هو الخالق

آشنایی با بیوانفورماتیک
و کـــــاربردهـــــای آن

سرشناسه : جان‌نثار، معصومه، ۱۳۶۳ -
عنوان و نام پدیدآور: آشنایی با بیوانفورماتیک و کاربردهای آن/ معصومه جان‌نثار، مریم معظم‌جزی، سیدمهدی سیدی؛ [برای] ریاست جمهوری معاونت علمی و فناوری ستاد توسعه زیست فناوری.
مشخصات نشر: تهران: ریاست جمهوری، معاونت علمی و فناوری، مرکز روابط عمومی و اطلاع‌رسانی، دانش بنیان فناور، ۱۳۹۶.
مشخصات ظاهری : ۱۳۰ ص.: مصور(رنگی).
شابک : 978-600-6844-48-0
وضعیت فهرست نویسی : فیپا
یادداشت : واژه‌نامه.
یادداشت : کتابنامه: ص. ۱۲۶.
موضوع : زیست‌انفورماتیک
موضوع : Bioinformatics
موضوع : زیست‌شناسی کامپیوتری
موضوع : Computational biology
شناسه افزوده : معظم جزی، مریم، ۱۳۶۲ -
شناسه افزوده : سیدی، سیدمهدی، ۱۳۴۳ -
شناسه افزوده : ایران. ریاست جمهوری. ستاد توسعه زیست فناوری
شناسه افزوده : ایران. ریاست جمهوری. معاونت علمی و فناوری. انتشارات دانش بنیان فناور
رده بندی کنگره : ۱۳۹۶ ۲آ۵ج/۲/QH۳۲۴
رده بندی دیویی : ۵۷۲/۸۰۲۸۵
شماره کتابشناسی ملی : ۴۸۳۹۸۷۲

طراح جلد: رقیه معارف وند
طراح گرافیک: رابعه فردوسی

پیشگفتار

پیشرفت های چند دهه ی اخیر درحوزه های مولکولی و زیست فناوری بسیار قابل توجه بوده است. نقطه عطف این جهش های بزرگ در سال ۲۰۰۳ میلادی با به اتمام رسیدن پروژه ی ژنوم انسان بود. پس از دستیابی به توالی DNA انسان، راه برای تعیین توالی سایر موجودات از میکروارگانیسم ها تا گیاهان و جانوران یکی پس از دیگری باز شد که حاصل آن تولید حجم بالایی از داده های زیستی بود. در این زمان ضرورت حضور دانش جدید و متفاوتی در کنار زیست شناسی برای ذخیره سازی و آنالیز این داده‌ها بوجود آمد. با بوجود آمدن چنین نیازی، و همزمان شدن آن با رشد رو به افزایش حجم داده های زیستی، دانش بیوانفورماتیک پا به عرصه دانش بشری نهاد. این علم با بهره گیری هم زمان از علوم کامپیوتر، ریاضیات و آمار به بررسی و تفسیر داده‌های زیست شناختی می پردازد.

تمرکز و سرمایه گذاری بر حوزه های بین رشته ای می تواند از طریق ارتباط موثر و کارآمدی که بین علوم مختلف برقرار می کند، راه را برای نتیجه گیری دقیق تر و گذر از دانش محض به سمت کاربردی، هموار ساخته و در نتیجه موجبات تسریع پیشرفت علم را در شاخه های مختلف فراهم آورد. لذا بیوانفورماتیک علمی بین رشته ای است که امروزه بعنوان یک دانش راهبردی توجه زیادی را به خود جلب کرده است. با توجه به آنچه گفته شد، آموزش، تربیت نیروی انسانی متخصص و آگاهی پایه از این حوزه می تواند در نیل به اهداف راهبردی در حوزه های زیست شناسی و زیست فناوری راهگشا باشد. در جمهوری اسلامی ایران بر اساس نقشه جامع علمی کشور، زیست فناوری بعنوان یکی از حوزه های اقتدار آفرین برای کشور لحاظ شده و در زمره

اولویت های اصلی (اولویت الف در نقشه جامع علمی کشور) قرار گرفته است، بنابراین توسعه ی آن به میزان زیادی مورد توجه قرار گرفته که از آن جمله می توان به ستاد توسعه زیست فناوری اشاره کرد. لذا، بیوانفورماتیک به عنوان یکی از ابزارها و زمینه های جدید و بسیار کارآمد مرتبط با علم زیست فناوری مورد توجه است.

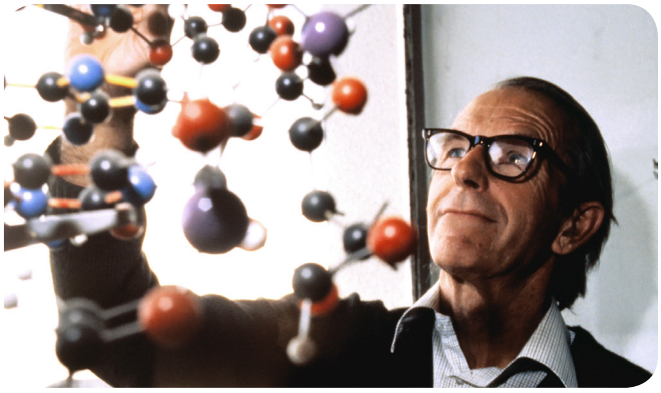
ستاد توسعه‌ی زیست فناوری بعنوان نقطه‌ی کانونی تمام فعالیت های زیست فناوری در کشور، مسئول سیاست گذاری و برنامه ریزی کلان فعالیت های زیست فناوری است. گروه آموزش، منابع انسانی و ترویج این ستاد بدنبال ایجاد زمینه هایی است تا با آموزش و ترویج زمینه های جدید، بویژه برای دبیران و علاقمندان، در راه توسعه زیست فناوری در کشور گام بردارد.

کتاب حاضر علاوه بر فراهم آوردن زمینه ی آشنایی با علم بیوانفورماتیک، کاربردهای آن را نیز در حوزه های مختلف مورد بررسی قرار داده است. این کتاب در زمینه ی آموزش کاربردی برخی از مهمترین پایگاه های داده های ژنتیکی و پروتئینی و همچنین نرم افزارهای مورد نیاز برای ویرایش و هم ردیف سازی توالی و نهایتاً طراحی پرایمر تهیه شده است. با توجه به کمبود جدی منابع فارسی که بتواند مفاهیمی پایه ای و کاربردی در حوزه ی بیوانفورماتیک را به زبانی ساده و قابل استفاده برای دایره ی وسیع تری از مخاطبین بیان نماید، نویسندگان این کتاب سعی نموده اند در ضمن حفظ زبان ساده، از غنای محتوای کتاب نکاهند.

امید که این کتاب بتواند زمینه آشنایی بیشتر کلیه علاقمندان بویژه دبیران عزیز کشور را با این حوزه از علم فراهم نماید.

مقدمـــه

اولین توالی آمینواسیدی شناخته شده، توالی هورمون انسولین انسانی با ۱۱۰ اسید آمینه بود که در سال ۱۹۵۵ توسط فردریک سنگر[۱] شناسایی شد (شکل ۱). حدود یک دهه ی بعد، اولین توالی نوکلئوتیدی که مربوط به tRNA آلانین مخمر با ۷۷ جفت باز بود، گزارش شد. در سال های بعد با پیشرفت هایی که در حوزه ی علوم ژنتیك و پروتئومیک (دانش بررسی ساختار و عملکرد پروتئین ها در مقیاس وسیع) حاصل شد، بتدریج مقدار اطلاعات و داده های زیستی افزایش یافت. داده های زیستی در آن زمان عمدتا از توالی های نوکلئوتیدی مربوط به مولکول های RNA و DNA و توالی های آمینو اسیدی متعلق به مولکول های پروتئینی تشکیل شده بود.



شکل ۱. فردریک سنگر بیوشیمی دان انگلیسی برنده دو دوره ی جایزه ی نوبل در شیمی

در چند دهه ی اخیر، تجهیزات مورد نیاز در تحقیقات مولکولی بطور گسترده ای پیشرفت های قابل توجهی یافت بطوری که امروزه تحقیقات مولکولی جزء مطالعات رایج آزمایشگاه های زیستی است. در حال حاضر، پروژه های تعیین توالی ژنوم موجودات مختلف از پروژه های بسیار رایج در سطح جهان به حساب می آیند. امروزه توالی ژنوم بسیاری از موجودات ساده مانند باکتری ها تا موجودات بسیار پیشرفته مانند یوکاریوت های بسیار تکامل یافته شناسایی شده است. پروژه ی شناسایی ژنوم انسان که از سال ۱۹۹۰ آغاز شد در بهار ۲۰۰۳ پایان یافت و به این ترتیب اطلاعات کامل مربوط به توالی کروموزوم های انسانی بدست آمد. این حجم وسیع داده های زیستی به دست آمده برای دانشمندان بسیار با ارزش بود، ولی استفاده از این حجم وسیع اطلاعات پردازش نشده، نمی توانست چندان مفید باشد. در این زمان با پیشرفت چشم گیر تکنولوژی اطلاعات و کاربرد آن در زمینه های مختلف، به نظر رسید که ادغام دو علم زیست شناسی و کامپیوتر می تواند در حل مشکل ذخیره و آنالیز این حجم گسترده از داده های تولید شده راهگشا باشد.

---

۱ Frederick Sanger

به این ترتیب، حدود اوایل سال ۱۹۷۵ بود که رشته ی بیوانفورماتیك با هدف استفاده ازتکنیك های مدیریت سیستم های داده از طریق ایجاد پایگاه داده و همچنین آنالیز داده ها با استفاده از نرم افزارهای مختلف در مطالعات زیست شناسی متولد شد. با پیشرفت بیوانفورماتیك دخالت سایر رشته ها در پیشبرد کار، امری اجتناب ناپذیر بود. حجم بالای اطلاعات و نیاز به پردازش آنها وجود کامپیوترهای پیشرفته تر را می طلبید. تحلیل داده ها و نتیجه گیری منطقی از آنها، دخالت موثر علم آمار را در این رشته رقم زد. به این ترتیب علم بیوانفورماتیك بعنوان یك تخصص میان رشته ای با ادغام زیست شناسی، ریاضیات (به ویژه آمار)، علوم کامپیوتر و فناوری اطلاعات ایجاد شد. در این علم با استفاده از سخت افزار کامپیوتر، نرم افزارهای کامپیوتری، و بانك های اطلاعاتی مسائل زیست شناختی به ویژه در زمینه های سلولی و مولکولی و همچنین تجزیه و تحلیل اطلاعات توالی، عملکرد و ساختار ملکول های پروتئین، DNA وRNA امکان پذیر شده است (شکل ۲).

شکل ۲- مفهوم بیوانفورماتیک

فهرست

پیشگفتار

مقدمه

فصل اول: **آشنایی با بیوانفورماتیک**

۱. سیر تاریخی علم بیوانفورماتیک ۱۰
۲. مروری بر اسیدهای نوکلئیک و پروتئین ۱۱
۳. قالب خواندن ۱۴
۴. توالی یابیDNA ۱۵
۴-۱. تعیین توالی به روش سنگر ۱۶
۴-۲. تعیین توالی به روش ماکسام-گیلبرت ۱۷
۴-۳. روش های جدید توالی یابی ۱۷
۴-۴. کاربردهای روش های جدید توالی یابی ۱۹
۴-۴-۱. توالی یابی مجدد ژنوم های شناخته شده ۱۹
۴-۴-۲. توالی یابی مناطق هدف ۱۹
۴-۴-۳. تعیین توالی از نو ۱۹
۴-۴-۴. بررسی تغییرات اپی ژنتیک ۲۰
۴-۴-۵. توالی یابی اسیدهای نوکلئیک آزاد موجود در خون ۲۰

فصل دوم: **کاربردهای بیوانفورماتیک**

۱. مقدمه ۲۲
۲. کاربرد بیوانفورماتیک در زیست شناسی ۲۲
۲-۱. تجزیه و تحلیل اطلاعات توالی ۲۳
۲-۲. تحلیل عملکرد ژنوم ۲۳
۲-۳. مشخص کردن کلیه پروتئین های موجود در سلول ۲۳
۲-۴. نقشه برداری برهمکنش های بین پروتئینی ۲۴
۲-۵. نقشه برداری آرایش های پروتئینی ۲۴
۲-۶. پیش بینی ساختار سه بعدی پروتئین ۲۴
۲-۷. مطالعه ی متابولوم ۲۴

۳. کاربرد بیوانفورماتیک در کشاورزی ۲۵
۳-۱. بهبود محصول ۲۶
۳-۲. بهبود کیفیت غذایی ۲۶
۴. کاربرد بیوانفورماتیک در مطالعه ی ژنوم میکروبی ۲۶
۴-۱. پاکسازی زباله ۲۷
۴-۲. تغییرات آب و هوا ۲۷
۴-۳. صنایع لبنی و پروبیوتیک ها ۲۸
۴-۴. انرژی جایگزین ۲۹
۵. کاربرد بیوانفورماتیک در پزشکی ۲۹
۵-۱. کشف دارو ۲۹
۵-۲. پزشکی فردی ۳۰
۵-۳. ژن درمانی ۳۲

فصل سوم: **پایگاه داده**

۱. مقدمه ۳۴
۲. پایگاه داده GenBank ۳۶
۲-۲. انواع داده های موجود در GenBank ۳۷
۲-۲-۱. اطلاعات مربوط به DNA ژنومی ۳۷
۲-۲-۲. اطلاعات مربوط به STS ۳۷
۲-۲-۳. اطلاعات مربوط به GSS ۳۷
۲-۲-۴. اطلاعات مربوط به EST ۳۸
۳. معرفی صفحه اصلیNCBI ۳۸
۴. آشنایی با پایگاه داده PubMed ۴۱
۴-۱. ذخیره نتایج حاصل از جستجو در پایگاه اطلاعاتیPubMed ۴۴
۵. آشنایی با پایگاه داده Bookshelf ۴۵
۶. آشنایی با پایگاه اطلاعاتیMeSH ۴۷
۷. آشنایی با پایگاه اطلاعاتی OMIM (Online Mendelian Inheritance in Man) ۴۸
۸. بازیابی اطلاعات مربوط به یک ژن معین در پایگاه داده NCBI ۵۰
۹. آشنایی با پایگاه داده RefSeq ۶۱
۱۰. آشنایی با پایگاه داده UniProt ۶۵
۱۱. یافتن توالی های پروتئینی یا نوکلئوتیدی مشابه با توالی مورد نظر با استفاده از برنامه BLAST ۷۴
۱۲. اجرا BLASTP در NCBI ۷۶

فصل چهارم: اصول آماده سازی، ویرایش و هم ردیف سازی توالی

۴. آماده سازی توالی ۸۴
۴. ویرایش و هم ردیف سازی توالی ۸۷
۴. آموزش کاربردی ویرایش توالی با نرم افزار Chromas 2.6.2 ۸۷
۴. آموزش کاربردی ویرایش و هم ردیف سازی توالی با نرم افزار BioEdit 7.2.5 ۹۵
۴-۱. ویرایش توالی با نرم افزار BioEdit 7.2.5 ۹۵
۴-۲. هم ردیف سازی توالی ها با نرم افزار BioEdit 7.2.5 ۹۸
۵. آموزش کاربردی هم ردیف سازی با نرم افزار MegAlign ۱۰۲

