

باسمه تعالی



پروژه درس بیوانفورماتیک

فاز دو

نیم سال اول سال تحصیلی ۱۴۰۲-۱۴۰۱

اعضای گروه: علی نظری، پرهام باطنی، سیدمحمدیوسف نجفی

شماره های دانشجویی:

علی نظری: ۹۹۱۰۲۴۰۱

پرهام باطنی: ۹۹۱۰۵۲۹۴

سیدمحمدیوسف نجفی: ۹۹۱۰۲۳۶۱

ایمیل ها:

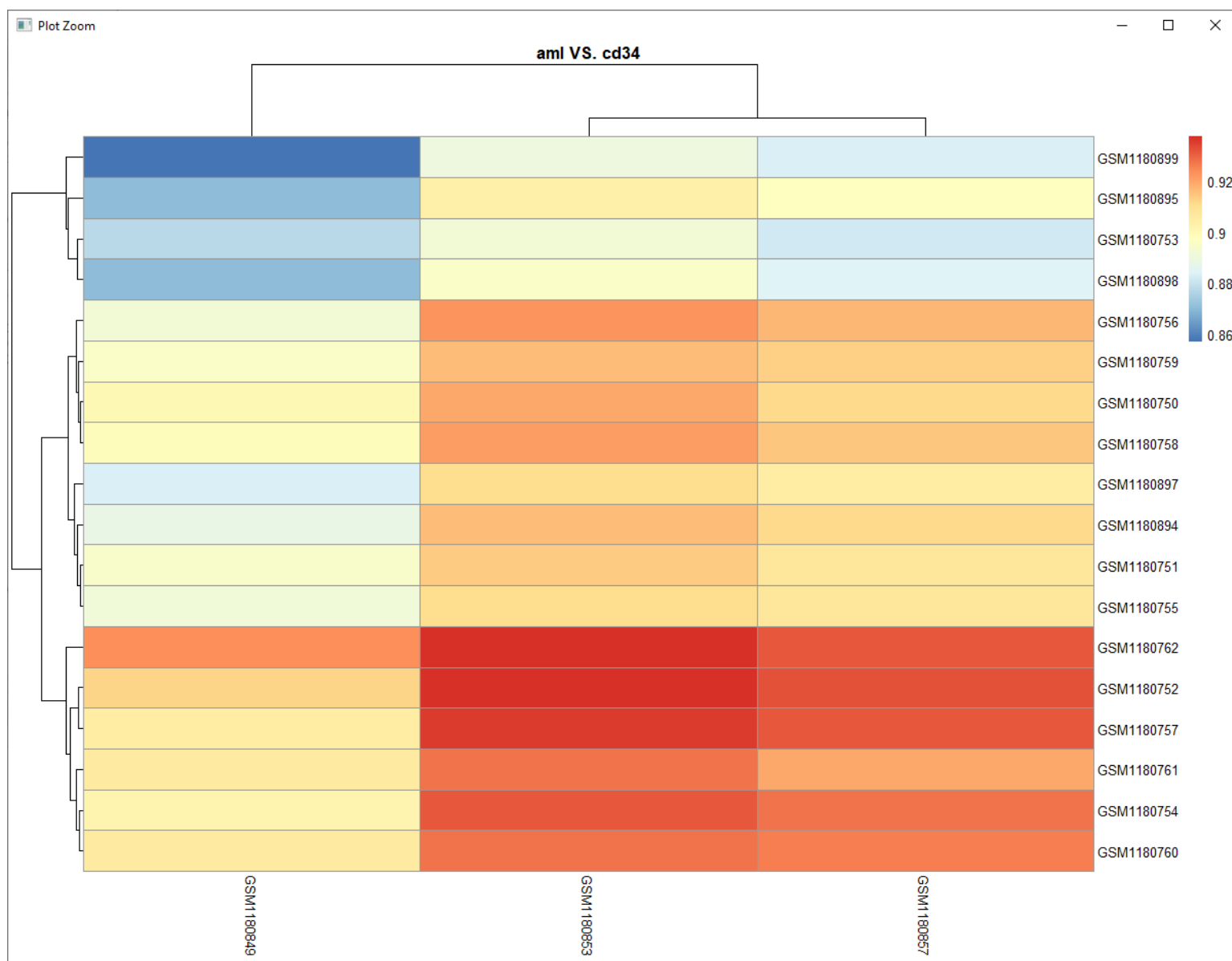
ali.nazari.8102@gmail.com

mp.bateni@gmail.com

najafim2002@gmail.com

۱. دسته ای از داده های سالم که در فاز اول متوجه شدید بیشترین همبستگی را با گروه بیمار دارند را در نظر بگیرید. بیان بعضی از ژن ها به طور معناداری بین این دسته و گروه بیمار متفاوت است، این ژن ها را بیابید. در یک جدول ژن هایی که در بیماران بیشتر بیان شده اند و در جدول دیگری ژن هایی که در بیماران کمتر بیان شده اند را گزارش کنید. (سطح معناداری را ۰,۰۵ در نظر بگیرید) (30)

در فاز نخست، با بررسی heatmap و correlation میان گروه های مختلف سالم و گروه سرطانی، متوجه شدیم که گروه CD34 از دسته سالم، بیشترین شباهت را با گروه سرطانی دارد و heatmap آن به شکل زیر است:



در حال حاضر سطرها اسم نمونه های AML است و ستون ها اسم نمونه های CD34. هر نمونه میزان بیان تعداد زیادی ژن که حدود ۳۲۳۲۱ تا ژن را شامل می شود را مشخص کرده است.

