فاز دوم

پروژه مقدمه‌ای بر بیوانفورماتیک

دکتر علی شریفی‌زارچی و دکتر سمیه کوهی

دانشکده مهندسی کامپیوتر

دانشگاه صنعتی شریف

نیم‌سال اول ۰۱-۰۲

امیرحسین باقری - ۹۸۱۰۵۶۲۱
مهدی مستانی - ۹۷۱۰۰۵۱۳
محمدرضا مفیضی - ۹۸۱۰۶۰۵۹

فهرست مطالب

۱ بیان ژن‌ها **۲**
۱.۱ انتخاب تمام نمونه‌های سالم ۲
۲.۱ انتخاب نمونه‌های سالم با بیشترین هم‌بستگی ۲

۲ بررسی pathway و gene ontology **۳**
۱.۲ ژن‌هایی که کمتر بیان شده‌اند ۳
۱.۱.۲ انتخاب تمام نمونه‌های سالم ۳
۲.۱.۲ انتخاب نمونه‌های سالم با بیشترین هم‌بستگی ۴
۲.۲ ژن‌هایی که بیشتر بیان شده‌اند ۸
۱.۲.۲ انتخاب تمام نمونه‌های سالم ۸
۲.۲.۲ انتخاب نمونه‌های سالم با بیشترین هم‌بستگی ۹

۳ جست‌وجو در مقالات **۹**

۱ بیان ژن‌ها

۱.۱ انتخاب تمام نمونه‌های سالم

در جدول ۱ ژن‌هایی که بیان آن‌ها به طرز معناداری بین گروه سالم و گروه بیماران متفاوت است را نمایش می‌دهیم [۱].

Gene Symbol	Gene ID	Adj. P-Val	logFC
MPO	4353	3.617e-19	5.563
FLT3	2322	4.835e-19	5.250
KIAA0101	9768	6.308e-19	4.559
BUB1B	701	1.664e-18	2.756
SUCNR1	56670	1.938e-18	2.996

جدول ۱: ۵ ژنی که بیان آن‌ها به طرز معناداری بین گروه سالم و گروه بیماران متفاوت است.

حالا ژن‌هایی که در بیماران از بقیه بیشتر بیان شده‌اند و همچنین ژن‌هایی که کمتر از بقیه بیان شده‌اند را در جدول ۲ نمایش می‌دهیم [۲].

Least Expressed	Most Expressed
STK38	MPO
CBX7	FLT3
PLCL2	KIAA0101
PECR	BUB1B
HLA-F	SUCNR1

جدول ۲: ۵ ژنی که در بیماران بیشتر/کمتر از بقیه بیان شده‌اند.

۲.۱ انتخاب نمونه‌های سالم با بیشترین هم‌بستگی

در جدول ۳ ژن‌هایی که بیان آن‌ها به طرز معناداری بین گروه سالم و گروه بیماران متفاوت است را نمایش می‌دهیم.

Gene Symbol	Gene ID	Adj. P-Val	logFC
MPO	4353	7.714e-16	5.73
PECR	55825	7.714e-16	-2.314
KIT	3815	7.126e-15	5.262
KIAA0101	9768	7.969e-15	4.308
FLT3	2322	7.969e-15	5.155

جدول ۳: ۵ ژنی که بیان آن‌ها به طرز معناداری بین گروه سالم و گروه بیماران متفاوت است.

حالا ژن‌هایی که در بیماران از بقیه بیشتر بیان شده‌اند و همچنین ژن‌هایی که کمتر از بقیه بیان شده‌اند را در جدول ۴ نمایش می‌دهیم.

[۱] جدول کامل به همراه این گزارش ضمیمه شده است.
[۲] جدول کامل این بخش هم به همراه گزارش ضمیمه شده است.