فصل پنجم: اصول طراحی پرایمر

۱. مقدمه ای بر واکنش زنجیره ایPCR ۱۰۹
۱-۱. ارزیابی محصولات PCR ۱۱۲
۱-۲. برخی از کاربرد های مهم PCR ۱۱۳
۲. پرایمر ۱۱۴
۳. فاکتورهـــای مهم در طراحی پرایمر ۱۱۵
۳-۱. طول پرایمر ۱۱۵
۳-۲. دمای ذوب (Tm) ۱۱۵
۳-۳. محتوایGC ۱۱۶
۳-۴. پایداری نوکلئوتیدهای انتهای ´۳ ۱۱۶
۳-۵. اجتناب از توالی های تکراری ۱۱۶
۳-۶. اجتناب از Cross homology ۱۱۶
۳-۷. اجتناب از تشکیل ساختارهای ثانوی ۱۱۶
۳-۷-۱. پرایمردایمر ۱۱۶
۳-۷-۲. ساختار سنجاق سری ۱۱۷
۴. نرم افزارهای طراحی پرایمر ۱۱۷
۵. آموزش کاربردی طراحی پرایمر با نرم افزار OLIGO 7 ۱۱۸
منابع ۱۲۶
واژه نامه ۱۲۸

فصل اول

آشنایی با بیوانفورماتیک

۱ سیر تاریخی علم بیو انفورماتیک

اولین تلاش ها در زمینه ی بیوانفورماتیك به دهه ی ۱۹۶۰ برمی گردد، اگر چه در آن زمان لغت بیوانفورماتیك وجود نداشت. مارگارت دایهوف[۱] برای اولین بار در سال ۱۹۶۵ نخستین پروژه ی بیوانفورماتیك را انجام داد (شکل ۱). در این پروژه همه ی توالی های پروتئینی در دسترس تا آن زمان جمع آوری و برای ساخت یک برنامه ی کامپیوتری در جهت ذخیره و ارائه این داده ها استفاده شد. در واقع دایهوف در این پروژه اولین پایگاه داده که مربوط به توالی های پروتئینی بود با نام اطلس توالی و ساختمان پروتئین را ایجاد کرد.

شکل۱. مارگارت دایهوف شیمی فیزیک دان پیشگام در علم بیوانفورماتیک

اولین الگوریتم همترازی در سال ۱۹۷۰ توسط نیدلمن[۲] و وونچ[۳] ایجاد شد که مرحله ای بنیادی درتوسعه ی حوزه بیوانفورماتیك بود و راه را برای مقایسه توالی های پروتئینی و جستجو در پایگاه های داده، هموار نمود. در دهه ۱۹۸۰ بانك ژن (Gene Bank)، بعنوان اولین بانک توالیDNA ساخته شد. واژه ی Bioinformatic که از ادغام دو کلمه ی Bio مخفف کلمه ی Biology به معنی زیست شناسی و Informatic به معنی علوم کامپیوتر، ساخته شده است برای اولین بار در دهه ی ۱۹۹۰ به کار رفت. در آغاز، این عبارت برای فرآیندهای وابسته به آنالیز داده های مربوط به توالی های RNA، DNA و پروتئین کاربرد داشت ولی امروزه واژه ی بیوانفورماتیك مفهوم گسترده تری پیدا کرده و برای فرآیندهای آنالیزی انواع بسیار زیادی از سایر داده های زیست شناسی نیز به کار می رود.

اولین دانش آموختگان فارغ التحصیل آموزش دیده در حوزه ی بیوانفورماتیک، مدرک تحصیلی در رشته های زیست شناسی یا علوم کامپیوتر داشتند. در هم آمیختن کامل این دو رشته ی علمی زمان زیادی به طول انجامید و نهایتاً این تلفیق موجب ایجاد رشته ای بنام بیوانفورماتیك در مقطع کارشناسی گردید. امروزه مراکز دانشگاهی زیادی در دنیا رشته ی بیوانفورماتیك را در گرایش های مختلف و تا سطح دکتری ارائه می کنند. با توجه به بین رشته ای بودن بیوانفورماتیك، گروه های آموزشی مختلف از قبیل علوم کامپیوتر، بیوشیمی، زیست شناسی مولکولی، میکروبیولوژی و مهندسی پزشکی می توانند چنین رشته ای را راه اندازی کنند. در ایران، در سال ۱۳۸۱ گروهی از اساتید رشته های علوم

۱ Margaret Dayhoff

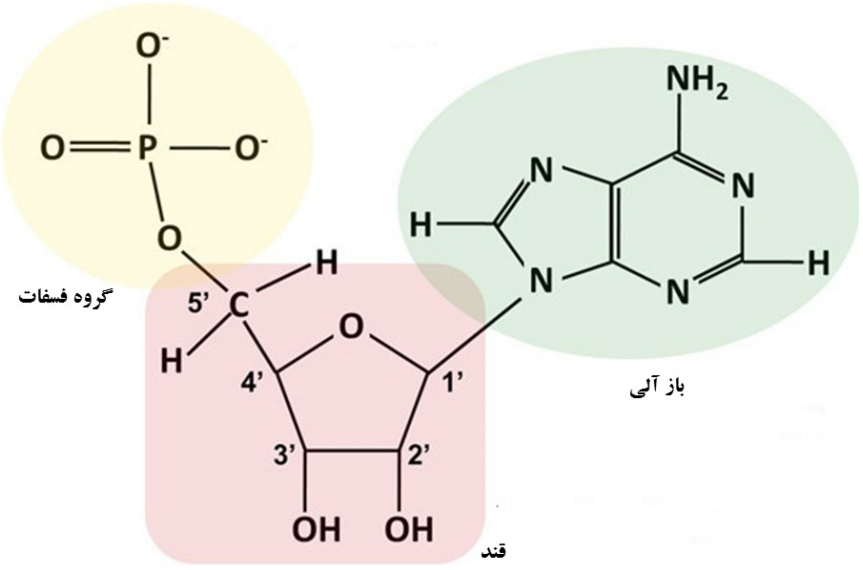
۲ Needleman

۳ Wunsch

کامپیوتر، ریاضیات، بیوفیزیك، ژنتیك، آمار و داروسازی این رشته را بنا نهادند و در سال ۱۳۸۳ اولین دوره ی دکتری بیوانفورماتیك پایه گذاری شد.

۲ مروری بر اسیدهای نوکلئیک و پروتئین

از آنجایی که اطلاعات مربوط به اسیدهای نوکلئیک و پروتئین ها مواد خام علم بیوانفورماتیك محسوب می شوند، در این بخش برخی از مفاهیم اصلی مربوط به این ماکرومولکول ها به اختصار بیان می شود. در تمام موجودات زنده از میکروارگانیسم های تک سلولی گرفته تا جانداران پیشرفته پر سلولی، اطلاعات ژنتیکی در ژنوم که از جنس DNA (یا RNA در برخی ویروس ها) است، ذخیره شده است و از نسلی به نسل دیگر منتقل می شود. واحد ساختاری اسید های نوکلئیک، نوکلئوتید ها هستند که شامل یک باز (پورینی یا پیریمیدینی) متصل به قند دئوکسی ریبوز در DNA یا ریبوز در RNA همراه با گروه فسفات متصل به قند می باشند (شکل ۲).



شکل۲. نوکلئوتیدها، واحد سازنده اسید های نوکلئیک

قرار گیری نوکلئوتید ها به دنبال هم در یک مولکول DNA یا RNA یک توالی نوکلئوتیدی را تشکیل می دهد. تمام نوکلئوتیدها در یک توالی نوکلئوتیدی دارای جهت مشخص و یکسان می باشند، به این ترتیب که نوکلئوتید انتهایی در یک سمت توالی دارای یگ گروه ′۵ آزاد و در سمت دیگر واجد یک گروه ′۳ آزاد است. بنابراین توالی نوکلئوتیدی دارای جهت بوده که این جهت را به صورت ′۳-′۵ نشان می دهند. از آنجایی که DNA دارای یک ساختار مارپیچی دو رشته است و این دو رشته مکمل یکدیگر هستند، در صورتی که جهت یک رشته ′۳→′۵ باشد، رشته دیگر در جهت ′۵→′۳خواهد بود (شکل ۳). در متون علمی، پایگاه های اطلاعاتی و اغلب نرم افزارها بطور قراردادی، توالی DNA همواره به صورت ′۳→′۵ نوشته می شود.

شکل ۳. ساختار دو رشته ای مکمل DNA

فردیک میشر[۴] برای اولین بار در سال ۱۸۶۰ در طی مطالعه بر روی گلبول های سفید خون و استخراج هسته این سلول ها دریافت که افزودن محلول قلیایی به هسته این سلول ها موجب تولید رسوب لزجی می شود که آن را نوکلئین نامید. این ماده ترکیبی از کربن، هیدروژن، اکسیژن، نیتروژن و فسفر است، و زمانی که ماهیت اسیدی این ماده مشخص شد، نام آن به اسید دئواکسی ریبونوکلئیک تغییر کرد. حدود یک دهه بعد اروین چارگاف[۵] و همکارانش با بررسی های بیوشیمیایی دریافتند که زیر واحد اصلی DNA ، نوکلئوتید می باشد که از سه بخش اصلی یعنی قند ۵ کربنه (پنتوز)، یک گروه ´۵ فسفات و یکی از چهار باز آلی تشکیل شده است. چارگاف و همکارانش به کمک کروماتوگرافی کاغذی، نسبت بازهای آلی DNA را در موجودات مختلف و در سلول های پیکری یک جاندار تعیین کردند و دریافتند که مقدار بازها در DNA گونه های مختلف جانداران متفاوت است و با تغییر رژیم غذایی، تغییر شرایط محیطی یا افزایش سن جاندار، تغییر نمی کند. اما در تمام نمونه ها، میزان آدنین با میزان تیمین و میزان سیتورین با میزان گوانین برابر است. در ادامه این پیشرفت ها، واتسون و کریک[۶] در دهه پنجاه میلادی تلاش کردند تا با استفاده از مدل های فیزیکی، ساختارهای احتمالی ممکن برای DNA را تعیین کنند، در همان زمان گروه دیگری متشکل از ویلکینز و فرانکلین[۷] نیز با استفاده از تصاویر حاصل از پراش اشعه ایکس از مولکول DNA، ساختار این مولکول را مشخص کردند. سرانجام تلاش دانشمندان در سال ۱۹۶۲ به ثمر نشست و واتسون و کریک و ویلکینز بطورمشترک جایزه نوبل دریافت کردند. بر اساس مدل پیشنهادی واتسون و کریک، DNA یک مارپیچ دو رشته ای است که رشته های آن به دور یک محور مرکزی، معمولاً به صورت راست گرد پیچ می خورند. طبق این مدل، ستون های قند - فسفات همانند نرده های پلکان به دو قسمت خارجی بازهای آلی پیچیده و به این ترتیب در معرض محیط آبی داخل سلول هستند و بازهای آلی که خاصیت آبگریزی دارند، همانند پله های نردبان و در داخل مارپیچ قرار می گیرند. بسیاری کشف DNA را با قدم گذاشتن روی ماه مقایسه کرده اند، قدمی کوتاه برای انسان و جهشی بزرگ برای بشریت، تحولی که هنوز در مرحله نوزادی است و راه زیادی در پیش دارد.

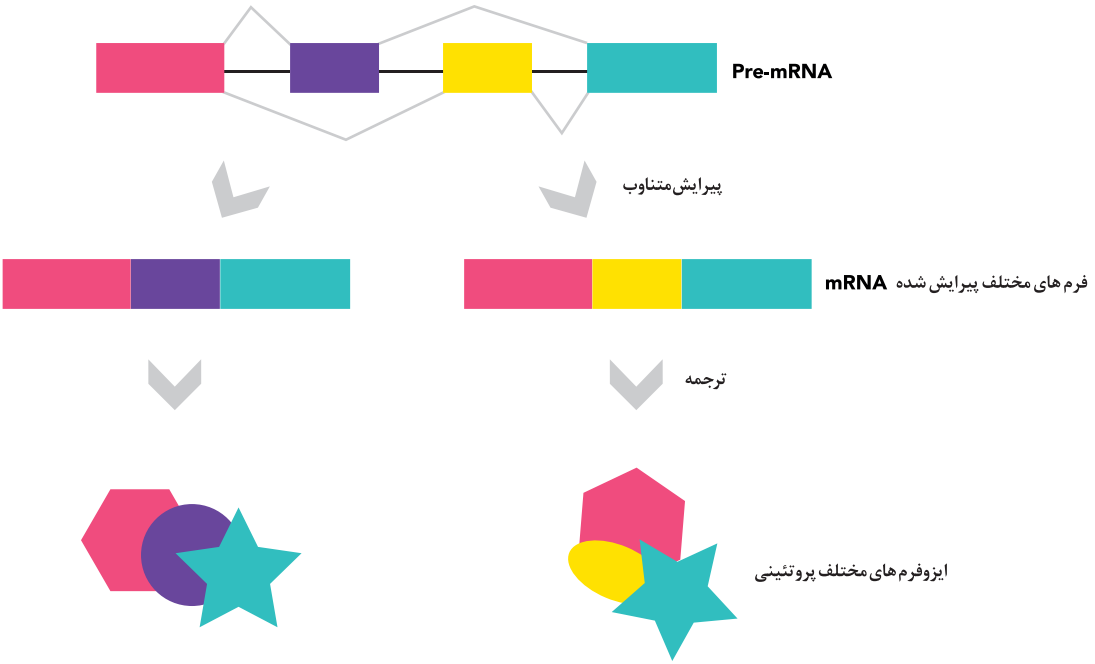
۴ Friedrich Miescher

۵ Erwin Chargaff

۶ Watson and Crick

۷ Franklin and Wilkins

در یک توالی نوکلئوتیدی، نوکلئوتید ها به صورت سه تایی خوانده می شوند، این نوکلئوتیدهای سه تایی همان کدون ها هستند که در طی فرایندهای رونویسی و ترجمه به آمینواسیدها تبدیل می شوند. در طی فرایند رونویسی مولکول RNA از روی مولکول DNA (الگو) به کمک کمپلکس آنزیمیRNA پلیمراز ساخته می شود (RNA اولیه)[۸] ، سپس در طی فرایند پیرایش متناوب، اینترون ها حذف و اگزون ها به یکدیگر متصل می شوند و به این ترتیبRNA پیامبر (mRNA)[۹] ایجاد می شود که توسط ماشین ترجمه در سیتوپلاسم به پروتئین ترجمه می شود. در طی فرایند پیرایش متناوب[۱۰] ، اگزون های مختلف یک نسخه RNA اولیه می توانند با ترتیب های متفاوتی به یکدیگر متصل شوند و نسخه های متفاوت mRNA را به وجود آورند که به پروتئین های مختلفی ترجمه می شوند (ایزوفرم)، به این ترتیب یک نسخه RNA اولیه می تواند ایزوفرم های مختلفی از یک پروتئین را تولید کند (شکل۴). از این رو، نظریه اولیه یک ژن، یک رشته پلی پپتید امروزه منسوخ شده است.



