آنالیز بیان ژن داده های ماکرواری بیماری لوکمیا با برنامه DAVID

حکیمه زالی[*1]، راضیه امینی[2]، رضا شیری هریس[3]

1) مرکز تحقیقات پروتئومیکس، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی تهران
2) گروه میکروبیولوژی، دانشکده علوم، دانشگاه آزاد اسلامی واحد همدان
3) گروه علوم آزمایشگاهی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی ایلام

تاریخ دریافت: 91/10/2 **تاریخ پذیرش: 92/1/17**

چکیده

مقدمه: سرطان خون یا لوکمیا بیماری پیشرونده و بدخیم اعضای خون ساز بدن است. ناهنجاری های ژنتیکی نقش بسیار مهمی در رشد لوکمی در بدن دارد. مطالعات زیادی پیرامون کشف عوامل ملکولی درگیر در این بیماری صورت گرفته است یکی از حوضه های دانش جدید در کشف بیان ژن ها در حالت بیماری استفاده از تکنولوژی ماکرواری است که یک تصویر کلی از میزان بیان ژن را در سایز بزرگ ژنومی ارائه می کند. در واقع با بررسی تعداد زیادی ژن در کنار یکدیگر انتظار می رود تغییرات حاصل از بیماری را دقیق تر بتوان در علم بیولوژیک مورد بررسی قرار داد. در این مطالعه داده های حاصل از ماکرواری بیماری لوکمیا توسط برنامه DAVID مورد آنالیز قرار می گیرد و هدف آن انجام آنالیز های عملکردی بر روی لیست های ژنومی و پروتئومی که داده های آن ها از مطالعات بیولوژیک با دستگاه های پیشرفته(high-throughput) حاصل می شود.

مواد و روش ها: ست ژنی ماکرواری لوکمیا از پایگاه داده http://www.biomedcentral.com/content/supplementary/1471-به دست آمد و با تجزیه و تحلیل یافته ها با نرم افزارهای بیوانفورماتیکی(DAVID)، ارتباطات بیان ژن در قالب کلاس های مختلف ژنی، چارت ها و خوشه بندی ژن ها مورد بررسی قرار گرفت.

یافته های پژوهش: لیست ژن های شناسایی شده در میکرواری با برنامه DAVID تهیه شد. از 615 ژن شناسایی شده ارتباط با کلاس بیماری های مختلف شناسایی شد. بیشترین ژن های دخیل در بیماری در دسته ژن هایی هستند که در انواع سرطان ها دخالت دارند. 23/7 درصد ژن های شناسایی شده(146 ژن) در بیماری لوکمیا در دسته ژن های موجود در انواع سرطان ها هستند. از 615 ژن شناسایی شده 70 چارت مسیرهای بیولوژیکی(موجود در پایگاه داده KEGG) مرتبط با بیماری شناسایی شد. از 615 ژن شناسایی شده 12 کلاستر بر اساس حاشیه نویسی عملکردی مرتبط با بیماری شناسایی شد.

بحث و نتیجه گیری: نتایج نشان می دهد که برنامه آنالیز ژنومDAVID قادر است کلاس های مهم ژنی و مسیرهای اصلی درگیر در بیماری را تعیین نماید و بهترین کاندیدا مارکر ژنی برای تشخیص بیماری، علت آن و درمان لوکمیا را معرفی نماید.

واژه های کلیدی: ماکرواری، بیماری لوکمیا، برنامه DAVID، بیان ژن

* **نویسنده مسئول:** مرکز تحقیقات پروتئومیکس، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی تهران

Email: Hakimehzali@yahoo.com

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

مقدمه

سرطان خون یا لوکمیا(لوسمی) بیماری پیشرونده و بدخیم اعضای خون ساز بدن است. این بیماری در اثر تکثیر و تکامل ناقص گویچه های سفید خون و پیش سازهای آن در خون و مغز استخوان ایجاد می‌شود. در این بیماری مغز استخوان به صورت غیر عادی، مقدار بسیار زیادی سلول خونی تولید می کند. این سلول ها با سلول های خون نرمال و عادی متفاوت هستند و درست عمل نمی‌کنند در نتیجه، تولید سلول های سفید خون طبیعی را متوقف کرده و توانایی فرد را در مقابله با بیماری ها از بین می برند. سلول های لوکمی هم چنین بر تولید سایر انواع سلول های خونی که توسط مغز استخوان ساخته می شود از جمله گویچه های قرمز خون که اکسیژن را به بافت های بدن می رسانند و پلاکت های خونی که از لخته شدن خون جلوگیری می کنند نیز اثر می گذارند،(3-1). لوسمی بر اساس نوع گویچه سفید خون به دو دسته لنفوئیدی یا لنفوبلاستی و میلوئیدی یا مغز استخوانی تقسیم بندی می شوند. هم چنین هر دو دسته لوسمی به دو نوع حاد و مزمن تقسیم می شوند لذا چهار نوع کلی لوسمی وجود دارد که شامل لوسمی لیمفوبلاستیک حاد(ALL)، لوسمی میلوسیتی حاد(AML)، لوسمی لیمفوبلاستیک مزمن(CLL) و لوسمی میلوسیتی مزمن(CML) هستند.(6-4)

علت دقیق ابتلا به لوسمی مشخص نیست و به نظر می رسد عوامل ژنتیکی و محیطی در به وجود آمدن این بیماری نقش داشته باشند. پژوهش های انجام شده روند بدخیمی بیماری لوسمی را به ژنتیک، قرار گرفتن در معرض پرتوهای یونیزه کننده و برخی مواد شیمیایی خاص و یا نارسایی سیستم ایمنی طبیعی بدن ارتباط می دهند،(۶،۷). مطالعات متعددی پیرامون کشف عوامل ملکولی درگیر در این بیماری صورت گرفته است.

