EDS-Bayes-Net

Casey developed a Probabilistic Graphical Model 'Adverse Outcome Pathway' for the effect of the COL5A1 genetic mutation as it leads to Chronic Fatigue and Brain Fog via the Ehlers-Danlos pathway, using the R bnlearn package. This is described below.

```
# The EDS BN
# A Bayesian Network for EDS AOP

# *** Libraries
library(bnlearn)
library(graph)
library(Rgraphviz)
library(RBGL)
library(gRain)
library(ggplot2)
library(RColorBrewer)
```

Below creates the Model String which describes the Bayesian Network nodes and parent-child relationships among them

```
modelStr <- paste(
  "[COL5A1_Mutation]",
  "[TypeV_Collagen_Function|COL5A1_Mutation]",
  "[Fibrils_Strength|TypeV_Collagen_Function]",
  "[Skin_Strength|Fibrils_Strength]",
  "[Tissue_Strength|Fibrils_Strength]",
  "[AAD|Tissue_Strength]", #AAD = Arthritis-Adrenaline Disorder
  "[Tendon_Strength|Fibrils_Strength]",
  "[Immunity|Tendon_Strength]",
  "[Wound_Healing|Skin_Strength:Tendon_Strength]",
  "[Chronic_Pain|Wound_Healing:Immunity]",
  "[Vein_Plasticity|Fibrils_Strength]",
  "[Blood_Flow_to_Brain|Vein_Plasticity]",
  "[GI_Efficiency|Vein_Plasticity]",
  "[POTS|Vein_Plasticity]",
  "[Brain_Fog|Blood_Flow_to_Brain:POTS]",
  "[Pharynx_Structure|Vein_Plasticity]",
  "[Sleep_Apnea|Pharynx_Structure]",
  "[Chronic_Fatigue|AAD:Sleep_Apnea]",
  sep=""
)
```

Below creates the directed acyclic graph (DAG) from the string

```
dag = model2network(modelStr)
dag # Details

# Plot the DAG
graphviz.plot(dag)
```

Below highlights the observed nodes. Teal highlighted nodes are observed (Mark's data is available, such is the case for his genomic data and symptoms) and white nodes are unobserved.

```
graphviz.plot(dag,
        highlight = list(nodes = c("COL5A1_Mutation","AAD",
                        "Immunity", "Wound_Healing","Chronic_Pain",
```

```
                    "GI_Efficiency", "POTS", "Brain_Fog",
                    "Sleep_Apnea", "Chronic_Fatigue"),

col = "black", fill = "#d6ffeb"))
```

Here is the resulting plot:

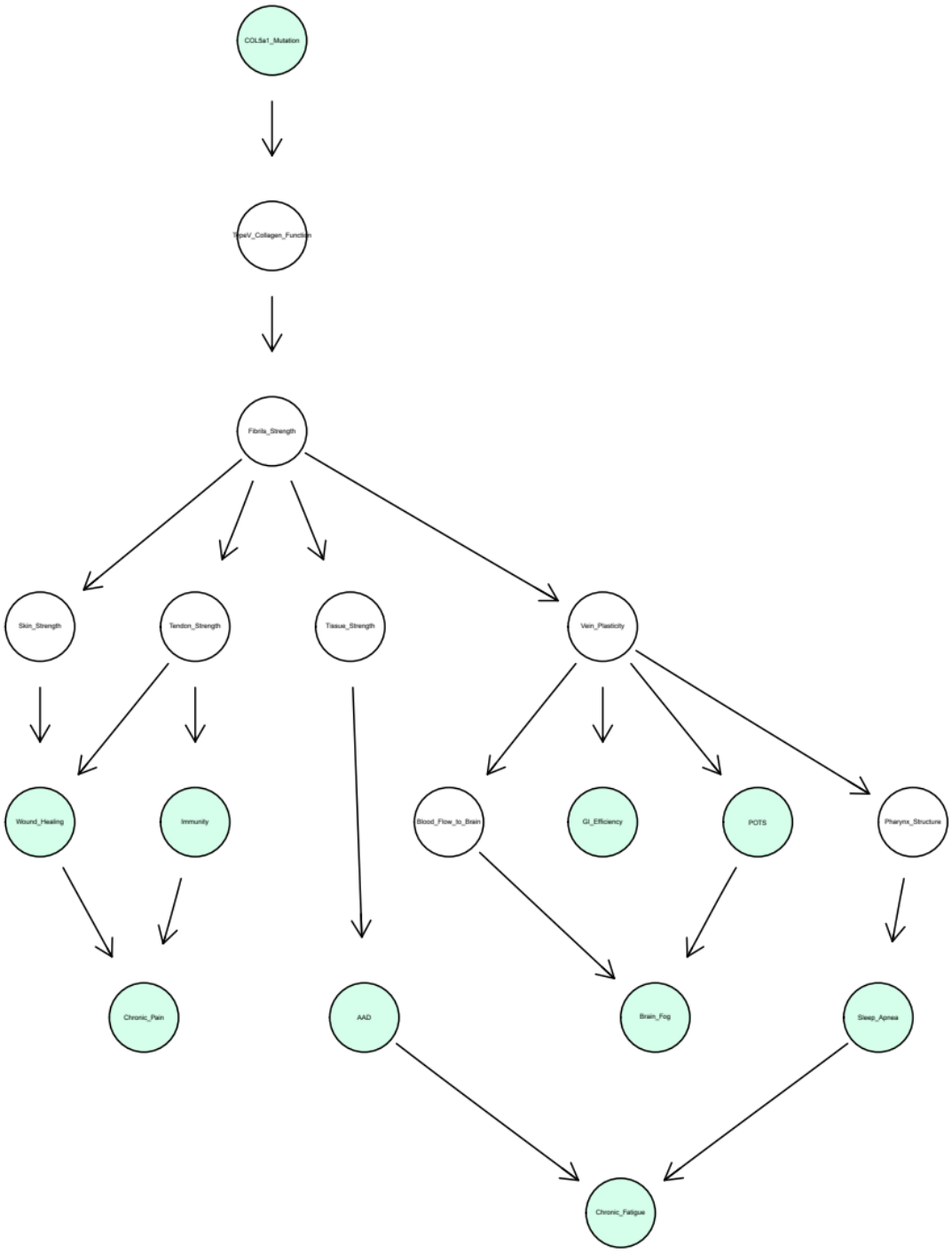

Figure 1: R bnlearn generated COL5A1 AOP

We can create a matrix for each patient we want to run through the AOP. For example, I created dummy data for Mark's observed nodes below. We know that he has the mutation and we know what his symptoms are. We could do the same for Kira's data.

```
#Mark
probsMark =
 matrix(data = c(
   0.0, 0.0, 1.0,  # COL5A1_Mutation ("None", "Heterozygous", "Homozygous")
   0.3, 0.7, 0.0,  # AAD ("None", "Mild", "Severe")
   0.9, 0.1, 0.8,  # Immunity ("Weak", "Ok", "Strong")
   0.6, 0.4, 0.9,  # Wound_Healing ("Weak", "Ok", "Strong")
   0.8, 0.2, 1.0,  # Chronic_Pain ("None", "Mild", "Severe")
   0.2, 0.8, 0.6,  # GI_Efficiency ("Weak", "Ok", "Strong")
   0.7, 0.3, 0.2,  # POTS ("None", "Mild", "Severe")
   0.0, 0.5, 0.5,  # Brain_Fog ("None", "Mild", "Severe")
   0.0, 0.8, 0.2,  # Sleep_Apnea ("None", "Mild", "Severe")
   0.0, 0.8, 0.2), # Chronic_Fatigue ("None", "Mild", "Severe")
   nrow = 10, ncol = 3, byrow = TRUE,
   dimnames = list(c("COL5A1_Mutation","AAD", "Immunity", "Wound_Healing",
                "Chronic_Pain", "GI_Efficiency", "POTS", "Brain_Fog",
                "Sleep_Apnea", "Chronic_Fatigue"),c("1","2","3"))
 )

# Set the input probabilities for each run (each run = each patient)
probsUse <- probsMark
numEnts=1 #because we only have Mark right now

# Convert the chosen input probabilities to a data frame
probs <- as.list(data.frame(t(probsUse)))
```