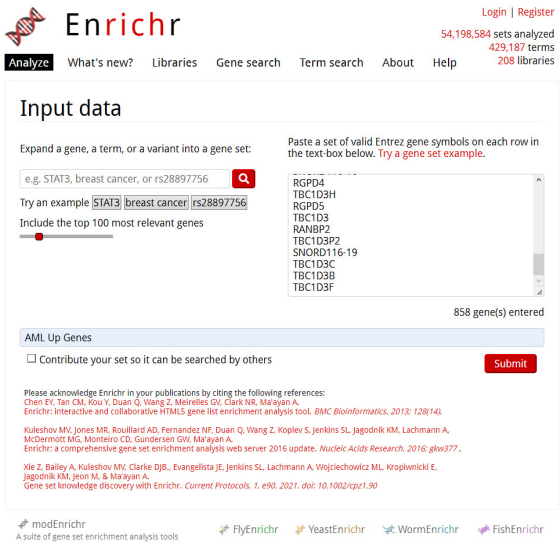
Least Expressed	Most Expressed
PECR	MPO
DDX58	KIT
CBX7	KIAA0101
STK38	FLT3
NR1D2	SUCNR1

جدول ۴: ۵ ژنی که در بیماران بیشتر/کمتر از بقیه بیان شده‌اند.

۲ بررسی pathway و gene ontology

۱.۲ ژن‌هایی که کمتر بیان شده‌اند

با رفتن به سایت enrichr و قرار دادن لیست ژن‌های کمتر بیان‌شده (شکل ۱) آن‌ها را بررسی می‌کنیم.