شکل۴. فرآیند پیرایش متناوب موجب ایجاد چندین نسخه mRNA و پروتئین از نسخه ی اولیه ی RNA می گردد

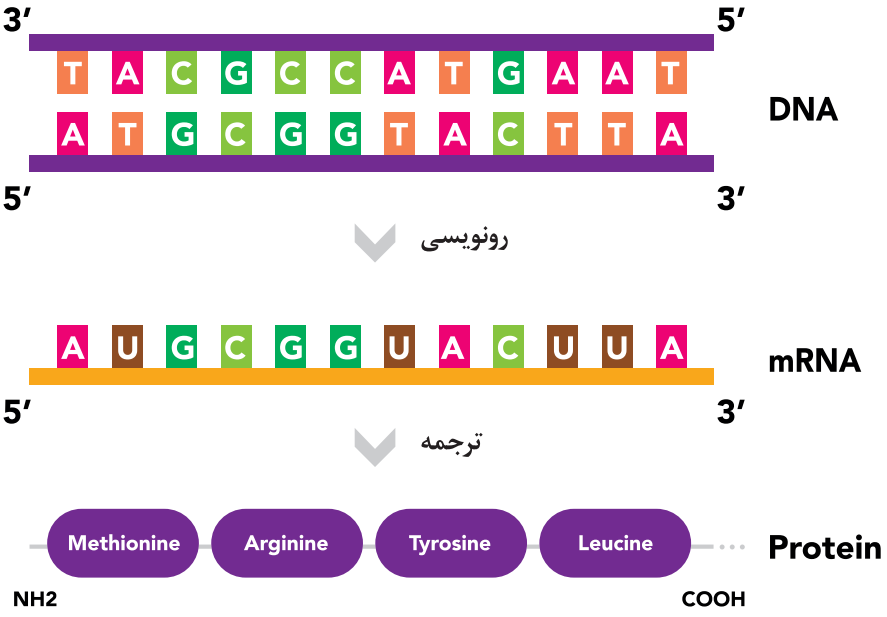
۸ Pre-mRNA

۹ Messenger RNA

۱۰ Alternative splicing

در طی ترجمه، به فرایند کد شدن آمینواسیدها از کدون های نوکلئوتیدی RNA، کد گذاری[11] می گویند. به این ترتیب کدون های سه تایی مختلف، ۲۰ نوع آمینواسید را کد می کنند که واحد ساختاری پروتئین ها را تشکیل می دهند. در یک سمت توالی آمینواسیدی پروتئین ها که در واقع معادل با انتهای ´۵ رشته RNA است، گروه کربوکسیل آزاد و در انتهای دیگر آن که معادل با انتهای ´۳ است، گروه آمین آزاد قرار گرفته است.

پروتئین های ایجاد شده پس از کسب ساختارهای اختصاصی، از نظر عملکردی فعال می شوند. بنابراین ژن ها به تنهایی عملکرد خاصی ندارند و در واقع محصول نهایی آنها، یعنی پروتئین ها، دارای عملکردهای معین و ویژه ای در سلول هستند (شکل۵).



شکل۵. نمای شماتیک مراحل رونویسی و ترجمه. به جهت رشته ها توجه نمایید

۳ قالب خواندن[12]

ناحیه ای از یک توالی نوکلئوتیدی که به منظور تولید آمینواسید های خاص به صورت کدون های سه تایی خوانده می شوند را قالب خواندن می نامند. از آنجایی که کدهای ژنتیکی سه تایی هستند، از نظر تئوری هر ناحیه از DNA دو رشته ای می تواند در شش قالب خوانده شود، قالب های خواندن متفاوت محصولات پروتئینی متفاوت ایجاد خواهند کرد. برای درک بیشتر این موضوع، توالی نوکلئوتیدی تک رشته ای زیر را نظر بگیرید:

5´...ATGTGCAGCCTAGCTGCCGTC...3´

همانطور که قبلاً ذکر شد، نوکلئوتیدها به صورت سه تایی خوانده می شوند، در صورتی که نقطه شروع خواندن، نوکلئوتید اول باشد، کدون های سه تایی کد کننده آمینواسید ها به صورت زیر خواهند بود که به ترتیب آمینو اسیدهای

۱۱ Coding

۱۲ Reading frame

متیونین، سیستئین، سرین، لوسین، آلانین، آلانین و والین را کد می کنند (شماره ۱).

۱) ATG TGC AGC CTA GCT GCC GTC

در حالی که اگر نوکلئوتید دوم یا سوم نقطه شروع خواندن باشد، کدون های سه تایی کد کننده آمینواسیدها به ترتیب مشابه شماره های ۲ و ۳ می باشند که آمینواسیدهای متفاوتی را کد می کنند.

۲) A TGT GCA GCC TAG CTG CCG TC

۳) AT GTG CAG CCT AGC TGC CGT C

به این ترتیب برای هر رشته DNA سه قالب خواندن متفاوت وجود دارد، از این رو یک مولکول DNA دو رشته ای می تواند در شش قالب متفاوت خوانده شود که در موجودات زنده، خوانش تنها در یکی از قالب های خواندن اتفاق می افتد و پروتئین مربوطه کد می شود. به ناحیه ای از DNA که واجد کدون آغاز ترجمه (ATG) و یکی از کدون های پایان ترجمه (UAG، UGA و UAA) باشد، قالب خواندن باز[۱۳] (ORF) می گویند.

در صورتی که شما یک توالی نوکلئوتیدی ناشناخته در اختیار داشته باشید و ندانید که آیا این توالی، یک توالی کد کننده پروتئین است یا خیر، یکی از متداول ترین و ساده ترین روش ها برای پاسخ به این سوال، شناسایی ORF های احتمالی در این توالی است. نرم افزارهای متعددی برای پیش بینی ORF در توالی های نوکلئوتیدی وجود دارند. همانطور که در بالا اشاره شد قالب های خواندن متفاوت می توانند باعث کد شدن پروتئین های مختلفی شوند. معمولاً اغلب نرم افزارها، ORF ای که طول بلندتری دارد را بعنوان ORF اصلی و کد کننده پروتئین در موجود زنده گزارش می کنند. به این ترتیب شما می توانید توالی آمینواسیدی پروتئین کد شونده توسط توالی نوکلئوتیدی خود را به کمک یک برنامه کامپیوتری به آسانی پیش بینی کنید.

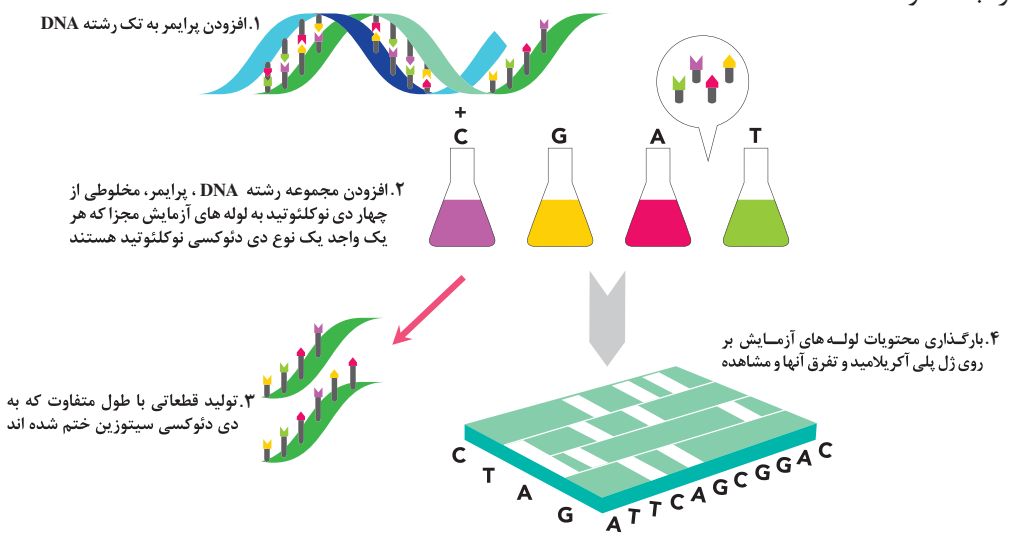
۴ توالی یابی DNA

همزمان با پیشرفت فناوری های DNA نوترکیب، کلون کردن ژن ها و توالی یابی DNA در دهه هفتاد و ابتدای دهه هشتاد میلادی، دانشمندان بحث درباره ی امکان توالی یابی ژنوم انسان را آغاز کردند. در سال ۱۹۹۰ که پروژه ژنوم انسان آغاز شد، پیش بینی می شد که با همکاری چندین شرکت مهم از سراسر جهان (آمریکا، انگلستان، ژاپن، فرانسه، آلمان، چین) این پروژه ۱۵ سال به طول انجامد. اما با ابداع روش های توالی یابی DNA، دو پیش نویس اولیه از توالی ژنوم انسان در فوریه ۲۰۰۱ و با هزینه ای بالغ بر ۳۰۰ میلیون دلار منتشر شد. با پیشرفت کار، توالی تقریباً کاملی از ژنوم انسان شامل ۹۹ درصد DNA یوکروماتیك در سال ۲۰۰۳منتشر و پس از آن تحولی شگرف در علم ژنتیک اتفاق افتاد.

۱۳ Open reading frame

۱-۴. تعیین توالی به روش سنگر[۱۴]

روش توالی یابی سنگر که به روش دی دئوکسی[۱۵] نیز معروف است، یکی از اولین و پر کاربردترین روش های تعیین توالی DNA می باشد که در سال ۱۹۷۷ توسط فردریک سنگر ابداع شد. در این روش توالی یابی، ابتدا توالی مورد نظر که می تواند بخشی از DNA ژنومی یا یک قطعه cDNA باشد دناتوره شده و به یک مولکول تک رشته ای تبدیل می شود و سپس یک قطعه اولیگونوکلئوتیدی کوتاه (معمولاً به طول ۲۰ جفت باز) و مکمل این تک رشته به آن افزوده می شود. در حضور آنزیم DNA پلیمراز و چهار نوع dNTP در چهار لوله آزمایش جداگانه، رشته دوم سنتز می شود. این واکنش با اضافه شدن دی دئوکسی نوکلئوتیدهای نشاندار شده با مواد رادیواکتیو متوقف خواهد شد. این نوع نوکلئوتیدها فاقد گروه هیدروکسیل متصل به کربن ′۳ هستند و لذا مانع اضافه شدن نوکلئوتید بعدی و طویل شدن رشته در حال سنتز می شوند. برای مثال، واکنش سنتز با اضافه شدن یک دی دئوکسی سیتوزین در یکی از لوله های آزمایش موجب تولید مجموعه ای از قطعات با طول های متفاوت که به نوکلوتید C ختم می شوند، می شود، از این رو این روش توالی یابی، روش خاتمه دهنده زنجیر نیز نامیده می شود. این فرایند برای سایر دی دئوکسی نوکلئوتیدها در هر لوله آزمایش بطور جداگانه انجام می‌گیرد، به این ترتیب هر یک از لوله های آزمایش واجد مجموعه ای از قطعات DNA هستند که به یک دی دئوکسی نوکلئوتید خاص ختم شده است. سپس محتویات هر یک از لوله های آزمایش بر روی یک ژل پلی اکریلامید بار گذاری[۱۶] می شود. این ژل قادر به تمایز مجموعه ای از قطعات DNA که اندازه ی آنها تنها یك نوکلئوتید با یکدیگر تفاوت دارند، می باشد. در نهایت قطعات مجزا شده با استفاده از فیلم اشعه ایکس مشاهده می شوند (شکل ۶). با پیشرفت علم به تدریج ستون های کوتاه جایگزین استفاده از ژل های پلی اکریلامید بزرگ شده و به کمك این ستون ها می توان قطعات DNA را در ۲ تا ۳ ساعت از یکدیگر جدا نمود.



شکل ۶. نمای شماتیک ساده توالی یابی به روش سنگر

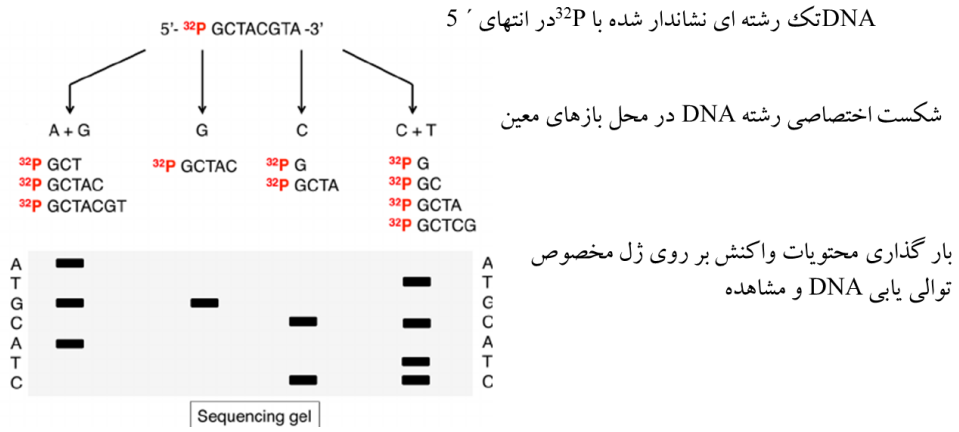
۱۴ Sanger sequencing

۱۵ Dideoxy

۱۶ Load

۲-۴. تعیین توالی به روش ماکسام-گیلبرت

این روش تعیین توالی با فاصله زمانی اندکی از روش سنگر توسط ماکسام و گیلبرت ابداع شد که مبتنی بر شکست شیمیایی در قطعات DNA تک رشته ای است. در این روش ابتدا انتهای ´۵ مولکول DNA دو رشته ای با استفاده از فسفر رادیواکتیو (۳۲P) نشاندار می شود، سپس این مولکول DNA تک رشته ای می شود و به چهار لوله آزمایش مختلف که هر یک واجد ماده شیمیایی خاصی است، منتقل می شوند. این مواد شیمیایی بطور اختصاصی رشته DNA را در محل بازهای پورینی و پیریمدینی می شکنند. غلظت این مواد به گونه ای تنظیم می شود که تنها یک شکستگی (در محل یکی از بازها) در هر رشته DNA ایجاد می شود. سپس محتویات هر لوله آزمایش در چاهک هایی مجزا در ژل پلی اکریلامید بارگذاری شده و از هم متمایز می شوند و با استفاده از فیلم اشعه ایکس قابل رویت می گردند (شکل ۷).