Now we need to set the CPTs, or Conditional Probability Tables, for each node as I do below for COL5A1. We can't do this until we determine the probability functions the edges represent. These will be determined by smaller networks (see Nikki's work later in this document).

```
COL5A1_Mutation.vals <- c("None", "Heterozygous", "Homozygous")
COL5A1_Mutation.cpt <-
 array(probs$COL5A1_Mutation,
      dim=3,
      dimnames = list(COL5A1_Mutation=COL5A1_Mutation.vals))

#need to create CPTs for each node
```

Below "fits" the dag and cpt together into a complete network model. This step means we have connected our probability functions to our network and can now run Mark and Kira's respective observed variables through the node.

```
bn = custom.fit(dag, cpt)
# or if any of the nodes are ordinal discrete, we can add them in a list
# bn = custom.fit(dag, cpt, ordinal=c(" Chronic_Fatigue ","Brain_Fog"))
```

Below plots nodes and chart probabilities to produce the probability of Brain Fog or CFS for each patient.

```
graphviz.chart(bn, type = "barprob", grid = TRUE, bar.col = "darkgreen", strip.bg = "lightskyblue")


# Turn data into a data frame
dfResults <- data.frame(Company=results[2,], Probability=as.numeric(results[1,]))
```

```
dfResults

# ****************************************************************
# Plot the Results

# Set up dynamic coloring for the bar chart
topVal <- max(dfResults$Probability,0.60) # Set top value for coloring, with 0.60 minimum
palette <- colorRampPalette(brewer.pal(9, 'RdYlGn'))
ncols <- topVal*1000
allColors <- palette(ncols)#apply the function to get 100 colours

colors <- array(dim=numEnts)
for(i in 1:numEnts) {
 colors[i] <- allColors[as.numeric(dfResults$Probability[i])*1000]
}

p<-ggplot(data=dfResults,
 aes(x=Company, y=Probability)) + # Order alphabetically
 geom_bar(stat="identity", color="black", fill=colors) +
 geom_text(aes(label=format(round(Probability, 3))), vjust=1.6, color="darkblue", size=3.5) +
 ggtitle(expression(atop("Investment Probabilities", atop(italic("Bayesian Network Analysis")))))

p
```

The next step will be to add treatment nodes as parents to the intermediary nodes. We can then run Mark and Kira's data through the AOP again, and see how their symptoms (primarily Brain Fog and Chronic Fatigue in this model) are affected. We can also propose experiments to transfer more of the unobserved nodes to observed nodes.

Target Mutations AOP:

Casey compiled the data from Mark's genomics report and researched potential downstream effects of each genetic mutation, using the Online Mendelian Database for Man (omim.org) database to determine most likely targets for modeling. This 'compilation' spreadsheet is included in the Dropbox, but a screenshot is pasted below for reference.

