Find potential cholesterol genes by using CHTC

Anne Huen Wai Wong

Hongwei Pan

Shuyang Chen

Xiyue Wang

Data Description

Two datasets come from archs4
:https://amp.pharm.mssm.edu/archs4/download.html

Human genes dataset：4.9GB

Mouse genes dataset： 4.8GB

Analyze liver samples（main organ to process cholesterols)

Purpose

To find potential cholesterol genes by CHTC

Measuring Similarity

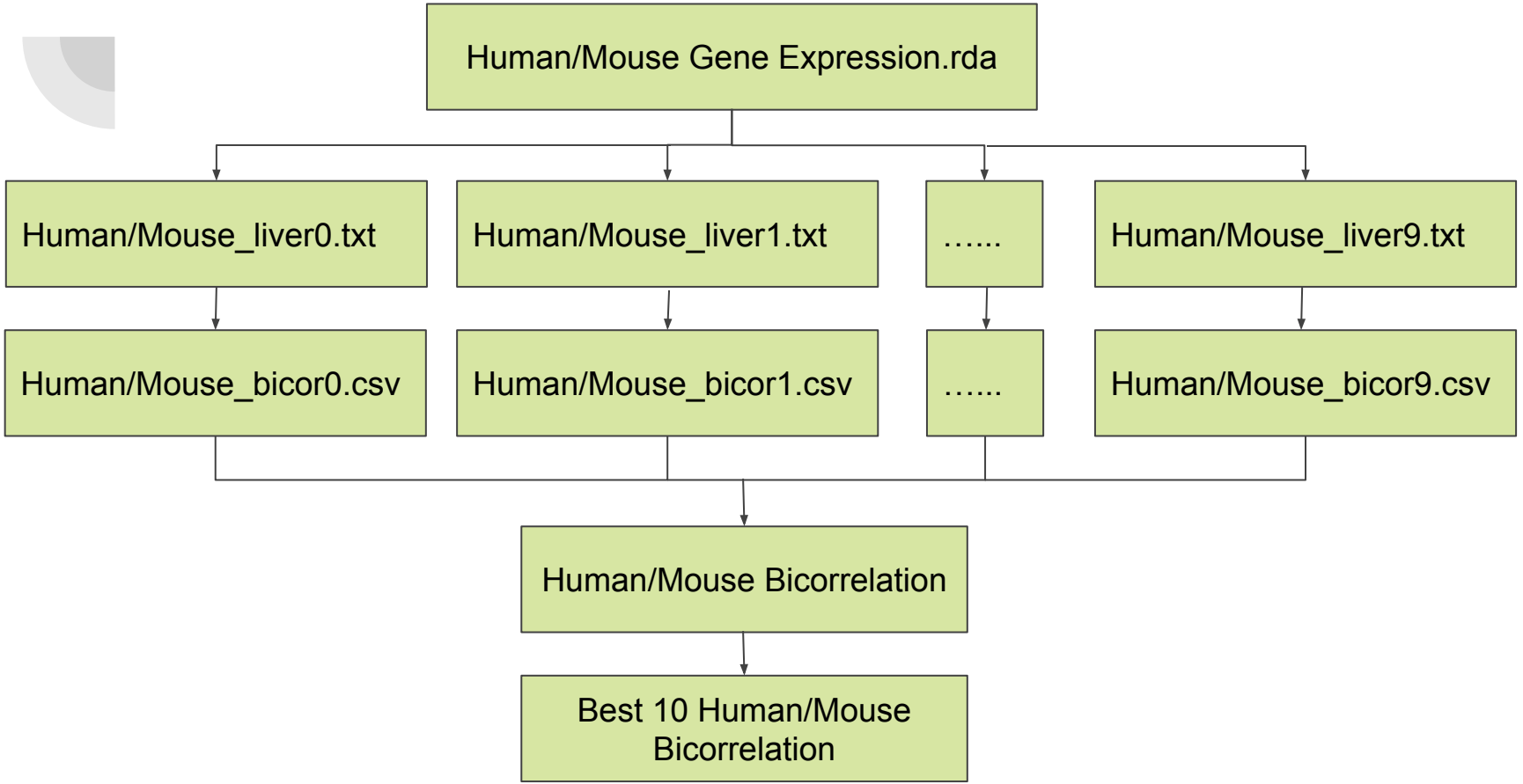
Assumption: With similar distributions may have similar functions.

