Processes in Phenome to Genome mapping

David Dorneau
Department of Engineering and Computer Science
Florida Atlantic University
Boca Raton, Florida
ddorneau2014@fau.edu

***Abstract* - Genome mapping is the locating and assigning of a specific gene to particular region of a chromosome and determining the location of and relative distances between genes on the chromosome. This mapping allows one to create links between genes and certain conditions, processes, and diseases.**

Keywords - phenome, GWAS, genome, angiogenesis, genetics, disease and mapping.

I. Introduction

A phenome is the set of all phenotypes expressed by a cell, tissue, organ, organism, or species. Just as the genome and proteome signify all of an organism's genes and proteins, the phenome represents the sum total of its phenotypic traits. Your genome is similar to a blueprint — containing all of the instructions needed to make and maintain the function of every cell, tissue and organ in your body.

Given a disease or condition, we seek to find genes, gene products, processes, and or other disease that may be linked to it. We split the phenome to genome mapping process into phases, where we use tools and resources such as:

1) Phase 1: Text mining with tools such as medie and literature reviews, with NIH bookshelf and google scholar articles to find major genes involved in a process or disease.
2) Phase 2: KEGG pathway search tool to identify significant pathways related to to a process or disease, as well as related genes. Princeton GO term finder tool to identify biological processes based on the genes found in the text mining and literature review stage.
3) Phase 3: GWAS catalog to identify a set of genes/gene products based on different factors including disease conditions, environment factors, biomarkers, measurements, and attributes.

II. Angiogenesis Phase 1

For the text mining and literature review phase, with angiogenesis as the process, we gathered roughly five recent papers that touched up on angiogenesis, through NIH bookshelf. Using the most basic form of text mining, the find feature (ctrl-f) of pdf readers and webpages, we found terms related to angiogenesis, which helped select the review papers. In addition to that, we read the review papers, and while reading we highlighted important areas that we want to include in the research, and noted specific genes and gene products. Depicted in tables 1 is a listing of the genes found from the review.

Genes	**Category**	**Major Function**
VEGF-A	inducer	Induction of EC proliferation
VEGF-B	inducer	Induction of EC proliferation
Angiopoietin 1	inducer	Induction of EC proliferation
Angiostatin	inhibitor	Inhibit EC proliferation
Endostatin	inhibitor	Inhibit EC proliferation
interstitial collagenase (MMP-1)	Proteolytic enzyme	Degrading ECM components

collagenase-3 (MMP-13)	Proteolytic enzyme	Degrading ECM components
Ephrins	inducer	Induction of EC proliferation
Integrins	inducer	Induction of EC proliferation
Vasostatin	inhibitor	Inhibit EC proliferation

Table 1: Angiogenesis related genes.

Along with the genes, we are able to label their purpose in angiogenesis; an inducer or inhibitor of the process.

III. Angiogenesis Phase 2

For the pathway and biological processes search phase, with angiogenesis as the process, we found a more extensive list of genes, pathways which are closely related to angiogenesis, as well as processes involved within angiogenesis. For the pathways and extensive gene list, we use the KEGG pathway tool, with the keyword “angiogenesis”, and come up with the pathways depicted in table 2. For the biological processes we use the GO term finder tool, with the list of genes that we found from the review and the new additions that are involved in different pathways. Using 0.01 as the p-value cutoff, the resulting significant biological processes are shown in table 3.

Pathways	**Description**	**Status**
Proteoglycans in cancer	Proteoglycans (PGs) have been shown to be key macromolecules that contribute to biology of various types of cancer, including angiogenesis in affecting tumor growth.	Previously known pathway
Rheumatoid arthritis	Can promote synovial angiogenesis with abnormal activation of the immune system elevates pro-inflammatory cytokines and chemokines level.	New pathway
Cytokine-cytokine receptor interaction	Cytokines are glycoproteins that are crucial intercellular regulators of cells engaged in angiogenesis.	New pathway
Viral carcinogenesis	Via expression of many potent oncoproteins, tumor viruses can promote an aberrant cell-proliferation.	New pathway
VEGF signaling pathway	VEGFR-2 is the major mediator of VEGF-driven responses in endothelial cells and it is considered to be a crucial signal transducer in both physiologic and pathologic angiogenesis	Previously known pathway
Apelin signaling pathway	Apelin is an endogenous peptide capable of binding the apelin receptor and is implicated in the process of angiogenesis.	New pathway

Table 2: Pathways related to Angiogenesis.