برای بخشی که خواسته شده تا ژن‌هایی که میزان بیان متفاوت به شکل معناداری دارند، مشخص شود، کد مربوط به آن در فایل R که همراه با گزارش ارسال می‌شود موجود است و Id مربوط به ژن‌هایی که این ویژگی را دارند، جدا شده است. در بخش بعد که خواسته شده تا آن ژن‌هایی که میزان بیانشان تفاوت معنادار دارد را به این شکل جدا کنیم که آن‌هایی که در گروه بیمار بیشتر بیان می‌شوند، در یک گروه باشند و آن‌هایی که در گروه سالم بیشتر بیان می‌شوند، در گروهی دیگر. برای این بخش از میانگین بیان ژن در sample های مختلف استفاده می‌کنیم تا بتوانیم تصمیم بگیریم، در کدام گروه بیشتر بیان شده‌اند و کد این بخش هم در فایل R که همراه با گزارش ارسال می‌شود موجود است. همانطور که مشاهده می‌شود، از ۷۳۴۲ ژنی که تفاوت میزان بیان به شکل معنادار داشتند، تعداد ۳۶۸۰ تا به تا به میزان بیشتر در نمونه‌های بیمار بیان می‌شوند و تعداد ۳۶۶۲ تا به میزان بیشتر در نمونه‌های سالم بیان می‌شوند.

۲. برای ژن هایی که دریافت شده اید تفاوت بیان معنادار دارند، pathway ها و gene ontology مرتبط را با استفاده از ابزار های موجود (مثال: سایت Enricher) بیابید و گزارش کنید. (30)

در این مرحله، ما Id مربوط به ژن هایی را داریم که تفاوت بیان معنی دار دارند. این Id برای هر ژن یکتا است. پس الان می خواهیم gene ontology pathway analysis را انجام دهیم. نخست سعی می کنیم pathway را به دست آوریم. پس نیاز است که در یک فایل Id تمام ژن های مشخص شده را داشته باشیم که در فایل R این مورد انجام شده است. در واقع از دستور unique استفاده شده است چون که در prob های آزمایش ممکن است ژن تکراری هم وارد دیتا شده باشد و با اینکار از ورود آنها به تحلیل های بعدی خود، جلوگیری می کنیم. ژن های مدنظر را در یک فایل می ریزیم که کد R اینکار را برای ما انجام می دهد. حال به سراغ سایت Enricher می رویم. نخست باید Id ژن ها را تبدیل به symbol کنیم که برای این منظور از سایت syngoportal استفاده می کنیم:

The screenshot shows the syngoportal website interface. On the left, under the heading "your protein/gene ID list", there is a text area containing a list of IDs: 8180397, 8180398, 8180399, 8180403, 8180406, 8180408, 8180409, and 8180411. On the right, under the heading "input ID type", a dropdown menu is set to "NCBI Entrez gene". Below that, under "input species", a dropdown menu is set to "human". At the bottom right, there is a yellow button labeled "start ID conversion" and a version number "version: 12-12-2020".

سپس این سایت symbol ها را برای ما به دست می آورد و می توانیم به Enricher بازگردیم. البته این سایت برخی Id ها را موفق نشد به symbol مربوطه تبدیل کند و در نتیجه تعداد symbol ها مقداری کمتر است. سپس symbol ها را به شکل زیر در box مربوطه وارد می کنیم و دکمه submit را می زنیم:

The screenshot shows a text box with a red border containing a list of gene symbols: VWF1, ZNF366, ZNF521, ZNF662, TTF2, TTK, WDHD1, WDR76, XRCC2, and ZGRF1. The text box has a scrollbar on the right side.

نتیجه gene list هایی از دیتابیس های مختلف است و چندین حالت برای ما نتیجه می آورد:

Description
Top 500 genes co-express with CD34 identified with ARCHS4 RNA-seq gene-gene co-expression matrix (396 genes)

ChEA 2022

- E2F4 17652178 ChIP-ChIP JURKAT Human
- IRF1 21803131 ChIP-Seq MONOCYTES Hum
- FOXM1 23109430 ChIP-Seq U2OS Human
- FOXM1 25889361 ChIP-Seq OE33 AND U2OS
- IRF1 21803131 ChIP-Seq PrimaryMonocytes

ENCODE and ChEA Consensus TFs from

- E2F4 ENCODE
- IRF8 CHEA
- FOXM1 ENCODE
- SIN3A ENCODE
- IRF1 ENCODE

ARCHS4 TFs Coexp

- KNTC1 human tf ARCHS4 coexpression
- HMGB2 human tf ARCHS4 coexpression
- SOX18 human tf ARCHS4 coexpression
- E2F8 human tf ARCHS4 coexpression
- DEPDC1 human tf ARCHS4 coexpression