LDLR: A commonly known cholesterol gene

> "The *LDLR* gene provides instructions for making a protein called a low-density lipoprotein receptor. "
>
> -- *Genetics Home Reference*

Measurement Criteria: Biweight Midcorrelation (median-based)

- Metric based on median rather than mean, more robust, less sensitive to outliers


Human/Mouse Gene Expression.rda
Human/Mouse_liver0.txt
Human/Mouse_liver1.txt
……
Human/Mouse_liver9.txt
Human/Mouse_bicor0.csv
Human/Mouse_bicor1.csv
……
Human/Mouse_bicor9.csv
Human/Mouse Bicorrelation
Best 10 Human/Mouse Bicorrelation

Highly Correlated Genes

Top 10 genes correlated with LDLR in human genes:

	gene	bicorrelation	p-value
1	SND1	0.6985117	8.43e-243
2	VCP	0.6932201	1.15e-237
3	PHB2	0.6928290	2.73e-237
4	FKBP4	0.6907002	2.93e-235
5	PITRM1	0.6884732	3.74e-233
6	AFG3L2	0.6878611	1.41e-232
7	IMMT	0.6861711	5.37e-231
8	FAF2	0.6859340	8.94e-231
9	PSMD3	0.6846405	1.42e-229
10	MAPKAP1	0.6843025	2.93e-229

VCP
-Valosin-containing protein
-Important for the cholesterol-accelerated degradation

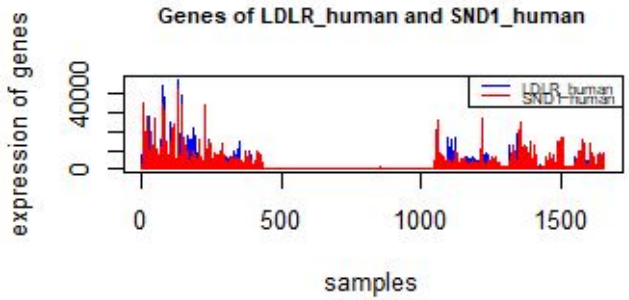
Highly Correlated Genes

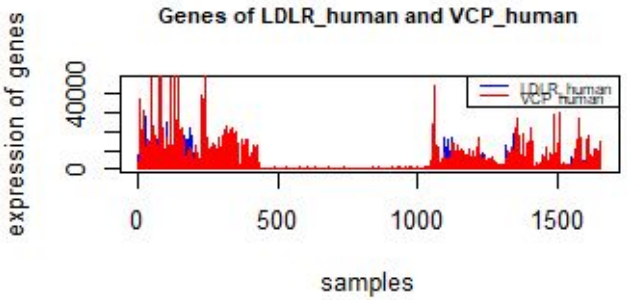
Top 10 genes correlated with LDLR in human genes:

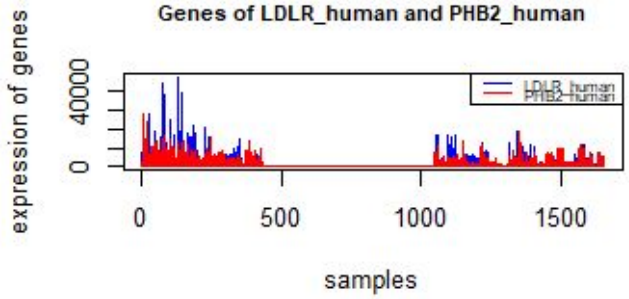
	gene	bicorrelation	p-value
1	SND1	0.6985117	8.43e-243
2	VCP	0.6932201	1.15e-237
3	PHB2	0.6928290	2.73e-237
4	FKBP4	0.6907002	2.93e-235
5	PITRM1	0.6884732	3.74e-233
6	AFG3L2	0.6878611	1.41e-232
7	IMMT	0.6861711	5.37e-231
8	FAF2	0.6859340	8.94e-231
9	PSMD3	0.6846405	1.42e-229
10	MAPKAP1	0.6843025	2.93e-229

