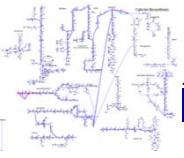


Tissue Specific Models



LEARNING OBJECTIVES

Each student should be able to:

- Describe the process of creating tissue-specific models.
- Describe the limitations of tissue-specific models.
- Describe the constraint-based red blood cell model.

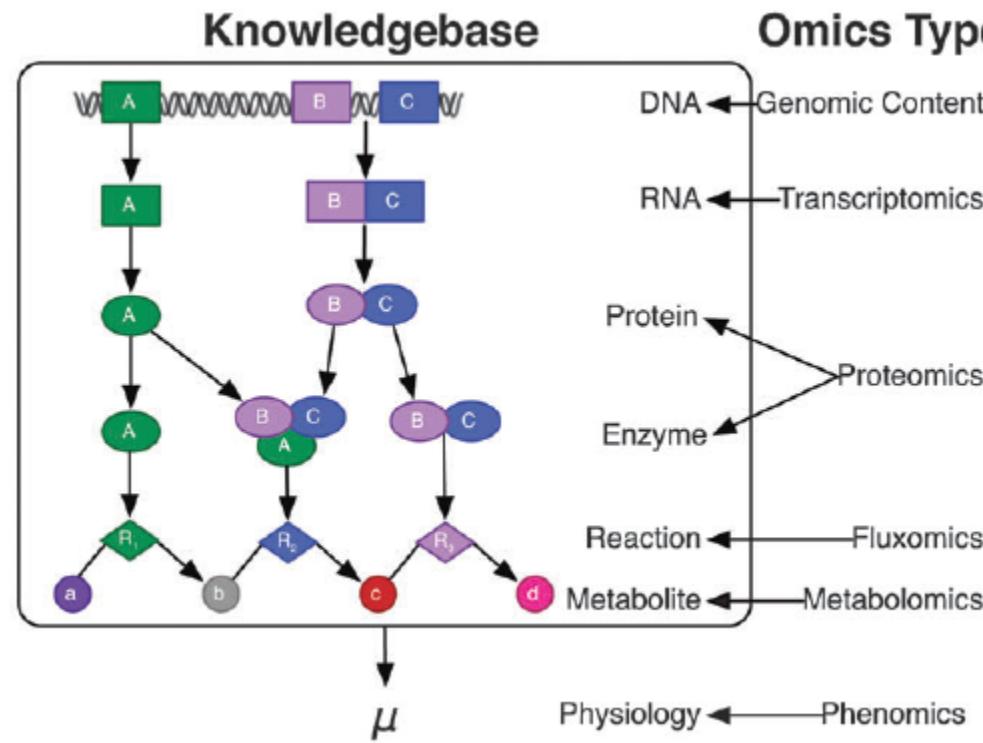
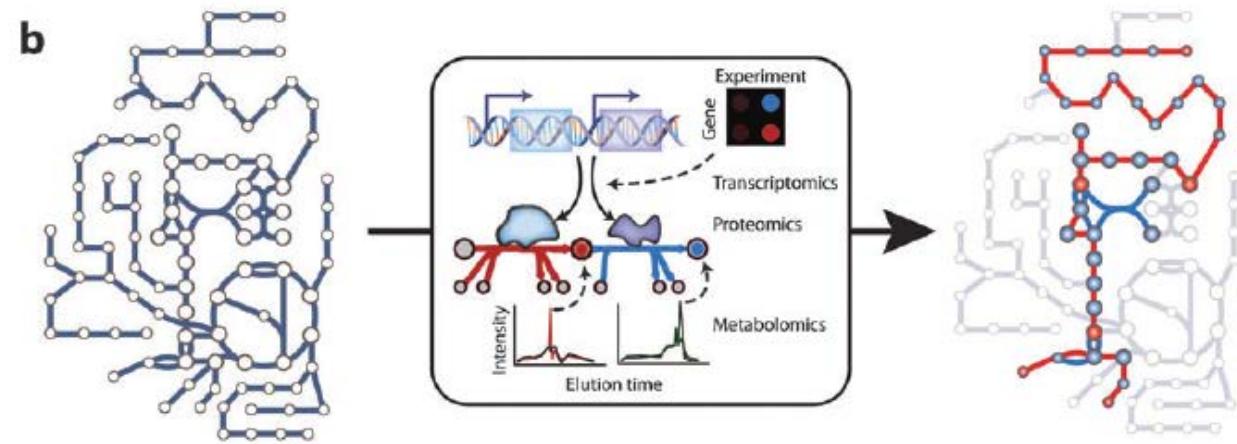
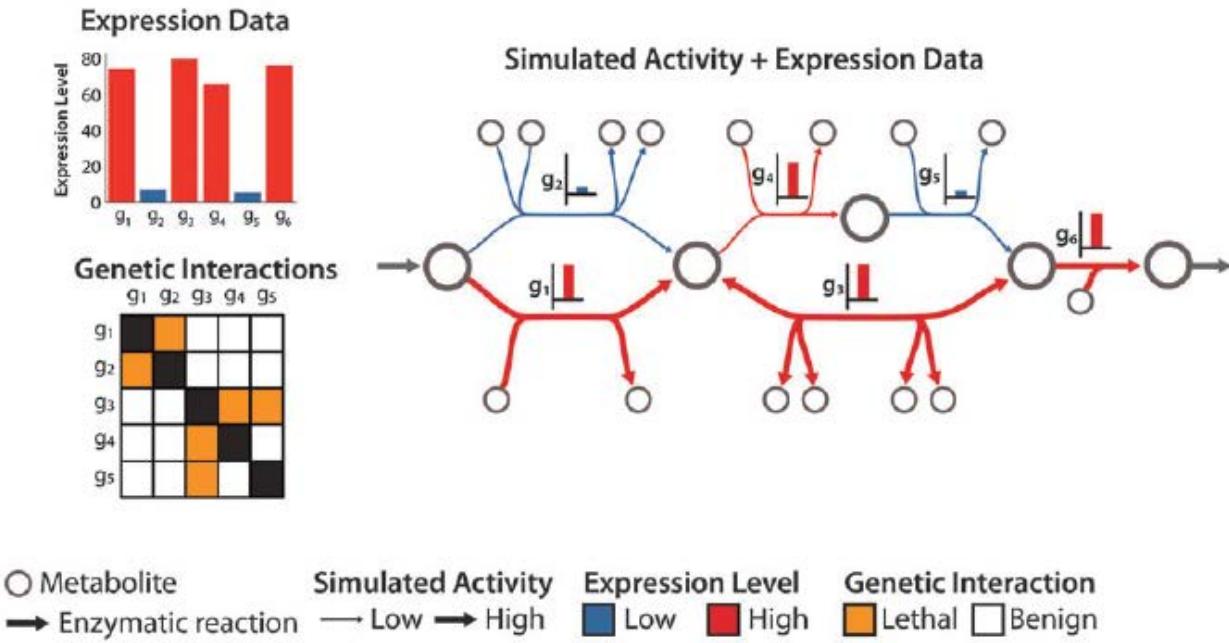


Lesson Outline

- Overview
- Creating Tissue-specific Models
- Tissue-specific Example
 - ✓ Red Blood Cells
 - ✓ Multi-tissue Modeling



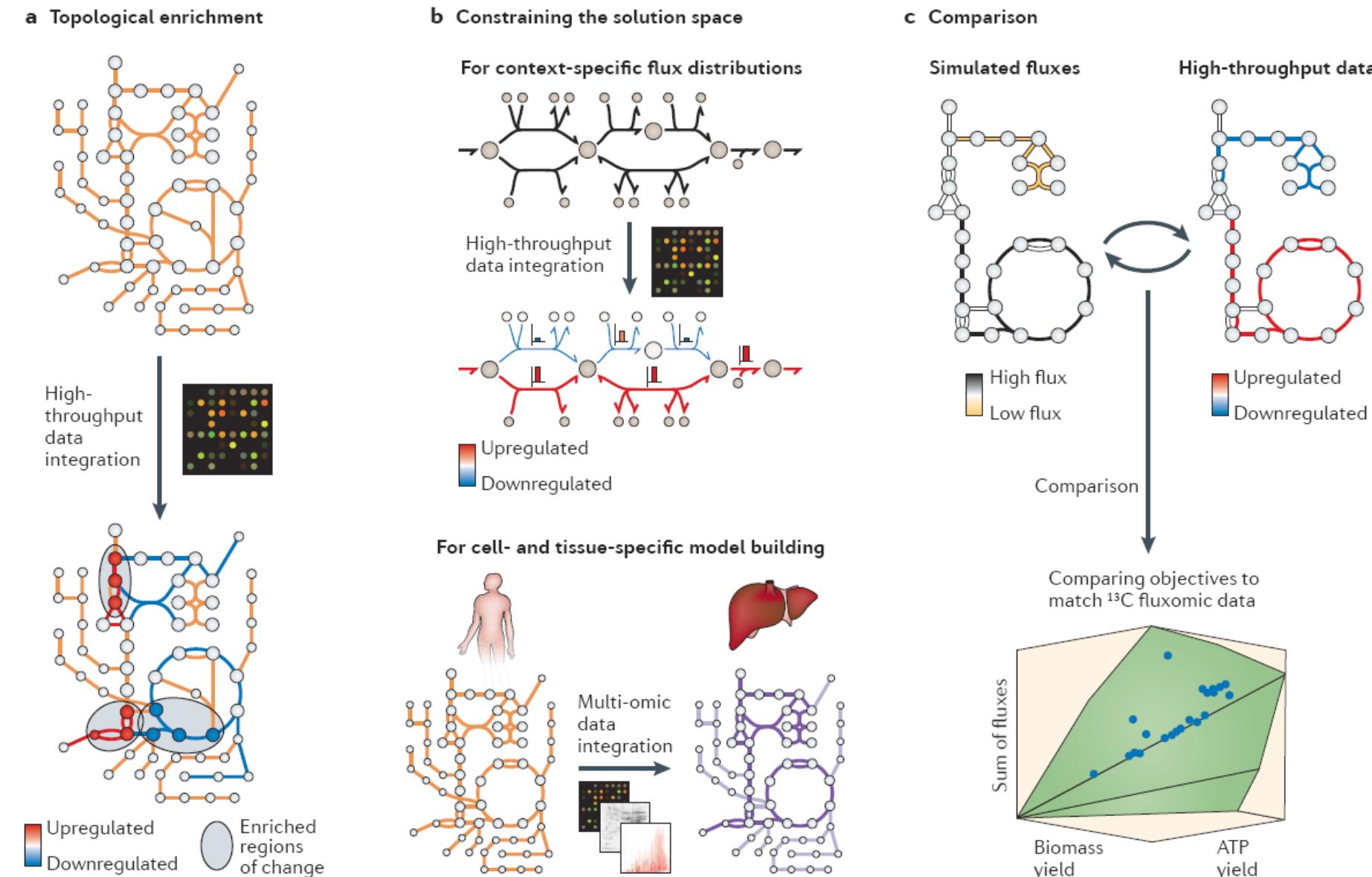
Omics Data and Genome-scale Models

a**b****c**

Hyduke, D. R., N. E. Lewis, et al. (2013). "Analysis of omics data with genome-scale models of metabolism." *Mol Biosyst* 9(2): 167-174.



The multiple uses of high-throughput omics data in constraint-based models



Bordbar, A., J. M. Monk, et al. (2014). "Constraint-based models predict metabolic and associated cellular functions." *Nature reviews. Genetics* 15(2): 107-120.

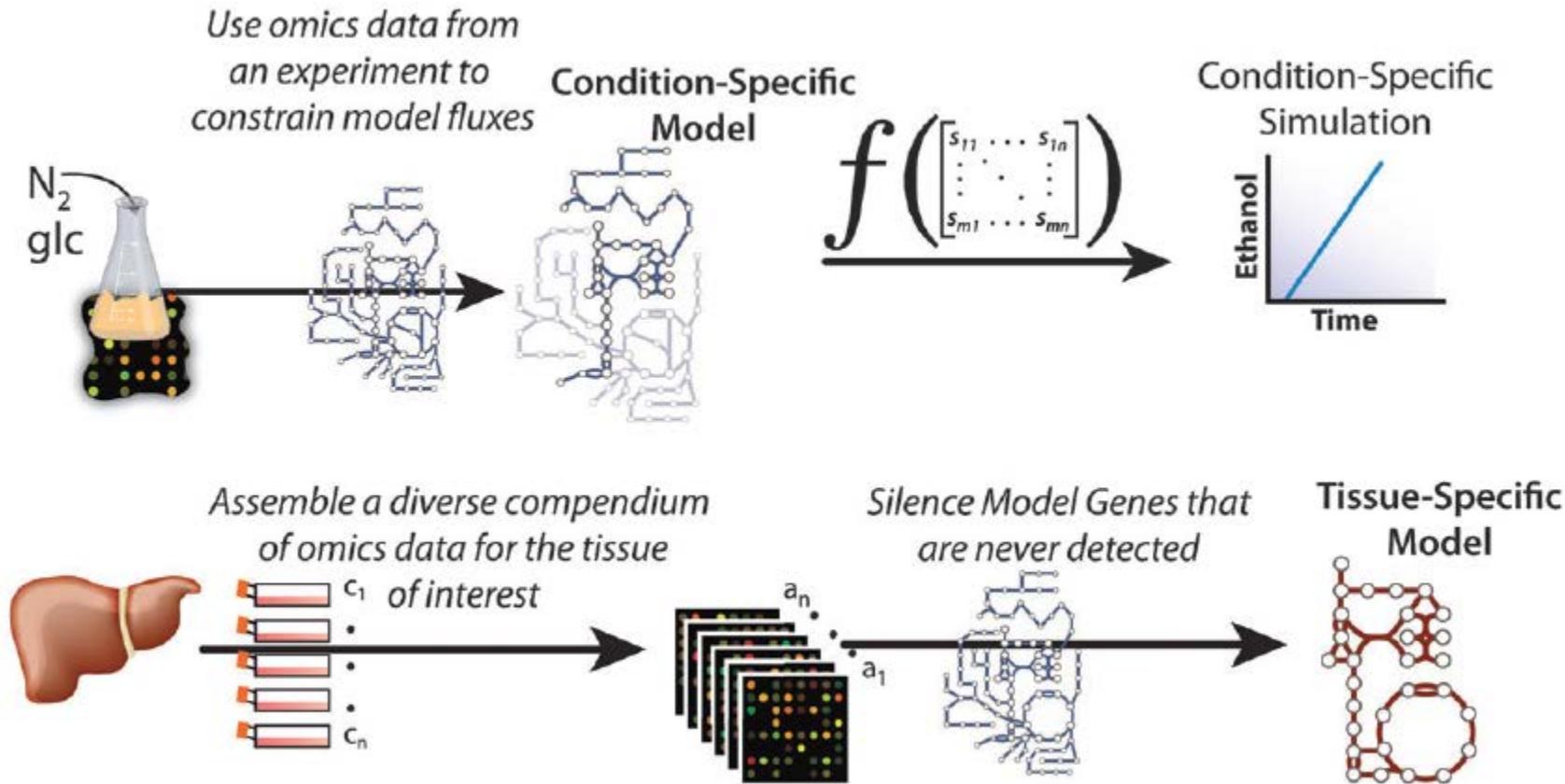


Lesson Outline

- Overview
- • Creating Tissue-specific Models
- Tissue-specific Example
 - ✓ Red Blood Cells
 - ✓ Multi-tissue Modeling



