



فاز دوم

پروژه مقدمه‌ای بر بیوانفورماتیک
دکتر علی شریفی زارچی و دکتر سمیه کوهی

دانشکده مهندسی کامپیووتر

دانشگاه صنعتی شریف

نیم‌سال اول ۰۲-۰۱

امیرحسین باقری - ۹۸۱۰۵۶۲۱

مهندی مستانی - ۹۷۱۰۰۵۱۳

محمد رضا مفیضی - ۹۸۱۰۶۰۵۹

فهرست مطالب

۱	بیان ژن‌ها
۱.۱	انتخاب تمام نمونه‌های سالم
۱.۲	انتخاب نمونه‌های سالم با بیشترین همبستگی
۲	بررسی gene ontology و pathway
۲.۱	ژن‌هایی که کمتر بیان شده‌اند
۲.۱.۱	انتخاب تمام نمونه‌های سالم
۲.۱.۲	انتخاب نمونه‌های سالم با بیشترین همبستگی
۲.۲	ژن‌هایی که بیشتر بیان شده‌اند
۲.۲.۱	انتخاب تمام نمونه‌های سالم
۲.۲.۲	انتخاب نمونه‌های سالم با بیشترین همبستگی
۳	جستجو 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 gene ontology و pathway

۱.۲ ژن‌هایی که کمتر بیان شده‌اند

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

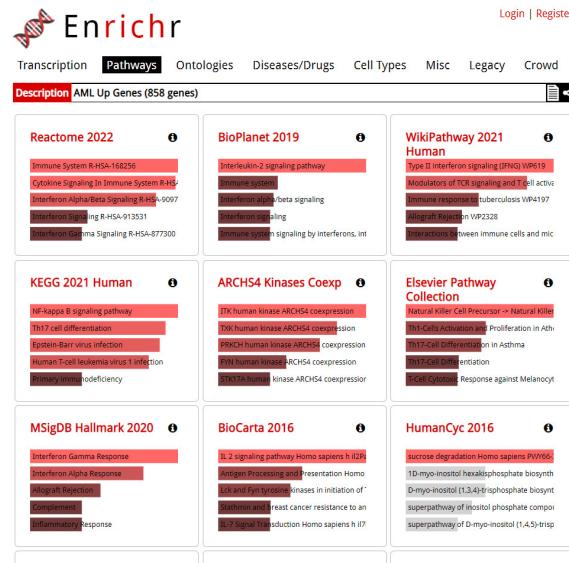
The screenshot shows the Enrichr web interface. At the top, there are links for 'Login | Register', '54,198,584 sets analyzed', '429,187 terms', and '208 libraries'. Below this is a navigation bar with 'Analyze' (selected), 'What's new?', 'Libraries', 'Gene search', 'Term search', 'About', 'Help', and 'Logout'. The main area is titled 'Input data' with a sub-section 'Expand a gene, a term, or a variant into a gene set:'. A text input field contains 'e.g. STAT3, breast cancer, or rs28897756' with a magnifying glass icon. Below it is a dropdown menu with 'Try an example' and 'STAT3 | breast cancer | rs28897756'. A checkbox 'Include the top 100 most relevant genes' is checked. To the right, a text box says 'Paste a set of valid Entrez gene symbols on each row in the text-box below. Try a gene set example.' A scrollable list of genes is shown, starting with RGPD4, RGPD3H, RGPDS, TBLCD3, BANBP2, TBC1D3P2, SNORD116-19, TBC1D3C, TBC1D3B, and TBC1D5F. A note at the bottom right says '858 gene(s) entered'. Below this is a section for 'AML Up Genes' with a checkbox 'Contribute your set so it can be searched by others' and a 'Submit' button. At the bottom, there is a note about citation requirements and a list of references. At the very bottom, there are links for 'modEnrichr', 'FlyEnrichr', 'YeastEnrichr', 'WormEnrichr', and 'FishEnrichr'.