یکی از حوضه های دانش جدید در کشف بیان ژن ها در حالت بیماری استفاده از تکنولوژی ماکروواری است که یک تصویر کلی از میزان بیان ژن را در سایز بزرگ ژنومی ارائه می کند. نتایج به دست آمده از داده های میکروواری، منجر به ایجاد دیدگاه های تازه ای در مورد نحوه شکل گیری، پیشرفت و پاسخ به درمان بیماری ها به ویژه بیماران سرطانی گردیده است. در این تکنیک میزان بیان هزاران ژن به طور هم زمان، جهت مطالعه رفتار و عملکرد سلول مورد بررسی قرار می گیرد و در نهایت با استفاده از روش های مناسب آماری، داده های حاصله تجزیه و تحلیل می شوند،(10-8). در واقع با بررسی تعداد زیادی ژن در کنار یکدیگر انتظار می رود تغییرات حاصل از بیماری را دقیق تر بتوان در علم بیولوژیک مورد بررسی قرار داد،(15-11). برقراری ارتباط منطقی بین حجم عظیم داده های حاصل از میکروواری و عوامل اصلی دخیل در ایجاد بیماری و پیشرفت آن از جمله چالش های آنالیز داده های حاصل از دستگاه های پیشرفته (high-throughput) است لذا برنامه های متعددی جهت تسهیل انجام آنالیزهای عملکردی بر روی لیست های ژنومی و پروتئومی که داده های آن ها حاصل مطالعات بیولوژیک با دستگاه های پیشرفته است طراحی شده است. از کارآمدترین برنامه های آنالیز ژنوم و پروتئوم برنامه DAVID است که توانایی جستجو و مطالعه ژن های مرتبط با هم را به صورت یک واحد در قالب شبکه های بیولوژیکی بزرگ تر نسبت به مطالعه ژن ها به صورت جداگانه برای محقق فراهم می نماید. اساس گروه بندی بر پایه عملکرد مشابه، کلاسترینگ ژن ها در گروه های مجزایی است که هر گروه دارای عملکردی یکسان هستند. این برنامه هم چنین دارای پایگاه اطلاعاتی است که می توان ارتباط بیان ژن ها با بیماری ها، مسیرهای بیولوژیک، نوع سکانس، هومولوژی و بیان بافتی ژن ها را مورد بررسی قرار داد،(21-16). در این مطالعه داده های حاصل از ماکروواری بیماری لوکمیا توسط برنامه DAVID مورد آنالیز قرار می گیرد تا علت بیولوژیک ایجاد بیماری در سطح ملکولی آشکار شود و مسیرهای بیولوژیکی اصلی دخیل در بیماری کشف شود.

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

مواد و روش ها

داده های لوسمی

لوسمی به طور عمده به چهار طبقه ALL، AML، CLL و CML تقسیم می شوند. گالوپ(1999) و همکاران،(22)، کوشیدند تا با ارائه روش های نوینی برخلاف روش های متداول و بسیار پرهزینه رایج، بر مبنای داده های بیان ژنی حاصل از میکروواری این سرطان را تشخیص دهند. نهایت 50 ژن را به عنوان ژن های تشخیص افراد استفاده نمودند. داده های لوسمی که در این جا مورد بررسی قرار گرفت از مطالعه 615 ژن مربوط به 38 فرد مبتلا به لوکمیای حاد استخراج شده است. ست ژنی ماکرواری لوسمی از سایت زیر به دست آمد.

[http://www.biomedcentral.com/content/supplementary/1471-2105-9-502-S5.pdf]

روش های تجزیه و تحلیل داده ها

DAVID پایگاه داده ای برای حاشیه نویسی و تصویر سازی ادغام شده از منابع بیوانفورماتیک است و متشکل از یک پایگاه جامع دانش زیستی و مجموعه ای جامع از ابزار تحلیلی برای استخراج ویژگی های بیولوژیکی از لیست های بزرگ ژنی و پروتئینی به دست آمده از مطالعات ژنومیکی و پروتئومیکی است. از مزایای این برنامه استفاده از 22 نوع ID ژنی مختلف و بیش از 40 حاشیه نویسی عملکردی شناخته شده از ده ها پایگاه داده عمومی می باشد. از جمله داده پایگاه های مهم مورد استفاده در برنامه شامل LocusLink، PIR، GeneCards، KEGG، Ensembl، و wiss-Prot هستند. هر کاربر برحسب نوع مطالعه و هدف تحقیق می تواند یکی از این پایگاه داده ها را انتخاب نماید و برنامه DAVID نیز فقط در آن پایگاه داده به جستجوی هدف بپردازد. با پیشرفت مجموعه ای از الگوریتم های جدید مانند ابزار کلاس بندی عملکردی ژنی، ابزار کلاسترینگ حاشیه نویسی عملکردی، ابزار جستجوی خطی، جستجوگر DAVID را قادر ساخته است تا در پایگاه داده های KEGG، BioCarta و غیره ارتباط بیولوژیکی ژن ها را به صورت پویا تصویرسازی نماید.(21-16)

برنامه DAVID به صورت رایگان از طریق سایت http://david.abcc.ncifcrf.gov/ قابل دسترس است. ID ژنی برای داده های میکروواری با پسوند at- مشخص می شود. از برنامه DAVID نتایج مختلفی را می توان به دست آورد که بر حسب معنی داری آماری(P و Benjamin) از آن ها جهت تفسیر ارتباط بیماری لوکمیا و ست های ژنی استفاده نمود.