	Level 1 Nodes		Data Values for Level 2 Nodes and Edges to Level 3 Nodes						Subsequent Levels of Nodes			
Category	Gene	SNP	Zygosity	Risk Alele	Genotype	Genotype	Interpretation	Where Gene is Expressed	Immediate Effect	Downstream Effects		
Glutamate	GRM7	rs2229902	Hetero	0.405			Potentially damaging	Brain	Could affect glutamate			
Glutamate	GRIN3A	rs62000403	Hetero	0.059			Potentially damaging	Brain	Could affect glutamate			
Glutamate	GRIN3A	rs10989589	Hetero	0.408			Potentially damaging	Brain	Could affect glutamate			
Serotonin, Dopamine	MAOA	rs6323	Hetero	0.427	0.3	0.555	Damaging	Brain	Reduced activity of MAOA	reduced degradation of dopamine,	behavioral problems, including	
Dopamine	COMT	rs4680	Hetero	0.528	0.512	0.784	Damaging	Brain	lowered levels of COMT	increase in extracellular	Fibromyalgia is a common	
CFS	NPAS2	rs356653	Hetero	0.285	0.431	0.501	Damaging	Brain			Chronic Fatigue Syndrome	
CFS	RNASEL	rs74315364	Hetero	0.003	0.006	0.006	Damaging	Everywhere	increased activity of Rnase L		Chronic Fatigue Syndrome	(also associated with
CFS	RNASEL	rs151296858	Hetero	0.006	0.011	0.011	Damaging	Everywhere	increased activity of Rnase L		Chronic Fatigue Syndrome	
Mitochondria		MT:302-A/ACC	Hetero	0.127			Damaging	--			Melanoma risk	
Mitochondria		MT:310-T/TC	Hetero	0.042			Damaging	--			Melanoma risk	
Serotonin	IKBKAP	rs1140064	Hetero	0.028	0.054	0.055	Damaging	Brain				
Serotonin	S251G	rs17853166	Hetero	0.028	0.054	0.055	Potentially Damaging	Brain				
Serotonin	SLC6A4, or 5-HTT	5'-HTTLPR	**Homo**	0.4	0.25	-	**Protective**	Brain				
POTS	DST	rs41271862	Hetero	0.057	0.107	0.11	Likely Benign	Cytoskeleton Figments			EDS, Skin blistering	
POTS	DST	rs45582036	Hetero	0.105	0.189	0.2	Likely Benign	Cytoskeleton Figments			EDS	
POTS	FLNC	rs1390516682	Hetero				Potentially Damaging	Cardiac & Skeletal Muscles		Affects collagen	EDS	
POTS	TTN	rs377571654	Hetero	0	0.001	0.001	Potentially Damaging	Skeletal Muscles	Reduced efficacy and	distal myopathy (heart, upper &		
Serotonin	HTR3E	rs5855015	Hetero	0.097	0.174	0.184	Unknown	Brain			cardiomyopathy, muscular	
Serotonin	HTR3D	rs6789754	Hetero	0.585	0.485	0.828	Unknown	Brain				
Serotonin	HTR3C	rs6807362	Hetero	0.524	0.277	0.772	Likely Benign	Brain				
Serotonin	HTR3B	rs1176744	Hetero	0.323	0.435	0.54	Unknown	Brain				
Serotonin	SLC18A1, or VMAT1	rs2270641	Hetero	0.41	0.533	0.677	Likely Benign	Brain				
Serotonin	HTR6	Chr 1:19992947	Hetero				Unknown	Brain				
Carrier Variant	PMM2	rs28936415	Hetero				Potentially Damaging	Everywhere	produced	glycosylation type 1a		
Carrier Variant	TPRN	rs387906221	Hetero				Potentially Damaging	Ear and other sensory skin	reduced production of	hearing loss		