شکل ۱: وارد کردن داده‌ها برای شروع آنالیز

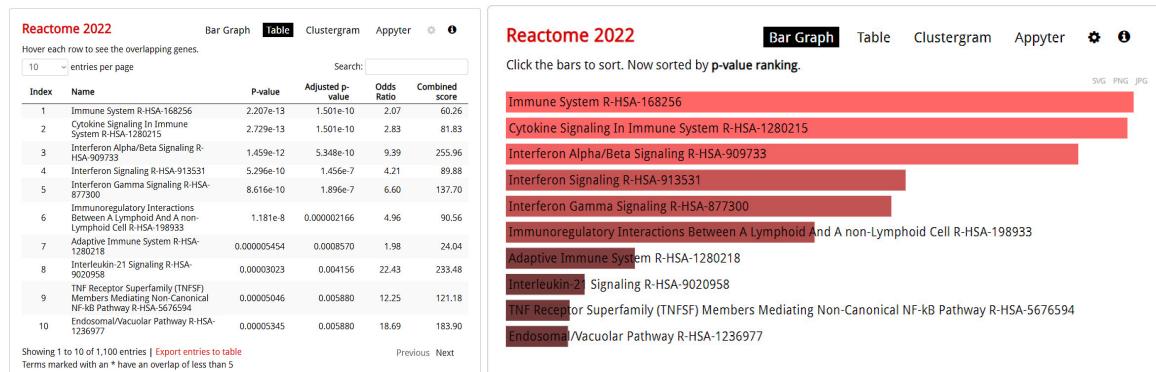
۱.۱.۲ انتخاب تمام نمونه‌های سالم

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

³Immune System



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



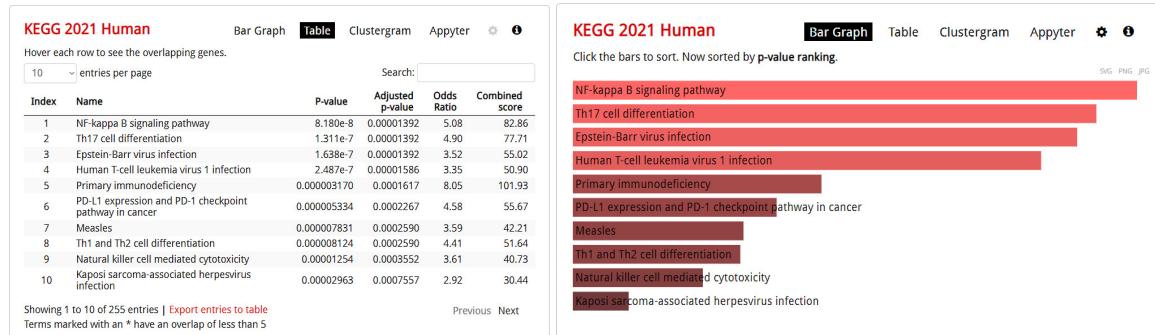
شکل ۳: Reactome پایگاهداده‌های pathway

و با مراجعه به سایت reactome و جستجوی 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).

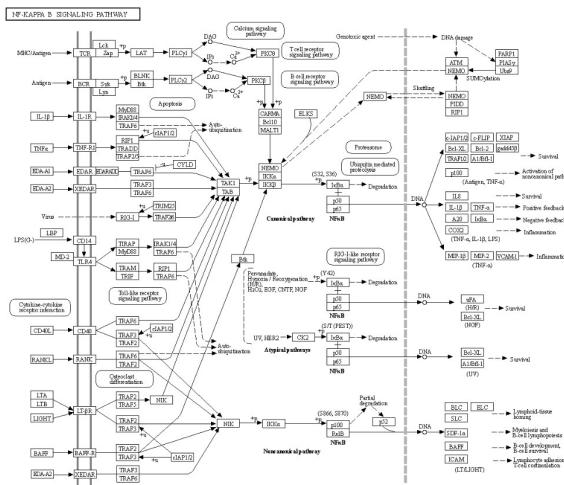
۲.۱.۲ انتخاب نمونه‌های سالم با بیشترین همبستگی

دوباره با رفتن به بخش pathways (شکل ۱۱) می‌توان پایگاهداده‌های pathway مختلف را این‌بار برای نمونه‌های جدید بررسی کرد.

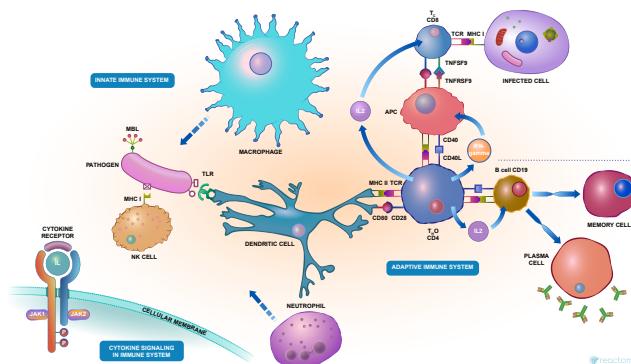
پایگاهداده Reactome 2022 و همچنین پایگاهداده KEGG 2021 در شکل ۱۲ نمایش داده شده است. در بخش ontologies می‌توان biological processes و cellular components و molecular functions را بررسی کرد (شکل ۱۳).