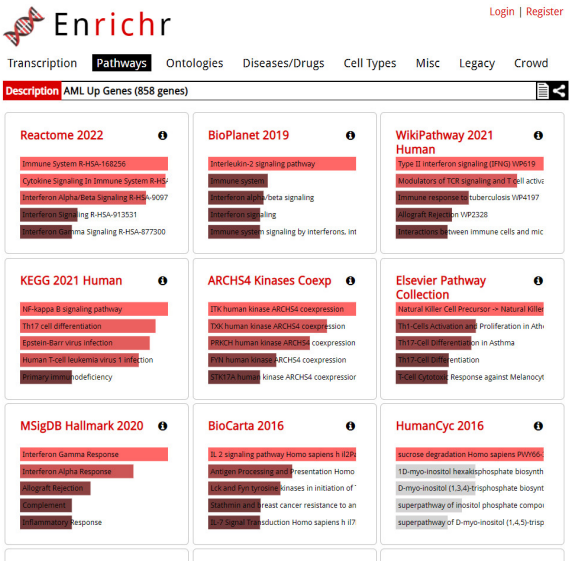


شکل ۱: وارد کردن داده‌ها برای شروع آنالیز

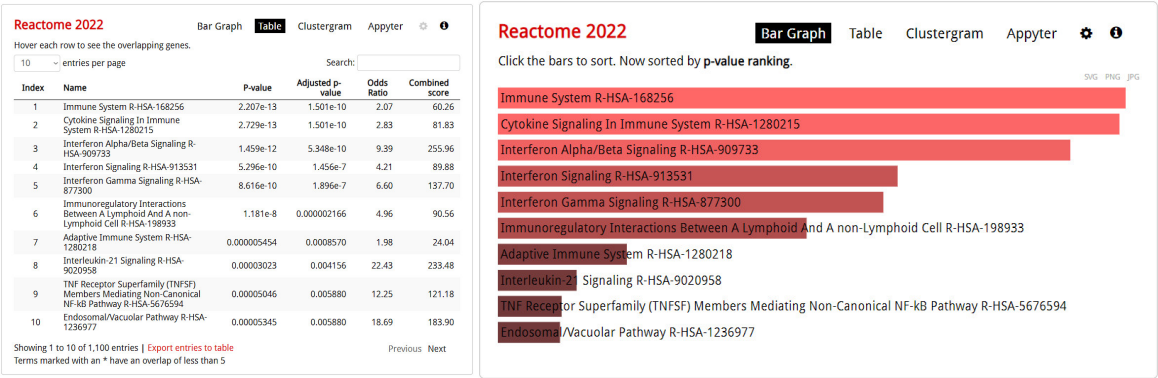
۱.۱.۲ انتخاب تمام نمونه‌های سالم

در بخش pathways (شکل ۲) می‌توان پایگاه‌داده‌های pathway مختلف را بررسی کرد.
به‌عنوان مثال در پایگاه‌داده Reactome 2022 (شکل ۳) تعداد زیادی از ژن‌هایی که میزان بیان کمتری داشتند مربوط به pathway سیستم ایمنی [3] هستند.
و همچنین پایگاه‌داده KEGG 2021 در شکل ۴ نمایش داده شده است.
با مراجعه به سایت KEGG و جست‌وجوی معنادارترین pathway یعنی NF-kappa B signaling pathway تصویر آن را در شکل ۵ نمایش می‌دهیم.

[3] Immune System



شکل ۲: بخش pathways در enrichr برای ژن‌های کمتر بیان‌شده و تمام نمونه‌ها

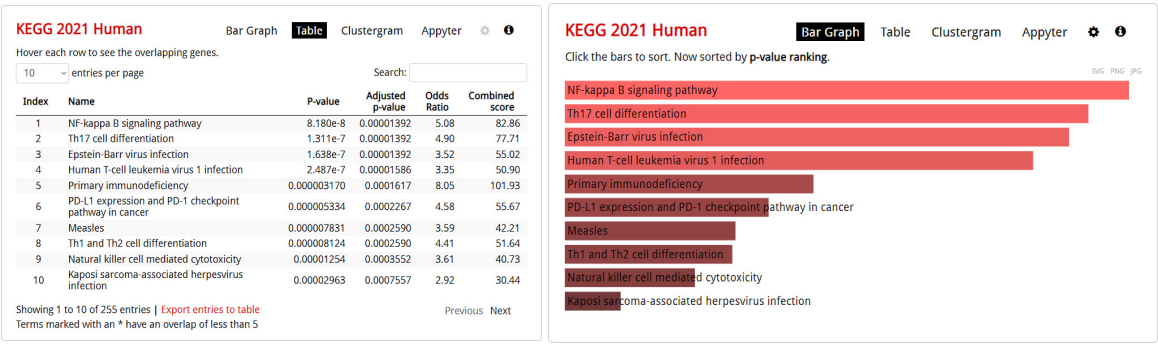


شکل ۳: pathwayهای پایگاه‌داده Reactome

و با مراجعه به سایت reactome و جست‌وجوی معنادارترین pathway یعنی Immune System R-HSA-168256 تصویر آن را در شکل ۶ نمایش می‌دهیم.
در بخش ontologies می‌توان biological processes (شکل ۸) و cellular components (شکل ۹) و molecular functions (شکل ۱۰) را بررسی کرد (شکل ۷).

۲.۱.۲ انتخاب نمونه‌های سالم با بیشترین هم‌بستگی

دوباره با رفتن به بخش pathways (شکل ۱۱) می‌توان پایگاه‌داده‌های pathway مختلف را این‌بار برای نمونه‌های جدید بررسی کرد.
پایگاه‌داده Reactome 2022 و همچنین پایگاه‌داده KEGG 2021 در شکل ۱۲ نمایش داده شده است.
در بخش ontologies می‌توان biological processes و cellular components و molecular functions را بررسی کرد (شکل ۱۳).



KEGG 2021 Human — 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	NF-kappa B signaling pathway	8.180e-8	0.00001392	5.08	82.86
2	Th17 cell differentiation	1.311e-7	0.00001392	4.90	77.71
3	Epstein-Barr virus infection	1.638e-7	0.00001392	3.52	55.02
4	Human T-cell leukemia virus 1 infection	2.487e-7	0.00001586	3.35	50.90
5	Primary immunodeficiency	0.000003170	0.0001617	8.05	101.93
6	PD-L1 expression and PD-1 checkpoint pathway in cancer	0.000005334	0.0002267	4.58	55.67
7	Measles	0.000007831	0.0002590	3.59	42.21
8	Th1 and Th2 cell differentiation	0.000008124	0.0002590	4.41	51.64
9	Natural killer cell mediated cytotoxicity	0.00001254	0.0003552	3.61	40.73
10	Kaposi sarcoma-associated herpesvirus infection	0.00002963	0.0007557	2.92	30.44