TF Perturbations Followed by Expression

- ZNF750 KD HUMAN GSE38039 CREEDSID GE
- ZNF750 KD HUMAN GSE38039 CREEDSID GE
- HMG2 KO MOUSE GSE55340 CREEDSID GEI
- SRF KO MOUSE GSE34545 CREEDSID GENE 2
- TVIST1 OE MOUSE GSE50002 CREEDSID GEI

TRRUST Transcription Factors 2019

- NFKB1 human
- RELA human
- IRF1 human
- EPAS1 mouse
- ERG human

FANTOM6 lncRNA KD DEGs

- RP11-422J8.1-ASO_G0233621_05-DEGs Down
- RP13-463N16.6-ASO_G0242147_07-DEGs Do
- RP11-139H15.1-ASO_G0225973_10-DEGs Do
- EMX2OS-ASO_G0229847_AD_04-DEGs Down
- RP11-38L15.3-ASO_G0231187_02-DEGs Dow

IncHUB lncRNA Co-Expression

- DIAPH3-AS1
- SGO1-AS1
- HID1-AS1
- TMPO-AS1
- LINC01985

Enrichr Submissions TF-Gene Cooccurrence

- WDHD1
- KNTC1
- DEPDC1
- DEPDC1B
- RAD51

TRANSFAC and JASPAR PWMs

- PRDM1 (mouse)
- POU1F1 (human)
- ETV4 (human)
- TBP (human)
- HINFP (human)

Epigenomics Roadmap HM ChIP-seq

- H2BK120ac IMR90
- H2BK20ac IMR90
- H2BK15ac IMR90
- H2BK12ac IMR90
- H3K4me1 Adipose Derived Mesenchymal St

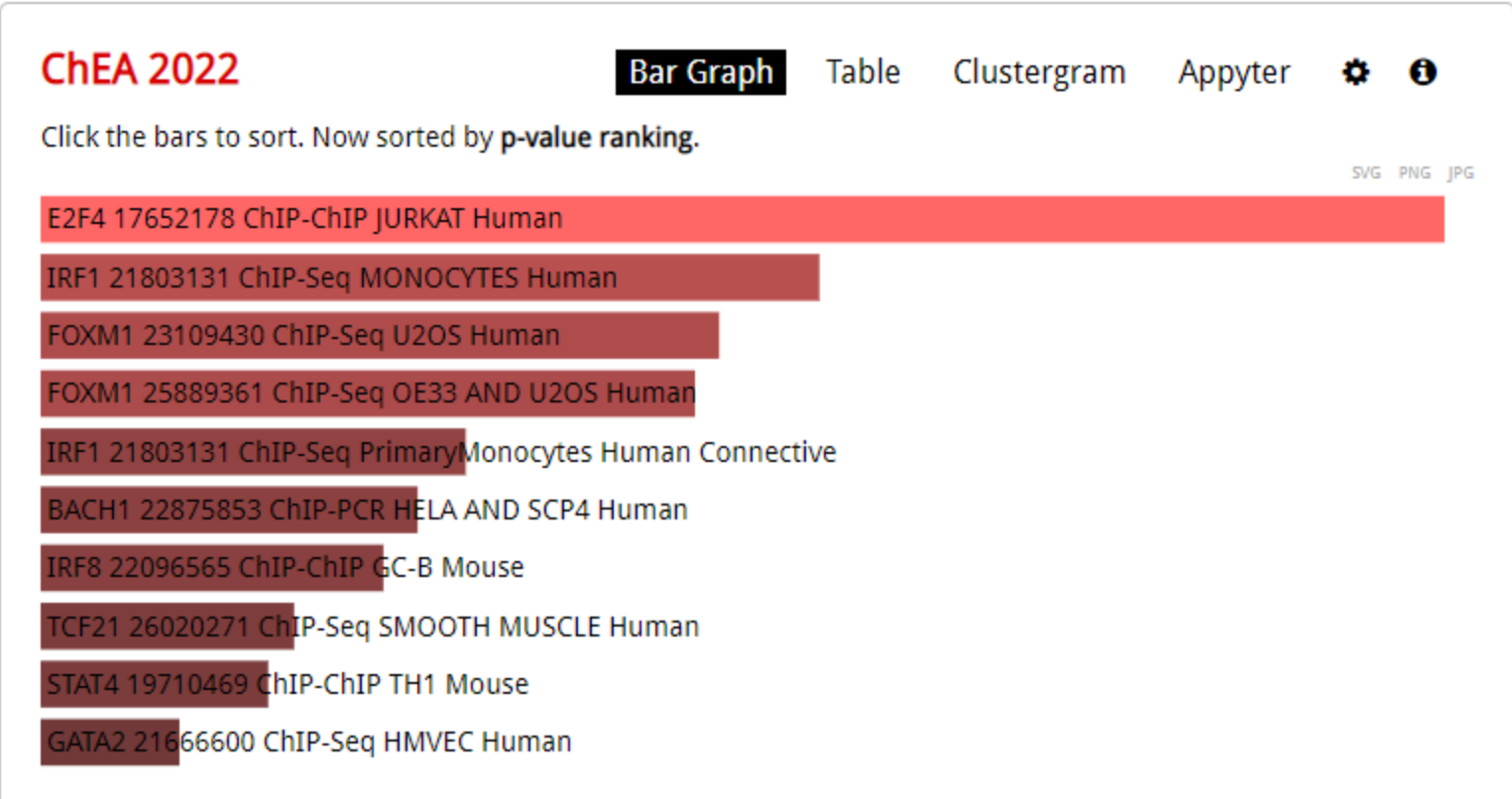
TargetScan microRNA 2017

- hsa-miR-196a
- hsa-miR-196b
- hsa-miR-3117-3p
- hsa-miR-3146
- hsa-miR-941

miRTarBase 2017

- hsa-miR-215-5p
- hsa-miR-192-5p
- hsa-miR-193b-3p
- mmu-miR-29b-1-5p
- hsa-miR-146a-5p