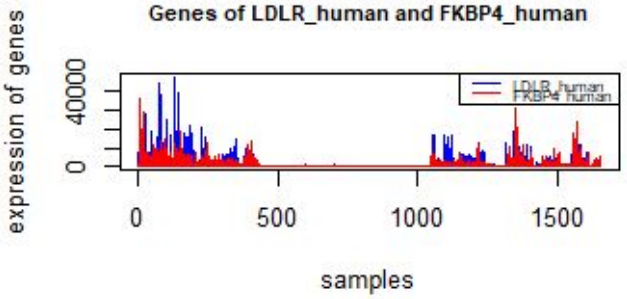
IMMT
- Encoded mitochondrial inner membrane protein.
- Oxidation helps the process of building up cholesterol.

Highly Correlated Genes

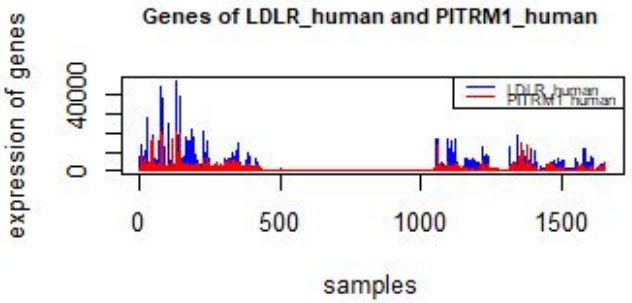


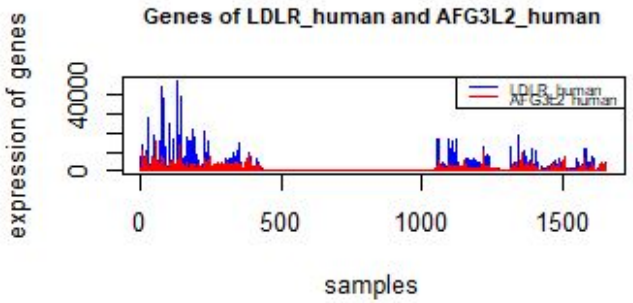


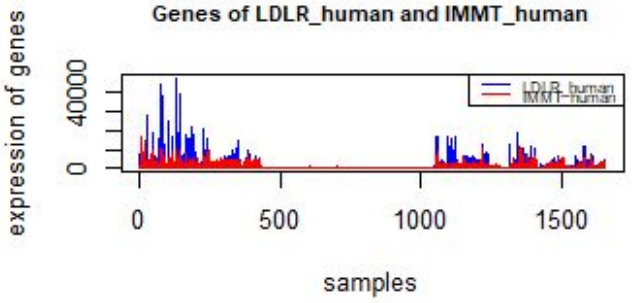


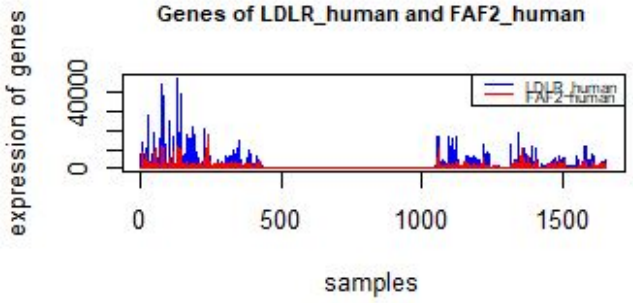


Highly Correlated Genes

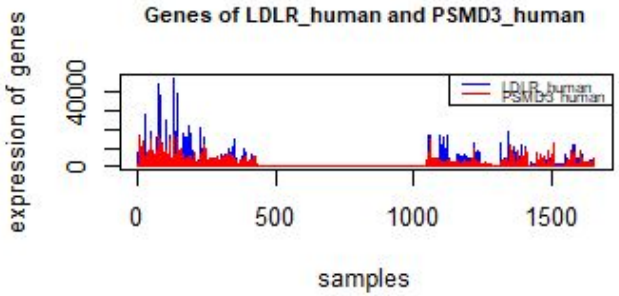


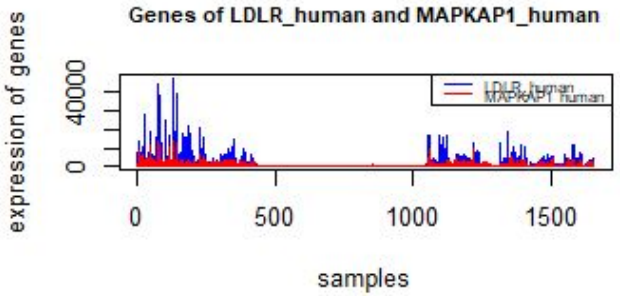






Highly Correlated Genes



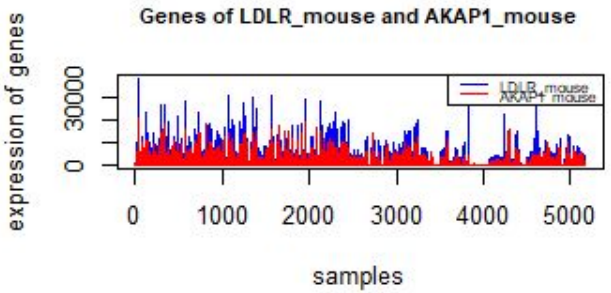