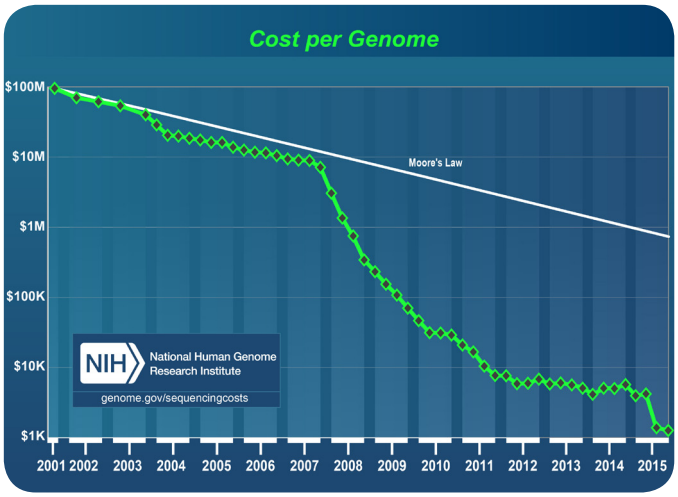
شکل ۷. نمای شماتیک ساده توالی یابی به روش ماکسام-گیلبرت

۳-۴. روش های جدید توالی یابی

همگام با پیشرفت علوم مختلف، روش های توالی یابی نیز تحول اساسی یافته اند به طوری که در حال حاضر استفاده از روش های توالی یابی قبلی تقریباً منسوخ شده است. در واقع هدف آن است تا فناوری توالی یابی به اندازه ای سریع، ساده و ارزان قیمت شود که امکان توالی یابی ژنوم افراد به منظور تشخیص های بالینی و درمان بیماری ها فراهم گردد. امروزه، ظهور روش های جدید توالی یابی موسوم به نسل جدید توالی یابی (NGS)[۱۷] یا توالی یابی در مقیاس وسیع[۱۸] انقلابی عظیم در ژنتیک ایجاد کرده اند. در حال حاضر این روش ها قادر به توالی یابی ژنوم انسان به قیمتی در حدود ۱۰۰۰ دلار و در طی چند روز با کیفیت بالا می باشند، لذا امروزه زمان و هزینه توالی یابی به شدت کاهش یافته است. شکل ۸ کاهش هزینه توالی یابی ژنوم در طی سال های ۲۰۱۵-۲۰۰۱ را نشان می دهد (شکل ۸).

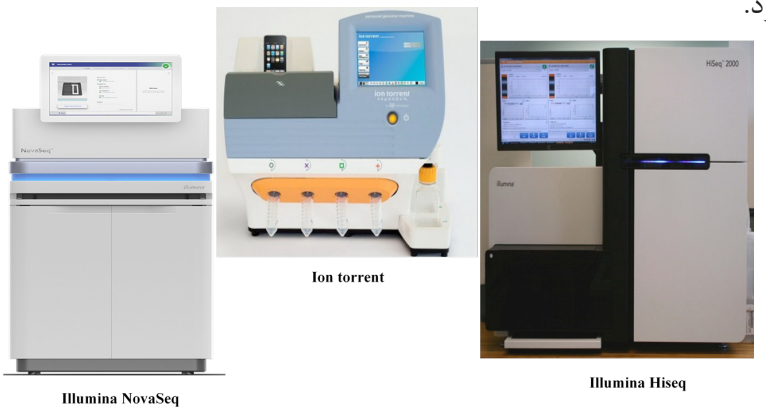
۱۷ Next-generation sequencing

۱۸ High throughput sequencing



شکل ۸. کاهش هزینه توالی یابی ژنوم

نسل جدید (دوم) دستگاه های توالی یابی دراواخردهه۱۹۹۰توسط شرکت های مختلف مانند Life ۴۵۴ / Roch Science، Illumina / Solexa وABI / SOLiD ایجاد و در طی سال های ۲۰۰۵ و ۲۰۰۶ تجاری شدند. به هر حال در کنار مزایای ارزشمند تکنیک NGS، اکنون آنالیز و تفسیر داده های عظیم حاصل از توالی یابی بزرگترین چالش در این زمینه می باشدکه نیازمند کامپیوترهای قدرتمند، نرم افزارهای خاص برای آنالیز داده ها و همچنین دانش و تخصص کافی است. درحال حاضر، جدیدترین سیستم توالی یابی، به توالی یابی نسل سوم معروف است که برای اولین بار توسط شرکت Helicos Bioscience معرفی شد و قادر به توالی یابی یک مولکول منفرد است. همچنین شرکت Ion Torrent در اواخر سال ۲۰۱۰ دستگاهی موسوم به Ion PGM (Ion personal genome machine) را روانه بازار کرد که نوع دیگری از دستگاه نسل سوم توالی یابی است. اخیراً شرکت Illumina جدیدترین دستگاه توالی یابی با نام Novaseq را روانه بازار کرده است و مدعی است که تنها با پرداخت ۱۰۰ دلار امکان توالی یابی ژنوم با استفاده از این سیستم وجود دارد (شکل ۹). با توجه به پیشرفت بسیار سریع علم در حال حاضر، در آینده نزدیک شاهد تحولات دیگری در زمینه توالی یابی خواهیم بود.



Illumina NovaSeq

Ion torrent

Illumina Hiseq

شکل ۹. دستگاه های جدید توالی یابی

۴-۴ . کاربردهای روش های جدید توالی یابی

روش های نسل جدید توالی یابی DNA به واسطه داشتن سرعت بالا و هزینه پایین، کاربردهای بسیاری دارند که از جمله مهمترین آنها می توان به موارد ذیل اشاره کرد:

۴-۴-۱. توالی یابی مجدد ژنوم های شناخته شده

توالی یابی مجدد[۱۹] کل ژنوم یا بخشی از آن یکی از کاربردهای رایج و مهم NGS است که برای شناسایی SNP ها، درج شدگی[۲۰] یا حذف شدگی[۲۱] ها، تعداد نسخه های ژنی و تغییرات ساختاری قابل استفاده است. با استفاده از این روش می توان آن دسته از تغییرات ژنتیکی که برخی از افراد را نسبت به سایر افراد جامعه مستعد ابتلا به یک بیماری خاص می کند شناسایی کرد. اولین پروژه توالی یابی مجدد ژنوم در سال ۲۰۰۵ و بر روی نمونه خون دکتر واتسون با استفاده از تکنولوژی ۴۵۴ انجام گرفته است.

۴-۴-۲. توالی یابی مناطق هدف[۲۲]

اگر چه توالی یابی کل ژنوم روش بهینه و جامع تری برای آنالیزهای ژنتیکی است، اما این روش در بسیاری از مراکز تحقیقاتی و بالینی هنوز در دسترس نیست. به همین دلیل، توالی یابی برخی از توالی های خاص که بروز جهش یا تغییرات دیگر در آنها سبب ایجاد بیماری می شود، مورد توجه قرار گرفته است. از آنجایی که اگزون ها نواحی کد کننده ژن ها محسوب می شوند و هر گونه تغییر در این نواحی حتی در سطح یک باز می تواند باعث ایجاد بسیاری از بیماری ها شود، در اغلب موارد تنها این نوع توالی ها (اگزون)، توالی یابی می شوند که به Exosome sequencing معروف است. برای مثال، برای بررسی تغییر طول ریزماهواره[۲۳] ها در سرطان کلورکتال از این روش استفاده شده است.

۴-۴-۳. تعیین توالی از نو[۲۴]

توالی یابی از نو به مفهوم تولید توالی اولیه از ژنوم یا ترانسکریپتوم (مجموعه ی رونوشت های (mRNA) موجود در یک سلول یا جمعیتی از سلولها) در موجوداتی که اطلاعات ژنتیکی اندکی از آنها در دسترس است و ژنوم آنها هنوز توالی یابی نشده است. این نوع توالی یابی در سطح ترانسکریپتوم به توالی یابی[۲۵] RNA یا RNA-seq معروف است. در سطح ژنوم، توالی ژنوم پیچیده و هتروزیگوت انگور در سال ۲۰۰۷ با استفاده از ترکیبی از روش سنگر و روش های جدید توالی یابی تکمیل شد. در حال حاضر، پروژه تعیین توالی ژنوم گیاه *Fritillaria* به کمك تکنولوژی NGS در حال انجام است. اندازه ژنوم در گونه های دیپلوئید این گیاه از ۳۰ Gb تا ۸۰ Gb می باشد که بزرگترین اندازه ژنوم گزارش شده برای یك گیاه دیپلوئید است. تصور کنید که اگر قرار بود چنین ژنوم هایی با روش سنگر توالی یابی شوند علاوه بر هزینه بالا به چند سال زمان نیاز بود.

یکی از کاربردهای مهم و در حال گسترش NGS، توالی یابی RNA کل است. با استفاده از این روش امکان بررسی

---

۱۹ Re-sequencing

۲۰ Insertion

۲۱ Deletion

۲۲ Targeted sequencing

۲۳ Microsatellite

۲۴ De novo sequencing

۲۵ RNA-sequencing

تمام انواع مختلف نسخه های موجود در سلول اعم از RNA های کد کننده و غیر کد کننده (ترانسکریپتوم) وجود دارد. تکنیك RNA-seq برای انجام مطالعات ژنتیکی و مولکولی به ویژه در موجودات غیر مدل که اطلاعات ژنتیکی بسیار اندکی از آنها موجود است، بسیار حائز اهمیت و پر کاربرد می باشد. با استفاده از این روش، تصویر تقریباً کاملی از ترانسکریپتوم از یك نمونه بیولوژیك حاصل می شود که اطلاعات قابل استخراج از آن، در برخی موارد تا حد زیادی می تواند جایگزین توالی یابی کل ژنوم گردد، در حالی که روش سنگر تنها قادر به شناسایی ۶۰ درصد از نسخه های موجود در سلول است.

۴-۴-۴. بررسی تغییرات اپی ژنتیک

یکی از برهم کنش های مهم در زیست شناسی، ارتباط بین پروتئین و DNA است که نقش مهمی در دسترس بودن توالی DNA و بیان ژن ها دارد. در سال های اخیر برای بررسی و مطالعه پروتئین های متصل شونده به DNA در کل ژنوم و همچنین تغییرات پروتئین های هیستون و نوکلئوزوم ها از روشی موسوم به chromatin immunoprecipitation sequencing یا ChIP-seq استفاده می شود. در این روش، تنها توالی های DNA متصل به پروتئین توالی یابی می شوند، به این ترتیب که پس از جداسازی و تفکیک پروتئین ها با استفاده از روش های رسوب دهی ایمونوکروماتینی[۲۶] از DNA، DNA به کمک روش NGS توالی یابی می گردد.

۴-۴-۵. توالی یابی اسیدهای نوکلئیک آزاد موجود در خون

توالی یابی مولکول های DNA یا RNA موجود در مایعات بدن به ویژه خون، یکی از روش های جدید غیر تهاجمی به منظور تشخیص بسیاری از بیماری های مانند سرطان ها و نقایص جنینی می باشد. یکی از مشکلات عمده بررسی اسیدهای نوکلئیک موجود در خون در گذشته، مقدار پایین آنها و امکان ردیابی تنها برخی از تغییرات و جهش های خاص در آنها بود، اما امروزه با ظهور روش های جدید توالی یابی می توان به راحتی کل DNA یا RNA موجود در مایعات بدن را تعیین توالی کرد.

۲۶ Immunoprecipitation

فصل دوم

کاربردهای بیوانفورماتیک

۱ مقدمه

امروزه بیوانفورماتیک تنها محدود به تحقیقات زیست شناسی مولکولی و ژنومی پایه نمی شود و تاثیر عمده ای بر حوزه های بیوتکنولوژیکی مختلف و علوم زیست پزشکی دارد. این حوزه ی علمی ابزارهای مناسبی برای تجزیه و تحلیل داده ها به منظور استخراج الگوهای مفید در شبیه سازی ساختارهای پیچیده مولکولی و سیستم های زیستی، به محققان ارائه می دهد. تمرکز امروزی بر بیوانفورماتیک، بیانگر آغاز یك تغییر اساسی و مهم در مطالعه سیستم های زیستی است که نتایج پایه ای و کاربردی در بر خواهد داشت. این دانش در تمام زمینه هایی که در سامانه های زیستی مختلف نقش ایفا می نمایند، مانند پزشکی، کشاورزی، محیط زیست، بیوتکنولوژی میکروبی، صنایع سبز و فرآیندهای زیستی تاثیر گذار است.

۲ کاربرد بیوانفورماتیک در زیست شناسی

هم اکنون بیوانفورماتیک بخش مهمی از زیست شناسی و همچنین تحقیقات وابسته به آن را تشکیل می دهد. بیوانفورماتیک در حوزه ی تحقیقات زیست شناسی بر روی تعیین توالی، مطالعه ترانسکریپتوم[۱] ، کشف ژن، محل یابی ژن ها، توالی یابی ژنوم، تعیین ساختار پروتئینی، پیشگویی ساختار پروتئینی، پیش‌بینی بیان ژن، برهمکنش پروتئین-پروتئین، مقایسه های فیلوژنتیکی، یافتن محل اتصال فاکتورهای رونویسی در ژن ها و مدل سازی تکامل تمرکز دارد. بیوانفورماتیک همچنین ابزارهای آنالیز کننده ای را برای داده های ریزآرایه یا میکروواری[۲] (روشی برای اندازه گیری سطح بیان تعداد زیادی از ژن ها بصورت همزمان) و همچنین داده های حاصل از نسل جدید تعیین توالی (NGS) فراهم می کند. در حوزه ی متابولومیکس[۳] (علم مطالعه ی فرآیندهای شیمیایی و متابولیت ها)، بیوانفورماتیک در مطالعه دینامیک های سلولی و همچنین شبیه سازی فعل و انفعالات سلولی کاربرد دارد. بیوانفورماتیک امکان فهم عمیق زیست شناسی سامانه ای[۴] شامل پاسخگوئی به مباحث پیچیده در مورد برهمکنش میان ژنومیک، ترانسکریپتومیک، پروتئومیک و متابولومیک و همچنین شناخت منشا شبکه های ژنتیکی را فراهم می نماید. وقتی در مورد کاربرد بیوانفورماتیک در زیست شناسی صحبت می شود، با توجه به معنی وسیعی که زیست شناسی دارا می باشد این کاربردها می تواند با سایر زیر مجموعه های کاربردی بیوانفورماتیک مانند پزشکی و کشاورزی نیز هم پوشانی داشته باشد (شکل ۱).

---

۱ Transcriptome

۲ Microarray

۳ Metabolomics

۴ Systems Biology

شکل ۱. آنالیز داده های زیستی حاصل از مطالعات امیکس توسط ابزارهای بیوانفورماتیکی

۲-۱. تجزیه و تحلیل اطلاعات توالی

یکی از وجوه مهم بیوانفوماتیک، زیست شناسی محاسباتی[۵] در زمینه ی تجزیه و تحلیل اطلاعات نهفته در توالی می باشد. تجزیه و تحلیل اطلاعات توالی شامل پیدا کردن ژن ها در توالی های RNA و DNA، توسعه ی روش های پیش بینی ساختار و تعیین عملکرد پروتئین های جدید و توالی های ساختاری RNA و صف بندی پروتئین های مشابه و ایجاد درخت های فیلوژنتیکی برای بررسی روابط تکاملی است. علاوه براین امروزه با ظهور و توسعه سریع تکنولوژی NGS و به تبع آن فراهم آمدن حجم وسیعی از اطلاعات مربوط به توالی های RNA و DNA، زیست شناسی محاسباتی همراه با نیازمند شدن به نرم افزارها و بانک های اطلاعاتی جدید وارد مرحله ی تازه و مهمی شده است.