در آن tab بالا هم که انواع حالت‌هایی که مدنظر هست را می‌توان استفاده کرد. روی یکی از موارد بالا کلیک می‌کنیم در همان بخش Transcription که خود بخش‌های مختلفی دارد.



همچنین برخی اطلاعات آماری مرتبط را هم برای ما می‌آورد.

Hover each row to see the overlapping genes.

10 entries per page

Search:

| Index | Name | P-value | Adjusted p-value | Odds Ratio | Combined score |
|-------|--|-------------|------------------|------------|----------------|
| 1 | E2F4 17652178 ChIP-ChIP JURKAT Human | 2.755e-21 | 2.049e-18 | 5.56 | 263.05 |
| 2 | IRF1 21803131 ChIP-Seq PrimaryMonocytes Human Connective | 4.382e-10 | 6.520e-8 | 11.02 | 237.39 |
| 3 | IRF1 21803131 ChIP-Seq MONOCYTES Human | 3.933e-14 | 1.463e-11 | 6.45 | 199.09 |
| 4 | FOXM1 23109430 ChIP-Seq U2OS Human | 5.507e-13 | 1.366e-10 | 6.82 | 192.44 |
| 5 | FOXM1 25889361 ChIP-Seq OE33 AND U2OS Human | 1.039e-12 | 1.932e-10 | 3.63 | 100.03 |
| 6 | E2F7 22180533 ChIP-Seq HELA Human | 0.00009368 | 0.001834 | 9.41 | 87.28 |
| 7 | IRF8 22096565 ChIP-ChIP GC-B Mouse | 3.828e-9 | 4.069e-7 | 3.77 | 73.05 |
| 8 | IRF8 22096565 ChIP-ChIP GC-B Human | 0.000001768 | 0.0001012 | 5.25 | 69.49 |
| 9 | BACH1 22875853 ChIP-PCR HELA AND SCP4 Human | 1.552e-9 | 1.924e-7 | 2.73 | 55.31 |
| 10 | BP1 19119308 ChIP-ChIP Hs578T Human | 0.005039 | 0.04687 | 9.97 | 52.74 |

Showing 1 to 10 of 744 entries | [Export entries to table](#)

[Previous](#) [Next](#)

Terms marked with an * have an overlap of less than 5

در این جدول مورد جالب این است که موس را روی هر کدام ببریم، ژنهای مرتبط با آن سطر را برای ما می آورد.

Hover each row to see the overlapping genes.

10 entries per page

| Index | Name | P-value | Adjusted p-value | Odds Ratio | Combined score |
|-------|--|-----------|------------------|------------|----------------|
| 1 | E2F4 17652178 ChIP-ChIP JURKAT Human | 2.755e-21 | 2.049e-18 | 5.56 | 263.05 |
| 2 | IRF1 21803131 ChIP-Seq PrimaryMonocytes Human Connective | 4.382e-10 | 6.520e-8 | 11.02 | 237.39 |

TOP2A, SUV39H2, NUP107, MCM8, BUB1B, MCM10, BRCA1, KIF11, TTF2, MKI67, KIF15, MTBP, EXO1, CHEK1, NUSAP1, KNTC1, FBXO5, GTSE1, B2M, HELLS, ESCO2, MASTL, WDR76, SENP1, ASPM, RBL1, FANCD2, PRR11, DTL, BIRC3, BLM, CDCA2, TTK, HMMR, RAD51AP1, BRIP1, CLSPN, BUB1, BARD1, PLK4, POLO, STIL, FANCM, NUP155, CEP152, XRCC2, ATAD2, PLXDC1, SHCBP1, CENPF, KIF18A, RFPWD3, POLE2, SAS56, RAD18

در بخش name نامهای transcription factor ها را داریم که مشخص می کند از لیست ما کدام ژن ها توسط این transcription factor ها تاثیر می پذیرد. بحث p.value را هم مشخص کرده است تا بتوانیم میزان معنی دار بودن اعداد و ارقام را بسنجیم. پس می توان transcription factor های مشخص شده را سرچ کرد و در مقاله ها جست و جو کرد که آیا واقعا ارتباطی بین آن ها وجود دارد یا خیر. حال به سراغ بخش pathway می رویم. وقتی که روی بخش pathway کلیک کنیم، باز هم تعدادی دیتاست می آورد که می توان جداگانه هم آن ها را سرچ کرد و اطلاعات آن ها را بررسی کرد.

