Exploring Genetics Through Genetic Disorders

Teacher Guide and Answer key

The information here will help you assign Allele Profiles and evaluate students' Lab Notebooks and other work. Within each disorder, the **alleles with lower numbers are easiest**, and difficulty goes up with higher numbers. The alleles for **hemophilia and hemoglobin disorders are more challenging** than the rest.

All the alleles are based on real ones described in the literature (references are listed at the end). We changed the names of the alleles, mainly so we could streamline the text for students. The tables show the names that are used by the scientific community and a brief description. We suggest the names of the protein symbols to help evaluate questions 5 and 6 in the Lab Notebooks, though there is often more than one reasonable answer.

Alpha-1 Antitrypsin Deficiency

Autosomal recessive inheritance pattern—except for D6, which is autosomal dominant.

Allele	Molecular changes	Difficulty and notes
D1 – M (Mineral Springs)	Single-base substitution. AA #91 changed from G to E	Easy. Non-working protein is degraded & not released to the blood stream. High risk of lung damage.
D2 – W (Bethesda)	Single-base substitution. AA #360 changed from A to T	Easy. Non-working protein is degraded and not released to the blood stream. High risk of lung damage.
D3 – Z	Single-base substitution. AA #366 changed from E to K	Medium. Non-working protein forms clumps that build up inside liver cells; proteins don't get to lungs. High risk of lung & liver damage.
D4 – M (Malton)	Three-base deletion. One AA is deleted at #75/76 (both are F)	Medium. Non-working protein forms clumps that build up inside liver cells; proteins don't get to lungs. High risk of lung & liver damage.
D5 – NULL (Mattawa)	Single-base insertion. Causes a frameshift beginning at AA #376 & premature termination at #400.	Medium; the DNA/amino acid data is a little harder to interpret. Non-working protein is degraded, so it isn't released into the blood stream. High risk for lung damage.
D6 – Pittsburgh	Single-base substitution. AA #382 changed from M to R	Bonus/Advanced. This gain-of-function allele does not cause lung or liver problems, but a bleeding disorder! Rather than targeting elastin, the modified protein targets the blood clotting protein thrombin. Having one allele causes the disorder.

The Z allele (aka D3) is by far the most common disorder-causing allele of SERPINA1. Another common allele, S, is not included in this activity. The S allele codes for a version of AAT protein with reduced function that differs from the healthy protein by one amino acid. The S allele is fairly common, but it generally causes AAT deficiency only when it is in combination with a Z or null allele.

Cystic Fibrosis

Autosomal recessive inheritance pattern

In the literature, *CFTR* alleles are usually referred to by a code that indicates the position of the change in the amino acid sequence and the type of change.

Allele	Molecular changes	Difficulty and notes
C1 – G542X	Single-base substitution. AA #542 changed from G to A, creating a premature termination codon.	Easy. Cells make very small amounts of non-working protein. Causes severe effects in the lungs and digestive organs.
C2 – F508del	Three-base deletion. AA #508 is deleted.	Easy. Non-working protein is degraded before it reaches the cell membrane. Causes severe effects in the lungs and digestive organs.
C3 – G551D	Single-base substitution. AA #551 changed from G to D.	Easy. Non-working protein is made and moved to the cell membrane, but it does not move chloride ions. Causes severe effects in the lungs and digestive organs.
C4 – R1070W	Single-base substitution. AA #1070 changed from R to W.	Medium. This mild allele codes for a partially working protein that moves fewer chloride ions. When it's combined with a severe allele, it sometimes affects the lungs. The digestive organs are usually unaffected.
C5 – A455E	Single-base substitution. AA #455 changed from A to E.	Medium. This moderate allele codes for a partially working protein that still moves chloride ions. But because the protein is degraded quickly, there is less of it than usual. When it's combined with a severe allele, it usually affects the lungs. Effects on the digestive organs are usually mild or absent.

The F508del allele is the most common disorder-causing allele of CFTR. It's present in >70% of people with the disorder. Most of the work done on R1070W and A455E is done with F508del as the second allele. To make the data more understandable, we estimated the sweat chloride levels for R1070W / R1070W and A455E / A455E, based on papers that studied protein function.