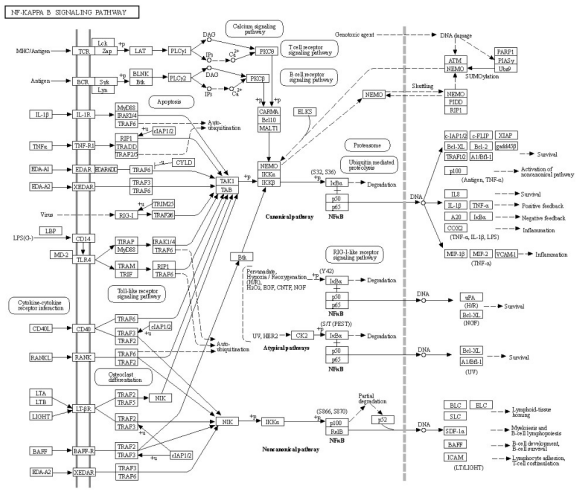
Showing 1 to 10 of 255 entries | Export entries to table — Previous Next

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

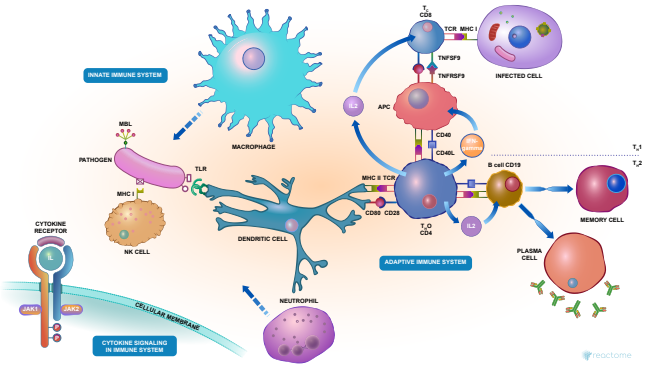
KEGG 2021 Human — Bar Graph | Table | Clustergram | Appyter

Click the bars to sort. Now sorted by **p-value ranking**.

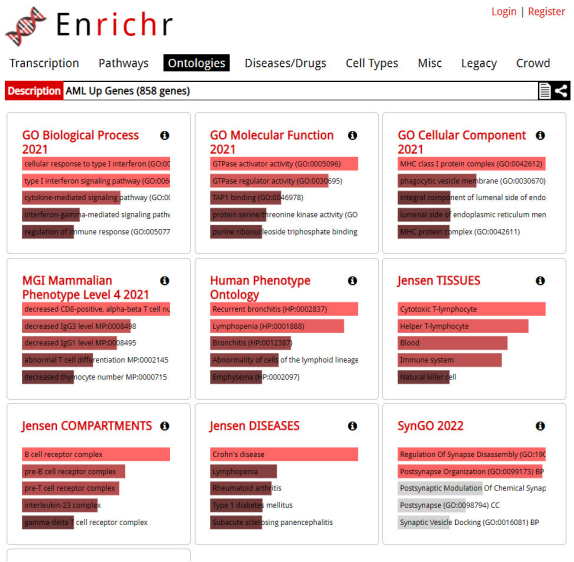
شکل ۴: pathwayهای پایگاه‌داده KEGG



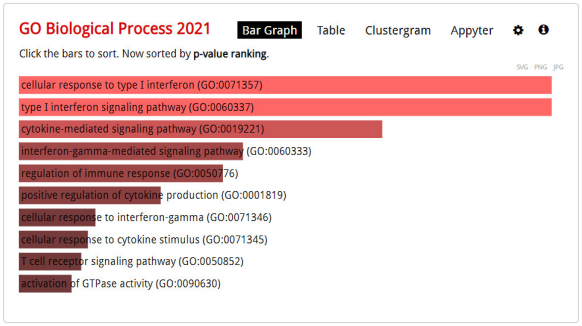
شکل ۵: شمای NF-kappa B signaling pathway