۲-۲. تحلیل عملکرد ژنوم

کارکرد شناسی ژنوم شامل به کارگیری روش های آماری پیشرفته، استفاده از پایگاه های داده و نرم افزارهای مناسب، خوشه بندی مسائلی چون بررسی همزمان میزان فعالیت هزاران ژن در سلول، تحلیل نحوه تعامل تعداد زیادی پروتئین و تحلیل خصوصیات هزاران سلول جهش یافته در آن واحد می باشد که می تواند پیش بینی نقش و کارکرد ژن ها در سلول را بدون نیاز به آنالیز داده های پروتئینی انجام دهد.

۲-۳. مشخص کردن کلیه پروتئین های موجود در سلول

در این بخش، کلیه پروتئین هایی که در سلول تحت یک شرایط معین (مانند، حالت استراحت، رشد، تمایز، بیماری، تأثیر دارو و...) وجود دارند، مشخص می شود. به این ترتیب می توان پروتئین هایی که در شرایط مختلف وجود دارند یا میزان آنها تغییر می کند را شناسایی کرد و از این طریق به عملکرد آنها پی برد. شناسایی این پروتئین ها در تشخیص و بررسی روند پیشرفت یا بهبودی بیماری و همچنین شناسایی داروهای جدید، مفید می باشد.

---

۵ Computational Biology

۴-۲. نقشه برداری برهمکنش های بین پروتئینی

پروتئین ها در سلول بصورت منفرد عمل نمی‌کنند و اغلب تأثیر خود را با همکاری پروتئین های دیگر و برهم کنش با آنها اعمال می کنند. نمونه بارز برهم کنش های پروتئینی، در مسیر های انتقال پیام و مسیرهای بیوسنتزی مشاهده می شود. با شناسایی این برهم کنش ها، بطور کارآمدتری می توان عملکرد و رفتار پروتئین ها را مشخص کرد.

۵-۲. نقشه برداری آرایش های پروتئینی

اغلب پروتئین ها پس از ترجمه متحمل آرایش های مختلفی مانند، گلیکوزیله شدن، متیله شدن، استیله شدن، فسفریله شدن و... می شوند. این آرایش ها بر فعالیت و عملکرد پروتئین ها، همچنین ساختار فضایی، پایداری و نیمه عمر آنها تأثیر گذار است. بسیاری از داروها گروه های الکتروفیلی دارند که از طریق آنها به پروتئین هدف متصل شده و اثر خود را اعمال می‌کنند. شناسایی این آرایش ها، تأثیر آنها بر عمکرد پروتئین ها و شرایطی که منجر به این آرایش ها می شود، به شناسایی رفتار و تفسیر عملکرد پروتئین ها کمک می کند.

۶-۲. پیش بینی ساختار سه بعدی پروتئین

عملکرد مولکول های بزرگ پروتئینی به شکل فضایی و ساختار سه بعدی آنها وابسته است. ژن ها با عملکرد پروتئین هایی که می سازند نقش خود را اعمال می‌کنند. بنابراین لازمه ی شناخت کامل ژن ها، شناخت کامل پروتئین ها است. پیش بینی ساختار پروتئین به معنای استنتاج ساختار سه بعدی پروتئین از روی دنباله آمینواسیدهای آن و یا به بیان دیگر، تعیین ساختار دوم و سوم از روی ساختار اولیه پروتئین است. تعیین ساختار پروتئین یکی از مسائل مهم در حوزه بیوانفورماتیک و شیمی تئوری است و اهمیت زیادی در پزشکی (برای مثال در طراحی دارو) و زیست فناوری (در طراحی آنزیم ها) دارد. دو اصل مهم برای تعیین ساختار سه بعدی پروتئین از روی توالی آن وجود دارد که هر کدام روش جداگانه ای ارائه می دهند:

۱. پروتئین ها با توالی نسبتاً مشابه، شکل فضایی شبیه به هم پیدا می کنند (جست و جو برای یافتن توالی های مشابه).

۲. شکل فضایی مولکول به گونه ای است که به حداقل سطح انرژی برسد (استفاده از قوانین شیمی، فیزیک و ترمودینامیک).

۷-۲. مطالعه ی متابولوم

متابولومیکس، به دانش مطالعه ی مولکول های کوچک درون سلولی یا مطالعه پروفایل متابولوم[۶] یک سیستم با ملکول هایی با وزن کمتر از ۱۰۰۰ دالتون[۷] و در یک واحد زمانی و تغییرات آنها در شرایط مختلف اطلاق می گردد. اندازه گیری کمّی و کیفی این متابولیت ها اطلاعات مفیدی از وضعیت بیوشیمیایی سلول و فعالیت ژن ها را فراهم می کند. امکان مطالعه هم زمان متابولیت ها فقط به کمک دستگاه های مدرن و جدید طیف سنجی جرمی و یا طیف

۶ Metabolomic Profile

۷ Dalton

سنجی تشدید مغناطیس هسته مانند LC / Ms / Ms و NMR با قدرت تفکیک پذیری بالا امکان پذیر است. مانند سایر مطالعات عملکردی ژنوم، متابولومیکس مقدار زیادی داده تولید می کند که آنالیز این حجم وسیع داده بعنوان یک چالش مطرح است و به آمار و ابزارهای بیوانفورماتیکی مناسب نیاز دارد. بیوانفورماتیک به متابولومیکس از طریق مدیریت داده ها و اطلاعات، پردازش داده های اولیه، آنالیز آماری و داده کاوی و نهایتاً ادغام داده ها و مدل سازی ریاضی شبکه های متابولیکی در چارچوب زیست شناسی سامانه ای کمک می کند.

۳ کاربرد بیوانفورماتیک در کشاورزی

گیاهان نقش های مهم و متنوعی را در حوزه های مختلف زندگی بشر از جمله تامین غذا، دارو، اقتصاد و محیط زیست بازی می‌کنند. با توجه به افزایش جمعیت، تامین مواد غذایی مورد نیاز بعنوان یک چالش برای زیست فناوری گیاهی مطرح شده است. با وجود اینکه بازده تولید محصولات زراعی در قرن گذشته افزایش پیدا کرده ولی تلاش ها برای ادامه بهبود روش های اصلاح نژاد و توسعه استراتژی های زیست فناورانه جدید به منظور افزایش تولید محصولات زراعی در حال انجام است. با شروع مطالعات ژنومیکس گیاهی اطلاعات بسیار زیادی در جهت بهبود صفات فنوتیپی گیاهان فراهم شده است. هدف مطالعات ژنومیکس شناخت ژنها و عملکرد آنها می باشد. داده های مربوط به ژنومیکس و ترانسکریپتومیکس گیاهان زراعی که با استفاده از تکنولوژی های جدید مانند نسل جدید تعیین توالی (NGS) فراهم شده است همراه با ابزارهای بیوانفورماتیکی توسعه یافته، تحول عظیمی را در شناخت ژن های مربوط به صفات با ارزش از نظر کشاورزی، مهندسی ژنتیک و اصلاح نژاد گیاهان زراعی موجب شده است. در شکل ۲ برخی کاربردهای مهم علم بیوانفورماتیک در حوزه ی کشاورزی نشان داده شده است.

شکل ۲. کاربردهای بیوانفورماتیک در کشاورزی

۳-۱. بهبود محصول

ژنتیک مقایسه ای ژنوم های گیاهی با استفاده از ابزارهای بیوانفورماتیکی نشان داده است که سازماندهی ژن های آنها در طی تکامل بیشتر از آن چیزی که تصور می شده، حفاظت شده باقی مانده است. این یافته ها پیشنهاد می کند که اطلاعاتی که از سیستم های گیاهی مدل بدست آمده است می تواند برای بهبود گیاهان به کار رود. ارابیدوپسیس تالیانا (*Arabidopsis thaliana*) اولین گیاه تعیین توالی شده است که بعنوان گونه ی گیاهی مدل در تحقیقات ژنتیکی و زیست شناختی به کار می رود. تعداد زیادی از ژن ها در تمام گیاهان یکسان هستند، بنابراین مطالعه ژن ها در گیاهان مدلی مانند *A. thaliana* و ایجاد پایگاه های داده ای بر این اساس می تواند دانش ما را درباره بیان و عملکرد ژن ها در تمام گیاهان ارتقا دهد. ارابیدوپسیس دارای کوچکترین ژنوم گیاهی است که این خصوصیت مهمترین دلیلی است که این گیاه را بعنوان مدل برای تعیین توالی ژنوم انتخاب کرده اند. DNA ارابیدوپسیس ۱۴۰ میلیون باز دارد که بین ۵ کروموزوم تقسیم شده است.

برنج (*Oryza sativa*) مهمترین محصول غذایی انسان بوده و غذای اصلی بیش از نیمی از جمعیت جهان را تامین می کند. این گیاه با ۴۳۰ میلیون نوکلئوتید دارای کوچکترین ژنوم در بین غلات است که این خصوصیت علاوه بر اهمیت غذایی آن، باعث شده برنج را بعنوان یک گیاه مدل تک لپه ای مطرح نمایند. از آنجایی که تحقیقات مختلفی بر تولید محصول، بررسی خصوصیات هیبریدها، مقاومت ژنتیکی به تنش های زیستی و غیر زیستی در برنج انجام گرفته، دانشمندان توانسته اند واریته های بسیاری از برنج را معرفی کنند که می توانند با شرایط محیطی متنوعی سازگاری داشته باشند.

۴ کاربرد بیوانفورماتیک در مطالعه ی ژنوم میکروبی

در چند دهه ی اخیر ظهور ابزارها و روش های علمی جدید مانند پیشرفت هایی که در حوزه ی تعیین توالی و بیوانفورماتیک حاصل شده است توانسته در شناسایی گونه های میکروبی جدید راهگشا بوده و دانش ما را درباره ی تنوع و فراوانی میکروب ها و برهم کنش آنها با محیط زیست افزایش دهد. اولین ژنوم باکتریایی پس از ۱۳ ماه کار تحقیقاتی در سال ۱۹۹۵ بطور کامل تعیین توالی شد. امروزه با ظهور تکنولوژی های جدید می توان با صرف زمان کم و هزینه ی بسیار پایین تعیین توالی کل ژنوم میکروبی را انجام داد. در شکل ۳ برخی کاربردهای مهم بیوانفورماتیک در مطالعه ی ژنوم میکروبی نمایش داده شده است.

شکل ۳. کاربردهای بیوانفورماتیک در مطالعه ی ژنوم میکروبی

۴-۱. پاکسازی زباله

همزمان با فراهم شدن امکان مطالعه ی ژنوم باکتری ها و میکروب ها و شناسایی ژن های مفید در پاکسازی زباله، ایده ی کاربرد این موجودات در پاکسازی محیط زیست مطرح گردید. باکتری *Deinococcus radiodurans* در کتاب گینس با نام «The world's toughest bacterium» به معنی سخت ترین باکتری جهان ثبت شده است (شکل ۴). این باکتری دارای توانایی ترمیم DNA آسیب دیده با جدا کردن بخش آسیب دیده در یک ناحیه ی متمرکز است. این باکتری به دلیل قدرت بقای بالا هدف مناسبی برای بیوتکنولوژی میکروبی با اهداف مختلف می باشد. دانشمندان چندین ژن را با هدف پاکسازی آلودگی های محیط زیست از سایر باکتری ها به *D. radiodurans* انتقال داده اند. این ژن ها برای شکستن ترکیبات آلی و فلزات سنگین در نواحی آلوده به مواد رادیواکتیو استفاده می شوند.

شکل ۴. تصویر حاصل از میکروسکوپ الکترونی روبشی (SEM) از باکتری *D. radiodurans*

۴-۲. تغییرات آب و هوا

تغییرات آب و هوایی ناشی از گردش های هوای اقیانوسی، تنوع در میزان و زاویه ی تابش خورشید، فوران های آتشفشانی و تغییرات وابسته به فعالیت انسان می باشد. میکروارگانیسم ها در راندن چرخه های حیاتی کربن، نیتروژن و سایر مواد غذایی نقش اساسی دارند. در طی این فرآیندها گازهای گلخانه ای مانند دی اکسی کربن و متان تولید می شوند. میکروارگانیسم ها هم در تولید و هم در مصرف گازهای گلخانه ای نقش بسیار مهمی را دارند، بنابراین مدیریت و بهره برداری از فرآیندهای میکروبی در آینده می تواند در کاهش تغییرات آب و هوایی وابسته به فعالیت های انسانی بسیار مفید باشد.

اطلاعات ما درباره ژنوم میکروارگانیسم ها از طریق کاربرد ابزارهای بیوانفورماتیکی و امروزه بوسیله ی به کارگیری روش NGS که قابلیت توالی یابی سریع کل ژنوم این موجودات را فراهم نموده است، افزایش قابل توجهی داشته است.

شناخت توالی ژنوم و مطالعه ی رفتار میکروارگانیسم ها در شرایط مختلف، امکان شناسایی و جداسازی ژن هایی که مسئول خصوصیات منحصر به فرد برای بقا در شرایط نامساعد هستند را فراهم می نماید.

Rhodopseudomonas palustris باکتری فتوتروفیک غیر سولفوره ارغوانی است که معمولا در خاک و آب یافت می شود. این باکتری بواسطه ی نور خورشید، دی اکسید کربن اتمسفری را تثبیت و تولید بیومس می کند. این میکروب می تواند همچنین در تجزیه و بازیافت ترکیبات آروماتیک مانند لیگنین نقش داشته باشد. *R. palustris* توسط میکروب شناسان بعنوان یک باکتری با سیستم متابولیکی چند کاره شناخته می شود. این باکتری نه تنها قادر به تبدیل گاز دی اکسید کربن بعنوان یکی از موثرترین و پر تولید ترین گازهای گلخانه ای، به مواد سلولی است بلکه قادر به تثبیت گاز نیتروژن و تبدیل آن به آمونیوم و تولید گاز هیدروژن نیز می باشد. این باکتری هم در حضور و هم در غیاب اکسیژن رشد می کند. درغیاب اکسیژن، این باکتری ترجیح می دهد تمام انرژی مورد نیازش را از نور بوسیله ی فتوسنتز تامین نماید. امکان استفاده از این باکتری برای کاهش تغییرات آب و هوایی وابسته به فعالیت های انسان در حال بررسی می باشد.