Description No description available (396 genes)



Reactome 2022

Bar Graph

Table

Clustergram

Appyter



Hover each row to see the overlapping genes.

10 entries per page

Search:

| Index | Name | P-value | Adjusted p-value | Odds Ratio | Combined score |
|-------|--|-------------|------------------|------------|----------------|
| 1 | Classical Antibody-Mediated Complement Activation R-HSA-173623 | 1.752e-8 | 6.856e-7 | 250.68 | 4477.11 |
| 2 | Defective B4GALT1 Causes B4GALT1-CDG (CDG-2d) R-HSA-3656244 | 1.582e-7 | 0.000004283 | 83.55 | 1308.34 |
| 3 | Defective CHST6 Causes MCDC1 R-HSA-3656225 | 1.582e-7 | 0.000004283 | 83.55 | 1308.34 |
| 4 | Defective ST3GAL3 Causes MCT12 And EIEE15 R-HSA-3656243 | 1.582e-7 | 0.000004283 | 83.55 | 1308.34 |
| 5 | Keratan Sulfate Degradation R-HSA-2022857 | 0.000002096 | 0.00004254 | 35.80 | 468.11 |
| 6 | Impaired BRCA2 Binding To PALB2 R-HSA-9709603 | 1.227e-8 | 5.534e-7 | 25.24 | 459.82 |
| 7 | Interleukin-9 Signaling R-HSA-8985947 | 0.00001763 | 0.0002807 | 40.00 | 437.81 |
| 8 | Defective HDR Thru Homologous Recombination (HRR) Due To BRCA1 Loss-Of-Function R-HSA-9701192 | 1.773e-8 | 6.856e-7 | 23.76 | 424.00 |
| 9 | Resolution Of D-loop Structures Thru Synthesis-Dependent Strand Annealing (SDSA) R-HSA-5693554 | 3.515e-8 | 0.000001241 | 21.25 | 364.78 |
| 10 | Interleukin-21 Signaling R-HSA-9020958 | 0.00002892 | 0.0004431 | 33.33 | 348.33 |

Showing 1 to 10 of 812 entries | [Export entries to table](#)

[Previous](#) [Next](#)

Terms marked with an * have an overlap of less than 5

در این بخش هم می‌بینیم که مشخص شده که هر ژن در کدام روند بدن نقش دارد و چه کاری انجام می‌دهد. حال اگر biological process ها را بخواهیم مشاهده کنیم، باید به بخش ontologies برویم و در کل در همین بخش‌ها می‌توانیم بیماری‌ها و داروهای مرتبط را هم ببینیم و مانند قبل با نگه داشتن موس روی هر سطر می‌توان مشاهده کرد که کدام ژن‌ها از لیست ما به آن سطر مرتبط است:

Description No description available (396 genes)**GO Biological Process 2021**

Bar Graph

Table

Clustergram

Appyter



Hover each row to see the overlapping genes.

10 entries per page

Search:

| Index | Name | P-value | Adjusted p-value | Odds Ratio | Combined score |
|-------|---|-------------|------------------|------------|----------------|
| 1 | complement activation, classical pathway (GO:0006958) | 3.502e-7 | 0.00004211 | 62.66 | 931.42 |
| 2 | retina vasculature development in camera-type eye (GO:0061298) | 0.00007479 | 0.002021 | 74.82 | 710.82 |
| 3 | positive regulation of glycoprotein metabolic process (GO:1903020) | 0.00007479 | 0.002021 | 74.82 | 710.82 |
| 4 | positive regulation of mitochondrial depolarization (GO:0051901) | 0.00007479 | 0.002021 | 74.82 | 710.82 |
| 5 | vasculature development (GO:0001944) | 1.855e-8 | 0.000003717 | 35.26 | 627.72 |
| 6 | meiotic sister chromatid cohesion (GO:0051177) | 0.000009950 | 0.0004659 | 50.00 | 575.90 |
| 7 | humoral immune response mediated by circulating immunoglobulin (GO:0002455) | 0.000001243 | 0.00009641 | 41.77 | 567.98 |
| 8 | keratan sulfate catabolic process (GO:0042340) | 0.000002096 | 0.0001400 | 35.80 | 468.11 |
| 9 | cell junction disassembly (GO:0150146) | 0.0001474 | 0.003222 | 49.88 | 440.02 |
| 10 | centromeric sister chromatid cohesion (GO:0070601) | 0.0001474 | 0.003222 | 49.88 | 440.02 |

Showing 1 to 10 of 2,405 entries | [Export entries to table](#)[Previous](#) [Next](#)

Terms marked with an * have an overlap of less than 5

به عنوان مثال در بخش ontology می‌توان مشاهده کرد که هر ژن روی کدام فعالیت زیستی تاثیر می‌گذارد و بر این اساس تحلیل‌هایی را می‌توان انجام داد. به عنوان مثال می‌توان بررسی کرد که خروجی تحلیل داده‌های این چینی با تحلیل‌های دستگاه microarray یکسان است یا نه.