ابزارهای مورد استفاده در این مطالعه شامل

1-وارد نمودن ست ژنی سرطان لوسمی به دست آمده از تکنیک میکروواری(615 ژن) به برنامه DAVID و گرفتن لیست ژنی از ID ژن های موجود در بیماری است.

2-داده پایگاه GENETIC_ASSOCIATION_DBدر بین پایگاه داده های مختلف در برنامه DAVID انتخاب می شود. این پایگاه داده حاوی همه ژن های مرتبط با انواع بیماری های انسانی شناخته شده است و DAVID را قادر می سازد که ارتباطی منطقی بین لیست ژنی و پروتئنی وارد شده با انواع خانواده های بیماری ها برقرار نماید.

3-هر کلاس از بیماری های مرتبط با لوکمیا را می توان انتخاب نمود و ژن های مشترک آن ها را مشاهده نمود.

4- پایگاه داده KEGG در بین پایگاه داده های موجود در DAVID جهت کشف مسیرهای بیولوژیکی انتخاب نموده و نتیجه آن شناسایی تعداد چارت های مسیرهای بیولوژیکی مرتبط با بیماری خواهد بود.

5-مسیرهای بیولوژیک را به صورت گرافیکی در پایگاه داده DAVID بر اساس نوع داده های ورودی و آنالیز آنجام شده بر روی داده ها ارائه می شود.

6-خوشه بندی بر اساس حاشیه نویسی عملکردی از ابزارهای اصلی برنامه DAVID است که خوشه بندی بر اساس داده های ورودی و نوع پایگاه داده انتخاب شده انجام می شود. داده پایگاه انتخاب شده برای لوکمیا GENETIC_ASSOCIATION_DB بود.

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

یافته های پژوهش

گالوپ و همکاران(1999) داده های بیان ژنی سرطان لوسمی را با تکنیک میکرواری ارائه نمودند که داده های آن قابل دسترس برای عموم است،(22). در این مطالعه داده های حاصل از ژن های شناسایی شده از آنالیز ماکروواری بیماری لوکمیا توسط برنامه آنالیز ژنوم مورد بررسی قرار گرفت. 615 ژن از ژن های شناسایی شده مورد آنالیز جهت کشف رابطه ژن و بیماری توسط برنامه DAVID مورد بررسی قرار گرفت. جدول شماره 1 تعدادی از ژن های موجود در بیماری را معرفی می کند.

با انتخاب داده پایگاه GENETIC_ASSOCIATION_DB_DISEASE_CLASS که حاوی همه ژن های مرتبط با انواع بیماری های انسانی شناخته شده است این برنامه را قادر می سازد که ارتباطی منطقی بین لیست ژنی و پروتینی وارد شده با انواع خانواده های بیماری ها برقرار نماید. داده های ژنی لوکمیا از 615 ژن شناسایی شده دارای ارتباط وسیعی با انواع بیماری ها می باشد نتایج حاصل از ارتباط بین انواع کلاس بیماری ها با بیماری لوکمیا در جدول شماره 2 به نمایش در آمده است.

23/7 درصد ژن های شناسایی شده در بیماری لوکمیا در کلاس ژن های موجود در انواع سرطان ها هستند. جدول شماره 3، 146 ژنی را که در دسته ژن های دخیل در سرطان هستند نشان می دهد.

مسیرهای بیولوژیک KEGG مجموعه ای از نقشه های دستی مسیرهای کشیده شده است که بیانگر دانش ما در اینترکشن های مولکولی و شبکه های اینترکشنی برای انواع هدف های مطالعات ژنومی، پروتئومی و متابولومی، پردازش اطلاعات ژنتیکی، پردازش اطلاعات زیست محیطی، فرایندهای سلولی و بیماری های انسان می باشد. با انتخاب پایگاه داده KEGG در بین پایگاه داده های موجود در DAVID جهت کشف مسیرهای بیولوژیکی، از 615 ژن شناسایی شده 70 چارت مسیرهای بیولوژیکی (موجود در پایگاه داده KEGG) مرتبط با بیماری شناسایی شد که نتایج آن در جدول شماره 4 آمده است. در پایگاه داده DAVID بر اساس نوع داده های ورودی و آنالیز انجام شده بر روی داده ها، مسیرهای بیولوژیک را به صورت گرافیکی ارائه می کند. شکل شماره 1 مسیر بیولوژیکی دخیل در سرطان را نشان می دهد که بر اساس مسیرهای بیولوژیک موجود در داده پایگاه KEGG. طراحی شده است. از 615 ژن شناسایی شده 12 کلاستر بر اساس حاشیه نویسی عملکردی مرتبط با بیماری لوکمیا شناسایی شد که نتایج آن در جدول شماره 5 آمده است.