Process	P-value
regulation of epithelial cell proliferation	0.01338
regulation of endothelial cell proliferation	0.01826
regulation of membrane protein ectodomain proteolysis	0.03718

Table 3: Biological processes related to Angiogenesis.

IV. Angiogenesis Phase 3

For the GWAS catalog search phase, with angiogenesis as the process, I used the keyword "artery", to come up with a list of diseases and processes because that term most closely related to angiogenesis. From there I selected coronary heart disease, and found genes, processes and pathways related to coronary artery disease. In table 4 is the list of genes found following the GWAS catalog search.

Genes/Gene products	Related trait	P-value
PHACTR1	Coronary heart disease	9 x 10^-26
ALDH2	Coronary heart disease	2 x 10^-34
CDKN2B-AS1	Large artery stroke	1 x 10^-59
TWIST1	Coronary heart disease	3 x 10^-12
SMARCA4	Coronary heart disease	2 x 10^-11
PSRC1, CELSR2	Large artery stroke	8 x 10^-17
MRAS	Coronary heart disease	7 x 10^-13
GRHL1	atrial fibrillation, heart failure	6 x 10^-7
RNU7-2P	atrial fibrillation, heart failure	5 x 10^-10
ITPK1	atrial fibrillation, heart failure	5 x 10^-16
THEMIS3P, AKR1B1P6	atrial fibrillation, heart failure	6 x 10^-64
LPA	Coronary artery disease	5 x 10^-39
PLG	Coronary artery disease	2 x 10^-32
MIA3	Coronary artery disease	1 x 10^-12
LIPC, LIPC-AS1	atrial fibrillation, coronary heart disease, diastolic blood pressure, heart failure	2 x 10^-35
HERPUD1, CETP	atrial fibrillation, coronary heart disease, diastolic blood pressure, heart failure	1 x 10^-149
THEMIS3P, AKR1B1P6	atrial fibrillation, coronary heart disease, diastolic blood pressure, heart failure	5 x 10^-67
APOC1, APOC1P1	atrial fibrillation, coronary heart disease, diastolic blood pressure, heart failure	1 x 10-22

Table 2: Angiogenesis post GWAS genes.

V. Results and Discussion

At the end of the text mining and literature review phase, we came away with clinical uses of angiogenesis being a possible game changer. Following the pathway and biological process phase and the GWAS catalog phase, and going into coronary heart disease, we ended up right back into the clinical uses of angiogenesis whether it be to induce, or inhibit EC proliferation, and the side effects of some of those manipulations. The clinical uses of angiogenesis can be used to create better conditions for those living with conditions that result from coronary heart disease. The most difficult part was figuring out where to focus because angiogenesis plays a role in a lot of biological processes. we

previously wanted to know whether angiogenesis clinically manipulated or not could repair damage caused by heart conditions, and as a result of the GWAS catalog extension, I found the answer to that.

VI. Conclusion

Using the three step process for phenome to genome mapping, we are able to connect angiogenesis to ischemic heart diseases. As a result of still being in the early stages of breakthroughs for clinically induced angiogenesis, we face the challenge of creating a positive outlook for therapy for coronary heart disease and other ischemic diseases. With every solution, there are repercussions and new factors that are discovered. Angiogenesis is already a complex process, so with every new factor, the complexity only increases. The three step process can still be repeated on our new findings, to narrow down even more and make more connections.

VII. References

[1] KEGG PATHWAY Database, www.genome.jp/kegg/pathway.html.

[2] “Generic Gene Ontology Term Finder.” Princeton University, The Trustees of Princeton University, go.princeton.edu/cgi-bin/GOTermFinder.

[3] GWAS Catalog, www.ebi.ac.uk/gwas/home.

[4] Seppo Ylä-Herttuala, Charles Bridges, Michael G. Katz, Petra Korpisalo, Angiogenic gene therapy in cardiovascular diseases: dream or vision?, European Heart Journal, Volume 38, Issue 18, 7 May 2017, Pages 1365–1371

[5] So Young Yoo and Sang Mo Kwon, “Angiogenesis and Its Therapeutic Opportunities,” Mediators of Inflammation, vol. 2013, Article ID 127170, 11 pages, 2013.

[6] Gupta K, Zhang J Angiogenesis: a curse or cure? Postgraduate Medical Journal 2005;81:236-242.