Hemoglobin Disorders

These disorders are a little on the tricky side. Variations in the HBB gene can cause several distinct disorders, and the symptoms and molecular mechanisms vary widely. To understand their alleles, students will need to read and process more information.

Most alleles of the *HBB* gene cause genetic disorders that follow an autosomal recessive inheritance pattern (AR), and others cause genetic disorders that follow an autosome dominant pattern (AD). Note that the first three alleles all affect the same codon, which is numbered 6 or 7 depending on the reference.

Allele	Molecular changes	Difficulty and notes
HB1 – HbS AR and AD inheritance	Single-base substitution. AA #7 changed from E to V.	Medium. Altered protein still carries oxygen but forms stiff fibers. Causes sickle cell disease (AR inheritance) and malaria resistance (AD inheritance). Sickled cells get caught in small blood vessels, causing pain, anemia, organ damage & enlarged spleen.
HB2 – HbC AR inheritance	Single-base substitution. AA #7 changed from E to K.	Medium. Altered protein still carries oxygen, but forms stiff fibers under some conditions. Protein also breaks down faster than usual so blood cells are shorter-lived. Causes sickle cell disease and resistance to severe malaria. Sickled cells can get caught in small blood vessels, causing mild anemia & enlarged spleen.
HB3 – Glu6FS Co-dominant inheritance	Single-base deletion in the codon for AA #7. Causes a frame shift, with a stop codon 12 codons later.	Medium. Non-working or no protein. Few red blood cells mature, leading to beta-zero thalassemia. Homozygotes require blood transfusions for life and are at risk for iron overload. Some heterozygotes have anemia.
HB4 – HbE AR inheritance	Single-base substitution. AA #27 changed from E to K.	Medium+: Partially working protein is slowly made and short-lived. Some homozygotes never have negative effects, others have mild beta-thalassemia. Conveys resistance to severe malaria.
HB5 – E121 to TER AD inheritance	Single-base substitution. AA #122 changed to a premature stop codon.	Advanced: Non-working protein cannot pair with alpha-globin. Unpaired globins clump together, and red blood cells do not mature. Causes beta-zero thalassemia. Two alleles would likely be lethal.
HB6 – Hemoglobin Denver AD inheritance	Single-base substitution. AA #42 changed from F to S, affecting oxygen binding.	Advanced: Partially working protein binds weakly to oxygen, and it is broken down more quickly than usual. Having one allele causes an oxygen transport disorder. Two alleles would likely be lethal.

HbS is the most common allele in sickle cell disease. HbC and HbE, because they also contribute to malaria resistance, are also quite common. The other alleles are rare.

Hemophilia

X-linked recessive inheritance pattern. In the literature, *F8* and *F9* alleles are referred to by a code that indicates the position of the change in the amino acid sequence and the type of change. The amino acid numbering system changed around the year 2000. We use the current system, though some publications still use "Legacy" numbering. See reference section for details.

Gene, Allele	Molecular changes	Difficulty and notes
H1 – *F8*, A415V	Single-base substitution. AA#415 is changed from A to V. (Legacy AA#396)	Easy. Partially working protein leads to a weakened clotting signal. Causes moderate to severe hemophilia. Risk of severe bleeding and joint damage.
H2 – *F9*, R449Q	Single-base substitution. AA#449 is changed from R to Q. (Legacy #449)	Easy. Partially working protein. Less protein is made, and some of it is degraded. A small amount of protein gets into the blood works and fairly well. Causes mild hemophilia with delayed blood clotting.
H3 – *F9*, L19F*1	Single-base deletion in the codon for AA#19. Causes a frame shift, with a stop 1 codon later.	Easy. Non-working or no protein. A very early stop codon causes only a tiny fragment of protein to be made. Causes severe hemophilia. High risk for spontaneous bleeding and joint damage.
H4 – *F8*, A303E	Single-base substitution. AA#303 is changed from A to E. (Legacy AA#284)	Medium. Partially working protein binds poorly to its partner and leads to a weakened clotting signal. Sometimes causes mild hemophilia with delayed blood clotting.
H5 – *F9*, F55I	Single-base substitution. AA#55 is changed from F to I. (Legacy #9)	Medium. Partially working protein. This allele affects a second interaction the protein needs to modify its target. Causes mild to moderate hemophilia with some risk of joint damage.
H6 – *F8*, intron 22 inversion	Chromosomal inversion. A piece of chromosome broke away, rotated 180 degrees, and fused back in place. The inverted portion includes the promoter & start co-don—but not the stop.	Bonus/Advanced. This is quite different than the molecular changes for the other alleles. There is a lot to read and process about the rearrangement. Understanding why the sequence of this particular inversion is read in the correct direction may require some knowledge of gene regulation. No working protein reaches the blood stream, causing severe hemophilia.