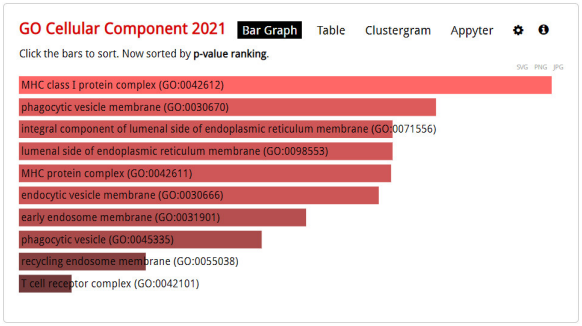
شکل ۶: شمای Immune System R-HSA-168256



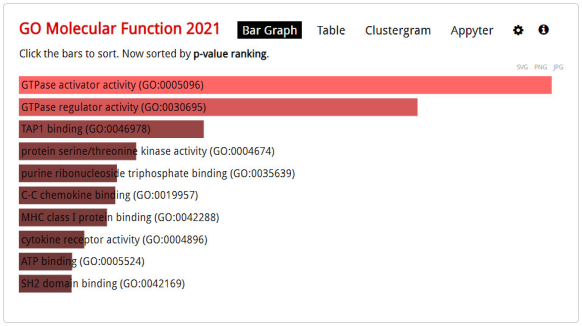
شکل ۷: بخش ontologies در enrichr برای ژن‌های کمتر بیان‌شده و تمام نمونه‌ها



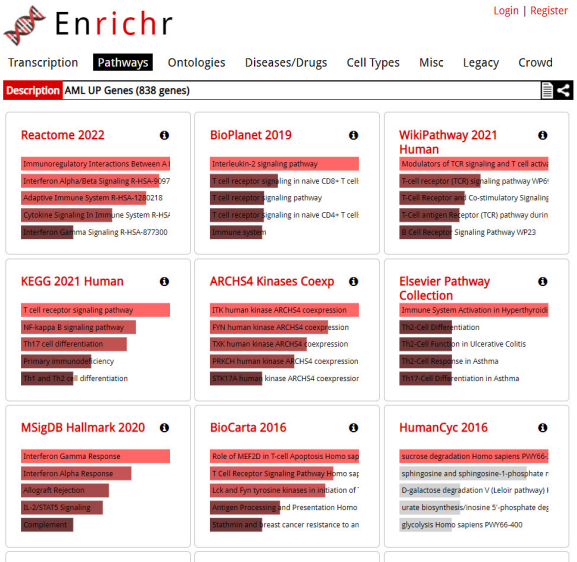
شکل ۸: biological processes در بخش ontologies



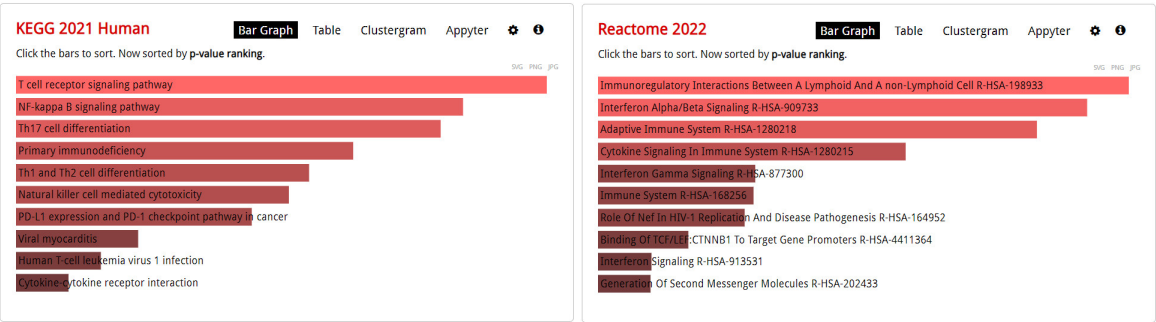
شکل ۹: cellular components در بخش ontologies