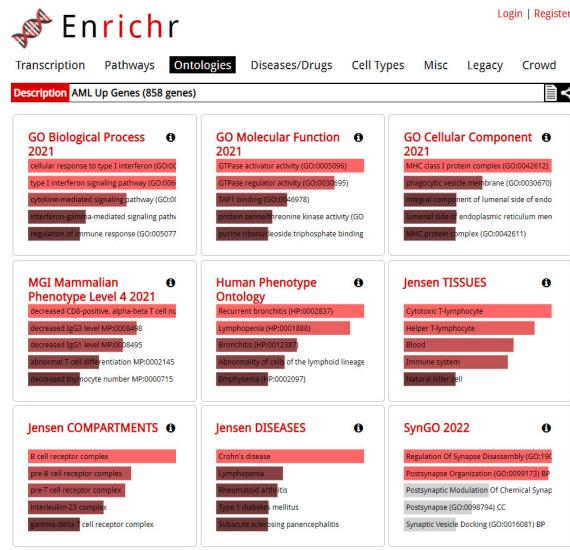
شکل ۴: پایگاهداده pathway‌های KEGG



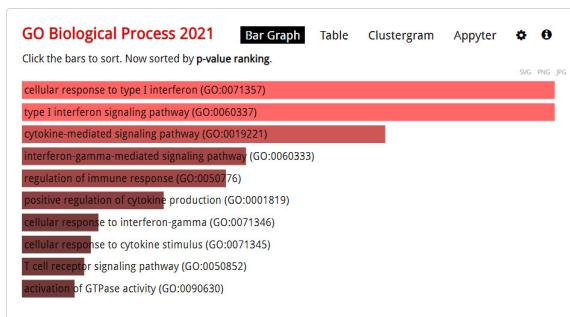
شکل ۵: شمای NF-kappa B signaling pathway



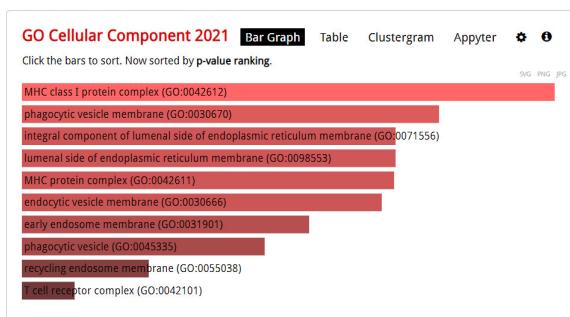
شکل ۶: شمای Immune System R-HSA-168256



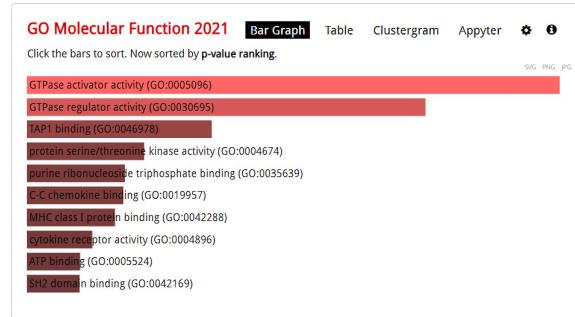
شکل ۷: بخش ontologies در enrichr برای ژن‌های کمتر بیان شده و تمام نمونه‌ها



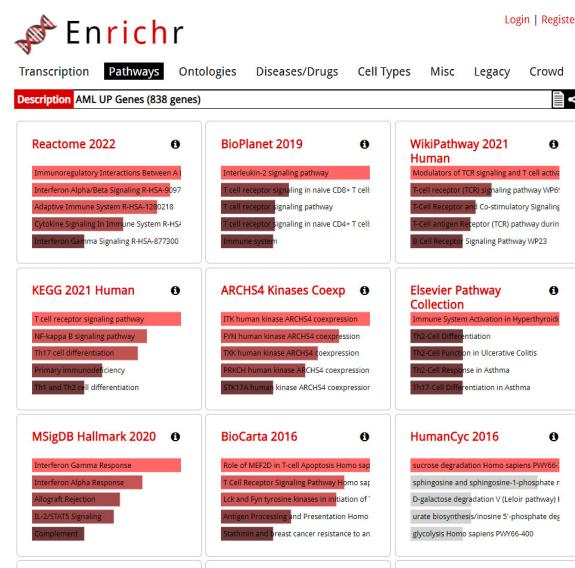
شکل ۸: ontologies در بخش biological processes