جدول شماره 1. لیست تعدادی از ژن های موجود در بیماری لوکمیا

AFFYMETRIX_3PRIME_IVT_ID	Gene Name
1002_f_at	cytochrome P450, family 2, subfamily C, polypeptide 19
1017_at	mutS homolog 6 (E. coli)
1021_at	interferon, gamma
1024_at	cytochrome P450, family 1, subfamily A, polypeptide 1
1025_g_at	cytochrome P450, family 1, subfamily A, polypeptide 1
1032_at	interleukin 8 receptor, beta
1033_g_at	interleukin 8 receptor, beta
1034_at	TIMP metallopeptidase inhibitor 3
1035_g_at	TIMP metallopeptidase inhibitor 3
1039_s_at	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

جدول شماره 2 . چارت های شناسایی شده در ارتباط بین خانواده های بیماری ها و لوکمیا

Category	Term	Genes	Count	%	P	Benjamini
GENETIC_ASSOCIATION_DB _DISEASE_CLASS	CANCER		146	23.7	6.3E-21	1.2E-19
	IMMUNE		119	19.3	2.8E-8	2.6E-7
	AGING		26	4.2	2.3E-5	7.2E-5
	PSYCH		66	10.7	4.6E-2	6.1E-2
	NEUROLOGICAL		67	10.9	6.7E-2	7.9E-2
	HEMATOLOGICAL		26	4.2	2.2E-3	4.2E-3
	CARDIOVASCULAR		86	14.0	3.4E-3	5.8E-3
	RENAL		27	4.4	1.8E-2	2.9E-2
	VISION		28	4.6	1.9E-2	2.8E-2

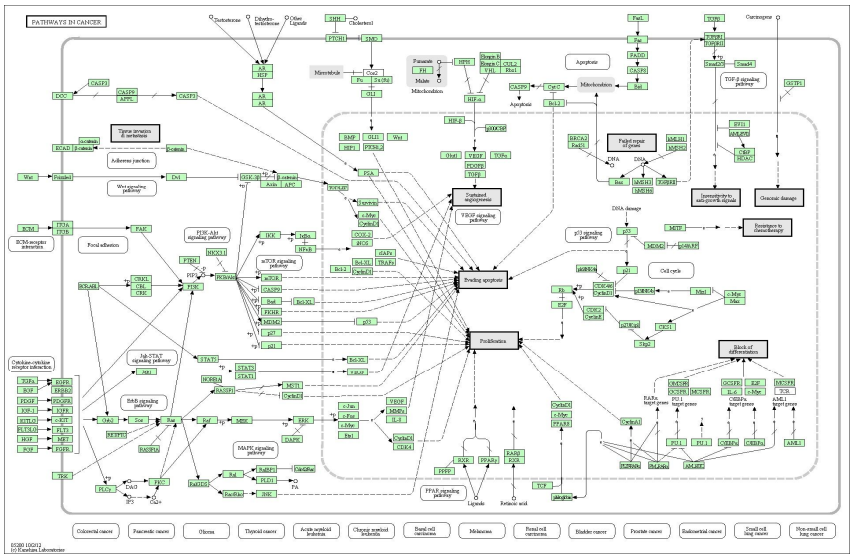
جدول شماره 3 . برخی از ژن های شناسایی شده در کلاس ژن های مرتبط با سرطان های مختلف

GENE NAME	GENE NAME
hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	ATP-binding cassette, sub-family B (MDR/TAP), member 1
insulin receptor	BCL2-antagonist/killer 1; BCL2-like 7 pseudogene 1
insulin-like growth factor 1 (somatomedin C)	Bloom syndrome, RecQ helicase-like
insulin-like growth factor 1 receptor	E2F transcription factor 1
insulin-like growth factor 2 (somatomedin A); insulin; INS-IGF2 readthrough transcript	Fanconi anemia, complementation group C
insulin-like growth factor 2 receptor	Fas (TNF receptor superfamily, member 6)
insulin-like growth factor binding protein 1	RAD50 homolog (S. cerevisiae)
insulin-like growth factor binding protein 3	RAD52 homolog (S. cerevisiae)
integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	SMAD family member 2

جدول شماره 4 . تعداد چارت های مسیرهای بیولوژیکی(موجود در پایگاه داده KEGG) مرتبط با بیماری لوکمیا

Category	Term	Genes	Count	%	P	Benjamini
KEGG_PATHWAY	Pathways in cancer		83	13.5	1.2E-23	1.6E-21
	Cytokine-cytokine receptor interaction		73	11.9	2.3E-23	1.6E-21
	MAPK signaling pathway		65	10.6	1.9E-17	8.6E-16
	Jak-STAT signaling pathway		48	7.8	2.8E-17	9.5E-16
	Neurotrophin signaling pathway		37	6.0	9.4E-13	2.5E-11
	Glioma		26	4.2	1.3E-12	3.0E-11
	Melanoma		26	4.2	3.1E-11	5.9E-10
	Non-small cell lung cancer		22	3.6	1.4E-10	2.4E-9
	Prostate cancer		28	4.6	2.5E-10	3.7E-9
	Toll-like receptor signaling pathway		29	4.7	1.2E-9	1.6E-8
	ErbB signaling pathway		26	4.2	4.1E-9	5.1E-8
	Renal cell carcinoma		23	3.7	5.7E-9	6.4E-8
	T cell receptor signaling pathway		29	4.7	6.1E-9	6.3E-8

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]