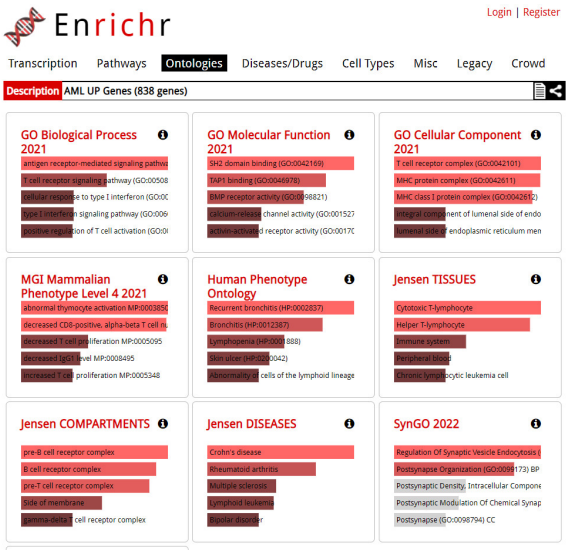
شکل ۱۰: molecular functions در بخش ontologies



شکل ۱۱: بخش pathways در enrichr برای ژن‌های کمتر بیان‌شده و نمونه‌های با هم‌بستگی بیشتر



شکل ۱۲: pathwayهای پایگاه‌داده Reactome و KEGG



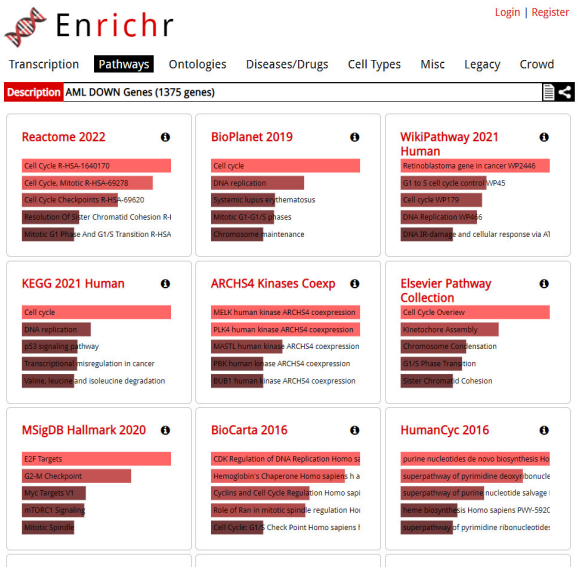
شکل ۱۳: بخش ontologies در enrichr برای ژن‌های کمتر بیان‌شده و نمونه‌های با هم‌بستگی بیشتر

۲.۲ ژن‌هایی که بیشتر بیان شده‌اند

حالا لیست ژن‌های بیشتر بیان‌شده را بررسی می‌کنیم.

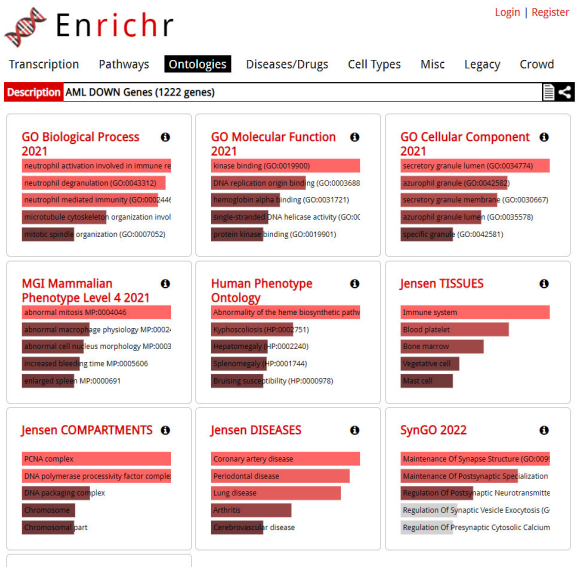
۱.۲.۲ انتخاب تمام نمونه‌های سالم

در بخش pathways (شکل ۱۴) می‌توان پایگاه‌داده‌های pathway مختلف را بررسی کرد.