The H6 allele is the most common allele in hemophilia. It's found in nearly 50% of all severe cases. Many of the other hemophilia alleles are the result of "founder effect," and a specific allele can be traced back to a common ancestor. In different parts of the world, new mutations occurred, creating new alleles that were passed on. This is why different hemophilia alleles tend to be more common in people with ancestors from different places. For example, the H2 allele is one of the most common alleles in people with French ancestry, but it is rare in other populations.

Marfan Syndrome

Autosomal dominant inheritance pattern

In the literature, *FBN1* alleles are usually referred to by a code that indicates the position of the change in the amino acid sequence and the type of change.

Allele	Molecular changes	Difficulty and notes
M1 – G1013R	Single-base substitution. AA #1013 changed from G to R	Easy. Partially working proteins don't link together very well and break down easily. Causes neonatal Marfan syndrome, the most severe form.
M2 – C2686F	Single-base substitution. AA #2686 changed from C to F	Easy. Partially working proteins are released from cells in small quantities. Altered proteins also interfere with the release of healthy proteins. High risk for all effects, especially lens dislocation.
M3 – I1892X	Five-base insertion. Causes a frameshift beginning at AA #1747, and a premature termination at #1892	Medium. Partially working proteins are made but rarely built into connective tissue. The altered proteins interfere with healthy proteins. High risk for effects in many different tissues.
M4 – R2726W	Single-base substitution. AA #2726 is changed from R to W	Medium. Non-working proteins are completely left out of microfibrils. Altered proteins do not interfere with healthy proteins, though there is less fibrillin protein than usual overall. Low risk for the most serious effects. Often only skeletal features.
M5 – C1564S	Single-base substitution. AA #1564 is changed from C to S	Medium. Partially working protein becomes part of connective tissue, but causes clumping. Microfibrils wrap around elastin incompletely, which stiffens stretchy connective tissue. Low risk for heart and skeletal effects. High risk for stiff skin syndrome (hard, thick skin).

Most people with Marfan syndrome have an allele that is unique to their family. In fact, only about 10% of alleles are shared by another family. With the exception of neonatal Marfan syndrome, few connections have been made between the type of allele a person has and the severity of the symptoms they experience.

Alpha-1 Antitrypsin Deficiency

Allele D1 – Codons 89–96

```
Healthy:    UCC CUG GGG ACC AAG GCU GAC ACU  mRNA
             S   L   G   T   K   A   D   T   protein
Disorder:   UCC CUG GAG ACC AAG GCU GAC ACU  mRNA
             S   L   E   T   K   A   D   T   protein
```

Allele D2 – Codons 355–362

```
Healthy:    AAG GCC GUG CAU AAG GCU GUG CUG  mRNA
             K   A   V   H   K   A   V   L   protein
Disorder:   AAG GCC GUG CAU AAG ACU GUG CUG  mRNA
             K   A   V   H   K   T   V   L   protein
```

Allele D3 – Codons 363–370

```
Healthy:    ACC AUC GAC GAG AAA GGG ACU GAA  mRNA
             T   I   D   E   K   G   T   E   protein
Disorder:   ACC AUC GAC AAG AAA GGG ACU GAA  mRNA
             T   I   D   K   K   G   T   E   protein
```