شکل شماره 1. مسیرهای ملکولی درگیر در سرطان

جدول شماره 5 . خوشه بندی بر اساس حاشیه نویسی عملکردی

Annotation Cluster 1	Enrichment Score: 7.4		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	colorectal cancer		53	3.9E-12	3.7E-9
GENETIC_ASSOCIATION_DB_DISEASE	lung cancer		43	1.6E-9	7.6E-7
Annotation Cluster 2	Enrichment Score: 4.38		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE_CLASS	UNKNOWN		45	4.7E-6	1.8E-5
Annotation Cluster 3	Enrichment Score: 3.95		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	preterm delivery		15	5.3E-6	1.1E-3
Annotation Cluster 4	Enrichment Score: 2.26		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	hypercholesterolemia; H. pylori infection; coagulation disorder		5	1.1E-3	5.7E-2
GENETIC_ASSOCIATION_DB_DISEASE	leukemia, myeloid		9	3.0E-3	1.1E-1
GENETIC_ASSOCIATION_DB_DISEASE	leukemia, adult acute		3	4.9E-2	6.3E-1
Annotation Cluster 5	Enrichment Score: 1.72		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	measles vaccine immunity		10	8.6E-7	2.0E-4
Annotation Cluster 6	Enrichment Score: 1.72		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	hypospadias		8	9.2E-4	4.9E-2
Annotation Cluster 7	Enrichment Score: 1.67		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	methadone levels		6	1.4E-3	6.7E-2
Annotation Cluster 8	Enrichment Score: 1.61		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	hypercholesterolemia; H. pylori infection; coagulation disorder		5	1.1E-3	5.7E-2
Annotation Cluster 9	Enrichment Score: 1.56		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	osteoporosis		8	6.0E-3	1.8E-1
Annotation Cluster 10	Enrichment Score: 1.21		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	bone mass		6	5.5E-3	1.7E-1
Annotation Cluster 11	Enrichment Score: 1.07		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	liver cancer; liver disease		4	7.7E-3	2.1E-1
Annotation Cluster 12	Enrichment Score: 0.72		Count	P	Benjamini
GENETIC_ASSOCIATION_DB_DISEASE	sex hormones		4	1.2E-1	8.7E-1

بحث و نتیجه گیری

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

پزشکان علت دقیق لوسمی را نمی دانند. به نظر می رسد عوامل ژنتیکی و محیطی در به وجود آمدن این بیماری نقش داشته باشند. مانند هر سرطان دیگری، عامل مستعد و پیشتاز در ایجاد لوسمی به هم خوردن نظم تقسیم سلولی است. پژوهش های انجام شده روند بدخیمی بیماری لوسمی را به ناهنجاری های ژنتیکی نسبت می دهند. برخی بیماری های ژنتیکی مثل سندرم داون، احتمال ابتلا به لوکمی را بالا می‌برد. قرار گرفتن در معرض پرتوهای یونیزه کننده و برخی مواد شیمیایی خاص، افرادی که در معرض پرتوهای شدید قرار گیرند مانند بازماندگان بعد از حملات اتمی یا رویدادهای رآکتورهای هسته‌ای، احتمال دچار شدن به لوکمی در آن ها بیشتر است. قرار گرفتن در معرض برخی مواد شیمیایی خاص مانند بنزین نیز می‌تواند احتمال ابتلا به لوکمی را افزایش دهد. نارسایی سیستم ایمنی طبیعی بدن نیز عامل دیگر ابتلا به این بیماری است. آمادگی ابتلا به لوسمی با افزایش سن ارتباط مستقیم دارد به طوری که گفته می شود افراد بالای 55 سال باید بیشتر مراقب علایم هشداردهنده این بیماری باشند با این حال لوسمی می‌تواند کودکان را هم در هر سنی گرفتار کند ولی بیشتر در سنین سه تا چهار سال اتفاق می‌افتد.(5-1)

گالوپ و همکاران(1999) در یک مقاله،(22)، کوشیدند تا با ارائه روش های نوینی برخلاف روش های متداول و بسیار پرهزینه رایج، بر مبنای داده های بیان ژنی، این سرطان را تشخیص دهند. آن ها دلیل انتخاب لوکمیا جهت این مطالعه را توانایی بررسی دقیق نتایج حاصله با نتایج روش های موجود بیان نمودند. آن ها عقیده داشتند که در حال حاضر لوکمیا توسط روش های موجود به خوبی قابل تشخیص است بنا بر این می توان میزان صحت نتایج روش های پیشنهادی را با دقت ارزیابی نمود. گالوپ و همکاران در نهایت 50 ژن را انتخاب و برای تشخیص افراد استفاده نمودند. این روش با دقت 100 درصد افراد را تشخیص داد. به طوری که توانست از 38 نفر بیمار مورد AML یا ALL مطالعه به درستی 36 نفر را به یکی از طبقات اختصاص دهد و برای 2 نفر باقی مانده ابراز عقیده ای ننمود. این بدان معناست که چنان چه نتیجه این روش به یک تشخیص منجر شود، بدون شک نتیجه آن درست خواهد بود. آن ها دلیل انتخاب 50 ژن جهت پیش بینی را تا حدودی اختیاری دانستند و بیان داشتند که این ژن ها با یکی از طبقات AML یا ALL هم بستگی بالایی دارند.

در این مطالعه به بررسی و آنالیز 615 ژن به دست آمده از داده های میکروواری بیماری لوکمیا توسط برنامه DAVID پرداخته می شود. این برنامه در سال 2003 نوشته شد و هدف آن انجام آنالیزهای عملکردی بر روی لیست های ژنومی و پروتئومی است که داده های آن ها از مطالعات بیولوژیک با دستگاه های پیشرفته(high-throughput) حاصل می شود. توانایی جستجو و مطالعه ژن های مرتبط با هم را به صورت یک واحد در قالب شبکه های بیولوژیکی بزرگ تر نسبت به مطالعه ژن ها به صورت جداگانه برای محقق فراهم می نماید. تا سال 2011 بیشتر از 2000 عنوان مقاله ارجاع به این برنامه داشتند،(23). برنامه تجزیه و تحلیل داده DAVID دارای مزایایی است از جمله این که تعداد معقولی از ژن های مختلف از صدها تا هزاران ژن به عنوان مثال، 100-2000 ژن) را مورد آنالیز قرار می دهد که می تواند حاوی ژن های مهم(ژن نشانگر) مورد انتظار برای مطالعه باشد. انتخاب ژن از طریق مقایسه بیان ژن بین کنترل و بیمار با آزمون آماری t-test و P انجام می شود که بخش قابل توجهی از افزایش و یا کاهش بیان ژن در فرایندهای بیولوژیکی ویژه ای به جای این که به طور تصادفی در طول همه فرایندهای بیولوژیکی ممکن گسترش یافته باشد مشخص می شود. تکرارپذیری بالا از مزایای دیگر این برنامه می باشد.(21-16)