Carrier Variant	SLC39A8	rs778210210	Hetero				Potentially Damaging	Brain	reduced uptake of manganese, zinc	reduced micronutrients to brain an	glycosylation type 2, Leigh syndrome	abnormal metabolism of carbs
Ion Channels	CLCA4	rs4001061	Hetero				Potentially Damaging	Esophagus & Intestine			cystic fibrosis in intestines	
Ion Channels	KCNJ12	rs112163749	Hetero				Potentially Damaging	Brain and muscles	gene encodes inward K+	contributes to cardiac inward	dilated cardiomyopathy	
Ion Channels	KCNJ12	rs782590454	Hetero				Potentially Damaging	Brain and muscles	gene encodes inward K+	contributes to cardiac inward	dilated cardiomyopathy	
Ion Channels	KCNS1	rs142180542	Hetero	0.01	0.02	0.02	Potentially Damaging	Gene is present in				
Ion Channels	CACNA1A	rs765169827	Hetero				Potentially Damaging	Brain	Reduced calcium channel	Reduced control of	episodic ataxia, hemiplegic	epilepsy, migraines,
Ion Channels	ITPR2	rs11626763	Hetero				Potentially Damaging	Highly expressed in brain,			anhidrosis	severe heat
Ion Channels	RYR3	rs3217346	Hetero				Potentially Damaging	Highly expressed in brain			myopathy with nemaline bodies	
Predictor Ion Channels	KCNJ11	rs5219	Hetero	0.626	0.468	0.86	Predictor & Potentially Damaging	Pancreas	Reduced K-ATP channels made	Reduced insulin release	Congenital hyperinsulinism, gestional diabettes, maturity-onset diabetes of the young	of glibenclamide, gliclazide, glimepirid glipizide, gliquidone, and sulfonamides; Risk factor for non-insulin dependent
Predictor Ion Channels	TPCN2	rs3829241	Hetero	0.389	0.488	0.633	Predictor		Predictor of blond hair (lol,			
Small Fiber Neuropathies	WNK1	rs141823469 -> rs35706572 -> rs11441897	**Homo**	0.397	0.163	0.63	Damaging		Reduced WNK1 proteins	Reduced blood pressure regulation and reduced pain sensation	Autonomic Neuropathy Type 2 (causes numbness and tingling in hands and feet, loss of	
Small Fiber Neuropathies	ABCA1	rs759114091	Hetero				Damaging	Highly expressed in liver and	prevent release of	drastic reduction of HDLs	Tangier's disease	early-onset
Small Fiber Neuropathies	PRNP	rs1799990	Hetero	0.335	0.446	0.557	Modifier	Highly expressed in brain	reduced prion proteins	reduced copper into neurons,	Pathogenic genetic prion disease	Alzheimer's disease,
EDS	COL3A1	rs111840783	Hetero				Damaging	Basically anywhere there's	Reduced type III collagen	reduced structure in skin, lungs,	EDS	
EDS	COL5A1	rs61735045	Hetero				Damaging	Skin, ligaments, bones,	Reduced type V collagen	reduced structure in skin,	EDS, Carpal tunnel, Keratoconus	
EDS	ELN	rs17855988	Hetero				Damaging	Connective tissue (Joints	Reduced tropoelastin	Reduced resilience and flexibility	Cutis laxa, supravalvular aortic	
EDS	HSPG2	rs143706338	Hetero				Damaging	Connective tissue (Joints	Reduced perlecan	Reduced production of new blood	Schwartz-Jampel sydnrome	