Allele D4 – Codons 73–80

```
Healthy:    AAU AUC UUC UUC UCC CCA GUG AGC  mRNA
             N   I   F   F   S   P   V   S   protein
Disorder:   AAU AUC UUC UCC CCA GUG AGC      mRNA
             N   I   F   S   P   V   S       protein
```

Allele D5 – Codons 376–380, 389–400

```
Healthy:    UUU UUA GAG GCC AUA CCC AUU GAA    mRNA
             F   L   E   A   I   P   I   E     protein
Disorder:   UUU UUU AGA GGC CAU ACC CAU UGA A  mRNA
             F   F   R   G   H   T   H  stp    protein
```

Allele D6 – Codons 378–385

```
Healthy:    GAG GCC AUA CCC AUG UCU AUC CCC  mRNA
             E   A   I   P   M   S   I   P   protein
Disorder:   GAG GCC AUA CCC AGG UCU AUC CCC  mRNA
             E   A   I   P   R   S   I   P   protein
```

Cystic Fibrosis

Allele C1 – Codons 537–544

Healthy:	GAC	AAU	AUA	GUU	CUU	GGA	GAA	GGU	mRNA
	D	N	I	V	L	G	E	G	protein
Disorder:	GAC	AAU	AUA	GUU	CUU	UGA	GAA	GGU	mRNA
	D	N	I	V	L	stp			protein

Allele C2 – Codons 504–511

Healthy:	GAA	AAU	AUC	AUC	UUU	GGU	GUU	UCC	mRNA
	E	N	I	I	F	G	V	S	protein
Disorder:	GAA	AAU	AUC	AUC.	GGU	GUU	UCC		mRNA
	E	N	I	I	. G	V	S		protein

Allele C3 – Codons 549–556

Healthy:	AGU	GGA	GGU	CAA	CGA	GCA	AGA	AUU	mRNA
	S	G	G	Q	R	A	R	I	protein
Disorder:	AGU	GGA	GAU	CAA	CGA	GCA	AGA	AUU	mRNA
	S	G	D	Q	R	A	R	I	protein

Allele C4 – Codons 1067–1074

Healthy:	GCC	UUC	GGA	CGG	CAG	CCU	UAC	UUU	mRNA
	A	F	G	R	Q	P	Y	F	protein
Disorder:	GCC	UUC	GGA	UGG	CAG	CCU	UAC	UUU	mRNA
	A	F	G	W	Q	P	Y	F	protein

Allele C5 – Codons 450–457

Healthy:	AGA	GGA	CAG	UUG	UUG	GCG	GUU	GCU	mRNA
	R	G	Q	L	L	A	V	A	protein
Disorder:	AGA	GGA	CAG	UUG	UUG	GAG	GUU	GCU	mRNA
	R	G	Q	L	L	E	V	A	protein

Hemoglobin Disorders

Allele HB1 — Codons 5–12

Healthy:	ACU	CCU	GAG	GAG	AAG	UCU	GCC	GUU	mRNA
	T	P	E	E	K	S	A	V	protein
Disorder:	ACU	CCU	GTG	GAG	AAG	UCU	GCC	GUU	mRNA
	T	P	V	E	K	S	A	V	protein

Allele HB2 — Codons 5–12

Healthy:	ACU	CCU	GAG	GAG	AAG	UCU	GCC	GUU	mRNA
	T	P	E	E	K	S	A	V	protein
Disorder:	ACU	CCU	AAG	GAG	AAG	UCU	GCC	GUU	mRNA
	T	P	K	E	K	S	A	V	protein

Allele HB3 — Codons 5–9, 18–20

Healthy:	ACU	CCU	GAG	GAG	AAG	AAG	GUG	AAC	mRNA
	T	P	E	E	K	K	V	N	protein
Disorder:	ACU	CCU	GGG	AGA	AGA	AGG	UGA	AC	mRNA
	T	P	G	R	R	R	stp		protein