حال سایت بسیار خوب دیگری که در تحلیل pathway می‌توان از آن استفاده کرد، reactome است. برای استفاده از این سایت، باید به جدول بخش pathway قبل برگشت و مشاهده کرد که در کدام pathway ها فعال بوده‌اند ژن‌ها و بر اساس آن جست‌وجو را انجام داد در reactome. به عنوان مثال دیدیم که در بخش Defective B4GALT1 ژن‌های زیادی بودند پس همین مورد را در reactome سرچ می‌کنیم. خروجی این بخش به شکل زیر است:

reactome 3.7

Pathways for: Homo sapiens

Event Hierarchy:

- Autophagy
- Cell Cycle
- Cell-Cell communication
- Cellular responses to stimuli
- Chromatin organization
- Circadian Clock
- Developmental Biology
- Digestion and absorption
- Disease
 - Diseases of signal transduction
 - Diseases of mitotic cell cycle
 - Diseases of cellular response to
 - Diseases of programmed cell de
 - Diseases of DNA repair
 - Disorders of transmembrane tra
 - Diseases of metabolism
 - Diseases of carbohydrate me
 - Phenylketonuria
 - Defects in vitamin and cofact
 - Metabolic disorders of biologi
 - Abnormal conversion of 2-ox
 - Diseases associated with sur
 - Diseases of glycosylation
 - Diseases associated wi
 - Defective SL26A2 cau
 - Defective PAPSS2 cau
 - Defective B4GALT7 cau
 - Defective B3GAT3 cau
 - Defective CHSY1 causi
 - Defective CHST3 cause
 - Defective CHST14 cau
 - Defective CHST6 cause
 - Defective EXT1 causes
 - Defective EXT2 causes

Search for a term, e.g. pten...

No interactors found in IntAct

Description Molecules Structures Expression Analysis Downloads

Defective B4GALT1 causes B4GALT1-CDG (CDG-2d) Id: R-HSA-3656244.2 Species: Homo sapiens

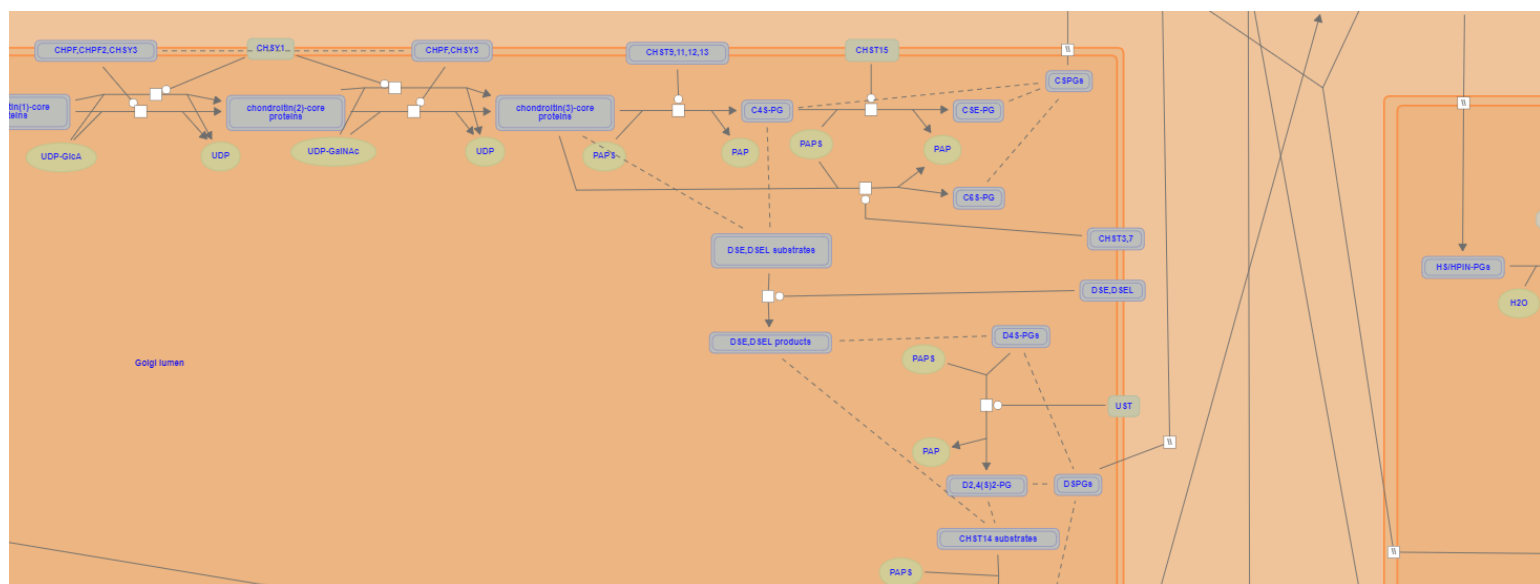
Summation

Congenital disorders of glycosylation (CDG, previously called carbohydrate-deficient glycoprotein syndromes, CDGSs), are a group of hereditary multisystem disorders. They are characterized biochemically by hypoglycosylation of glycoproteins, diagnosed by isoelectric focusing (IEF) of serum transferrin. There are two types of CDG, types I and II. Type I CDG has defects in the assembly of lipid-linked oligosaccharides or their transfer onto nascent glycoproteins, whereas type II CDG comprises defects of trimming, elongation, and processing of protein-bound glycans. Clinical symptoms are dominated by severe psychomotor and mental retardation, as well as blood coagulation abnormalities (Jaeken 2013). B4GALT1-CDG (CDG type II d) is a multisystem disease, characterized by dysmorphic features, hydrocephalus, hypotonia and blood clotting abnormalities (Hansske et al. 2002).