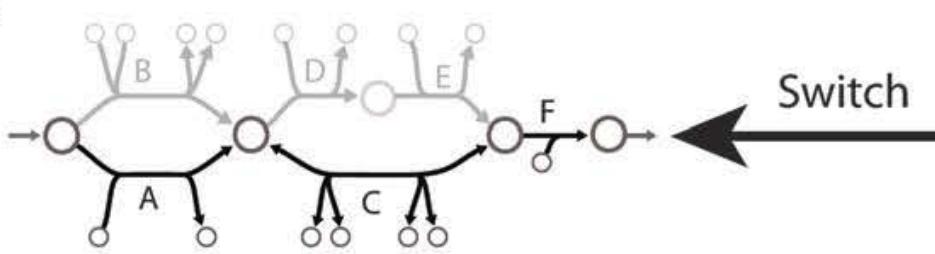
Creating Condition-specific Models



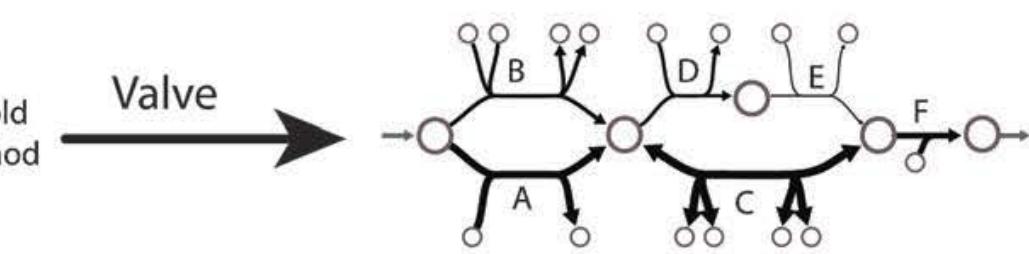
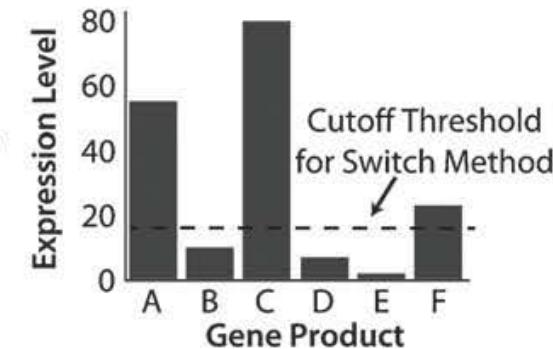
Hyduke, D. R., N. E. Lewis, et al. (2013). "Analysis of omics data with genome-scale models of metabolism." *Mol Biosyst* 9(2): 167-174.



The Switch or Valve Approach



Disable enzymes when associated gene products do not pass a specified threshold (intensity, p, etc.).



The maximum activity for an enzyme is limited by expression data for associated gene products.

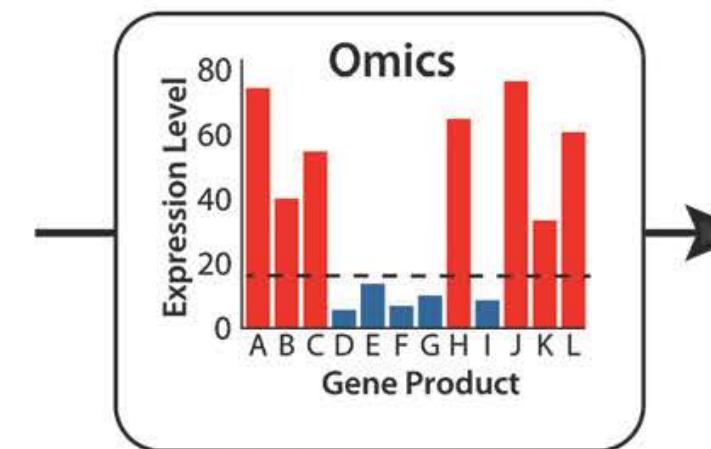
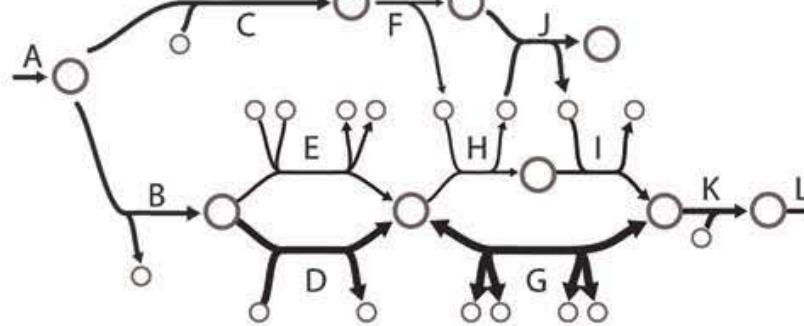
In the switch approach, omics data are used to identify which gene products should be included in the constrained model; here, the reactions catalyzed by gene products B, D, and E are disabled because their expression levels did not exceed a threshold. In the valve approach, omics data are used to limit the activities for the associated enzymes. Therefore, enzymes associated with weakly expressed genes are still able to participate in a simulation albeit to a notably reduced extent. Due to errors and noise inherent in omics data, it is possible that the model will no longer function after disabling enzyme activities; thus, it may be necessary to disregard a limited number of expression measurements when employing a switch style approach.

Hyduke, D. R., N. E. Lewis, et al. (2013). "Analysis of omics data with genome-scale models of metabolism." *Mol Biosyst* 9(2): 167-174.

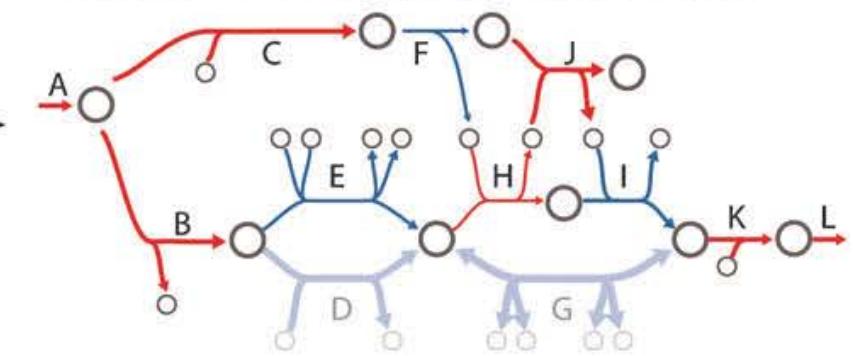


Unconstrained Simulation Used to Support Omics Constrained Models

Unconstrained Simulation



Condition- / Tissue-Specific Model

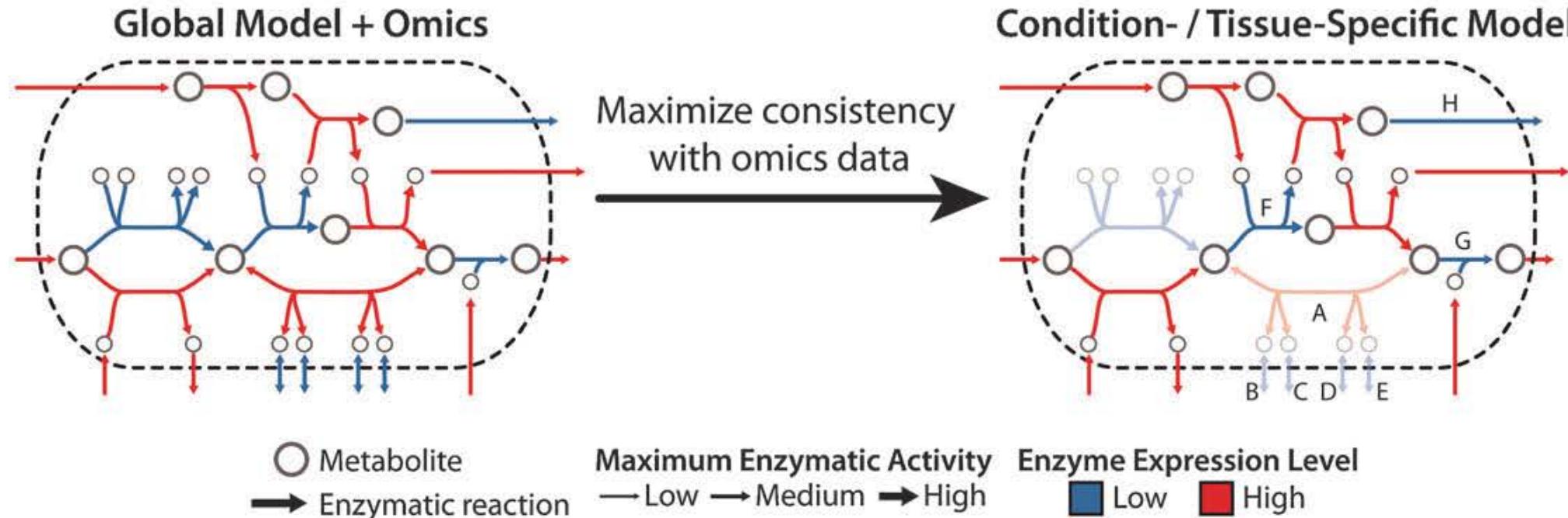


The simulation results from the unconstrained initial model were used to aid in identifying which expression measurements should be ignored. If an omics constrained model was unable to simulate a specified phenotype, here the production of L from A, then a set of enzymes were re-enabled to restore the model to a functional state. (GIMME)

Hyduke, D. R., N. E. Lewis, et al. (2013). "Analysis of omics data with genome-scale models of metabolism." Mol Biosyst 9(2): 167-174.



Maximally Consistent Models

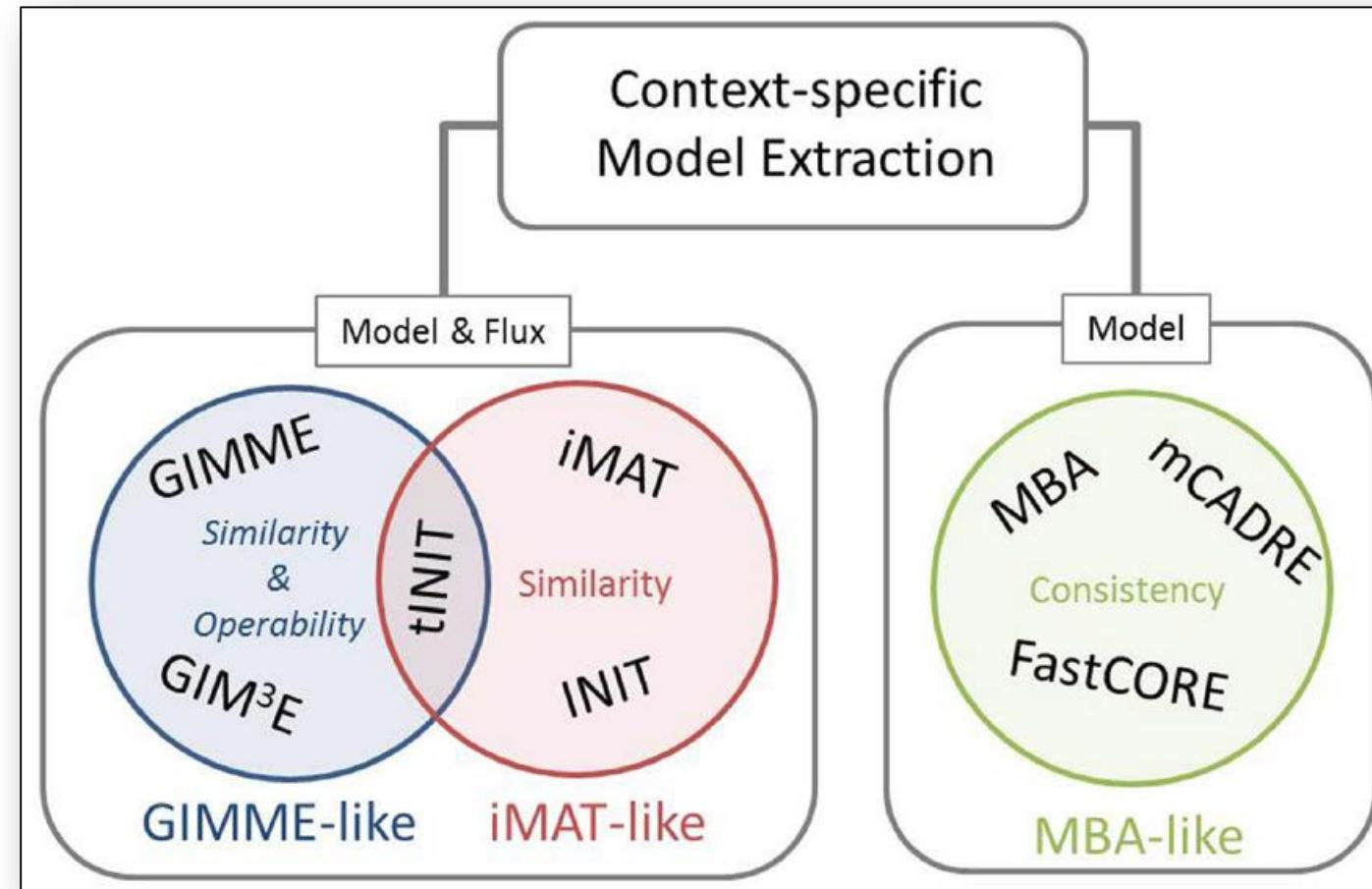


The goal was to construct the smallest model that was maximally consistent with the omics data and does not contain dead end metabolites. Enzyme A is disabled despite a high expression level because it would be necessary to enable enzymes B, C, D, and E all of which had low expression levels. In spite of low expression values, enzymes F, G, and H are enabled because their activities are required for a greater number of highly expressed enzymes to be connected. (iMAT)

Hyduke, D. R., N. E. Lewis, et al. (2013). "Analysis of omics data with genome-scale models of metabolism." Mol Biosyst 9(2): 167-174.