جدول شماره 1 لیست تعدادی از ژن های شناسایی شده را نشان می دهد که این نحوه بیان، جستجو و بررسی یک به یک ژن ها را میسر می سازد و سرعت جستجو و حاشیه نویسی در مورد ژن های مورد نظر را افزایش می دهد. ژن های مرتبط با انواع بیماری ها در این برنامه مورد بررسی قرار گرفته که 615 ژن شناسایی شده مطابق جدول شماره 2 با کلاس های متعددی از بیماری های انسانی نسبت داده شده است. از جمله کاربردهای این نوع جدول ها

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

و آنالیزها، شناسایی و نمایش حاشیه نویسی خوشه های به دست آمده در یک قالب جدولی، پیوندهای عمیق به اطلاعات بیشتر پیرامون ژن های مرتبط، جستجو سایر ویژگی ها و ارتباط با عملکرد ژنی و مشاهده ژن ها بر روی نقشه مسیر بیولوژیکی است. اطلاعات مربوط به ژن ها در این جدول بر اساس پایگاه داده انجمن ژنتیک(GENETIC_ASSOCIATION_DB_DISEASE_CLASS) است. این پایگاه داده بایگانی از مطالعات ژنتیکی انسانی در ارتباط با بیماری های پیچیده و اختلالات است که شامل اطلاعات خلاصه شده استخراج شده از مقالات منتشر شده در مجلات است که به بررسی بر روی ژن های نامزد بیماری ها می پردازند. بیشترین ژن های دخیل در بیماری در جدول شماره 2 دسته ژن هایی هستند که در انواع سرطان ها دخالت دارند. 23/7 درصد ژن های شناسایی شده در بیماری لوکمیا در دسته ژن های موجود در انواع سرطان ها هستند. بعد از بیماری سرطان به ترتیب بیماری های سیستم ایمنی(19/3 درصد)، پیری(4/2 درصد)، بیماری های روانی(10/7 درصد)، بیماری های نرولوژیک(10/9 درصد)، بیماری های خونی(4/2 درصد)، بیماری های قلبی-عروقی(14درصد)، بیماری های کلیوی(4/2 درصد) و بیماری های سیستم بینایی(4/3 درصد) در مرتبه بعدی ارتباط کلاس های بیماری های انسانی با لوکمیا قرار می گیرند. بنا بر این علت بیماری لوکمیا را می توان در ابتلا به هر کدام از بیماری های مذکور دانست یا می توان بیان نمود که عامل لوکمیا می تواند عاملی باشد که باعث بیان ژن های مشترک می شود. نقص سیستم ایمنی، بالا رفتن سن، اختلالات روانی و مشکلات سیستم عصبی می تواند از علل اصلی ابتلا به لوکمیا باشد.

از بیماری های اصلی مرتبط با لوکمیا کلاس ژن های بیماری سرطان بیشترین سهم ژنی مشترک را داشت که از 146 ژنی که در دسته ژن های مشترک دخیل در سرطان هستند تعدادی از آن ها در جدول شماره 3 نشان داده شده است که شامل گروه های مختلفی است.

از 615 ژن شناسایی شده 70 چارت مسیرهای بیولوژیکی(موجود در پایگاه داده KEGG) مرتبط با بیماری شناسایی شد که نتایج آن در جدول شماره 4 آمده است. پایگاه داده مسیر KEGG مجموعه ای از نمودارهای گرافیکی دستی کشیده شده، به نام نقشه مسیرKEGG است که بیانگر مسیرهای مولکولی برای سوخت و ساز بدن، پردازش اطلاعات ژنتیکی، پردازش اطلاعات زیست محیطی، سایر فرایندهای سلولی، بیماری های انسان، و توسعه دارو است. مسیرهای بیولوژیک شناسایی شده بر اساس پایگاه داده مسیر KEGG از ژن های لوکمیا با توجه به P به ترتیب شامل مسیرهای دخیل در سرطان، اینترکشن رسپتور سایتوکاین با سایتوکاین، مسیر سیگنالینگ Jak-STAT،MAPK و نوتروفین، گلیوما، ملانوما، سرطان ریه و پروستات است. نتایج نشان می دهد مسیرهای بیولوژیک نیز دارای تطابقی با کلاس های بیماری های مرتبط با لوکمیا هستند و داده های جدول شماره 2 و 4 با یکدیگر مطابقت دارند. پایگاه داده DAVID مسیرهای بیولوژیک را به صورت گرافیکی برای برخی از بیماری ها ارائه می نماید. همان طور که در شکل شماره 1 مشاهده می شود مسیر بیولوژیکی دخیل در سرطان را نشان می دهد که بر اساس مسیرهای بیولوژیک موجود در داده پایگاه KEGG. طراحی شده است. انواع ژن های دخیل در سرطان را نشان می دهد که ارتباط بین ژن ها نشان دهنده اینترکشن بین ژن ها از نظر فیزیکی یا فقط ارتباط عملکردی می باشد.