Allele HB4 — Codons 24–31

Healthy:	GUU	GGU	GGU	GAG	GCC	CUG	GGC	AGG	mRNA
	V	G	G	E	A	L	G	R	protein
Disorder:	GUU	GGU	GGU	AAG	GCC	CUG	GGC	AGG	mRNA
	V	G	G	K	A	L	G	R	protein

Allele HB5 — Codons 112–119

Healthy:	CAC	UUU	GGC	AAA	GAA	UUC	ACC	CCA	mRNA
	H	F	G	K	E	F	T	P	protein
Disorder:	CAC	UUU	GGC	AAA	UAA	UUC	ACC	CCA	mRNA
	H	F	G	K	stp				protein

Allele HB6 — Codons 38–45

Healthy:	UGG	ACC	CAG	AGG	UUC	CUU	GAG	UCC	mRNA
	W	T	Q	R	F	F	E	S	protein
Disorder:	UGG	ACC	CAG	AGG	UCC	CUU	GAG	UCC	mRNA
	W	T	Q	R	S	F	E	S	protein

Hemophilia

Allele H1 — Codons 410–417

Healthy:	GAG	GAC	UGG	GAC	UAU	GCU	CCC	UUA	mRNA
	E	D	W	D	Y	A	P	L	protein
Disorder:	GAG	GAC	UGG	GAC	UAU	GUU	CCC	UUA	mRNA
	E	D	W	D	Y	V	P	L	protein

Allele H2 — Codons 446–453

Healthy:	AAG	GUA	UCC	CGG	UAU	GUC	AAC	UGG	mRNA
	K	V	S	R	Y	V	N	W	protein
Disorder:	AAG	GUA	UCC	CAG	UAU	GUC	AAC	UGG	mRNA
	K	V	S	Q	Y	V	N	W	protein

Allele H3 — Codons 15–22

Healthy:	AUC	ACC	AUC	UGC	CUU	UUA	GGA	UAU	mRNA
	I	T	I	C	L	L	G	Y	protein
Disorder:	AUC	ACC	AUC	UGC	UUU	UAG	GAU	AUC	mRNA
	I	T	I	C	F	stp			protein

Allele H4 — Codons 300–307

Healthy:	CAU	CGC	CAG	GCG	UCC	UUG	GAA	AUC	mRNA
	H	R	Q	A	S	L	E	I	protein
Disorder:	CAU	CGC	CAG	GAG	UCC	UUG	GAA	AUC	mRNA
	H	R	Q	E	S	L	E	I	protein

Allele H5 — Codons 52–59

Healthy:	UUG	GAA	GAG	UUU	GUU	CAA	GGG	AAC	mRNA
	L	E	E	F	V	Q	G	N	protein
Disorder:	UUG	GAA	GAG	AUU	GUU	CAA	GGG	AAC	mRNA
	L	E	E	I	V	Q	G	N	protein

Allele H6 — Codons 2141–2143, 2156–2160

Healthy:	ACC	UUA	AUG	GUC	UUC	UUU	GGC	AAU	mRNA
	T	L	M	V	F	F	G	N	protein
Disorder:	ACC	UUA	AUG	UCC	UAC	CGC	UGG	UGA	mRNA
	T	L	M	S	Y	R	W	stp	protein

Marfan Syndrome

Allele M1 — Codons 1,010–1,017

Healthy:	AGA	GGA	CCC	GGA	UUU	GCC	ACA	AAA	mRNA
	R	G	P	G	F	A	T	K	protein
Disorder:	AGA	GGA	CCC	AGA	UUU	GCC	ACA	AAA	mRNA
	R	G	P	R	F	A	T	K	protein

Allele M2 — Codons 2,683–2,690

Healthy:	CAA	GGG	CAC	UGU	GUU	UCU	GGA	AUG	mRNA
	Q	G	H	C	V	S	G	M	protein
Disorder:	CAA	GGG	CAC	UUU	GUU	UCU	GGA	AUG	mRNA
	Q	G	H	F	V	S	G	M	protein