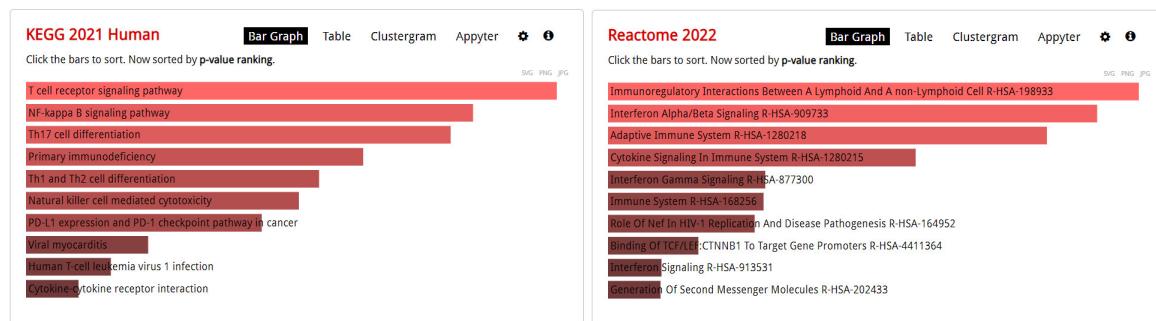
شکل ۹: ontologies در بخش cellular components



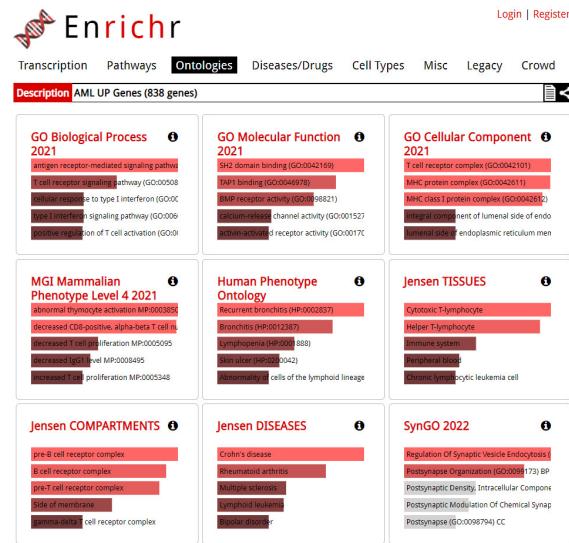
شکل ۱۰: ontologies در بخش molecular functions



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



شکل ۱۲: KEGG و Reactome پایگاه‌داده pathway



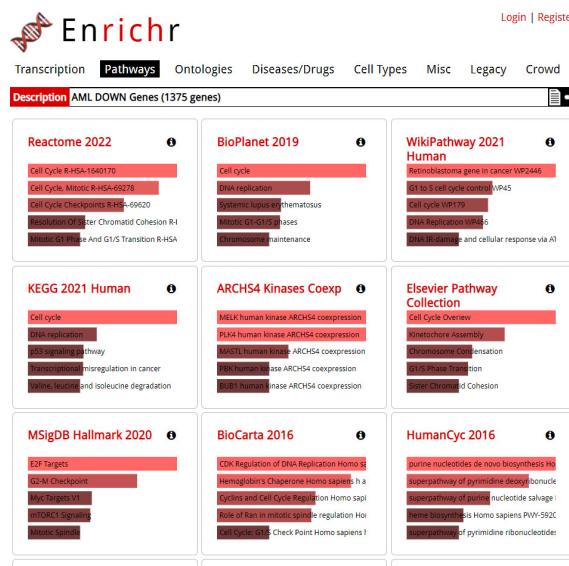
شکل ۱۳: بخش enrichr در ontologies برای ژن‌های کمتر بیان شده و نمونه‌های با همبستگی بیشتر

۲.۲ ژن‌هایی که بیشتر بیان شده‌اند

حالا لیست ژن‌های بیشتر بیان شده را بررسی می‌کیم.

