly 3 time increase in risk of history of cuncussions for thos with APOE TT genotype relative to GG genotype [Terrell et al, 2008]

CKM

a. Location 19q13.2-13.3, Creatine kinase enzyme bound specifically to the M line of the myofibril sub fragment. [Zhou et. al, 2006]

i. May change Ca2+ uptake and power of muscle, type 1 and 2 (slow and fast twitch muscle) have been shown to have differing CKM activity [Zhou et al, 2006]

ii. All variables measured decreased significantly in AG allele vs AA and GG after training. [Zhou et al, 2006]

PPAR-a

a. Peroxisome proliferator activated receptor α. Regulates genes responsible for myocardial fatty acid oxidation and is downregulated during cardiac hypertrophy, concomitant with the switch from fatty acid to glucose utilization

References:

Ahmetov II, Fedotovskaya ON. 2015. Current progress in sports genomics. Advances in Clinical Chemistry. 70:247.

BioDesign Studio,The Tech Museum of Innovation [Internet] [2/20/18]. Available from: https://localprojects.net/work/biodesign_studio .

Bolt P, Clerk AN, Luu HH, Kang Q, Kummer JL, Deng Z, Olson K, Primus F, Montag AG, He T, et al. 2007. BMP-14 gene therapy increases tendon tensile strength in a rat model of achilles tendon injury. Journal of Bone and Joint Surgery (American). 89(6):1315-20.

Cage J. 1969. Not wanting to say anything about marcel. Plexiglas.

Chapman K, Takahashi A, Meulenbelt I, Watson C, Rodriguez-Lopez J, Egli R, Tsezou A, Malizos KN, Kloppenburg M, Shi D, et al. 2008. A meta-analysis of european and asian cohorts reveals a global role of a functional SNP in the 5′ UTR of GDF5 with osteoarthritis susceptibility. Human Molecular Genetics. 17(10):1497-504.

Collins M, Posthumus M, Schwellnus MP. 2010. The COL1A1 gene and acute soft tissue ruptures. British Journal of Sports Medicine. 44(14): 1063-4.

Misleading Graphs: Real Life Examples [Internet]; c2014 [cited 2018 3/16/]. Available from: http://www.statisticshowto.com/misleading-graphs/ .

Griffiths A, Wessler SR, Carroll SB, Doebley J. 2015. Introduction to genetic analysis 11th ed. New York, NY: W.H. Freeman and Company.

Guth LM, Roth SM. 2013. Genetic influence on athletic performance. Current Opinion in Pediatrics. 25(6):653.

Hou Y, Mao Z, Wei X, Lin L, Chen L, Wang H, Fu X, Zhang J, Yu C. 2009. The roles of TGF-beta1 gene transfer on collagen formation during achilles tendon healing. Biochemical and Biophysical Research Communications. 383(2):235.

Levin G. 2013. Golan levin (US) - THE BIG PICTURE symposium - EN The Big Picture Symposium. .

Maffulli N, Margiotti K, Longo UG, Loppini M, Fazio VM, Denaro V. 2013. The genetics of sports injuries and athletic performance. Muscles, Ligaments and Tendons Journal. 3(3):173.

Miebach, Nathalie [Internet] 2/20/18]. Available from: http://nathaliemiebach.com/portfolio.html .

Mokone GG, Schwellnus MP, Noakes TD, Collins M. 2006. The COL5A1 gene and achilles tendon pathology. Scandinavian Journal of Medicine & Science in Sports. 16(1):19-26.

O'Connell K, Posthumus M, Schwellnus M, Collins M. 2013. Collagen genes and exercise-associated muscle cramping. Clinical Journal of Sport Medicine. 23(1):64-9.

Pinckert D. 2017. The genetics behind athletic injury. Connecticut College.

Posthumus M. 2009. The COL5A1 gene is associated with increased risk of anterior cruciate ligament ruptures in female participants. The American Journal of Sports Medicine. 37(11):2234-40.

Raleigh SM, van der Merwe L, Ribbans WJ, Smith RKW, Schwellnus MP, Collins M. 2009. Variants within the MMP3 gene are associated with achilles tendinopathy: Possible interaction with the COL5A1 gene. British Journal of Sports Medicine. 43(7):514-20.

Stinson E. 2016. Synthetic biology is … complex. but this exhibit makes it a blast. Wired.

Stolte, Christian. Paint my DNA [Internet]; c2016 2/20/18]. Available from: http://science-unseen.siggraph.org/stolte/ .

Rojo D, Mendez-Garcia C, Anna Raczkowska B, Bargiela R, Moya A, Ferrer M, Barbas C. 2017. Exploring the human microbiome from multiple perspectives: Factors altering its composition and function. Federation of European Microbiological Sciences. 41:453-78.

Synovitz R, Eshanova Z. 2014. Uzbekistan is using genetic testing to find future olympians. The Atlantic;.

Terrell T, Bostick R, Abramson R, Xie D, Barfield W, Cantu R, Stanek M, Ewing T. 2008. APOE, APOE promoter, and tau genotypes and risk for concussion in college athletes. Clinical Journal of Sport Medicine. 18(1):10-7.

Tsianos G, Sanders J, Dhamrait S, Humphries S, Grant S, Montgomery H. 2004. The ACE gene insertion/deletion polymorphism and elite endurance swimming. Eur J Appl Physiol. 92(3):360-2.

Tufte ER. 2013. Envisioning information. 14. printing ed. Cheshire, Conn: Graphics Press.

Tufte ER. 2010. Beautiful evidence. 3. print. ed. Chesire, Conn: Graphics Press.

Tufte ER. 2007. The visual display of quantitative information. 2. ed., 5. print. ed. Cheshire, Conn: Graphics Press.

Tufte ER. 2005. Visual explanations. 7. printing, with rev. ed. Cheshire, Conn: Graphics Press.

Williamson L. 2014. Two premier league clubs sign up with top genetics company to learn DNA profiles of players. MailOnline;.

Yang N, Garton F, North K. 2009. A-actinin-3 and performance. Genetics and Sports. 54:88.

Zhou DQ, Hu Y, Liu G, Gong L, Xi Y, Wen L. 2006. Muscle-specific creatine kinase gene polymorphism and running economy responses to an 18-week 5000-m training programme. British Journal of Sports Medicine. 40(12):988.