Disease

congenital disorder of glycosylation type II

همانطور که انتظار داشتیم، کلیه فعالیت‌های شناخته شده که در آن بخش انجام می‌شوند را در این حالت می‌توانیم ببینیم و به شکل بسیار دقیقی به بررسی مراحل اتفاق افتاده بپردازیم. به شکل زوم هم بخشی از آن مانند عکس زیر است:



سایت بسیار خوب دیگر KEGG pathway database است که بسیار شبیه به reactome است و از قرار دادن نتایج آن در گزارش کار، صرف نظر می‌کنیم. با بررسی بیشتر مشخص شد که بسیاری از ژن‌هایی که به دست آورده بودیم، در این بخش هم فعال هستند و تا حدی نتایج به دست آمده توسط ما، معتبر است.

۳. با جستجو در مقالات مرتبط با AML به سوالات زیر پاسخ دهید(40)

الف) آیا نتایجی که به دست آورده اید (ژن ها، pathway ها، ontology و ..) در منابع دیگر هم به عنوان عامل موثر یا مرتبط با بیماری ذکر شده اند؟ حداقل به سه مورد با توضیح مختصری از چگونگی ارتباط یافته های شما با بیماری اشاره کنید.

ب) در مقالاتی که برای این بیماری دارو پیشنهاد کرده اند جستجو کنید. حداقل دو مورد از داروهای پیشنهادی و اینکه چه ارتباطی با یافته های شما دارد را گزارش کنید. (مثال: دارویی که باعث میشود بیان ژنی در بیمار کم یا زیاد شود)

الف) به عنوان مثال از یکی از موارد transcription factor شروع می کنیم. همانطور که در بخش قبل مشاهده کردیم، تعداد زیادی transcription factor وجود دارد که با ژن های وارد شده توسط ما ارتباط دارند. در [این مقاله](#) که در سال ۲۰۲۰ نوشته شده است، ارتباط transcription factor با نام E2F4 با AML را مشخص کرده است و در آن توضیح داده است که با bind شدن به EZH2 جلوی MAPK Signaling را می گیرد. در abstract آن هم اشاره دقیق و واضحی به AML داشته است. در چکیده آن ذکر شده است که:

'در اینجا، هدف ما بررسی اختلال در تنظیم یک فاکتور رونویسی خاص، E2F4، و نقش آن در پیشرفت AML است.'

که به ارتباط این ۲ مورد به هم اشاره دارد. در این مطالعه، داده های بالینی انسانی از تجزیه و تحلیل تعاملی پروفایل بیان ژن (GEPIA) نشان داد که افزایش بیان E2F4 با پیش آگهی ضعیف در بیماران AML مرتبط است. علاوه بر این، نتایج تجربی نشان داد که E2F4 به طور نابجا در بیماران AML و رده های سلولی انسانی بیان می شود.

پس با اطمینان بالایی می توانیم بگوییم که این مورد درست بوده و ارتباط زیادی با AML داشته است. حال به سراغ نمونه ای در pathway می رویم. برای این بخش در enrichr دیدیم که Classical Antibody-Mediated Complement Activation به عنوان کاندیدی در ارتباط با ژن هایی که وارد کردیم، در بخش pathway معرفی شد. در [این مقاله](#)، ارتباط آن با AML بررسی شده است. این مقاله هم در سال ۲۰۲۰ منتشر شده است و در بخشی از چکیده آن آورده شده است که:

'مکمل، به عنوان یک سیستم نظارتی مرکزی سیستم ایمنی، می تواند در عرض چند ثانیه پس از تحریک فعال شود، در نتیجه عملکردهای موثر ایمنی متعدد را نشان می دهد. با این حال، در سناریوهای پاتولوژیک (مانند پیشرفت تومور)، مکمل فعال می تواند هم اثرات محافظتی برای کنترل رشد تومور نشان دهد و هم به طور غیر فعال رشد تومور را تقویت کند.'

پس همانطور که در این بخش هم دیده می شود، می تواند با زیاد فعالیت کردن و یا کم فعالیت کردن، میزان تومور و AML را تغییر دهد و در نتیجه بین این موارد هم رابطه ای وجود دارد. یعنی با ژن هایی که استخراج کردیم و به دست آوردن این pathway توانستیم ارتباط را مشخص کنیم و در یک فرآیند مرتبط با سیستم ایمنی بدن، نشانه های آن را ببینیم.