اساس گروه بندی بر پایه عملکرد مشابه با کمک تکنیک کلاسترینگ ژن هاست که در هر گروه ژن ها دارای عملکرد یکسان هستند. از 615 ژن شناسایی شده 12 کلاستر بر اساس حاشیه نویسی عملکردی مرتبط با بیماری شناسایی شد که نتایج آن در جدول شماره 5 آمده است. خوشه بندی بر اساس حاشیه نویسی عملکردی به ترتیب با امتیاز خوشه سازی هر گروه به نمایش در آمده است. امتیاز بیانگر رتبه ای است که معنی داری بیولوژیکی گروه های ژنی بر اساس p(EASE score) کلی از تمام شرایط حاشیه نویسی خوشه ها را محاسبه می کند. هر گروه بیان کننده عملکرد خاصی است و اعضای هر گروه بیشترین نزدیکی عملکرد بیولوژیکی

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

را دارند. در خوشه 1 سرطان کلورکتال دارای بیشترین معنی داری است به همین ترتیب در خوشه 2، 45 ژن با عملکرد ناشناخته قرار می گیرند که در صدر خوشه هستند. در خوشه 3، 15 ژن مربوط به زایمان زودرس دارای رتبه معنی داری بالاتری است. بیماری لوکمیا در خوشه شماره 4 قرار دارد. بنا بر این مطالعات پیشین ارتباط بین ژن های مختلفی را با لوکمیا بیان داشته اند. پروتئینی که از اتصال به هم قطعاتی از دو ژن ساخته می‌شود در اکثر بیمارانی که دچار لوسمی میلوژنز مزمن(CML) هستند، و در برخی از بیمارانی که به لوسمی لنفوبلاستیک حاد یا به لوسمی‌میلوژنز حاد دچار می باشند، این پروتئین یافت می شود. در داخل سلول های لوسمی، ژن ABL مربوط به کروموزوم 9 به ژن BCR متعلق به کروموزوم 22 متصل می شود تا ژن ترکیب شده BCR-ABL به وجود آید، که این ژن پروتئین ترکیب شده BCR-ABL را تولید می کند،(24). لوکوس INK4A/ARF هم چنین به نام CDKN2 نیز شناخته می شود بر روی کروموزوم q219 قرار دارد که دارای دو پروتئین به نام های P16 INK4A و P14 ARF می باشد. این لکوس ها در طیف وسیعی از تومورها شامل ملانوما، آدنوکارسینومای پانکراس، گلیوما، سرطان ریه، سر و گردن، پستان، مغز، تخمدان، لوسمی های خاص و کارسینومای مثانه حذف می شوند،(25). P16 INK4A یک پروتئین 16 کیلو دالتونی است و یک مهارکننده چرخه سلولی می باشد که عمل خود را از طریق کنترل مسیر رتینوبلاستوما(Rb) انجام می دهد؛ به نحوی که باعث جلوگیری از فسفوریلاسیون و غیرفعال شدن پروتئین Rb از طریق مهار CDK4/CDK6 می شود؛ که در نتیجه باعث توقف پیشروی از فازG1 به S می شود،(26). P14 ARF یک پروتئین 14 KD می باشد سرکوب کننده رشد سلولی از طریق چندین مسیر وابسته و غیر وابسته به p53 می باشد،(27). به طور مستقیم به MDM2 متصل می شود که هم p53 و هم MDM2 را پایدار می کند، که این امر توقف چرخه سلولی در هر دو فاز G1 و G2/M را القا می کند،(28). در حفظ هموستاز سلولی و جلوگیری از فرایند انکوژنیک نقش حیاتی دارند. ژن این پروتئین ها در اکثر سرطان های انسانی در اثر حذف ژنی، جهش نقطه ای یا متیلاسیون پروموتور، غیرفعال می شود.(29)

نتایج تحقیقاتی که تاکنون صورت گرفته است پروتئین های واحدی را جهت تشخیص و درمان معرفی نمودند و داروهای شیمی درمانی نیز بر همین اساس طراحی شده و مورد استفاده کلینیکی قرار می گیرد اما هم چنان علت اصلی این سرطان که همراه با تغییرات وسیعی در سطح ملکولی است به درستی شناخته نشده است. آنالیز ژنوم سرطان لوکمیا با برنامه DAVID منجر به شناسایی و تشخیص کلاس های مهم ژنی و مسیرهای بیولوژیکی اصلی درگیر در بیماری در بین انبوهی از ژن های کشف شده شد و بدین طریق مارکرهای ژنی کاندیدا برای کشف علت بیماری و هدف های درمانی لوکمیا را می توان معرفی نمود.

سپاسگزاری

این مطالعه حاصل طرح تحقیقاتی با کد 1391-1-159-8975 است که توسط کمیته پژوهشی دانشجویان دانشگاه علوم پزشکی شهید بهشتی تصویب گردیده است. از مرکز تحقیقات پروتئومیکس دانشگاه علوم پزشکی شهید بهشتی که در اجرای پروژه همکاری داشتند تشکر می گردد.

References

1-Leukemia, in Mosby’s Medical, Nursing & Allied Health Dictionary. 4th ed. Mosby-Year Book, Inc; 1994.P.903.