Existing Methods for Context-specific Model Extraction



Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." Frontiers in plant science 5.



Recon 1-based mCADRE Models

a549_cell_line.xml	cerebral_cortex.xml	hypothalamus.xml	parotid_gland.xml	stomach.xml
accumbens.xml	cervix.xml	jurkat_cells.xml	pbmc.xml	stomach_cardiac.xml
adipose_tissue.xml	cervix_tumor.xml	kidney.xml	penis.xml	substantia_nigra.xml
adrenal_cortex.xml	chorion_villus_cells.xml	kidney_tumor.xml	peritoneum_tumor.xml	subthalamic_nucleus.xml
airway_epithelial_cells.xml	colon.xml	liver.xml	pilocytic_astrocytoma_tumor.xml	superior_frontal_gyrus.xml
alveolar_macrophages.xml	colon_tumor.xml	liver_tumor.xml	pituitary_gland.xml	synovial_membrane.xml
amniotic_fluid.xml	colonic_mucosa.xml	lung.xml	pituitary_tumor.xml	temporal_lobe.xml
amygdala.xml	corpus_callosum.xml	lung_tumor.xml	placenta.xml	testes.xml
blood.xml	deltoid_muscle.xml	lymphocytes.xml	platelets.xml	thalamus.xml
bone_marrow.xml	dorsal_root_ganglia.xml	mcf-7_cells.xml	postcentral_gyrus.xml	thyroid.xml
breast.xml	endometrial_endothelial_cell.xml	medial_temporal_gyrus.xml	posterior_singulate.xml	tongue_squamous_cells.xml
breast_ductal_cells.xml	endometrium.xml	medulla.xml	primary_visual_cortex.xml	tongue_squamous_cells_tumor.xml
breast_lobular_cells.xml	endometrium_tumor.xml	midbrain.xml	prostate_gland.xml	tonsil.xml
breast_lobular_cells_tumor.xml	entorhinal_cortex.xml	monocyte.xml	putamen.xml	tonsil_epithelium.xml
breast_stroma.xml	entorhinal_cortex_layer_ii_stellate_island_cells.xml	monocyte-derived_dendritic_cells.xml	rectum.xml	trachea.xml
breast_stroma_tumor.xml	fallopian_tube_epithelium.xml	monocyte-derived_macrophage.xml	rectum_mucosa.xml	trigeminal_ganglia.xml
breast_tumor.xml	fallopian_tube_tumor.xml	myometrium.xml	rectum_mucosa_tumor.xml	urethra.xml
bronchial_epithelial_cells.xml	fetal_cartilage.xml	neutrophils.xml	salivary_gland.xml	uterus_tumor.xml
bronchus.xml	frontal_lobe.xml	nodose_nucleus.xml	sigmoid_colon_mucosa.xml	vagina.xml
cd3+_t_cells.xml	gingival_epithelium.xml	occipital_lobe.xml	sigmoid_colon_mucosa_tumor.xml	vastus_lateralis.xml
cd4+_cells.xml	glioblastoma_tumor.xml	omental_adipose_tissue.xml	skeletal_muscle.xml	ventral_tegmental_area.xml
cd14+_cells.xml	glioma_tumor.xml	omentum_tumor.xml	skin.xml	vestibular_nuclei_superior.xml
cd31+_cells.xml	head_and_neck_epithelial_cells.xml	ovary.xml	sperm.xml	vulva.xml
cd49a+_cells.xml	head_and_neck_epithelial_cells_tumor.xml	ovary_tumor.xml	spinal_cord.xml	
cecum.xml	head_and_neck_squamous_cell_carcinoma_tumor.xml	pancreas_tumor.xml	spleen.xml	
cerebellum.xml	hippocampus.xml	parietal_lobe.xml	stomach.xml	

Wang, Y., Eddy, J. A., and Price, N. D. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* 6:153. doi: 10.1186/1752-0509-6-153



Recon 2-based iMAT Models

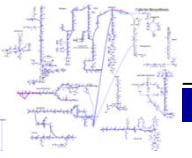
- adrenal_gland_glandular_cells.xml
- appendix_glandular_cells.xml
- appendix_lymphoid_tissue.xml
- bone_marrow_hematopoietic_cells.xml
- breast_glandular_cells.xml
- bronchus_respiratory_epithelial_cells.xml
- cerebellum_cells_in_granular_layer.xml
- cerebellum_cells_in_molecular_layer.xml
- cerebellum_Purkinje_cells.xml
- cerebral_cortex_glial_cells.xml
- cerebral_cortex_neuronal_cells.xml
- cervix_uterine_glandular_cells.xml
- cervix_uterine_squamous_epithelial_cells.xml
- colon_glandular_cells.xml
- duodenum_glandular_cells.xml
- epididymis_glandular_cells.xml
- esophagus_squamous_epithelial_cells.xml
- fallopian_tube_glandular_cells.xml
- gall_bladder_glandular_cells.xml
- heart_muscle_myocytes.xml
- hippocampus_glial_cells.xml
- hippocampus_neuronal_cells.xml
- kidney_cells_in_glomeruli.xml
- kidney_cells_in_tubules.xml
- lateral_ventricle_glial_cells.xml
- lateral_ventricle_neuronal_cells.xml
- liver_bile_duct_cells.xml
- liver_hepatocytes.xml
- lung_macrophages.xml
- lung_pneumocytes.xml
- lymph_node_germinal_center_cells.xml
- lymph_node_non_germinal_center_cells.xml
- nasopharynx_respiratory_epithelial_cells.xml
- oral_mucosa_squamous_epithelial_cells.xml
- ovary_ovarian_stroma_cells.xml
- pancreas_exocrine_glandular_cells.xml
- pancreas_islets_of_Langerhans.xml
- parathyroid_gland_glandular_cells.xml
- placenta_decidual_cells.xml
- placenta_trophoblastic_cells.xml
- prostate_glandular_cells.xml
- recon2.xml
- recon2_model.xml
- Recon2_spreadsheet.xls
- rectum_glandular_cells.xml
- salivary_gland_glandular_cells.xml
- seminal_vesicle_glandular_cells.xml
- skeletal_muscle_myocytes.xml
- skin_epidermal_cells.xml
- small_intestine_glandular_cells.xml
- smooth_muscle_smooth_muscle_cells.xml
- spleen_cells_in_red_pulp.xml
- spleen_cells_in_white_pulp.xml
- stomach_lower_glandular_cells.xml
- stomach_upper_glandular_cells.xml
- testis_cells_in_seminiferus_ducts.xml
- testis_Leydig_cells.xml
- thyroid_gland_glandular_cells.xml
- tonsil_germinal_center_cells.xml
- tonsil_non_germinal_center_cells.xml
- tonsil_squamous_epithelial_cells.xml
- urinary_bladder_urothelial_cells.xml
- uterus_post_menopause_cells_in_endometrial_stroma.xml
- uterus_post_menopause_glandular_cells.xml
- uterus_pre_menopause_cells_in_endometrial_stroma.xml
- uterus_pre_menopause_glandular_cells.xml
- vagina_squamous_epithelial_cells.xml
- vulva_anal_skin_epidermal_cells.xml

Thiele, I., N. Swainston, et al. (2013). "A community-driven global reconstruction of human metabolism." Nat Biotechnol 31(5): 419-425.



Lesson Outline

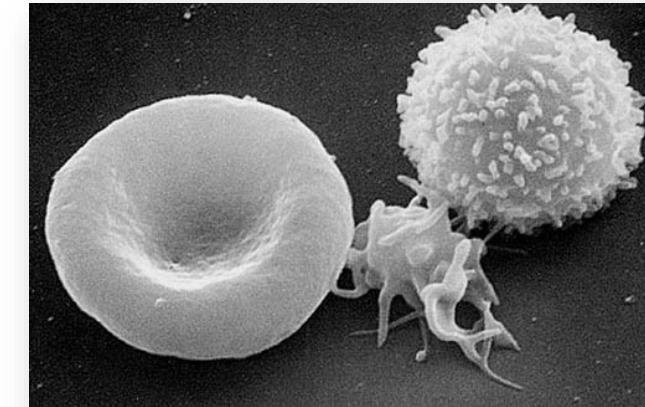
- Overview
- Creating Tissue-specific Models
- • Tissue-specific Example
 - ✓ Red Blood Cells
 - ✓ Multi-tissue Modeling



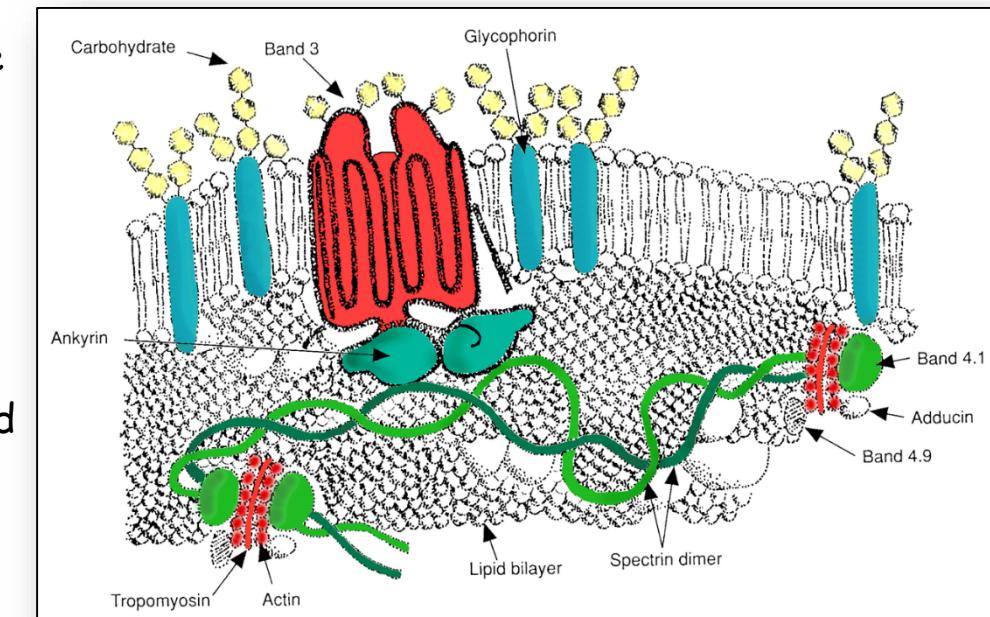
Red Blood Cells (erythrocytes)

- The most common type of blood cell and the vertebrate organism's principal means of delivering oxygen (O_2) to the body tissues—via blood flow through the circulatory system. Red blood cells (RBCs) take up oxygen in the lungs or gills and release it into tissues while squeezing through the body's capillaries.
- The cytoplasm of erythrocytes is rich in hemoglobin, an iron-containing biomolecule that can bind oxygen and is responsible for the red color of the cells. The cell membrane is composed of proteins and lipids, and this structure provides properties essential for physiological cell function such as deformability and stability while traversing the circulatory system and specifically the capillary network.
- In humans, mature red blood cells are flexible and oval biconcave disks. They lack a cell nucleus and most organelles, in order to accommodate maximum space for hemoglobin. Approximately 2.4 million new erythrocytes are produced per second in human adults. The cells develop in the bone marrow and circulate for about 100-120 days in the body before their components are recycled by macrophages. Each circulation takes about 20 seconds. Approximately a quarter of the cells in the human body are red blood cells.

https://en.wikipedia.org/wiki/Red_blood_cell



Scanning electron micrograph of blood cells.
From left to right: human erythrocyte,
thrombocyte (platelet), leukocyte.