در نهایت هم از بخش ontologies بخشی را به دست می آوریم که ارتباطی با AML داشته باشد. در این بخش، موردی را مشاهده می کنیم به نام positive regulation of glycoprotein metabolic process که به عنوان ontology که با ژن های معرف شده توسط ما همخوان است، وجود دارد. با بررسی ها و جست و جوی های انجام شده، به [این مقاله](#) رسیدیم که دقیقاً

ارتباط آن با AML را مورد بررسی قرار داده است. این مقاله که در سال ۲۰۱۹ منتشر شده است، در چکیده خود بیان کرده است که:

لوسمی حاد میلوئیدی (AML) یکی از شایع ترین و تهدید کننده ترین سرطان خون است. متابولیسم بسیار متنوع و انعطاف پذیر به تهاجمی بودن بیماری کمک می کند که هنوز درمان آن دشوار است. با استفاده از منابع مختلف مواد مغذی برای تامین انرژی و زیست توده، سلول های AML انعطاف متابولیکی پیدا می کنند و به سرعت از سلول های خونساز طبیعی پیشی می گیرند. هدف این بررسی رمزگشایی استراتژی های متابولیک متنوع و تغییرات انکوژنیک و محیطی است که رشد مداوم را حفظ می کنند، هموستاز ردوکس را واسطه می کنند و مقاومت دارویی را در AML القا می کنند.^۱

پس در این مورد هم مشاهده می کنیم که به درستی ارتباطی بین AML و این ontology بیان شده وجود دارد و ما با تحلیل داده ها توانستیم به ژن هایی برسیم که این ژن ها ما را به این ontology رساندند و می توان با اطمینان بالایی گفت که تاثیر زیادی روی AML دارد و توضیحات کاملی در مقاله گفته شده وجود دارد.

پس سه مورد را طبق صورت سوال مثال زدیم که در مقاله های معتبر هم به ارتباط آن ها با AML اشاره شده است و نشان از درستی حدودی تحلیل های ما دارد.

ب) یکی از مقاله هایی که در این مورد صحبت کرده است، [این مقاله](#) است که در آن تعداد زیادی موارد شیمیایی و ژن های درگیر شده را نام برده که در بخشی از آن داریم:

'SGN-CD33A یک ADC جدید با پیوندهای دی پپتیدی بسیار پایدار است که بارگذاری دارویی یکنواخت یک دایمر پیرولوبونزودیازپین را امکان پذیر می کند که DNA را به مرگ سلولی پیوند می دهد. یک تجزیه و تحلیل موقت از مرحله 1 مطالعه افزایش دوز SGN-CD33A در بیماران مبتلا به AML عود کننده CD33+ و آن دسته از بیماران که درمان فشرده را رد کردند، میزان 29% CRC بود. هفتاد و هفت درصد از بیماران که دوزهای 40 میکروگرم بر کیلوگرم یا بالاتر دریافت کردند، حداقل 50 درصد کاهش در بلاست های مغز استخوان داشتند. مطالعه ای در حال انجام بر روی SGN-CD33A به تنهایی و در ترکیب با عوامل هیپومتیلاتور وجود دارد.'

همانطور که می بینیم در این مقاله به موارد تاثیرگذار در روی AML اشاره داشته و دارویی را معرفی کرده که این موضوع را انجام می دهد. در این مورد به خانواده ژن های CD اشاره داشته که نمونه هایی از آن ها در تحلیل ما وجود داشتند مانند CDH5 و CD93 و CD300LG و C2CD4B و برخی موارد دیگر. پس با توجه به متن بالا، این دارو باعث کاهش میزان بیان ژن می شود و از این طریق جلوی پیشرفت را می گیرد.

مقاله دیگری که در این مورد نوشته است، [این مقاله](#) است که در آن در مورد داروهای جدید درمان AML صحبت کرده است. در بخشی از آن در مورد نحوه تغییر ژن ها صحبت شده است:

'جهش های مکرر در ژن های IDH1 و IDH2 به ترتیب در 7-14% و 8-19% از بیماران AML یافت می شود. جهش در این ژن ها منجر به از دست دادن فعالیت کاتالیزوری طبیعی ایزوسیترات دهیدروژناز (IDH) و ایجاد فعالیت آنزیم نئومورفیک می شود که باعث کاهش α -ketoglutarate به oncometabolite R-2-hydroxyglutarat می شود. این در نهایت باعث ایجاد تغییرات اپی ژنتیکی و ناتوانی سلول های خونساز در تمایز می شود.'

همانطور که مشاهده می‌شود به یکی از ژن‌های بسیار مهم در این بیماری اشاره شده است که با تحلیل‌های ما همخوانی دارد و نوع تغییر آن و در نتیجه بروز بیماری را شرح داده است. در همین مقاله باز هم به خانواده CD ها اشاره شده است و از اهمیت آن‌ها گفته است:

*اکثر سلول‌های AML متفاوت از سطحی CD33 را بیان می‌کنند (تخمین زده شده بیش از 80٪ از بیماران مبتلا به بیماری).
Gemtuzumab ozogamicin یک آنتی بادی مونوکلونال ضد CD33 انسانی است که به عامل سیتوتوکسیک N-acetyl
calicheamicin مرتبط است. تحقیقات اولیه نشان می‌دهد که درونی‌سازی برای اتصال به یکدیگر است و به دنبال آن آزادسازی
درون سلولی می‌تواند به ارائه عامل درمانی در سلول‌های لوسمی بیان‌کننده CD33 منجر شود.*
پس باز هم ارتباط آن با CD مشخص می‌شود و بخش‌هایی که داروی معرفی شده روی آن‌ها تاثیر می‌گذارد تا اینکه اثر بیماری را کمتر کند، مشخص است.