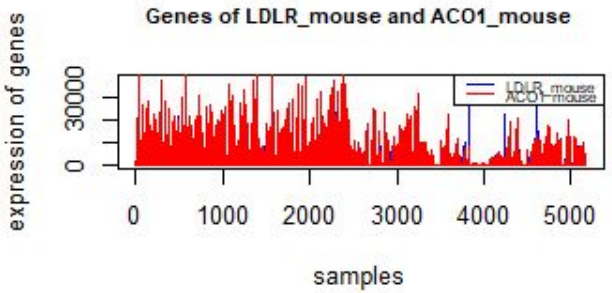
Highly Correlated Genes

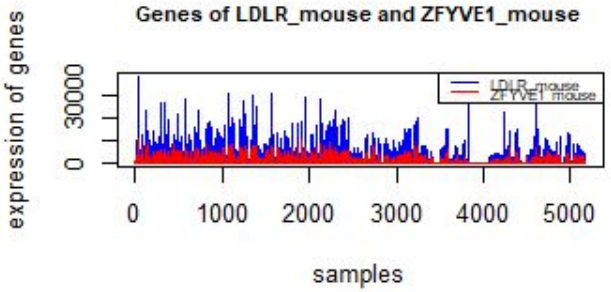
Top 10 genes correlated with LDLR in mouse genes:

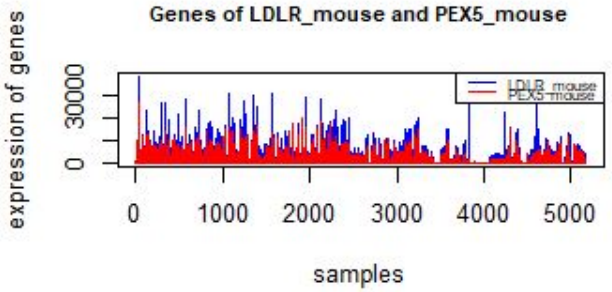
	gene	bicorrelation	p-value
1	AKAP1	0.8481386	<2.225074e-308
2	ACO1	0.8453033	<2.225074e-308
3	ZFYVE1	0.8439302	<2.225074e-308
4	PEX5	0.8427288	<2.225074e-308
5	CDS2	0.8336746	<2.225074e-308
6	HECTD3	0.8310280	<2.225074e-308
7	VAV2	0.8273318	<2.225074e-308
8	D17WSU92E	0.8265241	<2.225074e-308
9	RAB5B	0.8263013	<2.225074e-308
10	TRAK1	0.8260291	<2.225074e-308