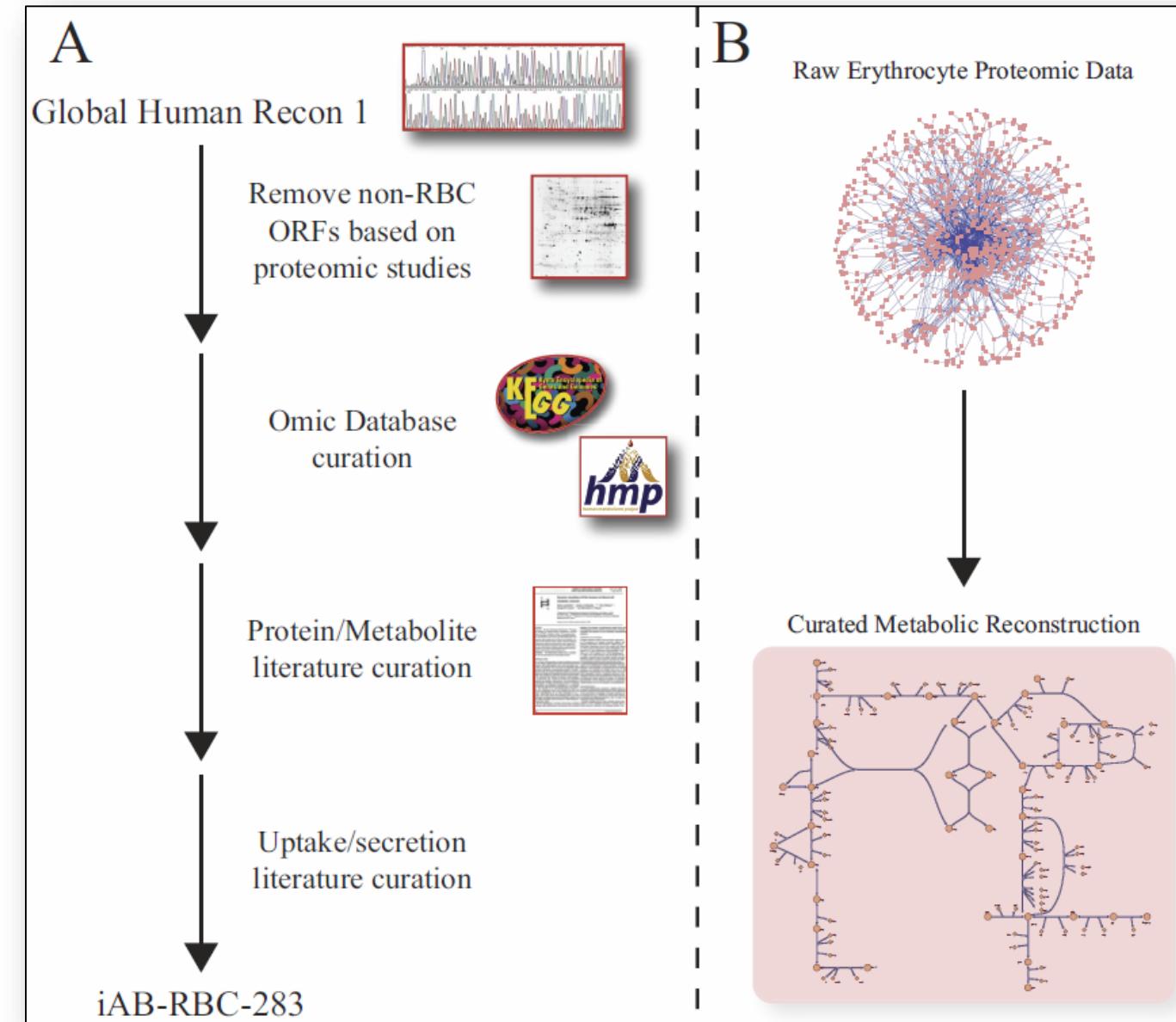
Red Blood Cell membrane major proteins

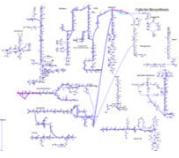


Building an *in silico* Red Blood Cell (Erythrocyte) Metabolic Network

- The three major required data types are: the human genome sequence, high-throughput data (specifically, proteomics for an enucleated cell), and primary literature
- To build the erythrocyte network, iAB-RBC-283, proteomics was used to remove non-erythrocyte related open reading frames (ORFs) or genes from Recon 1.
- Detailed curation utilizing protein, metabolite, and transport experimental literature was needed to build a high-quality metabolic reconstruction.
- Without network reconstruction and rigorous curation, the experimentally generated proteomic data is raw and difficult to interpret.

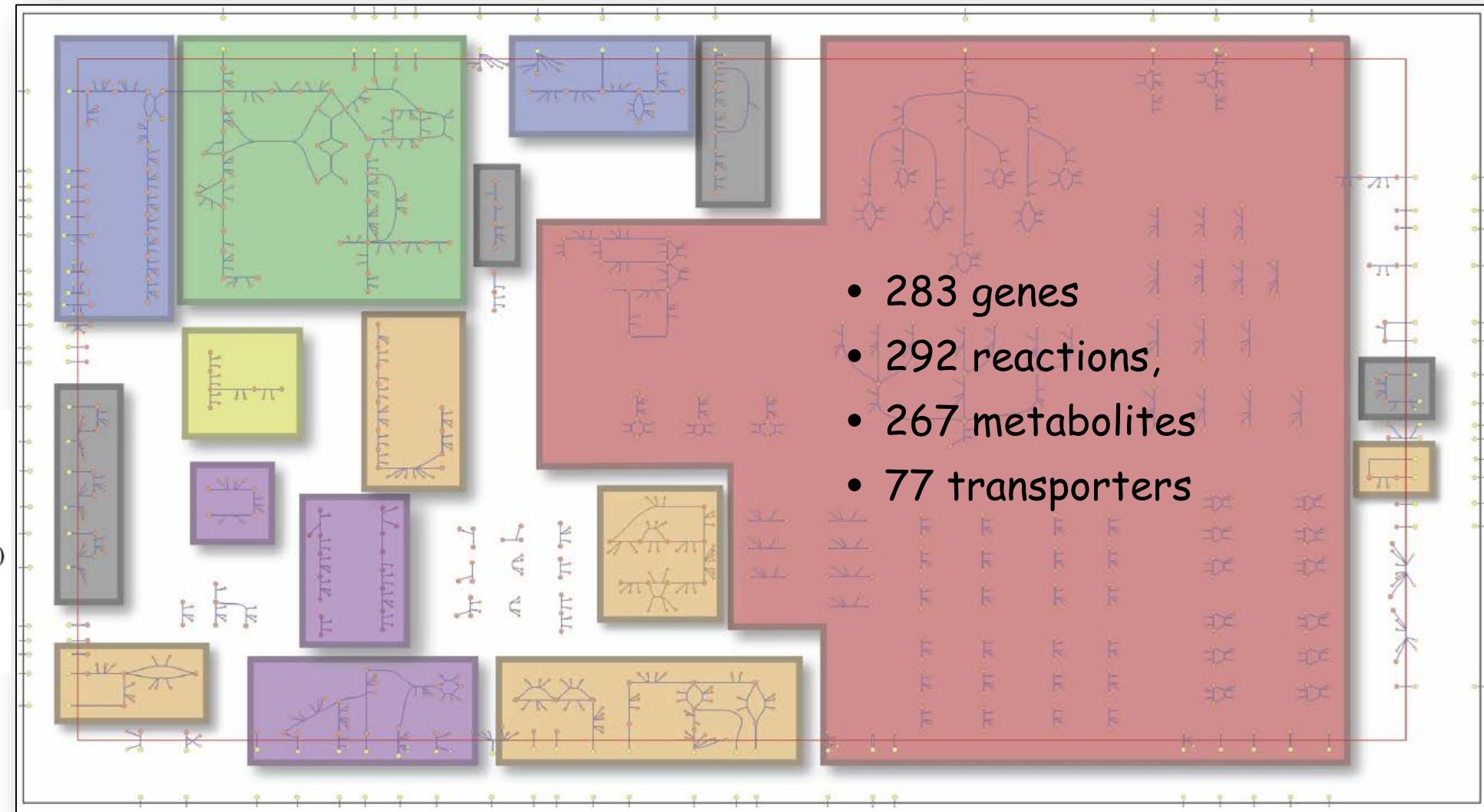
Bordbar, A., N. Jamshidi, et al. (2011). "iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states." *BMC systems biology* 5: 110.



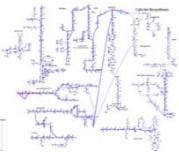


Topological Map of the Human Erythrocyte Metabolic Network

- █ Historic Erythrocyte Models
(Glycolysis/PPP/AMP Metabolism)
- █ Additional Carbohydrate Metabolism
(Galactose, Fructose, Mannose)
- █ Additional Nucleotide Metabolism
(XMP, UMP, GMP)
- █ Amino Acid Metabolism
(Arg, Glutathione, Homocysteine,
Polyamine, Met)
- █ Cofactor Metabolism
(NAD, Protopheme, Riboflavin, Thiamine
Vitamin B6, C)
- █ Lipid Metabolism
(Diacylglycerol Production, Phospholipid)
- █ Other
(Acetaldehyde, Aminosugar,
Catecholamine, Methylglyoxal)



Bordbar, A., N. Jamshidi, et al. (2011). "iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states." *BMC systems biology* 5: 110.