شکل ۱۴: بخش pathways در enrichr برای ژن‌های بیشتر بیان‌شده و تمام نمونه‌ها

در بخش ontologies می‌توان biological processes و cellular components و molecular functions را بررسی کرد (شکل ۱۵).



شکل ۱۵: بخش ontologies در enrichr برای ژن‌های بیشتر بیان‌شده و تمام نمونه‌ها

۲.۲.۲ انتخاب نمونه‌های سالم با بیشترین هم‌بستگی

دوباره با رفتن به بخش pathways (شکل ۱۶) می‌توان پایگاه‌داده‌های pathway مختلف را این‌بار برای نمونه‌های جدید بررسی کرد.
در بخش ontologies می‌توان biological processes و cellular components و molecular functions را بررسی کرد (شکل ۱۷).

۳ جست‌وجو در مقالات

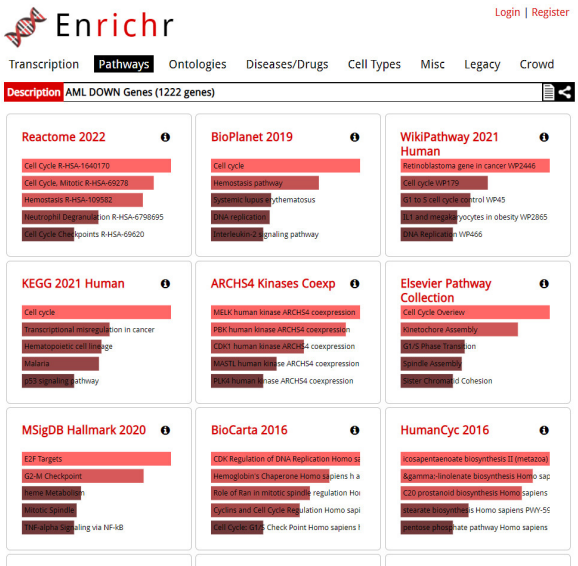
الف)

در مقاله [۱] و بخش نتیجه‌گیری بیان شده است که ژن E2F4 (که در پایگاه‌داده TRANSFAC and JASPAR PWMs به‌عنوان معنادارترین ژن با بیان زیاد انتخاب شده است) بیش از حد بیان شده است [۴].
در مقاله‌های [۵، ۴] گفته شده که ژن IRF8 (که در پایگاه‌داده ENCODE and ChEA Consensus TFs from آمده است) کمتر بیان شده است [۵].
در مقاله [۲] اشاره شده است که Interferon Alpha/Beta Signaling R-HSA-909733 در بیماران مبتلا به AML به میزان قابل توجهی کم بیان شده است (مطابق شکل ۳) [۶].

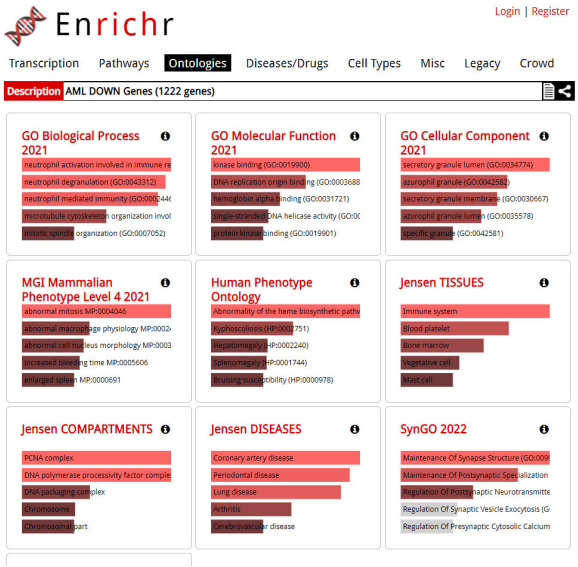
[4] E2F4 is aberrantly overexpressed in AML and is associated with poor prognosis

[5] We validated the critical role of the transcription factor IRF8 and demonstrated that it modulates the function of the cells by regulating important signaling molecules.