۳-۴. صنایع لبنی و پروبیوتیک ها[۸]

در شاخه بیوانفورماتیک، ارگانیسم ها و میکروارگانیسم هایی که می توانند در صنایع لبنی و تولید مواد غذایی مفید باشند، شناسایی و مطالعه می شوند. *Lactococcus lactis* یک باکتری غیر بیماری زای گرم مثبت میله ای شکل است که بعنوان عامل تولید کننده اسید لاکتیک برای تولید محصولات لبنی مانند دوغ، ماست و پنیر ضروری است، همچنین این باکتری برای تهیه ی ترشی سبزیجات، بعضی انواع نان، سوسیس، کالباس و سایر محصولات غذایی تخمیری مورد استفاده قرار می گیرد. این باکتری علاوه بر اینکه یکی از مهمترین میکروارگانیسم های وابسته به صنعت لبنیات است، بعنوان اولین ارگانیسم تغییر ژنتیکی یافته که بصورت زنده برای درمان بیماری انسان به کار برده شده نیز مطرح است. محققان پیش بینی کرده اند شناسایی فیزیولوژی و آرایش ژنتیکی این باکتری برای توسعه ی صنایع تولید مواد غذایی و همچنین صنعت دارو می تواند بسیار ارزشمند باشد.

واژه پروبیوتیک به معنای «برای زندگی»، برگرفته از زبان یونانی است. پروبیوتیک ها ارگانیسم های زنده ای هستند که به عنوان مکمل های غذایی مورد استفاده قرار می گیرند. این ارگانیسم ها اثرات سودمندی در بدن به جای می گذارند که این امر به واسطه تنظیم و تعادل میکروب های ناحیه گوارشی (روده ای) صورت می گیرد. امروزه با توجه به آگاهی و تمایل روزافزون مصرف کنندگان به خرید محصولات سالم تر و مفیدتر، توسعه فراورده های پروبیوتیکی در صنعت از اهمیت زیادی برخوردار است در این میان محصولات لبنی حامل های بسیار مطلوبی برای پروبیوتیک ها محسوب می شوند. امروزه شیرهای تخمیری مانند ماست بیشترین فراورده های لبنی حاوی باکتری های پروبیوتیک را تشکیل می دهند.

باکتری های اسید لاکتیک مانند *Weissela sp.*، *Lactococcus sp.* و *Lactobucillus sp.* در جهت بهبود سیستم ایمنی بدن و افزایش ماندگاری مواد غذایی، تحت عنوان باکتری های پروبیوتیک کاربرد دارند. این باکتری ها، در طبیعت پراکنده بوده و بالطبع در طیف وسیعی از غذاها نیز حضور دارند. بنا به اهمیت این میکروارگانیسم ها در

---

۸ Probiotic

سلامت انسان، شناسایی، طبقه بندی مولکولی و مطالعه ی ژنوم آنها می تواند گامی مؤثر در معرفی پروبیوتیک های بومی با خصوصیات عملکردی ویژه و به کارگیری آنها در محصولات لبنی صنعتی باشد.

۴-۴. انرژی جایگزین

باکتری *Chlorobium tepidum* دارای ظرفیت بسیار زیادی برای تولید انرژی از نور است، از این رو دانشمندان برای بررسی منشا و مکانیسم فتوسنتز ژنوم این میکروارگانیسم را مورد مطالعه قرار داده اند. این باکتری بعنوان مدل باکتریایی مطرح بوده و امکان استفاده از آن برای تولید انرژی جایگزین در حال بررسی می باشد. *C. tepidum* باکتری سبز گوگردی گرم منفی ترموفیل[۹] جدا شده از چشمه های آب گرم نیوزلند است که فتوسنتز را بصورت متفاوتی نسبت به گیاهان و سایر باکتری ها انجام می دهد و برخلاف گیاهان، در طی فتوسنتز اکسیژن تولید نمی کند و در طی فرآیند فتوسنتز نیتروژن و سولفور را متابولیزه می کند. بر اساس نظر برخی محققان، ممکن است بتوان ریشه تکاملی فتوسنتز را در میکروارگانیسم هایی مانند *C. tepidum* جستجو کرد. این باکتری قادر به رشد در حضور اکسیژن نمی باشد.

۵ کاربرد بیوانفورماتیک در پزشکی

بیوانفورماتیک نقش مهمی را در پیشرفت علم پزشکی ایفا کرده است. در شرایط حاضر که دانش مولکولی به سرعت در حال پیشرفت است، بیوانفوماتیک می تواند به پزشکی از طریق فراهم آوردن ابزراهای مورد نیاز برای آنالیز داده هایی با حجم وسیع کمک قابل توجهی نماید. اهداف تحقیقاتی در حوزه ی بیوانفورماتیک پزشکی شامل رسیدن به درک عمیقی از عملکرد مولکولی ژن ها و پروتئین ها تا مسیرهای فعالیت آنها و در نهایت دستیابی به نحوه ی عملکرد و تنظیم کل سیستم از طریق استفاده از استراتژی های محاسباتی پیشرفته می باشد. امروزه به کمك علم بیوانفورماتیك می توان حجم وسیع اطلاعات حاصل از امیکس[۱۰] را با اطلاعات مربوط به بیماران که بصورت الکترونیکی ثبت شده اند ترکیب کرد. ترکیب موفق داده های زیست پزشکی و اطلاعات بالینی در نهایت به کشف داروها و راه های موثرتری برای درمان می انجامد.

۵-۱. کشف دارو

استفاده از روش های بیوانفورماتیکی، تحقیقات و مطالعاتی که عملاً از نظر آزمایشگاهی انجام آنها امکان پذیر نیست را میسر می سازد. کاربرد کریستالوگرافی اشعه X برای کشف دارو بیش از ۳۰ سال پیش ظهور پیدا کرد، زمانی که اولین ساختار سه بعدی پروتئین تعیین شد. در طی یک دهه بعد، زمانی که دانش مربوط به ساختار سه بعدی پروتئین ها با فرایند های طراحی دارو ترکیب شد، تغییرات مهمی در طراحی داروهای جدید اتفاق افتاد. طراحی دارو روش محاسباتی می باشد که می تواند برهم کنش بین دو مولکول را پیشگویی کند. این روش بطور عمده شامل الگوریتم های مانند دینامیک مولکولی، شبیه سازی مونت کارلو[۱۱]، روش جستجو براساس بررسی قطعات و.. می باشد.

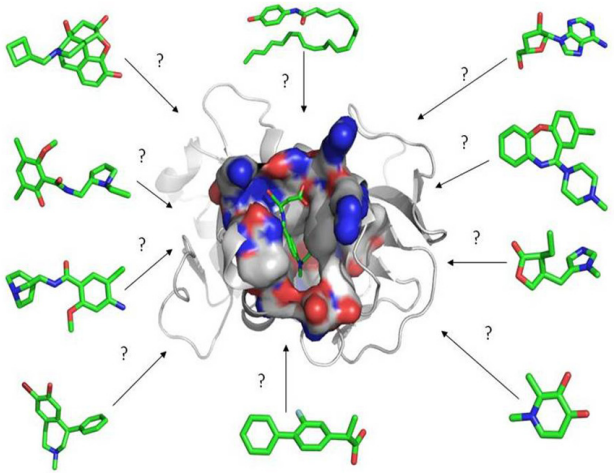
---

۹ Thermophile

۱۰ OMICs

۱۱ Monte Carlo

مولکولار داکینگ[۱۲] یکی از روش های زیر مجموعه ی مدلینگ مولکولی در علم بیوانفورماتیک می باشد و روشی پرکاربرد در طراحی دارو است. داکینگ مولکولی در تعیین برهم کنش بین دو مولکول مثل پروتئین با پروتئین یا پروتئین با DNA برای یافتن بهترین جهت گیری یک لیگاند در یک کمپلکس با حداقل انرژی به کار برده می شود. با پیشرفت علوم و تکنولوژی، استفاده از علوم بین رشته ای مانند بیوانفورماتیک کاربرد روز افزونی داشته است. به کمك این علم، امکان انجام محاسبات نرم افزاری و نهایتاً غربالگری قبل از مرحله آزمایشگاهی فراهم گردیده است که این امر می تواند موجبات صرفه جویی در وقت و هزینه شود. برای مثال لیگاند در مبحث داکینگ می تواند یک دارو باشد، داروهای زیادی وجود دارند که می توانند به پروتئین ها متصل شوند و آنها را مهار یا تحریک کنند. با استفاده از تکنیک داکینگ می توان تعداد زیادی دارو را بررسی کرد و بر اساس نتایج حاصل، به منظور صرفه جویی در هزینه و زمان تنها تعداد اندکی از آنها را برای مطالعات بیشتر در فاز آزمایشگاهی انتخاب کرد. نتایج حاصل از داکینگ توسط یک تابع درجه بندی آماری تجزیه و تحلیل می شود. این تابع درجه بندی آماری برای محاسبه انرژی برهم کنش، آن را به مقادیر عددی به نام درجه داکینگ تبدیل می کند. نتایج بدست آمده از داکینگ شامل اشکال ۳ بعدی از لیگاند متصل شده به ماکرومولکول بوده که با استفاده از نرم افزارهایی مانند Pymol و Rasmol قابل مشاهده می باشد و می تواند در بدست آوردن بهترین حالت از لیگاند برای برهم کنش با ماکرومولکول به ما کمک نماید (شکل ۵).



شکل ۵. نمایش جایگاه فعال یک پروتئین همراه با مولکول های کوچک بعنوان لیگاند در اطراف مولکول پروتئین. پس از انجام داکینگ بین مولکول لیگاند و پذیرنده، امتیاز دهی با استفاده از برنامه های غربالگری مجازی مانند Pymol انجام می گیرد.

۵-۲. پزشکی فردی

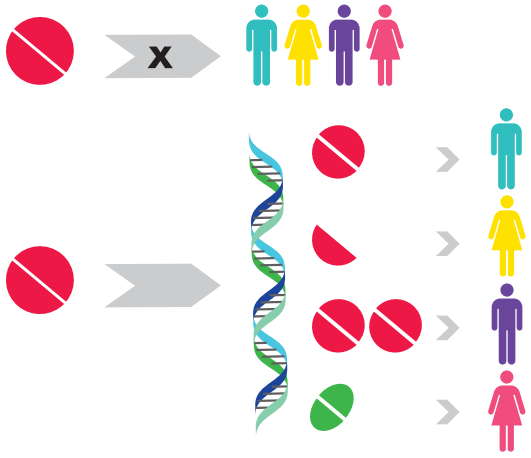
پزشکی فردی[۱۳] شاخه ای از پزشکی است که یک مجموعه ای از توصیه ها و روش های بهداشتی و درمانی را با توجه به ژنتیک و شرایط زندگی، منحصراً برای یک فرد در جهت پیشگیری یا درمان بیماری پیشنهاد می کند.

۱۲ Molecular Dockinging

۱۳ Personalized medicine

در حال حاضر، دارویی که برای درمان یك فرد (بیمار) تجویز می شود برای تمام مبتلایان به آن بیماری یکسان است. این در حالی است که واکنش بیمار به دارو و همچنین اثرات مطلوب و نامطلوب و یا اصطلاحا ناسازگاری دارویی می تواند در افراد مختلف متفاوت باشد. از طرفی بروز و شدت بیماری در افراد مختلف با توجه به دخالت عوامل ژنتیکی، محیطی و اپی ژنتیکی متفاوت است و درمان های استاندارد و معمول برای همه بیماران مفید واقع نمی شود. در راستای حل این مشکلات، پزشکی فردی تلاش می کند تا با استفاده از تکنیك های جدید توالی یابی، ژنوم تمام افراد را توالی یابی و آنالیز کرده و اطلاعات نهفته در ژنوم را مانند یك کارت شناسایی در اختیار هر فرد قرار دهد. در این صورت، با توجه به اطلاعات ژنومی هر فرد، مسیرهای سلولی پاسخ فرد به بیماری و داروهای مختلف و همچنین حساسیت های فرد به بیماری های خاص، بتواند با اقدامات پیشگیرانه مناسب از بروز بیماری هایی که فرد مستعد ابتلا به آنها است جلوگیری و یا در صورت ابتلا داروی مناسب آن فرد تجویز شود. امید است که این علم بتواند راه های جدید و موثرتری برای ساخت دارو، شناسایی پتانسیل ابتلا به بیماری های مختلف در هر شخص در جهت پیشگیری از بروز بیماری و راه های تشخیص درست و زود هنگام آن بر اساس تکیه بر خصوصیات فردی را ارائه نماید (شکل ۶).

ژنوم انسان ها علی رغم شباهت های بسیار زیاد، دارای تفاوت های جزئی با یکدیگر هستند که آنرا منحصر به فرد می نماید. امروزه پیشرفت های چشمگیری در تشخیص و درمان بیماری بر اساس تفاوت های ژنومی افراد در حوزه پزشکی فردی ایجاد شده است. در این میان دانش بیوانفورماتیك با آنالیز، تفسیر اطلاعات و ارتباط دادن داده های امیکس به یکدیگر در کنار آزمایش های بالینی قادر به شناخت تفاوت های افراد و معرفی مارکرهای زیستی مناسب می باشد. مارکرهای زیستی بعنوان یک ابزار کاربردی و از نتایج بسیار مهم این فن آوری قادر به شناسایی پتانسیل فردی برای ابتلا به یک بیماری خاص و یا تشخیص بیماری در یک فرد مبتلا می باشد. در واقع علاوه بر تجهیزات پیشرفته پزشکی برای تعیین فنوتیپ مولکولی و زمینه ژنتیکی افراد، شاخص های زیستی یا مارکرهای زیستی در پزشکی فردی کاربرد بسیاری دارند و امکان تشخیص زود هنگام و غربالگری افراد دارای پتانسیل ابتلا به یک بیماری خاص را فراهم می کنند.

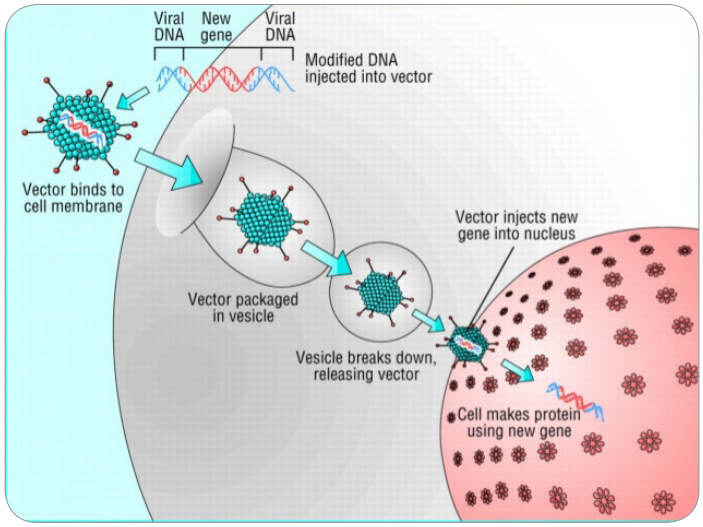


شکل ۶.الف)- برای افراد مختلف نباید از روش های درمانی و داروهای یکسان استفاده شود. ب)- بر اساس تفاوت های ژنتیکی و شرایط محیط زندگی هر فرد، احتمال ابتلا به بیماری های خاص و پاسخ به نوع و دوز دارو در افراد مختلف متفاوت است.

۵-۳. ژن درمانی

ژن درمانی شکل جدیدی از عرضه ی دارو به بدن است. در این روش از سلول های بدن بیمار برای تولید دارو استفاده می شود. این روش شامل انتقال ژن سالم و عملگر به سلول های مناسب در بدن بیمار و تولید مقدار کافی از پروتئین های محصول ژن به منظور درمان دائم بیماری که حاصل از نقص ژنتیکی است، می باشد (شکل ۷).