Highly Correlated Genes

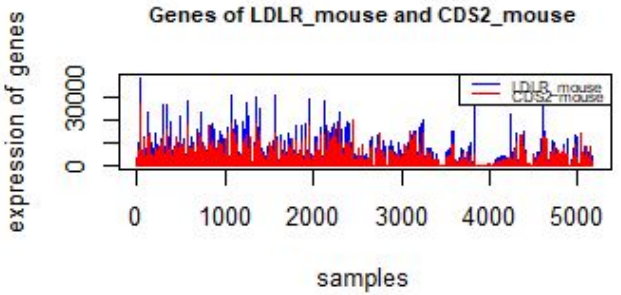


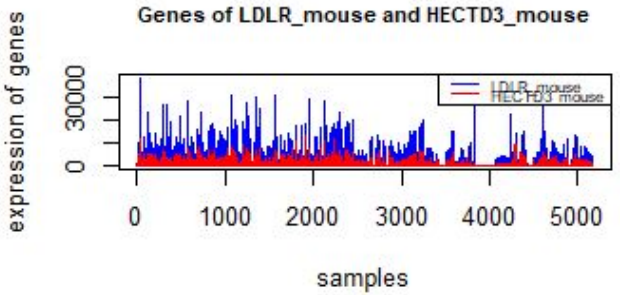


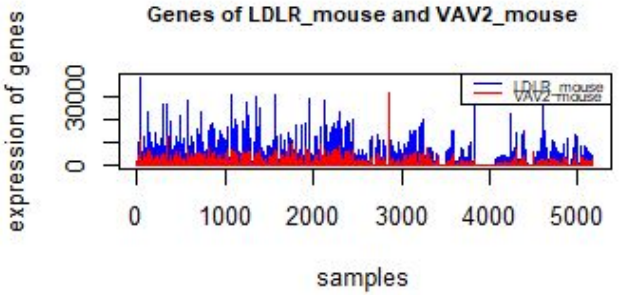


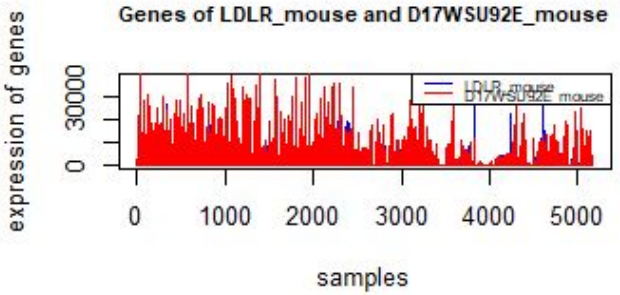


Highly Correlated Genes

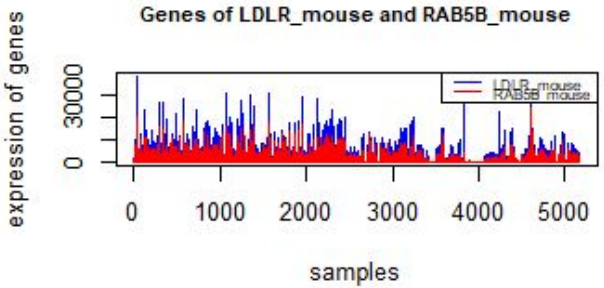


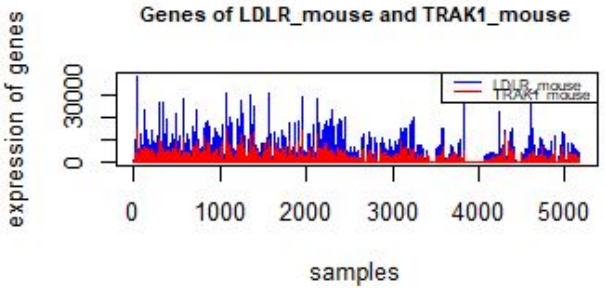






Highly Correlated Genes





Conclusion

- There might be probability to narrow relevant genes down by finding highly correlated genes.
- There is obvious difference between the LDLR gene of human and mouse.

Weakness

- Huge amounts of samples might reduce p-value.
- There might be other influence factors that are able to decide the expression of genes.

Thanks !