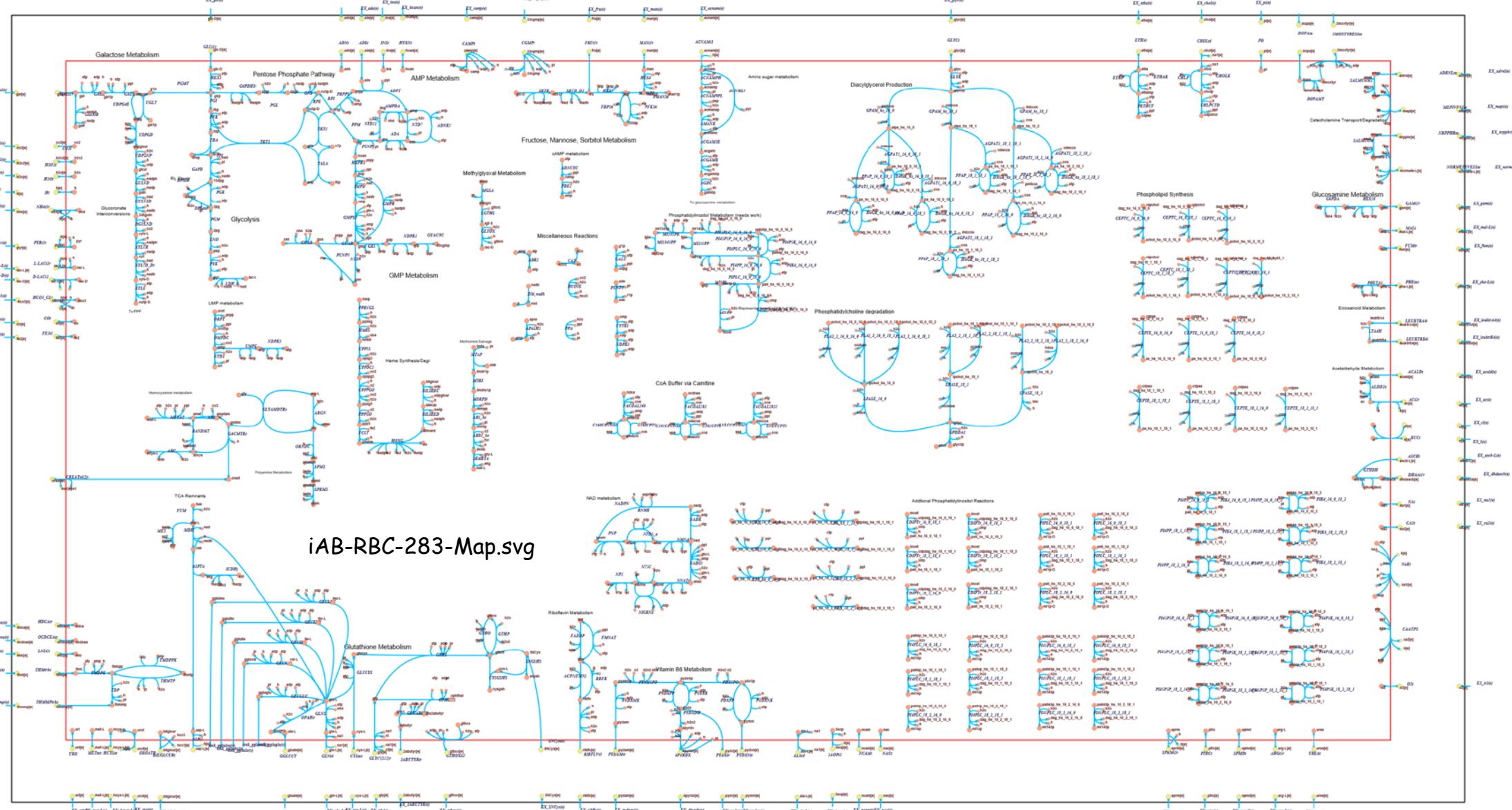


Constraint-based Metabolic Reconstructions & Analysis

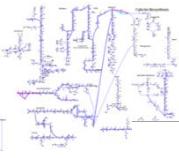
H. Scott Hinton, 2017

-18-

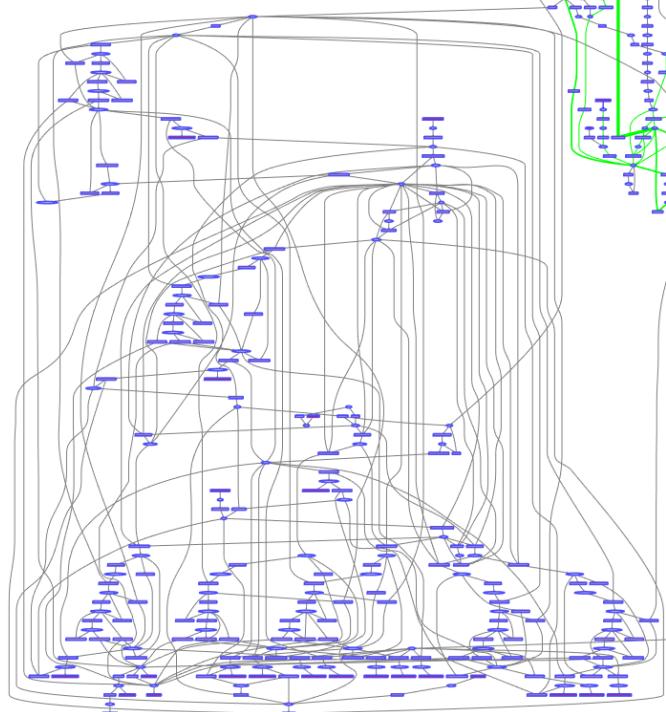
Red Blood
Cell Model
iAB-RBC-283



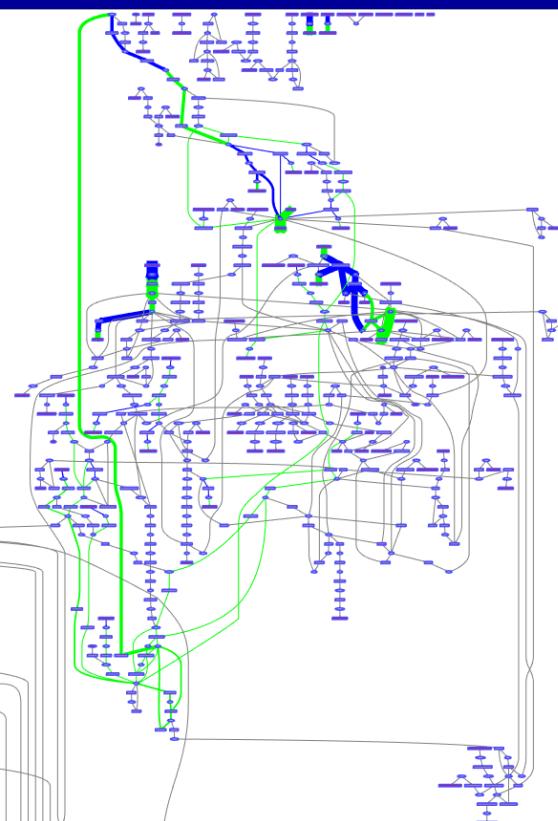
iAB-RBC-283-Map.svg



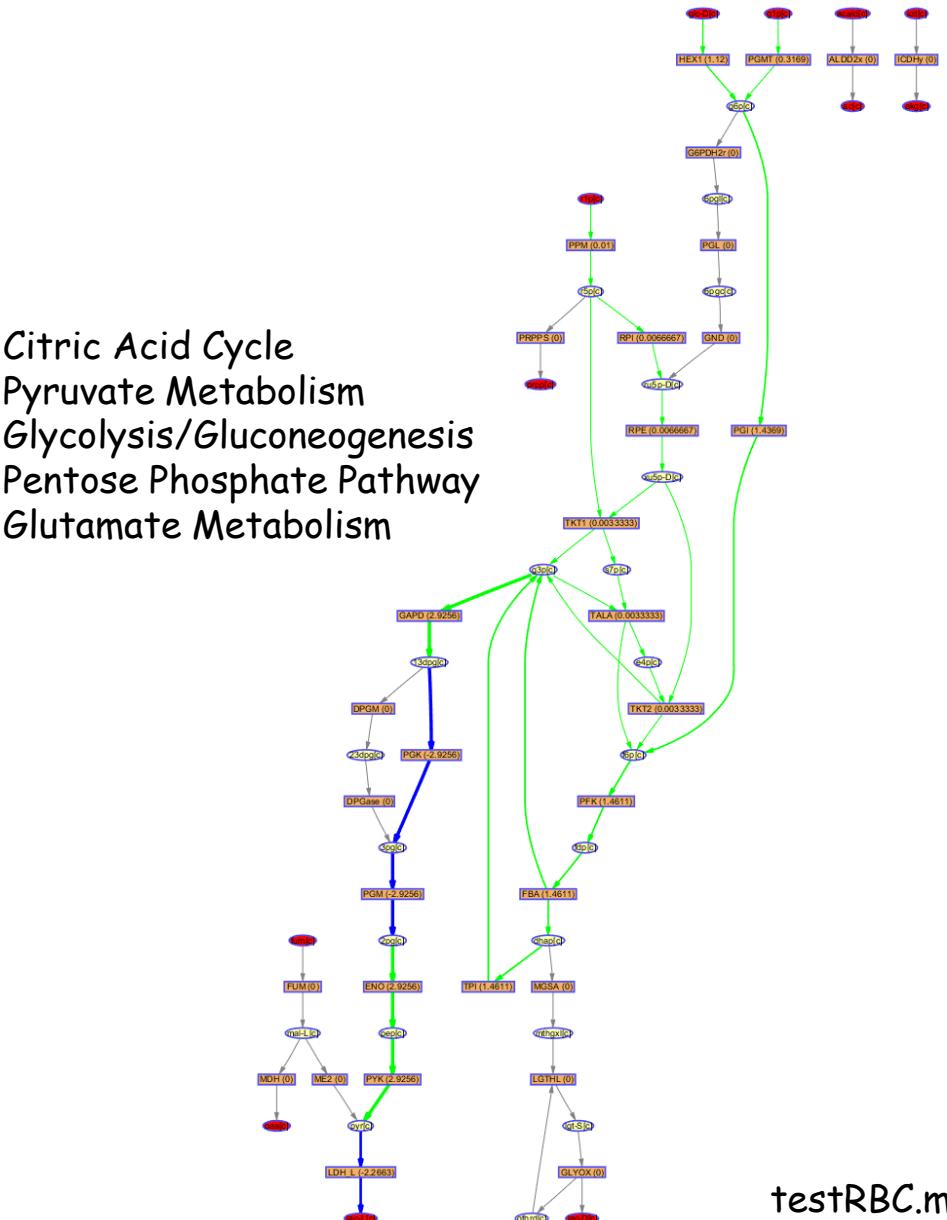
RBC Metabolic Map



Entire RBC Network



Citric Acid Cycle
Pyruvate Metabolism
Glycolysis/Gluconeogenesis
Pentose Phosphate Pathway
Glutamate Metabolism





RBC Flux Balance Analysis (Default, No Loops)

EX_ac(e)	3.74e-05	ACGAM2E	-3.74e-05	GLCt1r	1.12	PGI	1.4369
EX_acnam(e)	-3.74e-05	ACGAMK	3.74e-05	H2O2t	10	PGK	-2.92556
EX_ade(e)	0.01	ACNAMt2	3.74e-05	H2Ot	-4.00359	PGM	-2.92556
EX_adn(e)	-0.01	ACNMLr	3.74e-05	HCO3E	5.87112	PGMT	0.3169
EX_arg-L(e)	-0.1152	ACT2r	-3.74e-05	HCO3_CLt	-5.87112	PPM	0.01
EX_co2(e)	-5.75592	ADEt	-0.01	HEX1	1.12	PTRCT	-0.1152
EX_fru(e)	-0.0075	ADNt	0.01	HEX10	1e-05	PUNP1	0.01
EX_gal(e)	-0.3169	AGDC	3.74e-05	HEX4	0.01	PYK	2.92556
EX_gam(e)	-1e-05	ARGN	0.1152	HEX7	0.0075	PYRT2r	-0.659295
EX_glc(e)	-1.12	ARGt5r	0.1152	Ht	7.07445	RPE	0.00666667
EX_h(e)	10	CAT	5	KCCt	-5.87112	RPI	0.00666667
EX_h2o(e)	4.00359	CO2t	5.75592	L-LACt2r	-2.2663	TALA	0.00333333
EX_h2o2(e)	-10	DM_nadh	0.659258	LDH_L	-2.2663	TKT1	0.00333333
EX_hco3(e)	5.87112	ENO	2.92556	MAN6PI	0.01	TKT2	0.00333333
EX_lac-L(e)	2.2663	FBA	1.46111	MANt1r	0.01	TPI	1.46111
EX_man(e)	-0.01	FRUt1r	0.0075	NAt	8.80668	UDPG4E	-0.3169
EX_nh4(e)	4.74e-05	G6PDA	4.74e-05	NH4t3r	4.74e-05	UGLT	0.3169
EX_o2(e)	5	GALK	0.3169	NaKt	2.93556	UREAt	-0.1152
EX_ptrc(e)	0.1152	GALt1r	0.3169	O2t	-5		
EX_pyr(e)	0.659295	GAMt1r	1e-05	ORNDC	0.1152		
EX_urea(e)	0.1152	GAPD	2.92556	PFK	1.46111		

Objective Function
(ATP Production)

testRBC.m



Functional Assessment of Model

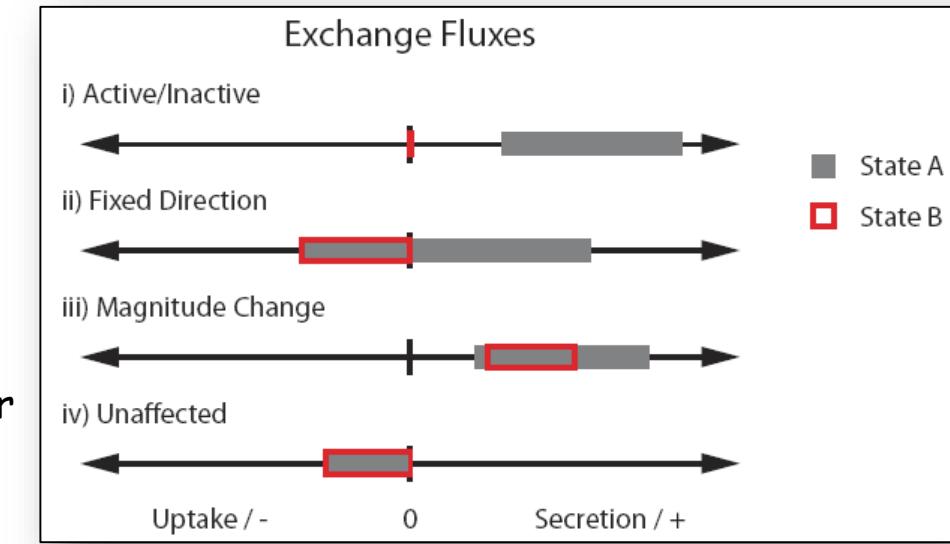
- Flux variability analysis (FVA) was utilized to determine the functional metabolic pathways of the erythrocyte network.
- FVA is used to define the bounding box on network capabilities (FVA determines the minimum and maximum allowable flux through each metabolic reaction).
 - ✓ Reactions with a calculated non-zero maximum or minimum have the potential to be active and have a potential physiological function. For a reaction to have a non-zero flux, the reaction must be linked to other metabolic reactions and pathways and plays a functional role in the system. Thus, potentially active reactions are deemed as functional.
 - ✓ Thus, FVA is used to determine the capability/capacity of the network reactions to determine metabolic functionality.
- After determining which reactions were functional, the reaction list was perused to determine pathway and subsystem functionality in the network.

Bordbar, A., N. Jamshidi, et al. (2011). "iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states." *BMC systems biology* 5: 110.



Analyzing iAB-RBC-283 as a Functional Biomarker

- The Morbid Map from the Online Mendelian Inheritance in Man (OMIM) and the DrugBank were downloaded from their respective databases. The enzyme names in iAB-RBC-283 were cross-referenced against the database entries to determine morbid single nucleotide polymorphisms (SNPs) in erythrocyte proteins and drugs with protein targets in the erythrocyte. The morbid SNPs that did not have sole pathological effects in the erythrocyte were classified using the Merck Manual.
- Just as FVA can be used to assess the function of a network under a particular set of constraints, it can also be used to assess the changes in function and thus has applications for characterizing disease states and identifying biomarkers. When simulating a morbid SNP or a drug inhibited enzyme, the lower and upper bound constraints on the affected reaction is set to zero. FVA is then used to characterize the exchange reactions under morbid SNP or drug treated conditions and then compared to the normal state.
- A reaction was considered to be confidently altered if the change in the minimum or maximum flux was 40% of the total flux span. The flux span is defined as the absolute difference between the original (unperturbed) maximum and minimum fluxes.



There are four major differences that can occur for an exchange reaction in two different states:
i) the reaction is either active (non-zero minimum or maximum flux) or inactive (zero minimum and maximum flux), ii) the exchange becomes fixed in one direction (uptake or secretion only), iii) there is a magnitude change in exchange, iv) the reaction is unaffected and is the same for both states.

Bordbar, A., N. Jamshidi, et al. (2011). "iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states." BMC systems biology 5: 110.



OMIM Database

OMIM® Online Mendelian Inheritance in Man®
An Online Catalog of Human Genes and Genetic Disorders
Updated 12 June 2015

Advanced Search : OMIM, Clinical Synopses, Gene Map
Need help? : Example Searches, OMIM Search Help, OMIM Tutorial
Mirror sites : us-east.omim.org, europe.omim.org

<http://omim.org/>

Amberger J, Bocchini CA, Scott AF, Hamosh A: McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res 2009, , 37 Database: D793-6.

OMIM Advanced Search

Search OMIM...

Sort by: Relevance Date updated Date created

Entries per page:

Search in:

- Mim Number
- Title
- Text
- Allelic Variants
- Clinical Synopsis
- Contributors

Only Records With:

- Allelic Variants
- Clinical Synopsis
- Gene Map Locus

MIM Number Prefix:

- * gene with known sequence
- + gene with known sequence and phenotype
- # phenotype description, molecular basis known
- % mendelian phenotype or locus, molecular basis unknown
- none - other, mainly phenotypes with suspected mendelian basis

Only Records without:

- Gene Map Locus

Chromosome:

- 1 2 3 4 5 6 7 8 9 10 11 12
- 13 14 15 16 17 18 19 20 21 22 X Y
- Mitochondrial Autosomal Unknown

Created From: To:

Updated From: To:

Note: Entries created before June 2, 1986 have a creation date of June 2, 1986.



Additional File #2: Detected SNPs and FVA Results for SNP Perturbations

OMIM	Location	Pathology	Gene Protein Name
115500	11p13	Acatalasemia (3)	CAT
108961	9p21-p12	Acromesomelic dysplasia, Maroteaux type, 602875 (3)	NPR2, ANPRB, AMDM
608958	20q13.11	Adenosine deaminase deficiency, partial, 102700 (3)	ADA
609712	1q21	Adenosine triphosphate, elevated, of erythrocytes, 102900 (3)	PKLR, PK1
180297	6p21.1-p11	Anemia, hemolytic, Rh-null, regulator type, 268150 (3)	RHAG, RH50A
606224	7p15-p14	Anemia, hemolytic, due to UMPH1 deficiency, 266120 (3)	NT5C3, UMPH1, PSN1
608313	6q23	Argininemia, 207800 (3)	ARG1
173335	6q22-q23	Arterial calcification, generalized, of infancy, 208000 (3)	ENPP1, PDNP1, NPPS, M6S1, PCA1, ARHR2
311850	Xq22-q24	Arts syndrome, 301835 (3)	PRPS1, CMTX5, DFNX1, DFN2
600650	1p32	CPT II deficiency, lethal neonatal, 608836 (3)	CPT2
600528	11q13	CPT deficiency, hepatic, type IA, 255120 (3)	CPT1A
600650	1p32	CPT deficiency, hepatic, type II, 600649 (3)	CPT2
154550	15q22-qter	Carbohydrate-deficient glycoprotein syndrome, type Ib, 602579 (3)	MPI, PMI1
182500		OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh. Its official home is omim.org.	
311850			
607672			
600179			
601785			
612732			
609414	2q35	Corneal fleck dystrophy, 121850 (3)	PIKFYVE, PIP5K3
191740	2q37	Crigler-Najjar syndrome, type I, 218800 (3)	UGT1A1, UGT1, GNT1, BILQLTL1
191740	2q37	Crigler-Najjar syndrome, type II, 606785 (3)	UGT1A1, UGT1, GNT1, BILQLTL1
311850	Xq22-q24	Deafness, X-linked 1, 304500 (3)	PRPS1, CMTX5, DFNX1, DFN2

Bordbar, A., N. Jamshidi, et al. (2011). "iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states." BMC systems biology 5: 110.

Affected Protein	ADA	AHCY	AK1
Affected Reaction	ADA	AHC	ADK1
Varying Exchanges	EX_ade(e)	EX_3moxtyr(e)	EX_35cgmp(e)
	EX_hxan(e)	EX_adrln(e)	EX_3moxtyr(e)
	EX_nh4(e)	EX_dopa(e)	EX_adrln(e)
		EX_hcyst-L(e)	EX_camp(e)
		EX_mepi(e)	EX_chol(e)
		EX_met-L(e)	EX_dopa(e)
		EX_normete-L(e)	EX_etha(e)
		EX_nrpphr(e)	EX_glyc(e)
			EX_hcyst-L(e)
			EX_hdca(e)
			EX_Inlc(e)
			EX_mepi(e)
			EX_met-L(e)
			EX_nac(e)
			EX_ncam(e)
			EX_normete-L(e)
			EX_nrpphr(e)
			EX_ocdcea(e)
			EX_orot(e)
			EX_spmd(e)
			EX_sprm(e)
			EX_thm(e)
			EX_thmmp(e)
			EX_uri(e)

A reaction was considered to be confidently altered if the change in the minimum or maximum flux was 40% of the total flux span. The flux span is defined as the absolute difference between the original (unperturbed) maximum and minimum fluxes.