استراتژی های ژن درمانی شامل انتقال ژن، حذف ژن مضر بوسیله ی توالی های آنتی سنس و کنترل بیان ژن می باشد. امروزه بیماران اندکی از اثرات مثبت ژن درمانی بهره مند شده اند، ولی آینده آن بسیار دلگرم کننده است. تحقیقات چند دهه اخیر در جهت بهبود نواقص ژن درمانی با هدف استفاده از حامل های بی خطر و موثرتر، هدف قرار دادن انواع بیشتری از سلول ها و کاهش پاسخ سیستم ایمنی بیمار بر علیه ژن انتقال یافته می باشد. ژن درمانی به دانش دقیق از ژنوم انسان و تجزیه و تحلیل آن با استفاده از ابزارهای بیوانفورماتیکی نیاز دارد. در نهایت این ابزارها داده های وسیع مربوط به ژنوم را به اطلاعات پروتئوم تبدیل می کنند که شناخت پروتئوم منجر به درک بهتری از فعالیت هورمون ها، ایمونوگلوبولین ها و آنزیم های کلیدی در بدن انسان می شود. امروزه با ظهور تکنولوژی NGS و توالی یابی هدف دار ژن یا ژن های معیوب، کار تشخیص ژن جهش یافته ی عامل بیماری و یا حتی یافتن جهش های جدیدی که می توانند عامل بیماری باشند آسان تر و دقیق تر شده است. ظهور تکنولوژی های جدید همگام با شکل گیری و رشد سریع ابزارها و نرم افزارهای بیوانفورماتیکی مورد نیاز برای آنالیز این حجم وسیع از داده های تولید شده، تحول عظیمی در عرصه ژن درمانی ایجاد کرده است.



شکل ۷. تصویر نشان دهنده ی انتقال ژن جدید از طریق حامل ویروسی به سلول و سپس به هسته ی سلول و نهایتاً بیان ژن جدید و ساخت محصول پروتئینی در طی فرآیند ژن درمانی می باشد.

فصل سوم
پایگـــــاه داده

۱ مقدمه[۱]

اولین توالی آمینواسیدی شناخته شده، توالی هورمون انسولین انسانی بود که در سال ۱۹۵۵ توسط فردریک سنگر مشخص شد. این توالی که متشکل از ۱۱۰ اسید امینه می باشد به صورت زیر است:

MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

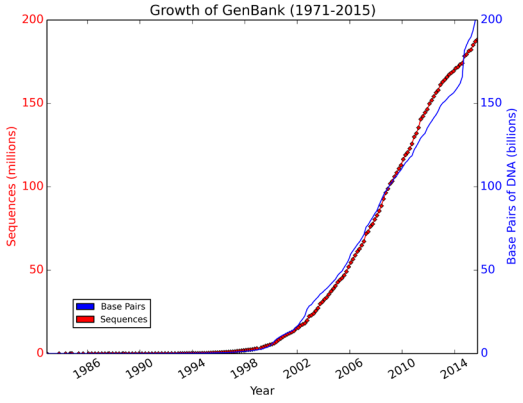
با شناسایی این توالی پروتئینی دوران جدیدی در زیست شناسی ساختاری و مولکولی آغاز شد. در اوایل دهه ۱۹۶۰ تعدادی از توالی های پروتئینی شناسایی شدند و چون کامپیوترهای مناسب برای پردازش آنها هنوز توسعه نیافته بود، اطلاعات توالی ها به صورت دستی جمع آوری و آنالیز می شد. در واقع توالی های پروتئینی و اطلاعات مربوط به آنها بر روی کاغذ نوشته می شد و در کنار هم بر روی دیوار آزمایشگاه ها چسبانده می شدند. با ظهور اولین کامپیوترها (کامپیوترهایی با حافظه ۸ کیلو بایتی)، زیست شناسان شروع به وارد کردن داده ها در کامپیوتر و ذخیره دیجیتالی آنها کردند و به این ترتیب اولین تفکرات برای ایجاد پایگاه های داده یا بانک های اطلاعاتی شکل گرفت. در طی چند دهه گذشته، پیشرفت های قابل توجه در زیست شناسی مولکولی و روش های توالی یابی[۲] منجر به رشد سریع و روزافزون داده های زیستی مانند مقالات، توالی های نوکلئوتیدی و آمینواسیدی، ساختارهای سه بعدی ماکرومولکول ها و غیره شد. این داده ها اطلاعات بسیار ارزشمندی هستند، از این رو دانشمندان برای حفظ و سازمان دهی آنها پایگاه های داده را ایجاد کردند.

پایگاه های داده زیستی یا بانک های اطلاعاتی محل نگهداری و ذخیره بسیار منظم انبوهی از اطلاعات توسط برنامه های کامپیوتری هستند که با روش های بسیار سریع قادر به جستجوی اطلاعات بر اساس کلید واژه می باشند. بسیاری از پایگاه های داده به گونه ای طراحی شده اند که امکان افزودن اطلاعات به آنها یا به روز کردن اطلاعات توسط کاربران وجود دارد. بانک اطلاعات پروتئین در سال ۱۹۷۲ با جمع آوری ساختارهای پروتئینی حاصل از کریستالوگرافی اشعه X شکل گرفت، و پایگاه داده توالی های پروتئینی SWISSPROT در سال ۱۹۸۷ ساخته شد. یکی دیگر از اولین پایگاه های داده زیستی GenBank می باشد که در سال ۱۹۸۲ راه اندازی شد و تا سال ۱۹۸۳ تنها ۲۰۰۰ توالی در

۱ Database

۲ Sequencing

آن ذخیره شده بود ولی به تدریج میزان توالی های موجود در این پایگاه داده بطور تصاعدی افزایش یافت بطوری که در سال ۲۰۱۵ تعداد آن به حدود ۱۸۱ میلیون توالی رسید. شکل ۱ میزان توالی ها[۳] و جفت بازهای[۴] ذخیره شده در پایگاه داده GenBank در طی سال های ۱۹۸۲ تا ۲۰۱۵ نشان می دهد (شکل ۱).



شکل ۱. رشد اطلاعات در پایگاه داده GenBank در طی سال های ۲۰۱۵-۱۹۷۱. رنگ آبی جفت بازها و رنگ قرمز تعداد توالی های موجود در این پایگاه داده را نشان می دهد.

سه پایگاه داده اصلی جهت حفظ و بازیابی اطلاعات وجود دارند که عبارتند از GenBank که توسط مرکز ملی اطلاعات بیوتکنولوژی آمریکا (NCBI)[۵] واقع در انستیتو ملی سلامت (NIH)[۶] اداره می شود، پایگاه داده EMBL[۷] که توسط انستیتو بیوانفورماتیک اروپا (EBI)[۸] اداره می شود و پایگاه داده DNA (DDBJ)[۹] که توسط انستیتو ملی ژنتیک ژاپن مدیریت می شود. داده های جدید هر ۲۴ ساعت بین این سه پایگاه به اشتراک گذاشته و مبادله می شوند. بطور کلی پایگاه‌های داده زیستی به دو دسته پایگاه های اولیه[۱۰] و ثانویه[۱۱] طبقه بندی می شوند، پایگاه های اطلاعاتی اولیه حاوی توالی های نوکلئوتیدی و آمینواسیدی هستند و پایگاه های اطلاعاتی ثانویه واجد اطلاعات استخراج شده از پایگاه های اطلاعاتی اولیه هستند.

---

۳ Sequence

۴ Base pair

۵ National Center for Biotechnology Information

۶ National Institute of Health

۷ European Molecular Biology Laboratory

۸ European Bioinformatics Institute

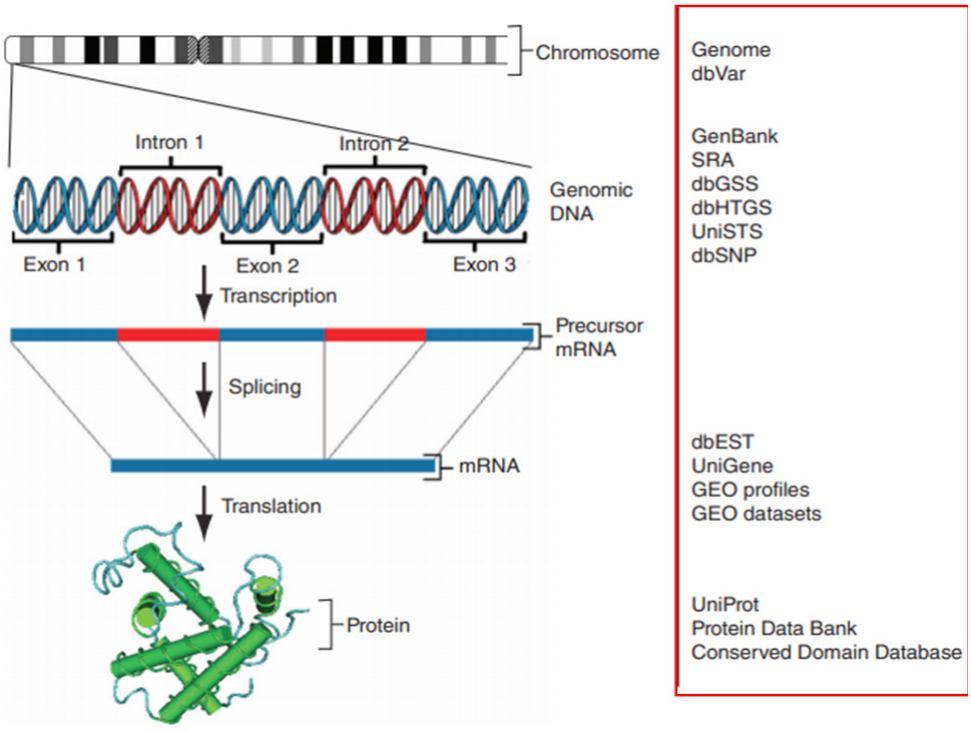
۹ DNA Database of Japan

۱۰ Primary database

۱۱ Secondary database

پایگاه داده چند شکلی های تک نوکلئوتیدی (SNP)[۱۲]، پایگاه داده ساختارهای سه بعدی پروتئین ها و پایگاه داده مسیرهای متابولیسمی مثال هایی از پایگاه های اطلاعاتی ثانویه می باشند.

انواع داده های ذخیره شده در پایگاه های داده مختلف یادآور اصل مرکزی[۱۳] زیست شناسی است که بر اساس آن، DNA ژنومی سازمان یافته در کروموزوم ها به پیش ساز[۱۴] mRNA رونویسی می شود. سپس این نسخه RNA به نسخه mRNA تبدیل شده و در نهایت به پروتئین ترجمه می شود. اطلاعات بسیار زیادی در سطوح متفاوت (از ژن تا پروتئین) در پایگاه های داده مختلف نگهداری می شوند (شکل ۲).



شکل ۲. انواع داده های ذخیره شده در پایگاه های داده مختلف (ستون راست) یاد آور اصل مرکزی زیست شناسی است.

۲ پایگاه داده GenBank

یکی از کامل ترین و شناخته شده ترین پایگاه های داده زیستی، پایگاه داده GenBank یا NCBI است که حاوی اطلاعات متنوعی از اسید های نوکلئیک و پروتئین ها است. علاوه بر ذخیره داده، GenBank حاوی انبوهی از مقالات علمی و کتاب های مرجع در علم زیست شناسی و رشته های مرتبط نیز می باشد.

۱۲ Single nucleotide polymorphism

۱۳ Central dogma

۱۴ precursor mRNA

۲-۲. انواع داده های موجود در GenBank

۲-۲-۱. اطلاعات مربوط به DNA ژنومی

توالی نوکلئوتیدی تمام ژنوم های توالی یابی شده موجودات مختلف در پایگاه داده Genome که زیر مجموعه ای از GenBank است، وجود داد. همچنین انواع تغییرات ساختاری در سطح ژنوم مانند درج شدگی[۱۵]، حذف[۱۶]، مضاعف شدگی ها [۱۷]، وارونگی[۱۸] و بازآرایی های کروموزومی[۱۹] در پایگاه داده ای به نام dbVar (db مخفف database است و var از variation به معنای تغییرات گرفته شده است) که در مجموعه GenBank قرار دارد، جمع آوری شده است.

۲-۲-۲. اطلاعات مربوط به STS[۲۰]

STSها بخش های کوتاه (معمولاً به طول ۵۰۰ جفت باز) و شناخته شده در سطح DNA هستند که تنها یک بار در ژنوم وجود دارند و به دلیل همین یکتایی بعنوان جایگاه های ویژه[۲۱] در نظر گرفته شده اند و در تهیه نقشه های ژنتیکی و فیزیکی ژنوم به کار می روند. این نواحی ژنومی به راحتی با استفاده از پرایمرهای اختصاصی و واکنش PCR قابل شناسایی هستند. توالی های STS در گذشته در پایگاه داده STS (db STS) که زیرمجموعه ای از GenBank است، نگهداری می شد ولی در حال حاضر این اطلاعات از طریق جستجو در پایگاه داده نوکلئوتید (در ادامه توضیح می دهیم) قابل دسترس هستند.

۲-۲-۳. اطلاعات مربوط به GSS [۲۲]

توالی های GSS توالی های کوتاهی هستند که حاصل یک بار[۲۳] توالی یابی انتهای کلون[۲۴] های کاسمید[۲۵]، BAC[۲۶] و یا YAC[۲۷] هستند. توالی هایی مانند AFLP[۲۸] و RFLP[۲۹] مثال هایی از توالی های GSS می باشند. بطور کلی GSSها مشابه با توالی های EST[۳۰] هستند با این تفاوت که برخلاف ESTها که منشا آنها مولکول mRNA است، منشا GSSها DNA ژنومی است. این توالی ها در پایگاه داده dbGSS که زیر مجموعه ای از GenBank است، نگهداری می شوند.