Allele M3 — Codons 1,746–1,749 and 1,889–1,892

Healthy:	CUC	UGU	GGA	AGU	UUG	GAC	AUA	AAU	mRNA
	L	C	G	S	L	D	I	N	protein
Disorder:	CUA	CAC	UCU	GUG	UGG	ACA	UAA	AUG	mRNA
	L	H	S	V	W	T	stp		protein

Allele M4 — Codons 2,723–2,730

Healthy:	UAC	CCC	AAA	CGG	GGC	AGG	AAA	CGG	mRNA
	Y	P	K	R	G	R	K	R	protein
Disorder:	UAC	CCC	AAA	UGG	GGC	AGG	AAA	CGG	mRNA
	Y	P	K	W	G	R	K	R	protein

Allele M5 — Codons 1,561–1,568

Healthy:	UCC	UGC	UGC	UGU	UCU	CUG	GGU	AAA	mRNA
	S	C	C	C	S	L	G	K	protein
Disorder:	UCC	UGC	UGC	UCU	UCU	CUG	GGU	AAA	mRNA
	S	C	C	S	S	L	G	K	protein

Key — mRNA and protein sequences

Students will get the most from this unit if they have the correct information in the Mutation & Alleles section of their Lab Notebooks. Print the following pages, and give each student a strip of paper with the information for their allele. They can use it to check their answers and correct any errors.

References

Alpha-1 Antitrypsin Deficiency

Brode, S. K., Ling, S. C., & Chapman, K. R. (2012). Alpha-1 antitrypsin deficiency: a commonly overlooked cause of lung disease. *Canadian Medical Association Journal, 184*(12), 1365-1371

Crystal (1990)?

Ferrarotti, I., Thun, G. A., Zorzetto, M., Ottaviani, S., Imboden, M., Schindler, C., ... & Probst-Hensch, N. M. (2012). Serum levels and genotype distribution of α1-antitrypsin in the general population. *Thorax, 67*(8), 669-674.

Hua, B., Fan, L., Liang, Y., Zhao, Y., & Tuddenham, E. G. (2009). α1-antitrypsin Pittsburgh in a family with bleeding tendency. haematologica, 94(6), 881-884.

Lewis, J. H., Iammarino, R. M., Spero, J. A., & Hasiba, U. (1978). Antithrombin Pittsburgh: an alpha1-antitrypsin variant causing hemorrhagic disease. *Blood, 51*(1), 129-137.

Owen, M. C., Brennan, S. O., Lewis, J. H., & Carrell, R. W. (1983). Mutation of antitrypsin to antithrombin: α1-antitrypsin Pittsburgh (358 Met –> Arg), a fatal bleeding disorder. *New England Journal of Medicine, 309*(12), 694-698.

Information about the SERPINA1 alleles came from the following sources (accessed March 2018):

- Online Mendelian Inheritance in Man (OMIM), entry 107400
- DNA and amino acid sequences were accessed through UniProt, entry P01009

  Note that the first 24 amino acids make up a signal peptide that is later cleaved to make the mature protein. Some sources number the amino acids according to their position in the mature protein (excluding the signal peptide). We have numbered them here according to their position relative to the translation start codon (including the signal peptide).

Cystic Fibrosis

Bompadre, S. G., Sohma, Y., Li, M., & Hwang, T. C. (2007). G551D and G1349D, two CF-associated mutations in the signature sequences of CFTR, exhibit distinct gating defects. *The Journal of General Physiology, 129*(4), 285-298.

Cebotaru, L., Rapino, D., Cebotaru, V., & Guggino, W. B. (2014). Correcting the cystic fibrosis disease mutant, A455E CFTR. *PloS One, 9*(1), e85183.

Farrell, P. M., & Koscik, R. E. (1996). Sweat chloride concentrations in infants homozygous or hetero-

zygous for F508 cystic fibrosis. *Pediatrics*, 97(4), 524-528.

Krasnov, K. V., Tzetis, M., Cheng, J., Guggino, W. B., & Cutting, G. R. (2008). Functional studies of rare missense mutations in CFTR facilitate interpretation of genotype-phenotype relationships. *Human Mutation, 29*(11), 1364.