DrugBank Database

**DrugBank Version 4.2**

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 7759 drug entries including 1600 FDA-approved small molecule drugs, 160 FDA-approved biotech (protein/peptide) drugs, 89 nutraceuticals and over 6000 experimental drugs. Additionally, 4282 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 200 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

<http://www.drugbank.ca/>

Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M: DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008, , 36 Database: D901-6.

DrugBank ID	Name	Weight	Structure	Categories	Therapeutic Indication
DB01048 136470-78-5	Abacavir	286.3323 <chem>C14H18N6O</chem>		Anti-HIV Agents / Reverse Transcriptase Inhibitors / Nucleoside and nucleotide reverse transcriptase inhibitors	For the treatment of HIV-1 infection, in combination with other antiretroviral agents.
DB05812 154229-19-3	Abiraterone	349.509 <chem>C24H31NO</chem>		Other hormone antagonists and related agents	Used in combination with prednisone for the treatment of metastatic, castration-resistant prostate cancer.
DB00659 77337-76-9	Acamprosate	181.21 <chem>C5H11NO4S</chem>		Alcohol Deterrents / Drugs used in alcohol dependence	For the maintenance of abstinence from alcohol in patients with alcohol dependence who are abstinent at treatment initiation
DB00284 56180-94-0	Acarbose	645.6048 <chem>C25H43NO18</chem>		Alpha glucosidase inhibitors	For treatment and management of diabetes type II (used in combination therapy as a second or third line agent)
DB01193 37517-30-9	Acebutolol	336.4259 <chem>C18H26N2O4</chem>		Antihypertensive Agents / Sympathomimetics / Adrenergic beta-1 Receptor Antagonists / Anti-Arrhythmia Agents / Beta blocking agents, selective	For the management of hypertension and ventricular premature beats in adults.



Additional File #3: Detected Drug Targets and FVA Results for Drug Effect Perturbations

RBC Protein	Associated Drug Bank Id	
ABCC4	DB01327	
ABCG2	DB00619, DB01054, DB01204, DB00622, DB01030, DB01098,	
ADA	DB00552, DB00242, DB00975, DB01073	
ADCY7	DB01200, DB01497	
AHCY	DB03216, DB02325	
AK1	DB01717	
AKR1B1	DB01689, DB02020, DB02021, DB02101, DB02132, DB02712,	
ALAD	DB02878	
ALDH1A1	DB00755	
ALDH1A2	DB00755	
AMD1	DB03375, DB03754	
AMPD3	DB00993, DB01033	
CA1	DB00311, DB00381, DB00703, DB00819, DB01144, DB01194,	
CA2	DB00869, DB01031, DB01671, DB01748, DB01784, DB01964,	
CAT	DB01213, DB04184	
CLC	DB02967, DB02983, DB02687	
COMT	DB00323, DB00494, DB01235	
CPT1A	DB01074	
CPT2	DB01074	
DCYD	Describes relationship between RBC genes and drugs in the drug bank	
ENPF		
FECO		
G6PD	DB03085, DB03461	
GALE	DB03095, DB04355, DB01867	
GALK1	DB04395	
GCK	DB02379	
GLO1	DB03130, DB03330, DB03602, DB03345, DB04132	
GMPR	DB00993, DB01033	

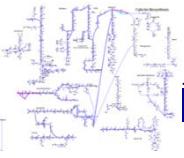
Drugs that affect RBC Proteins	Name	Drug Type
DB04899	Nesiritide	Approved,Biotech
DB01031	Ethinamate	Withdrawn,SmallMolecule
DB00755	Tretinoin	Approved,Nutraceutical
DB00180	Flunisolide	Approved,SmallMolecule
DB00203	Sildenafil	Approved,SmallMolecule
DB00242	Cladribine	Approved,SmallMolecule
DB00273	Topiramate	Approved,SmallMolecule
DB00494	Entacapone	Approved,SmallMolecule
DB00552	Pentostatin	Approved,SmallMolecule
DB00588	Fluticasone Propionate	Approved,SmallMolecule
DB00591	Fluocinolone Acetonide	Approved,SmallMolecule
DB00688	Mycophenolate mofetil	Approved,SmallMolecule
DB00727	Nitroglycerin	Approved,SmallMolecule
DB00806	Pentoxifylline	Approved,SmallMolecule
DB00820	Tadalafil	Approved,SmallMolecule
DB00909	Zonisamide	Approved,SmallMolecule
DB01030	Topotecan	Approved,SmallMolecule
DB01047	Fluocinonide	Approved,SmallMolecule
DB01204	Mitoxantrone	Approved,SmallMolecule
DB01232	Saquinavir	Approved,SmallMolecule
DB04880	Enoximone	Approved,SmallMolecule
DB00175	Pravastatin	Approved,SmallMolecule
DB00201	Caffeine	Approved,SmallMolecule

Affected Protein	ABCC4	ABCG2	ADA	ADCY(2-7)
Affected Reactions	CAMPt/CGMPt	RIBFLVt3o	ADA	ADNCYC
Varying Exchanges	EX_camp(e) EX_35cgmp(e)		EX_ade(e) EX_hxan(e) EX_nh4(e)	EX_camp(e)

A reaction was considered to be confidently altered if the change in the minimum or maximum flux was 40% of the total flux span. The flux span is defined as the absolute difference between the original (unperturbed) maximum and minimum fluxes.

Spreadsheet also includes drug descriptions like: Fluocinonide - A topical glucocorticoid used in the treatment of eczema. [PubChem]

Bordbar, A., N. Jamshidi, et al. (2011). "iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states." BMC systems biology 5: 110.



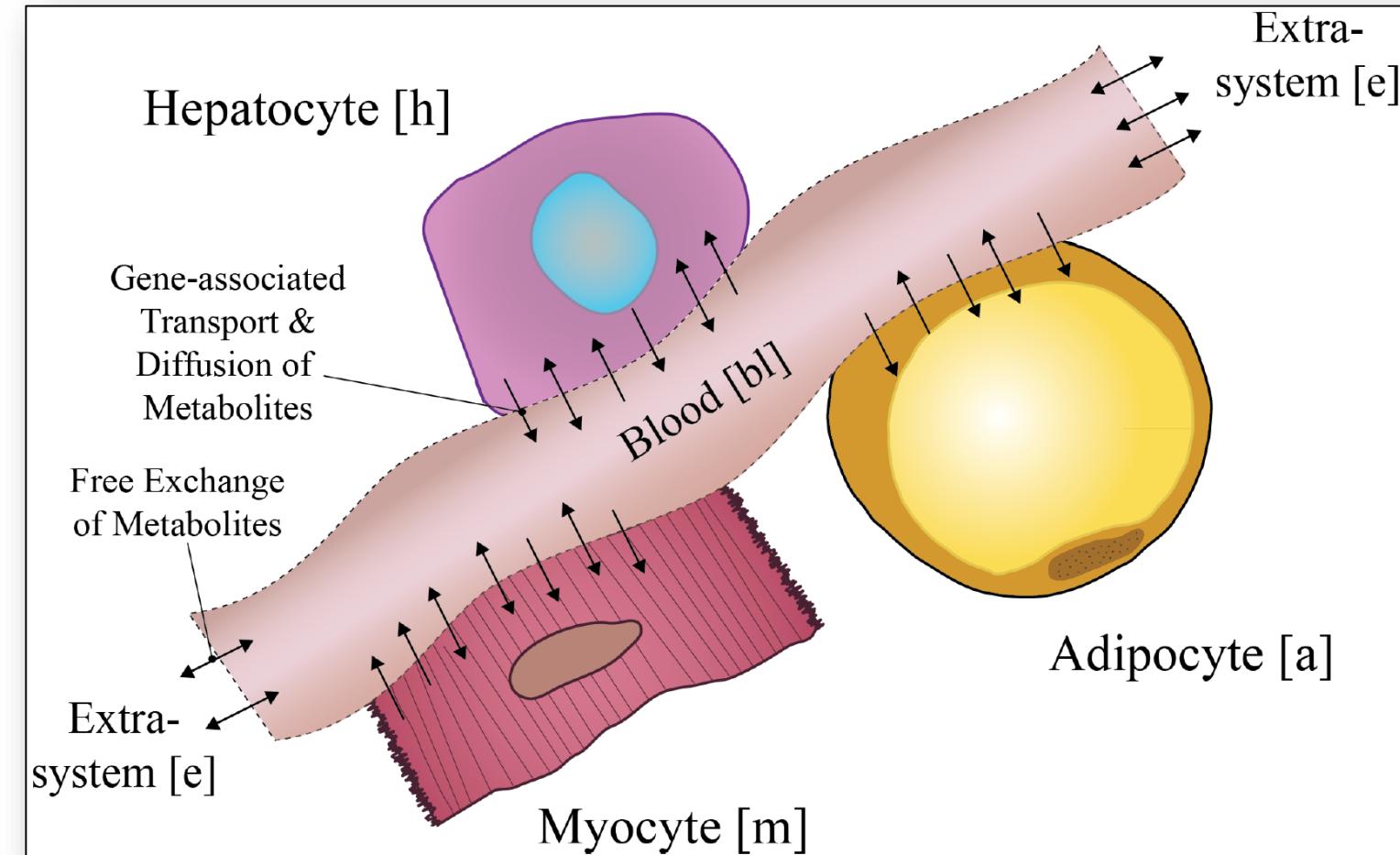
Lesson Outline

- Overview
- Creating Tissue-specific Models
- Tissue-specific Example
 - ✓ Red Blood Cells
 - ✓ Multi-tissue Modeling

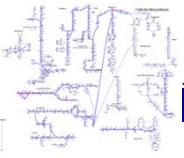


Multi-Tissue Modeling

- Three cell-specific reconstructions (Recon 1 & GIMME) for human cells including liver (hepatocyte), fat (adipocyte), and skeletal muscle (myocyte) are combined into a multi-tissue model by connecting them all to a new blood compartment.
- Metabolites enter the model through the extra-system through exchange reactions.
- Metabolites are then imported into the different cells through gene associated intercellular transporters and/or free diffusion.
- Compartment Reaction Notation
 - ✓ A:ATPM, H:ATPM, M:ATPM

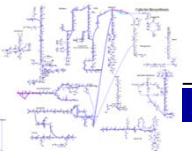


Bordbar et al.: A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. BMC Systems Biology 2011 5:180.



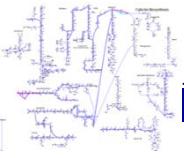
Lesson Outline

- Overview
- Creating Tissue-specific Models
- Tissue-specific Example
 - ✓ Red Blood Cells
 - ✓ Multi-tissue Modeling



Reflective Questions

- What is the process to create a context-specific metabolic reconstruction?
- What are the different algorithms that can be used to automatically create context-specific tissues?
- What is the OMIM database?
- What are SNP perturbations?
- What are drug targets?
- What is the DrugBank database?

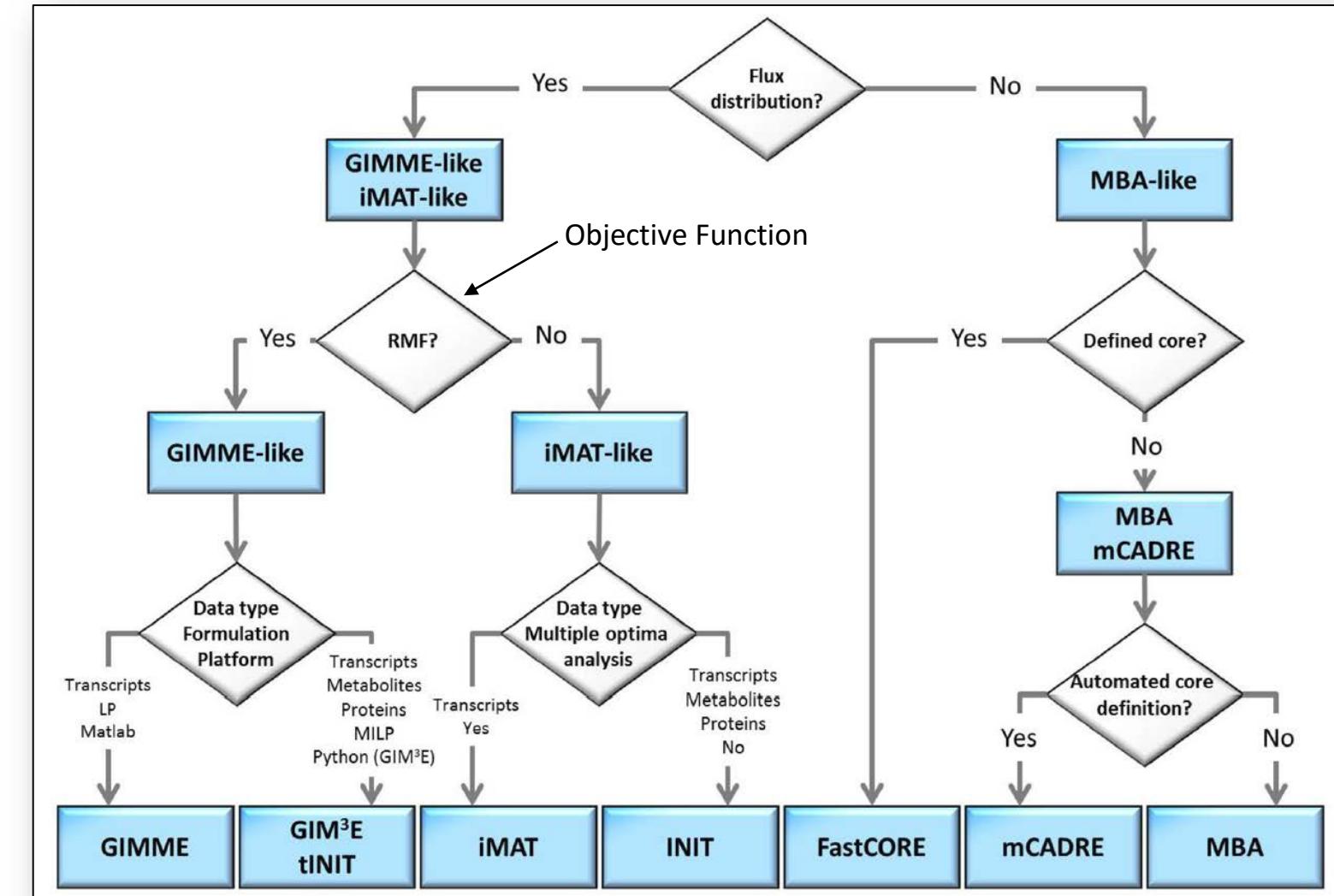


Appendix



Flowchart Context-specific Reconstruction Methods

Optimal choice of methodologies when tackling a context-specific reconstruction problem. The choice can be made by answering a few questions, in a flowchart manner, related to: demand of model extraction and flux prediction, knowledge on a required metabolic functionality, the type of experimental data available or the computational platform.



Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." Frontiers in plant science 5.



GIMME-like Family

- The objective employed by the GIMME-like family corresponds to the *similarity of the flux phenotype to data*, which is to be maximized while guaranteeing a given Required Metabolic Functionality (RMF), such as: growth or ATP production.
- This family reconstructs a context-specific model in two steps:
 - ✓ First, it optimizes an objective function, the RMF, by using the classical linear programming (LP) formulation of flux balance analysis which imposes mass balance and thermodynamic constraints. This objective function is assumed to be the main cellular task in the investigated condition.
 - ✓ It then involves solving a second LP that minimizes a penalty function, corresponding to the discrepancies between *flux values* and the *respective transcript levels*, with the additional constraint that the flux through the previous RMF must be above a given lower bound (e.g. a fraction of the optimum value found by flux balance analysis).
- The methods included in this family mainly differ in the way the discrepancies are minimized in the second step, the type of high-throughput data used, and in the treatment of reversible reactions

```
1. function GIMME-like (S,RMF,k)
2. maxv RMF
   s.t.
   SV = 0
   vmin ≤ v ≤ vmax
3. minv IS
   s.t.
   SV = 0
   vmin ≤ v ≤ vmax
   RMF = k * RMFopt
   k ∈ [0,1]
4. end function
```

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." Frontiers in plant science 5.



GIMME-like Family

GIMME

Becker, S. A. and B. O. Palsson (2008). "Context-specific metabolic networks are consistent with experiments." *PLoS computational biology* 4(5): e1000082.

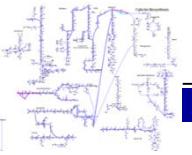
- The penalty function is termed inconsistency score.
- This function penalizes flux values of reactions whose associated expression levels are below a user-defined cut-off (i.e., threshold). More specifically, the inconsistency score is given by the dot product of the flux distribution and the reaction penalty, defined as the vector difference of the associated expression values from the threshold.
- The reaction associated expression level is obtained following the standard GPR rules (Becker and Palsson, 2008), which take into account the presence of isoenzymes and protein complexes.
- Although transcript profiles were used in the original formulation, a variant called GIMMEm allows for the integration of proteomic data (Bordbar, et al. (2012). "Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation." *Molecular Systems Biology*).
- The result of applying this algorithm is a flux distribution which ensures that a given RMF can be carried out and is as consistent as possible to the employed data.

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." *Frontiers in plant science* 5.

GIM³E

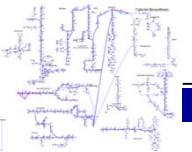
Schmidt, B. J., et al. (2013). "GIM³E: condition-specific models of cellular metabolism developed from metabolomics and expression data." *Bioinformatics* 29(22): 2900-2908.

- GIM³E introduces several modifications to the original GIMME.
 - First, it allows integration of metabolomics data, imposing a nonzero flux value to reactions involving a metabolite for which there is evidence of being synthesized in an investigated condition.
 - Second, it modifies the definition of the reaction penalty; here, the penalties for all reaction-associated genes are determined separately and are then mapped to the reaction following the GPR rules. Moreover, the penalties are calculated as the distance between each transcript and the maximum expression level of the set. Consequently, after mapping transcript penalties all reactions obtain a penalty value, rather than only the set below the threshold which is the case in GIMME.
 - Finally, GIM³E takes into account directionality of reversible reactions by constraining them to operate in only one direction, which is modeled by introducing a binary variable for the direction of choice. As a result, GIM³E is formulated as a mixed integer linear program (MILP), which is more computationally challenging than the LP formulation of GIMME.



Create a Tissue Model of the Eye

- Step #1 - Collect and format the gene expression data for the non-treated ARPE-19 cells
 - Gene Expression Omnibus data (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5741>)
 - GSM133871.CEL, GSM133872.CEL, GSM133873.CEL (Non-treatment)
- Step #2 - Convert the Affymetrix data to a format that can be used by the "createTissueSpecific()" Cobra Toolbox function
 - A script has been created by Farhad Farjood using the language R to get the Absence/Presence (AP.txt) data and Entrez IDs (EID.txt).
 - R is open source and can be downloaded for free at <http://cran.rstudio.com/>
 - A function call getExpD() will convert AP.txt and EID.txt to expression data that can be used by the Cobra Toolbox
- Step #3 - Create a tissue-specific model with the Cobra Toolbox
 - Use an update version of "createTissueSpecific()" called "createTissueSpecificRec()" that has been modified by Farhad Farjood to work with Recon 1 and Recon 2.
- Step #4 - Create a specific objective function based on the cellular functions of reabsorption and secretion.
- Step #5 - Use manual assessment with primary literature to validate the physiological functions for accuracy.

**createTissueSpecificModel****Create draft tissue specific model from mRNA expression data**

```
[tissueModel,Rxns] = createTissueSpecificModel(model,expressionData,proceedExp,orphan,exRxnRemove,solver,options,funcModel)
```

INPUTS

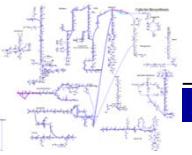
model global recon1 model
expressionData mRNA expression Data structure
Locus Vector containing GeneIDs
Data Presence/Absence Calls
 Use: (1 - Present, 0 - Absent) when proceedExp = 1
 Use: (2 - Present, 1 - Marginal, 0 - Absent)
 when proceedExp = 0
Transcript RefSeq Accession (only required if proceedExp = 0)

OPTIONAL INPUTS

proceedExp 1 - data are processed ; 0 - data need to be processed (Default = 1)
orphan 1 - leave orphan reactions in model for Shlomi Method 0 - remove orphan reactions (Default = 1)
exRxnRemove Names of exchange reactions to remove (Default = [])
solver Use either 'GIMME' or 'Shlomi' to create tissue specific model (Default = 'GIMME')
options If using GIMME, enter objectiveCol here
Default: objective function with 90% flux cutoff, written as: [find(model.c) 0.9]
funcModel 1 - Build a functional model having only reactions that can carry a flux (using FVA), 0 - skip this step (Default = 0)

OUTPUTS

tissueModel Model produced by GIMME or Shlomi, containing only reactions carrying flux
Rxns Statistics of test
 ExpressedRxns - predicted by mRNA data
 UnExpressedRxns - predicted by mRNA data unknown - unable to be predicted by mRNA data
 Upregulated - added back into model
 Downregulated - removed from model
 UnknownIncluded - orphans added



Build a draft tissue-specific human macrophage model from the global human metabolic network and omics data

Download the MAT file "testTissueModel.mat" from "<https://github.com/opencobra/cobratoolbox/tree/master/testing/testTissueModel>." It contains the global human metabolic network model and a formatted expressionData structure. The model is the version of the human metabolic network reconstruction Recon 1 that was used to create an alveolar macrophage model¹ using expression data from Kazerlos et al.²

```
>> load('testTissueModel.mat')
```

The GIMME algorithm retains reactions from Recon 1 that are orphans or are present in the high-throughput data. The reactions with no detected expression are minimized and those not required to retain flux through the objective reaction are removed.

```
>> [tissueModel,Rxns] = createTissueSpecificModel(model,expressionData);
```

Where tissueModel is the GIMME algorithm-derived draft model; and Rxns is a structure with lists of all the reactions. The reactions fall into the following categories:

- Expressed—1,769 potentially active reactions based on transcriptome data;
- UnExpressed—497 reactions without requisite gene products based on transcriptome data;
- Unknown—41 reactions unable to be predicted by transcriptome data;
- Upregulated—52 UnExpressed reactions added back into model;
- Downregulated— 0 Expressed reactions removed from model; and
- UnknownIncluded—1,476 orphan reactions included.

1. Bordbar, A., Lewis, N.E., Schellenberger, J., Palsson, B.O. & Jamshidi, N. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol. Syst. Biol.* 6, 422 (2010).
2. Kazerlos, A. et al. Overexpression of apoptotic cell removal receptor MERTK in alveolar macrophages of cigarette smokers. *Am. J. Respir. Cell Mol. Biol.* 39, 747-757 (2008).

Schellenberger, J., R. Que, et al. (2011). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0." *Nature protocols* 6(9): 1290-1307.



Step #1 - Collect Gene Expression Files for the Non-treated ARPE-19 Cells

NCBI

GEO Gene Expression Omnibus

HOME | SEARCH | SITE MAP

NCBI > GEO > Accession Display

Not logged

Scope: Self Format: HTML Amount: Quick GEO accession: GSE5741 GO

Series GSE5741 Query DataSets for GSE5741

Status	Public on Sep 01, 2006
Title	Expression data from ARPE-19 cells treated with LDL or ox-LDL
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	LDL or Ox-LDL 200ug/ml, which showed no loss of viability after a 48 hour exposure, induced a physiological and pathological transcriptional response, respectively. LDL induced a downregulation of genes associated with cholesterol biosynthesis while ox-LDL induced transcriptional alterations in genes related to inflammation, matrix expansion, lipid metabolism and processing, and apoptosis. Pentraxin-3 was secreted into the culture medium after RPE cells were stimulated with ox-LDL, and immunohistochemically evident in Bruch's membrane of human macular samples with age-related macular degeneration. ARPE-19 cells exposed to 200?g/ml ox-LDL had a 38% apoptosis rate compared to less than 1% when exposed to LDL or untreated controls ($p<0.0001$). While LDL induced a physiologic response by RPE cells, a pathological phenotypic response was seen after treatment with oxidatively modified LDL. The transcriptional, biochemical, and functional data provide initial support of a role for the hypothesis that modified LDLs are one trigger for initiating events that contribute to the development of age-related macular degeneration.
Keywords:	treatment with non-treatment control

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5741>

Contributor(s) Yamada Y, Tian J, Yang Y, Cutler R, Wu T, Handa JT

Citation(s) Yamada Y, Tian J, Yang Y, Cutler RG et al. Oxidized low density lipoproteins induce a pathologic response by retinal pigmented epithelial cells. *J Neurochem* 2008 May;105(4):1187-97. PMID: [18182060](#)

Submission date Sep 01, 2006

Last update date May 14, 2015

Contact name Yanqin Yang

Organization name The Johns Hopkins University

Street address 600 N. Wolfe Street

City Baltimore

State/province MD

ZIP/Postal code 21287

Country USA

Platforms (1) [GPL570](#) [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array

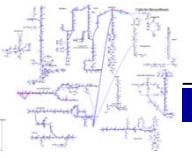
Samples (9)

[GSM133871](#) non-treatment, biological rep1
[GSM133872](#) non-treatment, biological rep2
[GSM133873](#) non-treatment, biological rep3

[GSM133874](#) LDL-treatment, biological rep1
[GSM133875](#) LDL-treatment, biological rep2
[GSM133876](#) LDL-treatment, biological rep3
[GSM133877](#) OX-LDL-treatment, biological rep1
[GSM133878](#) OX-LDL-treatment, biological rep2
[GSM133879](#) OX-LDL-treatment, biological rep3

Relations

BioProject [PRJNA97055](#)



R Filename = AP_HSH.R

```
source("http://bioconductor.org/biocLite.R")
biocLite("affy")
biocLite("simpleaffy")
biocLite("annotate")
biocLite("hgu133plus2.db")
```

```
library('affy')
library('simpleaffy')
library('annotate')
library('hgu133plus2.db')
```

setwd("/... /GSE5741_RAW") **Set working directory**

```
affy.data = ReadAffy()
```

```
data.AP = mas5calls(affy.data)
data.AP.calls = exprs(data.AP)
```

```
Affymetrix_ID <- row.names(data.AP.calls)
Entrez_ID <- getEG(Affymetrix_ID, annotation(affy.data))
```

```
write.table(data.AP.calls[,1:3], file="AP1.txt", row.names=FALSE,col.names=F, quote=F, sep="\t")
write.table(Entrez_ID, file="EID.txt", row.names=FALSE,col.names=F, quote=F, sep="\t")
```

Step #2 - Convert the Affymetrix Data to a Format That can be Used by the "createTissueSpecific()" Cobra Toolbox Function

P	P	P
P	P	P
A	A	M
P	P	P
A	A	A
A	A	A
P	P	P
A	A	A
A	A	M
A	A	A
A	A	A
P	P	P
P	P	P
P	P	P
A	A	A
A	A	A
P	P	P
P	P	P
A	A	A
A	A	A
A	A	A
A	A	A

AP.txt

NA
5982
3310
7849
2978
NA
7067
11099
6352
1571
2049
2101
1548
949
23170
112597
10406
5594
5594
203102
128153
163154

EID.txt



Step #3 - Create a Tissue-specific GIMME-based Model with the Cobra Toolbox

```
% CreateARPE19GIMMEmodel.m  
  
clear;  
  
load('Recon2.mat');  
  
ExpressionData = getExpD('EID.txt','AP.txt');  
  
[ARPE19_GIMME,Rxns] = createTissueSpecificRec(Recon2_2,ExpressionData,1,1,[],'GIMME',options,1);  
  
writeCbModel(ARPE19_GIMME,'sbml','ARPE19_GIMME_Model');
```

Variables - ExpressionData	
ARPE19_GIMME ExpressionData	
1x1 struct with 2 fields	
Field ▲	Value
Data	54675x1 double
Locus	54675x1 double

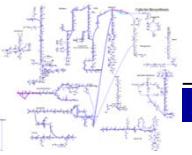
Expression Data

Variables - Rxns	
Rxns	
1x1 struct with 6 fields	
Field ▲	Value
Expressed	3268x1 cell
UnExpressed	929x1 cell
unknown	55x1 cell
UpRegulated	0x1 cell
DownRegulated	1129x1 cell
UnknownIncluded	1685x1 cell

Rxns

Variables - ARPE19_GIMME	
ARPE19_Shlomi ARPE19_GIMME	
1x1 struct with 36 fields	
Field ▲	Value
S	2121x3824 sparse double
rxns	3824x1 cell
Ib	3824x1 double
ub	3824x1 double
rev	3824x1 double
c	3824x1 double
rxnGeneMat	3824x2194 sparse double
rules	3824x1 cell
genes	2194x1 cell
grRules	3824x1 cell
subSystems	3824x1 cell
rxnNames	3824x1 cell
rxnKeggID	7440x1 cell
rxnConfidenceEcoLDA	7440x1 cell
rxnConfidenceScores	7440x1 cell
rxnsboTerm	7440x1 cell
rxnReferences	3824x1 cell
rxnECNumbers	3824x1 cell
rxnNotes	3824x1 cell
mets	2121x1 cell
b	2121x1 double

ARPE19 Model



iMAT-like Family

- The iMAT-like family comprises three methods, iMAT (Shlomi et al., 2008), INIT (Agren et al., 2012) and its extension, tINIT (Agren et al., 2014).
- The iMAT-like family does not assume a RMF achieved by the cell. More specifically, these methods maximize the number of matches between reaction states (i.e., active or inactive) and corresponding data states (i.e., expressed or not non-expressed).
- The mathematical formulation results in a MILP, in which the value of the binary variable denotes the most concordant reaction state for a given (data) context.
- Although sharing the general strategy, iMAT, INIT and tINIT differ considerably respecting to how they deal with data:
 - iMAT integrates data in the constraints,
 - INIT and tINIT do so directly in the objective function.