Table 1: Mark Genomic Data with Downstream Effects

Casey identified potentially damaging suspect genes a created an AOP in PowerPoint (PPT) to determine parent-child relationships. The genomic variants of greatest interest based on her research are highlighted in red.

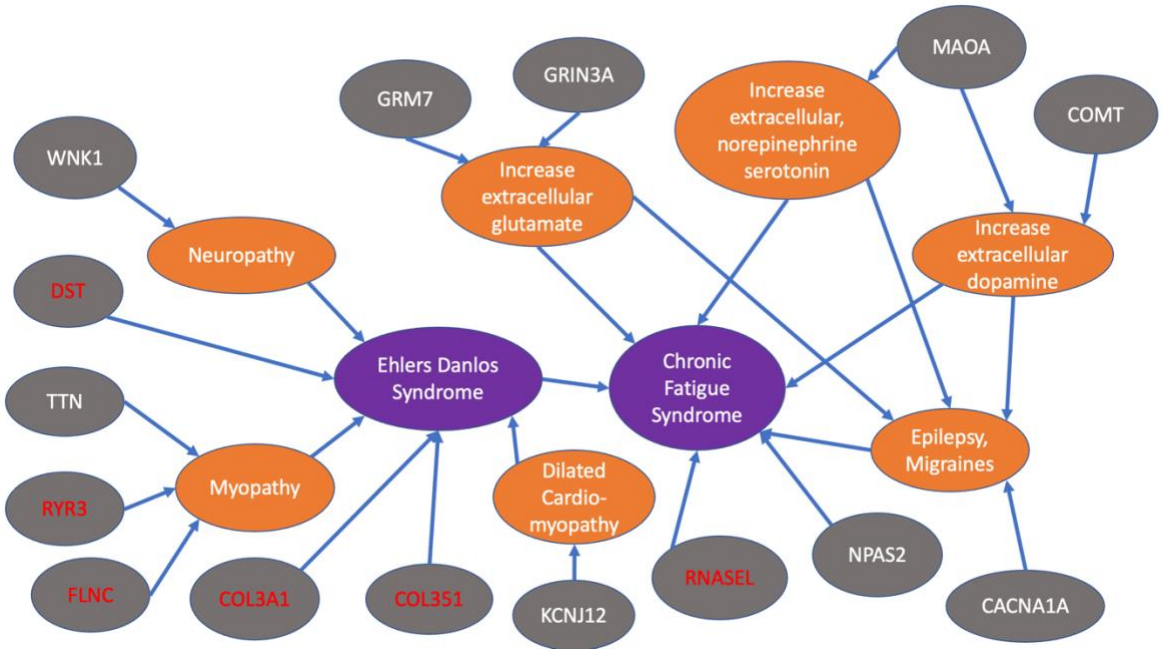

Figure 2: PPT generated Mark Genomic Targets of Interest AOP

She then created this AOP in the R bnlearn package using the same code as described above, in which teal highlighted nodes are observed (Mark's data is available) and white nodes are unobserved. This was the resulting plot:

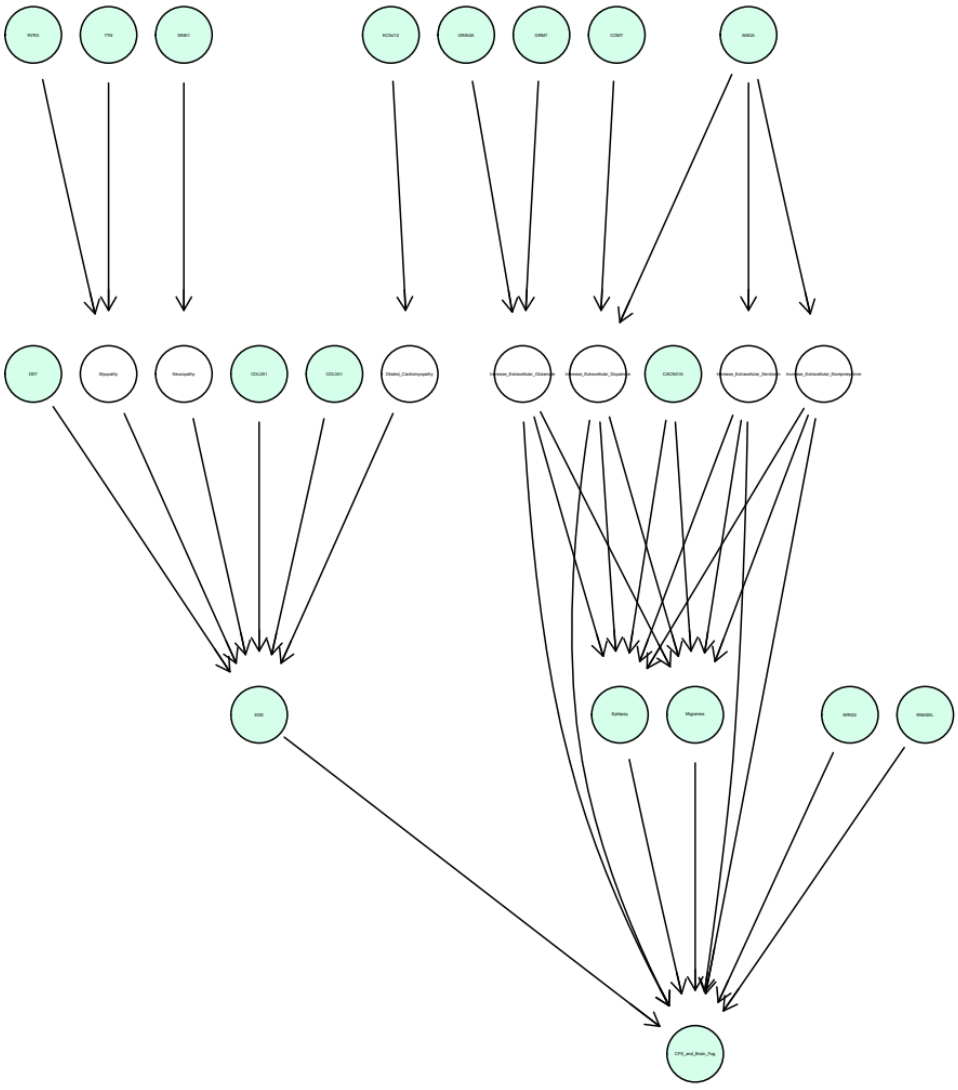

Figure 3: R bnlearn generated Mark Genomic Targets of Interest AOP

These AOPs can only be computationally useful once the probability functions (edges) are determined and justified. We plan to do so with biochemical modeling. Further, any node currently marked as 'unobserved' should be examined to determine if it *can* be observed via experimentation. This will strengthen the model.