[6] The absence of monocyte and pDC activation by IFNα ex vivo could explain the lack of an in vivo anti-leukemic effect, and the therapeutic effect of IFNα may potentially be enhanced by removing this inherent block of activation in healthy immune subsets in AML patients



شکل ۱۶: بخش pathways در enrichr برای ژن‌های بیشتر بیان‌شده و نمونه‌های با هم‌بستگی بیشتر



شکل ۱۷: بخش ontologies در enrichr برای ژن‌های بیشتر بیان‌شده و نمونه‌های با هم‌بستگی بیشتر

در مقاله [۳] نیز اشاره شده که Cell cycle در کنترل بیماری موثر است (شکل ۱۴) [۷].

[7] Cell cycle control in acute myeloid leukemia

ب)

در مقاله [۵] به دارویی اشاره شده است که برای درمان AML باید مقدار بیان ژن IRF8 را افزایش داد [۸].
داروی Arsenic trioxide که یکی از موثرترین داروهای در درمان بیماری AML است در [۶] بیان شده است که NF-kappa B signaling pathway از اهداف اصلی این دارو است (شکل ۵ و ۱۸).

Mechanism of action

The mechanism of action of Arsenic Trioxide is not completely understood. Arsenic trioxide causes morphological changes and DNA fragmentation characteristic of apoptosis in NB4 human promyelocytic leukemia cells *in vitro*. Arsenic trioxide also causes damage or degradation of the fusion protein PML/RAR-alpha. It is suspected that arsenic trioxide induces cancer cells to undergo apoptosis.

TARGET	ACTIONS	ORGANISM
Inhibitor of nuclear factor kappa-B kinase subunit beta	inducer	Humans
Thioredoxin reductase 1, cytoplasmic	inhibitor	Humans
Transcription factor AP-1	inducer	Humans
G1/S-specific cyclin-D1	antagonist	Humans
Mitogen-activated protein kinase 3	inducer	Humans
Mitogen-activated protein kinase 1	inducer	Humans
RAC-alpha serine/threonine-protein kinase	inducer	Humans
Cyclin-dependent kinase inhibitor 1	Not Available	Humans
Histone deacetylase 1	Not Available	Humans
Protein PML	Not Available	Humans

شکل ۱۸: مشاهده NF-kappa B signaling pathway در هدف‌های موثر از داروی Arsenic trioxide

[8] As an inhibitor for the related protein IRF5 was recently described, IRF8 may also be a suitable drug target.

مراجع

[1] Feng, Y, Li, L, Du, Y, Peng, X, Chen, F. (2020) E2F4 functions as a tumour suppressor in acute myeloid leukaemia via inhibition of the MAPK signalling pathway by binding to EZH2, Cell Mol Med.

[2] Forthun, R.B., Hellesøy, M., Sulen, A. et al. (2019) Modulation of phospho-proteins by interferon-alpha and valproic acid in acute myeloid leukemia, Cancer Res Clin Oncol.

[3] Schnerch D, Yalcintepe J, Schmidts A, et al. (2012) Cell cycle control in acute myeloid leukemia, Am J Cancer Res.

[4] Sharma A, Yun H, Jyotsana N, Chaturvedi A, Schwarzer A, Yung E, Lai CK, Kuchenbauer F, Argiropoulos B, Görlich K, Ganser A, Humphries RK, Heuser M. (2015) Constitutive IRF8 expression inhibits AML by activation of repressed immune response signaling. Leukemia, Epub.

[5] Liss F, Frech M, Wang Y, et al. (2021) IRF8 Is an AML-Specific Susceptibility Factor That Regulates Signaling Pathways and Proliferation of AML Cells, Cancers (Basel).

[6] Trisenox. Arsenic trioxide, Drug Bank `https://go.drugbank.com/drugs/DB01169`.