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." Frontiers in plant science 5.



iMAT-like Family

iMAT (Shlomi)

Shlomi, T., M. N. Cabili, et al. (2008). "Network-based prediction of human tissue-specific metabolism." *Nat Biotechnol* 26(9): 1003-1010.

- The algorithm first classifies reactions into two groups based on a previously defined threshold for the corresponding expression data; this results in the groups of reactions with a high and low associated expression values. It then maximizes the number of matches between a reaction state, defined through a minimum flux value, and the group to which the reaction belongs. Thus, if a reaction is included in the highly expressed group, the aim is to obtain a flux value over the minimum.
- iMAT tackles this issue through an adapted flux variability analysis (FVA): First, it forces each reaction to be active and evaluates the similarity, and then repeats the process in a similar way by forcing each reaction to be inactive. The final outcome is computed by comparing the two obtained similarities. A reaction is termed active if its inclusion results in higher similarity to data, and it is termed as inactive, if its inclusion decreases this similarity. In the case that both similarities are equal, iMAT categorizes the reaction as undetermined.

INIT

Agren, R., S. Bordel, et al. (2012). "Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT." *PLoS computational biology* 8(5): e1002518.

- INIT was optimized to integrate evidences from the Human Protein Atlas, although expression data are integrated when proteomic evidences are missing.
- INIT does not group reactions in categories in contrast to iMAT. Instead, it adopts experimental data to weight the binary variable of the corresponding reaction, whereby the weight is a function of experimental data (e.g., gene expression profiles) or a set of arbitrary numbers that quantify the color code of the entries of the Human Protein Atlas.
- INIT imposes a positive net production of metabolites for which there is experimental support for that context or tissue. Hence, when a metabolite is experimentally determined to be present, its net production is forced to comply with a given lower bound. As a result, INIT allows the integration of metabolomics data in a qualitative way.
- This method has been applied to generate a human metabolic reaction database ("Human Metabolic Atlas2") where several tissue-specific model reconstructions can be examined.

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." *Frontiers in plant science* 5.



Step #3 - Create a Tissue-specific Shlomi-based Model with the Cobra Toolbox

```
% CreateARPE19ShlomiModel.m
clear;
load('Recon2.mat');
ExpressionData = getExpD('EID.txt','AP.txt');
[ARPE19_ShloMi,Rxns] = createTissueSpecificRec(Recon2_2,ExpressionData,1,1,[],'Shlomi',options,1);
writeCbModel(ARPE19_ShloMi,'sbml','ARPE19_ShloMi_Model');
```

Variables - ExpressionData	
ARPE19_GIMME ExpressionData	
1x1 struct with 2 fields	
Field	Value
Data	54675x1 double
Locus	54675x1 double

Expression Data

Variables - Rxns	
Rxns Rxns.solution	
1x1 struct with 7 fields	
Field	Value
solution	1x1 struct
Expressed	3268x1 cell
UnExpressed	929x1 cell
unknown	55x1 cell
UpRegulated	38x1 cell
DownRegulated	923x1 cell
UnknownIncluded	2874x1 cell

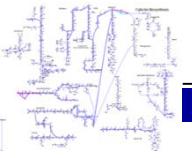
Rxns

Variables - Rxns.solution	
Rxns Rxns.solution	
Rxns.solution	
Field	Value
cont	7440x1 double
int	7465x1 double
obj	3.2360e+03
solver	'gurobi5'
stat	1
origStat	'OPTIMAL'
time	3.6410
full	14905x1 double

Rxns.solution

Variables - ARPE19_ShloMi	
ARPE19_ShloMi	
1x1 struct with 36 fields	
Field	Value
S	3709x5257 sparse double
rxns	5257x1 cell
Ib	5257x1 double
ub	5257x1 double
rev	5257x1 double
c	5257x1 double
rxnGeneMat	5257x2194 sparse double
rules	5257x1 cell
genes	2194x1 cell
grRules	5257x1 cell
subSystems	5257x1 cell
rxnNames	5257x1 cell
rxnKeggID	7440x1 cell
rxnConfidenceEcoliDA	7440x1 cell
rxnConfidenceScores	7440x1 cell
rxnsboTerm	7440x1 cell
rxnReferences	5257x1 cell
rxnECNumbers	5257x1 cell
rxnNotes	5257x1 cell
mets	3709x1 cell
b	3709x1 double

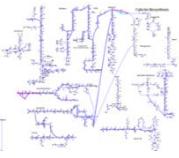
ARPE19 Shlomi Model



MBA-like Family

- The MBA-like family is composed of MBA,mCADRE and FastCORE.
- While previous methods perform both a flux prediction and a context-specific model reconstruction, MBA-like methods only return a context-specific model as output.
- This family a priori categorizes reactions in two sets, the core and the non-core.
 - The core set includes those reactions with positive evidences (e.g., high throughput data and/or well-curated biochemical knowledge) of being active in a certain context. Once these sets are defined, the MBA-like methods prune the GEM by eliminating non-core reactions that are unnecessary to ensure consistency in the core set, i.e., no blocked reaction is allowed in the final model. Thereby, all reactions must carry non-zero flux in at least one feasible solution.
- Checking model consistency is a crucial part of these methods and also the main difference in comparison to the other methods.
- FVA have been used to pinpoint blocked reactions, but it is computationally expensive since it requires solving two optimization problems per reaction. Thus, the major changes in formulation are due to finding faster alternatives to perform the same task.
- Other differences arise when defining the core set and during the pruning process.

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." *Frontiers in plant science* 5.



MBA-like Family

MBA

- MBA divides the core set in two subcores: a set with high likelihood to be present in the context-specific model (*CH*), if evidence comes from well-curated biochemical knowledge in that particular context, and a set with moderate likelihood (*CM*) if evidence comes from context-specific high-throughput data.
- The algorithm performs the pruning iteratively and randomly by selecting a non-core (*NC*) reaction to be eliminated, and checking consistency at each step: if *CH* and a user-defined fraction of *CM* remain unblocked, MBA removes the reaction out of the model along with *CM* and *NC* corresponding blocked reactions. This routine is repeated until no reaction is left in *NC*.
- The topology of the final model clearly depends on the order in which non-core reactions are eliminated. Therefore, to remove artifacts due to the order, the algorithm is repeated a number of times (1000) to obtain a population of context-specific models. Later, reactions are ranked according to their occurrence in the population and added up to *CH* until a consistent model is obtained.
- MBA proposes an alternative to FVA to check consistency in a more efficient way: First, it solves a LP problem which maximizes the total sum of fluxes. It then removes active reactions (i.e., carrying non-zero flux) and repeats the LP over the remaining set of reactions. If no reaction is found to be active, FVA is applied to each reaction to determine whether it is blocked. The process is repeated until all reactions have been classified either as blocked or unblocked.

mCADRE

- A prominent characteristic of mCADRE lies in ranking reactions of the genome-scale reconstruction according to three scores: expression-, connectivity-, and confidence-level-based.
- This ranking determines the core set of reactions as well as the order by which non-core reactions are eliminated. The core is determined by fixing a threshold value to the expression based score; therefore, reactions whose values are above the threshold are included in the core, and the rest constitute the non-core reactions.
- Unlike other methods, the expression-based score does not directly consider the levels of expression. Instead, it calculates the frequency of expressed states over a battery of transcript profiles in the same context, and, thus, requires a previous binarization of the expression data.
- Reactions outside the core are then ranked according to the connectivity-based score, which assesses the connectedness of adjacent reactions, and the confidence level-based score, which accounts for the type of evidences supporting a reaction in the genome-scale reconstruction.

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." *Frontiers in plant science* 5.



MBA-like Family

mCADRE

- A prominent characteristic of mCADRE lies in ranking reactions of the genome-scale reconstruction according to three scores: expression-, connectivity-, and confidence-level-based.
- This ranking determines the core set of reactions as well as the order by which non-core reactions are eliminated. The core is determined by fixing a threshold value to the expression based score; therefore, reactions whose values are above the threshold are included in the core, and the rest constitute the non-core reactions.
- Unlike other methods, the expression-based score does not directly consider the levels of expression. Instead, it calculates the frequency of expressed states over a battery of transcript profiles in the same context, and, thus, requires a previous binarization of the expression data.
- Reactions outside the core are then ranked according to the connectivity-based score, which assesses the connectedness of adjacent reactions, and the confidence level-based score, which accounts for the type of evidences supporting a reaction in the genome-scale reconstruction.

- Non-core reactions are in turn sequentially removed according to the previous ranking, and consistency is evaluated.
- Here, mCADRE presents two other innovations: it defines a set of key metabolites, with positive evidences of appearing in the context specific model reconstruction, and relaxes the stringent condition of including all core reactions in the final model. More specifically, a reaction can only be eliminated if it does not prevent the production of a key metabolite and if it is unnecessary to ensure core consistency. However, if evidence exists for the respective transcript to be unexpressed in any of the context-specific samples, mCADRE allows the elimination of the reaction even if it blocks some of the core reactions.
- To this end, two conditions have to be satisfied: (1) production of key metabolites is not impaired and (2) the relation between the number of blocked core and non-core reactions matches a predefined ratio.
- To check model consistency, mCADRE maintains the procedure proposed in MBA, although adapted to use FastFVA instead of maximizing the total sum of flux values.

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." *Frontiers in plant science* 5.



MBA-like Family

MBA

- MBA divides the core set in two subcores: a set with high likelihood to be present in the context-specific model (*CH*), if evidence comes from well-curated biochemical knowledge in that particular context, and a set with moderate likelihood (*CM*) if evidence comes from context-specific high-throughput data.
- The algorithm performs the pruning iteratively and randomly by selecting a non-core (NC) reaction to be eliminated, and checking consistency at each step: if *CH* and a user-defined fraction of *CM* remain unblocked, MBA removes the reaction out of the model along with *CM* and NC corresponding blocked reactions. This routine is repeated until no reaction is left in NC.
- The topology of the final model clearly depends on the order in which non-core reactions are eliminated. Therefore, to remove artifacts due to the order, the algorithm is repeated a number of times (1000) to obtain a population of context-specific models. Later, reactions are ranked according to their occurrence in the population and added up to *CH* until a consistent model is obtained.
- MBA proposes an alternative to FVA to check consistency in a more efficient way: First, it solves a LP problem which maximizes the total sum of fluxes. It then removes active reactions (i.e., carrying non-zero flux) and repeats the LP over the remaining set of reactions. If no reaction is found to be active, FVA is applied to each reaction to determine whether it is blocked. The process is repeated until all reactions have been classified either as blocked or unblocked.

FastCore

- While FastCORE aims also at obtaining a minimal consistent model containing all core reactions, typical for this family of methods, it differs principally from MBA and mCADRE in the algorithmic strategy.
- Instead of eliminating one non-core reaction followed by consistency evaluation at each step, FastCORE solves two LPs: The first LP maximizes the cardinality of the core set of reactions, computed as the number of reaction values above a small positive constant. On the other hand, the second LP minimizes the cardinality outside the core set by minimizing the L1-norm of the flux vector, under the constraint that the entire core set must remain active.
- These two LPs are repeatedly applied in alternating fashion until the core set is consistent, whereby activation of all core reactions is ensured while including a minimum set of non-core reactions in the final model. To deal with reversible reactions, FastCORE evaluates both directions by changing the sign of the corresponding column of the stoichiometric matrix.

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods." *Frontiers in plant science* 5.



Summary of Methods for Context-specific Metabolic Model Extraction.

	Parameters	Formulation	Implementation	Omics data	RMF	Flux distribution
GIMME	c, k, V_{max}, V_{min}	LP	COBRA (Matlab)	Transcripts	Required	Yes
GIM ³ E	k, V_{max}, V_{min}	MILP	COBRA (Python)	Transcripts, metabolites	Required	Yes
iMAT	Data discretization*, V_{max}, V_{min}	○ MILP	COBRA (Matlab)	Transcripts, proteins	Unrequired	Yes
INIT/tINIT	Data discretization*, $\epsilon, \delta, V_{max}, V_{min}$	MILP	RAVEN (Matlab)	Transcripts, proteins, metabolites	Optional	Yes
MBA	Data discretization*, $k, \epsilon, V_{max}, V_{min}$	○ LP	-	Curated biochemical knowledge, transcripts, proteins, metabolites, fluxes	Unrequired	No
mCADRE	Data discretization*, $k, \epsilon, V_{max}, V_{min}$	○ LP	Matlab	Transcripts, metabolites	Unrequired	No
FastCORE	$\epsilon, V_{max}, V_{min}$	○ LP	COBRA(Matlab)	-	Unrequired	No

*These methods discretize data following a heuristic approach without any concrete parameter. ○ stands for iteratively repeated.