Will, K., Dörk, T., Stuhrmann, M., Hardt, H. V. D., Ellemunter, H., Tümmler, B., & Schmidtke, J. (1995). Transcript analysis of CFTR nonsense mutations in lymphocytes and nasal epithelial cells from cystic fibrosis patients. *Human Mutation, 5*(3), 210-220.

Information about the *CFTR* alleles came from the following sources (accessed August 2018):

- https://cftr2.org/
- DNA and amino acid sequences were accessed through UniProt, entry P13569.

Hemoglobin Disorders

Forget, B. G., & Bunn, H. F. (2013). Classification of the disorders of hemoglobin. *Cold Spring Harbor perspectives in medicine, 3*(2), a011684.

Gerald, P. S., & Efron, M. L. (1961). Chemical studies of several varieties of Hb M. *Proceedings of the National Academy of Sciences, 47*(11), 1758-1767.

Kohne, E. (2011). Hemoglobinopathies: clinical manifestations, diagnosis, and treatment. *Deutsches Ärzteblatt International, 108*(31-32), 532.

Lohani, N., Bhargava, N., Munshi, A., & Ramalingam, S. (2018). Pharmacological and molecular approaches for the treatment of β-hemoglobin disorders. *Journal of Cellular Physiology, 233*(6), 4563-4577.

Origa, R. (2017). β-Thalassemia. *Genetics in Medicine, 19*(6), 609.

Patrinos, G. P., Giardine, B., Riemer, C., Miller, W., Chui, D. H., Anagnou, N. P., ... & Hardison, R. C. (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic acids research, 32*(suppl_1), D537-D541.

Thein, S. L., Hesketh, C., Taylor, P., Temperley, I. J., Hutchinson, R. M., Old, J. M., ... & Weatherall, D. J. (1990). Molecular basis for dominantly inherited inclusion body beta-thalassemia. *Proceedings of the National Academy of Sciences, 87*(10), 3924-3928.

Information about HBB alleles came from the following sources (accessed January 2019):

- Online Mendelian Inheritance in Man, entry 141900. https://www.omim.org/entry/141900
- HbVar: A Database of Human Hemoglobin Variants and Thalassemias. http://globin.bx.psu.edu/hbvar/menu.html
- DNA and amino acid sequences were accessed through UniProt, entry P68861. https://www.uniprot.org/uniprot/P68871

  Note that there are two schemes for numbering the amino acids. One counts the first methionine,

which is not part of the mature protein, as position one. The other counts the first amino acid in the mature protein as one. We used the former.

Hemophilia

Grant, M. A., Baikeev, R. F., Gilbert, G. E., & Rigby, A. C. (2004). Lysine 5 and phenylalanine 9 of the factor IX ω-loop interact with phosphatidylserine in a membrane-mimetic environment. *Biochemistry*, 43(49), 15367-15378.

Hakeos, W. H., Miao, H., Sirachainan, N., Kemball-Cook, G., Saenko, E. L., Kaufman, R. J., & Pipe, S. W. (2002). Hemophilia A mutations within the factor VIII A2-A3 subunit interface destabilize factor VIIIa and cause one-stage/two-stage activity discrepancy. T*hrombosis and Haemostasis*, 87(05), 781-787.

Ivaskevicius, V., Jurgutis, R., Rost, S., Müller, A., Schmitt, C., Wulff, K., ... & Oldenburg, J. (2001). Lithuanian haemophilia A and B registry comprising phenotypic and genotypic data. *British Journal of Haematology*, 112(4), 1062-1070.

Kurachi, S., Pantazatos, D. P., & Kurachi, K. (1997). The carboxyl-terminal region of factor IX is essential for its secretion. *Biochemistry*, 36(14), 4337-4344.

Lassalle, F., Marmontel, O., Zawadzki, C., Fretigny, M., Bouvagnet, P., & Vinciguerra, C. (2018). Recurrent *F8* and *F9* gene variants result from a founder effect in two large French haemophilia cohorts. *Haemophilia*, e213-e221.