Literature

AOP Articles

Foran, C. M., Rycroft, T., Keisler, J., Perkins, E. J., Linkov, I., & Garcia-Reyero, N. (2019). A modular approach for assembly of quantitative adverse outcome pathways. ALTEX-Alternatives to animal experimentation, 36(3), 353-362.

Halappanavar, S., van den Brule, S., Nymark, P. et al. Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale. Part Fibre Toxicol 17, 16 (2020). https://doi.org/10.1186/s12989-020-00344-4

Knapen, D., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta-Casaluci, L., ... & Villeneuve, D. L. (2018). Adverse outcome pathway networks I: development and applications. Environmental toxicology and chemistry, 37(6), 1723-1733.

Masison, J., Beezley, J., Mei, Y., Ribeiro, H. A. L., Knapp, A. C., Sordo Vieira, L., ... & Laubenbacher, R. (2021). A modular computational framework for medical digital twins. Proceedings of the National Academy of Sciences, 118(20), e2024287118.

Villeneuve, D. L., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta-Casaluci, L., ... & Knapen, D. (2018). Adverse outcome pathway networks II: network analytics. Environmental toxicology and chemistry, 37(6), 1734-1748.

Clinical Articles

EDS

Johnston, J. M., Connizzo, B. K., Shetye, S. S., Robinson, K. A., Huegel, J., Rodriguez, A. B., ... & Soslowsky, L. J. (2017). Collagen V haploinsufficiency in a murine model of classic Ehlers–Danlos syndrome is associated with deficient structural and mechanical healing in tendons. Journal of Orthopaedic Research, 35(12), 2707-2715.

Laguette, M. J., Abrahams, Y., Prince, S., & Collins, M. (2011). Sequence variants within the 3'-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries. Matrix Biology, 30(5-6), 338-345.

Malfait, F., Wenstrup, R. J., & De Paepe, A. (2010). Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. Genetics in medicine, 12(10), 597-605.

Peltonen, L., Halila, R., & Ryhänen, L. (1985). Enzymes converting procollagens to collagens. Journal of cellular biochemistry, 28(1), 15-21.

Royer, S. P., & Han, S. J. (2022). Mechanobiology in the comorbidities of Ehlers Danlos syndrome. Frontiers in Cell and Developmental Biology, 710.

Teti, A. (1992). Regulation of cellular functions by extracellular matrix. Journal of the American Society of Nephrology, 2(10), S83.

Wilson, G. N., Tonk, S. S., Tonk, V. S., & Lampe, R. (2020). Complement Gene Mutation and Ehlers-Danlos Syndrome. Journal of Biosciences and Medicines, 8(06), 28.

Graph Theory Articles

Bertsimas, D., Dunn, J., Gibson, E. *et al.* Optimal survival trees. *Mach Learn* **111**, 2951–3023 (2022). https://doi.org/10.1007/s10994-021-06117-0

Dimitris Bertsimas, Jack Dunn, Nishanth Mundru (2019) Optimal Prescriptive Trees. INFORMS Journal on Optimization 1(2):164-183. https://doi.org/10.1287/ijoo.2018.0005

Behrouz, A., Lécuyer, M., Rudin, C., & Seltzer, M. (2022). Fast optimization of weighted sparse decision trees for use in optimal treatment regimes and optimal policy design. arXiv preprint arXiv:2210.06825.

Liu, J., Zhong, C., Li, B., Seltzer, M., & Rudin, C. (2022). FasterRisk: Fast and Accurate Interpretable Risk Scores. arXiv preprint arXiv:2210.05846.