۱.۲.۲ انتخاب تمام نمونه‌های سالم

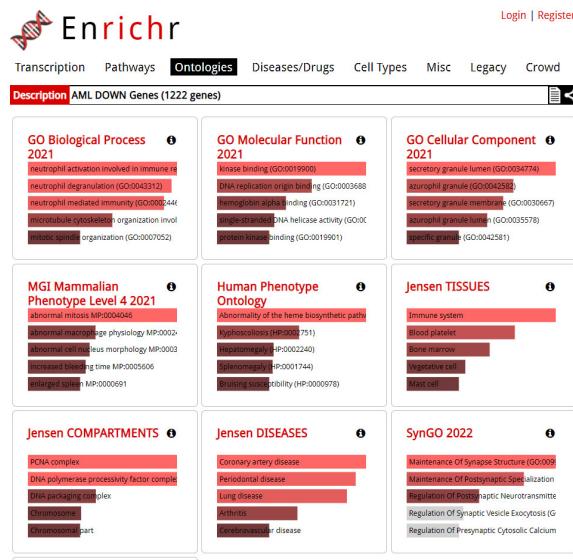
در بخش pathways (شکل ۱۴) می‌توان پایگاهداده‌های مختلف pathway را بررسی کرد.



شکل ۱۴: بخش enrichr برای ژن‌های بیشتر بیان شده 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 بیان شده 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 جستجو در مقالات

(الف)

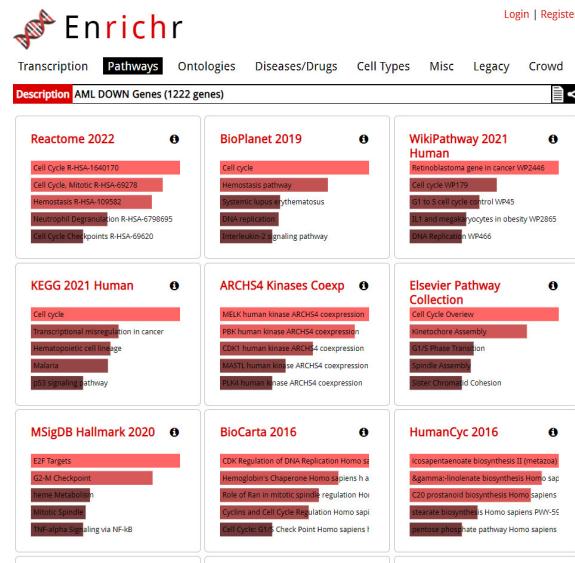
در مقاله [۱] و بخش نتیجه‌گیری بیان شده است که ژن E2F4 (که در پایگاه‌داده TRANSFAC and JASPAR به عنوان معنادارترین ژن با بیان زیاد انتخاب شده است) بیش از حد بیان شده است^۴. در مقاله‌های [۴، ۵] گفته شده که ژن IRF8 (که در پایگاه‌داده ENCODE and ChEA Consensus TFs from Interferon Alpha/Beta Signaling R-HSA-909733 در بیماران مبتلا به AML به میزان قابل توجهی کم بیان شده است (مطابق شکل ۳)^۵. در مقاله [۳] نیز اشاره شده که Cell cycle در کنترل بیماری موثر است (شکل ۱۴)^۶.

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

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

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

⁷Cell cycle control in acute myeloid leukemia



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



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

(ب)

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

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



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 <i>in vitro</i> . Arsenic trioxide also causes damage or degradation of the fusion protein PML/RARalpha. It is suspected that arsenic trioxide induces cancer cells to undergo apoptosis.		
TARGET	ACTIONS	ORGANISM	
① inhibitor of nuclear factor kappa B kinase subunit beta	Inducer	Humans	
② thioredoxin reductase 1, cytoplasmic	Inhibitor	Humans	
③ transcription factor AP-1	Inducer	Humans	
④ G1/S-specific cyclin-D1	antagonist	Humans	
⑤ Mitogen-activated protein kinase 3	Inducer	Humans	
⑥ Mitogen-activated protein kinase 1	Inducer	Humans	
⑦ RAC1alpha serine/threonine-protein kinase	Inducer	Humans	
⑧ Cyclin-dependent kinase inhibitor 3	Not Available	Humans	
⑨ Histone deacetylase 1	Not Available	Humans	
⑩ Protein PML	Not Available	Humans	

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



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>.