Rudzki, Z., Duncan, E. M., Casey, G. J., Neumann, M., Favaloro, E. J., & Lloyd, J. V. (1996). Mutations in a subgroup of patients with mild haemophilia A and a familial discrepancy between the one‐stage and two‐stage factor VIII: C methods. *British Journal of Haematology*, 94(2), 400-406.

Sauna, Z. E., Lozier, J. N., Kasper, C. K., Yanover, C., Nichols, T., & Howard, T. E. (2015). The intron-22–inverted *F8* locus permits factor VIII synthesis: explanation for low inhibitor risk and a role for pharmacogenomics. *Blood*, 125(2), 223-228.

Information about the *F8* and *F9* alleles came from the following sources (accessed December 2018):

- http://www.factorviii-db.org/index.php
- http://www.factorix.org/
- DNA and amino acid sequences were accessed through UniProt, entries P00451 (*F8*) and P00740 (*F9*).

  Note that the first amino acids make up a signal peptide that is later cleaved to make the mature protein. This stretch is 19 amino acids long for coagulation factor VIII and 46 amino acids long for coagulation factor IX. Some sources use a "Legacy" designation that numbers amino acids starting after the signal peptide. We numbered them according to their position relative to the translation start codon, which includes the signal peptide.

Marfan Syndrome

Jensen, S. A., Aspinall, G., & Handford, P. A. (2014). C-terminal propeptide is required for fibrillin-1 secretion and blocks premature assembly through linkage to domains cbEGF41-43. *Proceedings of*

the National Academy of Sciences, *111*(28), 10155-10160.

Jensen, S. A., Iqbal, S., Bulsiewicz, A., & Handford, P. A. (2015). A microfibril assembly assay identifies different mechanisms of dominance underlying Marfan syndrome, stiff skin syndrome and acromelic dysplasias. *Human Molecular Genetics*, *24*(15), 4454-4463.

Kirschner, R., Hubmacher, D., Iyengar, G., Kaur, J., Fagotto-Kaufmann, C., Bromme, D., ... & Reinhardt, D. P. (2011). Classical and neonatal Marfan syndrome mutations in fibrillin-1 cause differential protease susceptibilities and protein function. *Journal of Biological Chemistry*, jbc-M111.

Milewicz, D. M., Grossfield, J., Cao, S. N., Kielty, C., Covitz, W., & Jewett, T. (1995). A mutation in FBN1 disrupts profibrillin processing and results in isolated skeletal features of the Marfan syndrome. *Journal of Clinical Investigation*, *95*(5), 2373-2378.

Raghunath, M., Superti-Furga, A., Godfrey, M., & Steinmann, B. (1993). Decreased extracellular deposition of fibrillin and decorin in neonatal Marfan syndrome fibroblasts. *Human Genetics, 90*(5), 511-515.

Schrijver, I., Liu, W., Odom, R., Brenn, T., Oefner, P., Furthmayr, H., & Francke, U. (2002). Premature termination mutations in FBN1: distinct effects on differential allelic expression and on protein and clinical phenotypes. *American Journal of Human Genetics, 71*(2), 223-237.

Schrijver, I., Liu, W., Brenn, T., Furthmayr, H., & Francke, U. (1999). Cysteine substitutions in epidermal growth factor–like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes. *American Journal of Human Genetics, 65*(4), 1007-1020.

Information about the FBN1 alleles came from the following sources (accessed August 2018):

- Protein function data for M1 is primarily based off Kirshner 2011. We used data from a number of papers that looked at neonatal Marfan syndrome (such as Raghunath 1993) to estimate the amount of fibrillin-1 protein incorporated into microfibrils.
- Protein function data for M4 combines the patient data from Milewicz 1995 (allele shown) with functional studies performed on a similar allele in Jensen 2014. Both alleles disrupt the same cut site on the fibrillin-1 protein.
- DNA and amino acid sequences were accessed through UniProt, entry P35555

Funding

This work was supported by a Science Education Partnership Award (1R25GM021903) from the National Institute of General Medical Sciences of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.