2-Matutes E. T-cell prolymphocytic leukemia, a rare variant of mature post-thymic T-cell leukemias, has distinct clinical and

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

laboratory characteristics and a poor prognosis. Cancer Control J 1998;5:141-8.
3-Valbuena JR, Herling M, Admirand JH, Padula A, Jones D, Medeiros LJ. T-cell prolymphocytic leukemia involving extramedullary sites. Am J Clin Pathol 2005; 123:456-64.
4-Ross JA, Kasum CM, Davies SM, Jacobs DR, Folsom AR, Potter JD. Diet and risk of leukemia in the Iowa Women's Health Study. Cancer Epidemiol Biomarkers Prev 2002;11:777-81.
5-Kinzler KW, Vogelstein B. Genetic basis of human cancer. McGraw-Hill; 2002.
6-Wiernik PH. Adult leukemias. New York: B. C. Decker; 2001.P.3-15.
7-Robinette MS, Cotter S, Van de W. Quick look series in veterinary medicine: Hematology; Teton New Media 2001.P. 105.
8-Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 1996;93:10614-9.
9-Baggerly KA, Coombes KR, Hess KR, Stivers DN, Abruzzo LV, Zhang W. Identifying differentially expressed genes in cDNA microarray experiments. J Comput Biol 2001;8:639-59.
10-Satagopan JM, Panageas KS. Tutorial in biostatistics, a tatistical perspective on gene expression data analysis. Stat Med 2003; 22:481-99.
11-Kong SW, Pu WT, Park PJ. A multivariate approach for integrating genome-wide expression data and biological knowledge. Bioinformatics 2006;22:2373-780.
12-Liu Q, Dinu I, Adewale AJ, Potter JD, Yasui Y. Comparative evaluation of gene-set analysis methods. BMC Bioinformatics 2007;8:431-9.
13-Goeman JJ, Bühlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics 2007;23:980-7.
14-Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 2004;5:478-81.
15-Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:3.
16-Da WH, Brad TS, Richard AL. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 2009;1:47-53.
17-Sherman BT. DAVID knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinform 2007;8:426-9.
18-Dennis GT. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:3-8.
19-Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE Genome Biol 2003;4:R70.
20-Huang da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 2007;35:W169-75.
21-Huang da W, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 2007;8:R183.
22-Golub TR, Slonim DK, Tamayo P, Gaasenbeek CHM, Mesirov JP, Coller H, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531-7.
23-Xiaoli J, Brad T, Sherman D, Wei H, Robert S, Michael W, et al. DAVID-WS: a stateful web service to facilitate gene/ protein list analysis. Bioinformatics 2012;28:1805-6.
24-Franca P. Long-term remission in BCR/ABL-positive AML-M6 patient treated with Imatinib Mesylate. Leuk Res 2007;31:563-7.
25-Kim WY, Sharpless NE. The regulation of INK4/ARF in cancer and aging. Cell 2006;127:265-75.
26-Ruas M, Peters G. The p16INK4a/ CDKN2A tumor suppressor and its relatives. Biochim Biophys Acta 1998; 1378: F115-77.
27-Cividin M, Ayrault O, Sorel N, Séité P, Brizard F, Blanchet O, et al. Expression of the cell cycle regulators p14ARF and p16INK4a in chronic myeloid leukemia. Leuk Res 2006;30:1273-8.

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]

28-Lee YK, Park JY, Kang HJ, Cho HC. Overexpression of p16INK4A and p14ARF in haematological malignancies. Clin Lab Haematol 2003;25:233-7.
29-Hernández-Boluda JC, Cervantes F, Colomer D, Vela MC, Costa D, Paz MF. Genomic p16 abnormalities in the progression of chronic myeloid leukemia into blast crisis: A sequential study in 42 patients. Exp Hematol 2003;31:204-10.

Gene Expression Analysis of Leukemia Microarray Data By DAVID Program

Zali H[*1], *Amini R*[2], *Shiri haris R*[3]

(Received: 22 Dec. 2012 Accepted: 6 Aprl. 2013)

Abstract

Introduction: Leukemia is a progressive and malignant disease of hematopoietic organs of the body. Genetic abnormalities play an important role in the development of leukemia in the body. Many studies have been accomplished on the molecular factors that involved in the disease. DNA microarray technology provides a general picture of gene expression in whole genome and applied for the exploring of candidate genes that lead to diseases. In fact, having analyzed a large number of genes together along with the expected changes provides more closely examination the disease under study. In the present study, the DNA microarray data of leukemia disease were analyzed by bioinformatics software (DAVID). The aim of the study was to functionally analyze the genomic and proteomic lists of data that have been obtained with high-throughput tools during biological studies.

Materials & Methods: Microarray leukemia gene sets were obtained from the database http://www.biomedcentral.com/ content/supplementary/1471 and analyzed with the bioinformatics software (DAVID). The communication gene expression in different classes, chart and clustered genes were examined. The list of genes was identified by the DAVID analysis program.

Findings: A chart consisting of 615 identified genes associated with various diseases was detected. Most genes involved in the disease were those genes that were also involved in cancers. 23.7% of the identified genes (146 genes) were cancer genes. Of 615 genes, 70 charts of the identified biological pathways (the database KEGG) were associated with the disease. Of 615 genes identified, 12 clusters were associated with the disease based on the functional annotation.

Discussion & Conclusion: The results showed that the program, DAVID, is capable of analyzing genome. Also, the program was capable to evaluate the main classes of genes and pathways involved in the disease to determine the best candidate gene markers for the diagnosis and treatment of leukemia disease.

Keywords: microarray, leukemia, program DAVID, gene expression

1. Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Dept of Microbiology, Faculty of Sciences, Hamadan Islamic Azad University, Hamadan, Iran
3. Dept of Laboratory, Sciences, Faculty of Allied Medical Sciences, Ilam University of Medical Sciences, Ilam, Iran
** (corresponding author)*

[Downloaded from sjimu.medilam.ac.ir on 2026-11-08]