---

۱۵ Insertion
۱۶ Deletion
۱۷ Duplication
۱۸ nversion
۱۹ Chromosomal rearrangements
۲۰ Sequence-Tagged Sites
۲۱ Landmark
۲۲ Genome Survey Sequences
۲۳ Single-Pass
۲۴ Clone
۲۵ Casmid
۲۶ Bacterial artificial chromosome
۲۷ Yeast artifical chromosome
۲۸ Amplification fragment length polymorphism
۲۹ Restriction fragment length polymorphism
۳۰ Expressed sequence tag

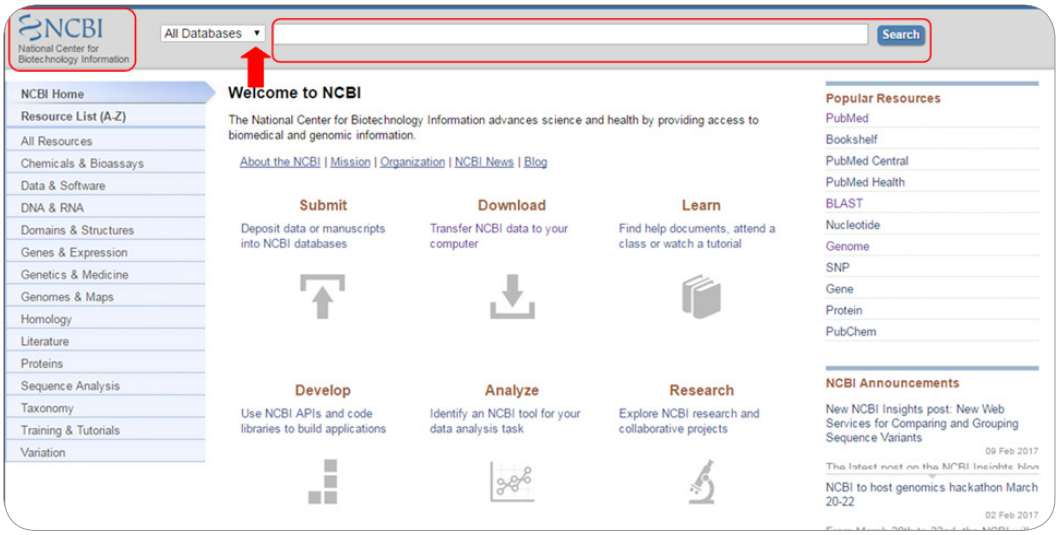
۲-۲-۴. اطلاعات مربوط به EST [۳۱]

ESTها توالی های نوکلئوتیدی کوتاهی هستند (معمولاً به طول ۳۰۰ تا ۸۰۰ جفت باز) که حاصل توالی یابی یک یا هر دو انتهای کلون های cDNA می باشند. برای ایجاد این توالی ها، مولکول های mRNA به فرم پایدارتر یعنی cDNA تبدیل می شوند که cDNA را می توان به راحتی کلون و تعیین توالی کرد. بنابراین ESTهای یک بافت معین نماینده ژن های بیان شده در آن بافت هستند و از این رو از آنها می توان جهت شناسایی ژن های جدید استفاده کرد. در صورتی که ESTها حاصل توالی یابی ابتدای کلون cDNA باشند 5´-EST و در صورتی که محصول توالی یابی انتهای کلون های cDNA باشند، 3´-EST نامیده می شوند. ESTهای موجود در GenBank در حال حاضر در سه گروه اصلی انسان، موش و سایر موجودات طبقه بندی شده اند.

علاوه بر این پایگاه های داده، اطلاعات مربوط به ژن ها، SNPها، پروتئین های مختلف و همچنین اطلاعاتی بیانی حاصل از آزمایشات مختلف انجام شده در دنیا و ... در پایگاه داده بزرگ GenBank حفظ و ذخیره شده اند که توسط کاربران قابل دسترس می باشند.

۳ معرفی صفحه اصلی NCBI [۳۲]

برای مشاهده این صفحه کافی است کلمه NCBI را در موتور جستجوی google وارد کنید و یا به آدرس https://www.ncbi.nlm.nih.gov بروید. صفحه ای مشابه با شکل ۳ مشاهده خواهید کرد، به لوگوی NCBI در بالای سمت چپ صفحه توجه کنید. با وارد کردن کلمات کلیدی مناسب، نام ژن یا پروتئین در کادر مستطیل شکل مشخص شده در شکل می توانید در این پایگاه داده به جستجو بپردازید.

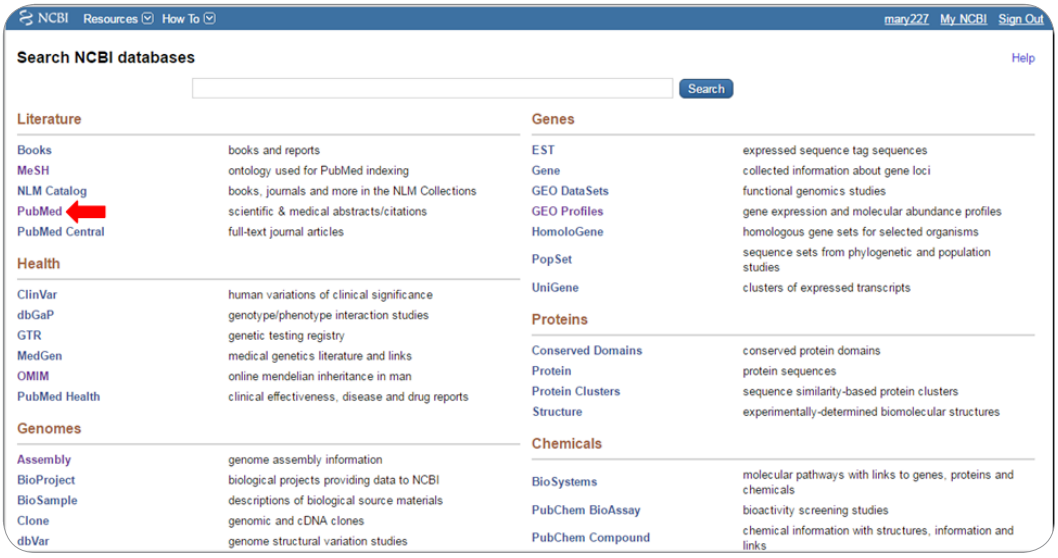


شکل ۳. صفحه اصلی پایگاه داده NCBI

۳۱ Expressed sequence tag

۳۲ Home page

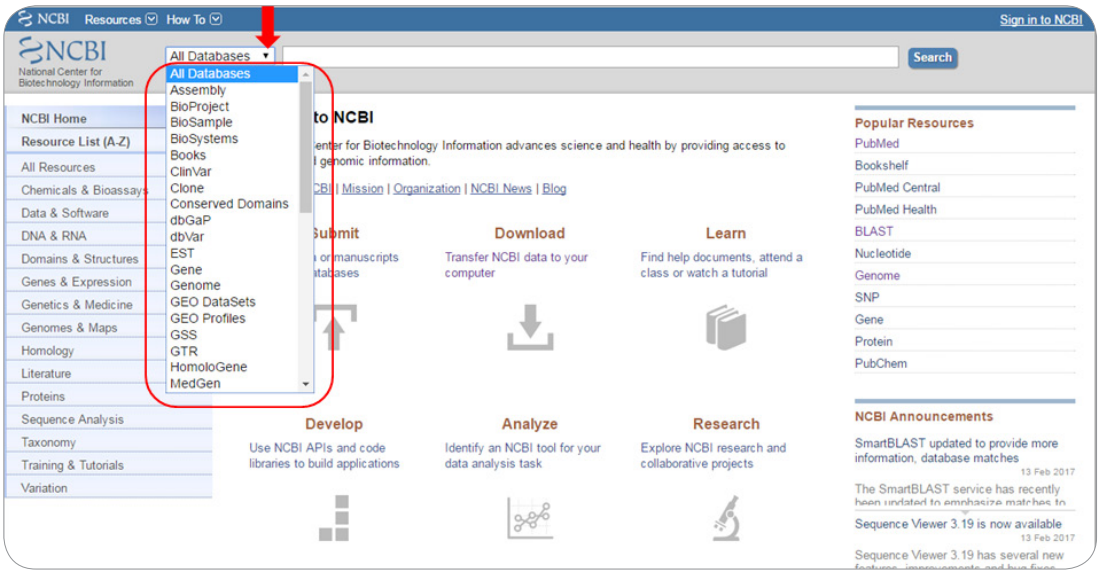
در صورتی که بر روی فلش رو به پایین (با پیکان در شکل ۳ مشخص شده است) کلیک کنید می توانید انواع پایگاه های داده‌ای که زیرمجموعه NCBI یا همان GenBank هستند را مشاهده، انتخاب کرده و به صورت اختصاصی در آنها جستجو کنید. در صورتی که هیچ موردی را در این قسمت انتخاب نکنید، NCBI به صورت پیش فرض[۳۳] در تمام پایگاه های داده (All database) جستجو را انجام می دهد. NCBI از موتور جستجوی Entrez برای جستجو استفاده می کند. برای دسترسی به صفحه اصلی این موتور جستجو، کافی است در صفحه اول NCBI (شکل ۳)، بدون وارد کردن کلمه کلیدی بر روی Search (در مقابل کادر مستطیل شکل مشخص شده در شکل ۳)کلیک کنید، صفحه ای مشابه با شکل ۴ باز خواهد شد که تمام پایگاه های داده ای که NCBI جستجو را در آن انجام می دهد را می توانید مشاهده کنید. این پایگاه های داده به صورت هایپرلینک هستند و با کلیک بر روی آنها می توانید وارد صفحه اصلی آن پایگاه داده شوید (شکل ۴).



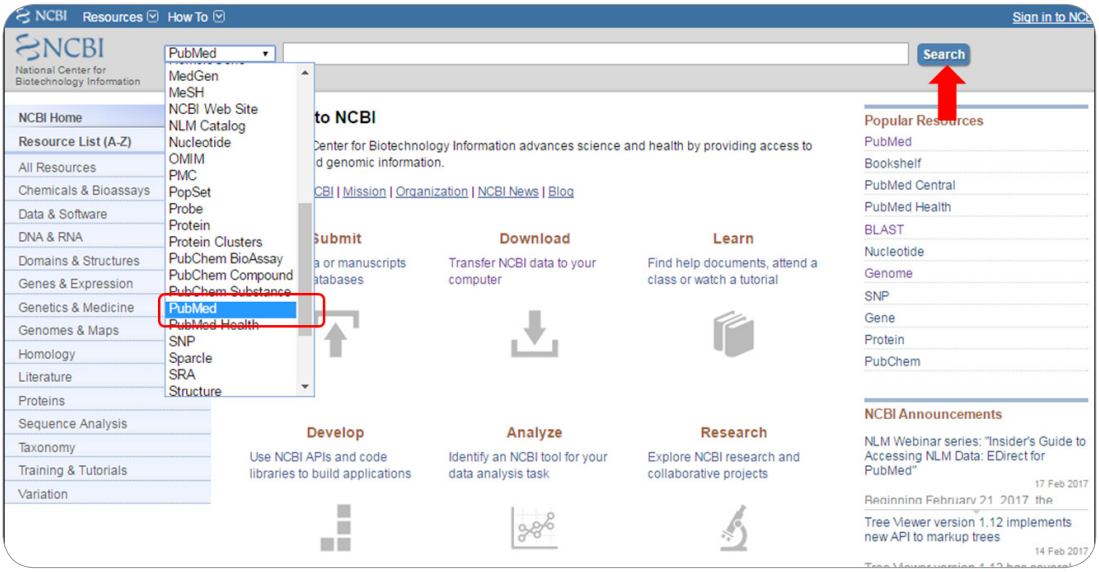
شکل ۴. صفحه اصلی موتور جستجوی Entrez، موتور جستجوی NCBI

برای مثال در صورتی که بر روی PubMed (با پیکان قرمز رنگ در شکل ۴ مشخص شده) کلیک کنیم وارد صفحه اصلی این پایگاه می شویم. علاوه بر این روش دسترسی به پایگاه های داده مختلف، در صورتی که در صفحه اصلی NCBI بر روی فلش کوچک رو به پایین در کنار گزینه All database کلیک کنید، منویی برای شما باز خواهد شد (شکل ۵) که در آن می توانید پایگاه داده مورد نظر را انتخاب کنید و بطور اختصاصی در پایگاه داده انتخاب شده به جستجو بپردازید. برای مثال ما در اینجا پایگاه داده PubMed را از این منو انتخاب کردیم و بر روی search کلیک کردیم (شکل ۶).

۳۳ Default



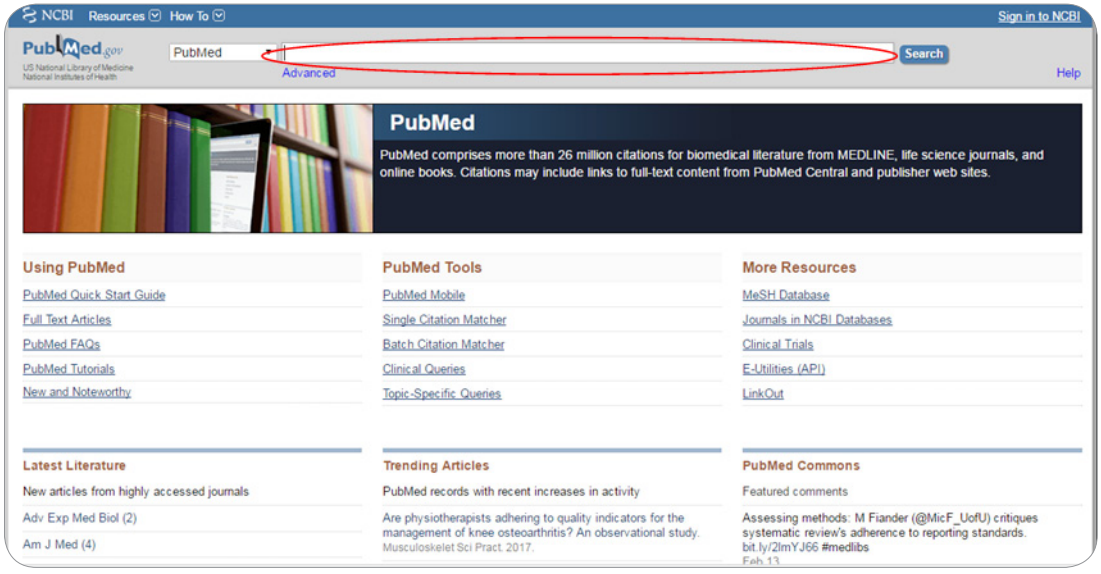
شکل ۵. کادر مشخص شده در شکل انواع پایگاه های داده موجود در NCBI را نشان می دهد



شکل ۶. انتخاب پایگاه داده PubMed از طریق صفحه اصلی NCBI

۴ آشنایی با پایگاه داده PubMed

پایگاه داده PubMed سرویس جستجو مقالات علمی است و متعلق به کتابخانه ملی پزشکی آمریکا (NLM)[۳۴] می باشد که مقالات یا خلاصه ی مقالات چاپ شده در مجلات علمی مختلف در آن ذخیره شده و قابل جستجو می باشد. روش های ورود به یک پایگاه داده از جمله PubMed در بالا توضیح داده شد، پس از ورود به این پایگاه داده با صفحه اصلی این پایگاه مواجه خواهید شد (شکل ۷). با وارد کردن کلمات کلیدی مناسب در کادر مستطیل شکل مشخص شده در شکل ۷ می توانید در این پایگاه داده مقالات مورد نظر را جستجو کنید.



شکل ۷- صفحه اصلی پایگاه داده PubMed

قبل از اینکه در این پایگاه داده جستجو کنیم، مروری بر روش های جستجو خواهیم داشت. به منظور انجام یک جستجو مناسب و اختصاصی لازم است تا از برخی عملگرها[۳۵] و توصیف گرها[۳۶] در جستجوی خود استفاده کنید. به عباراتی که میان فیلدهای جستجو به کار می رود، عملگر می گویند که به عملگرهای بولن[۳۷] نیز معروف هستند. سه نوع عملگر وجود دارد که عبارتند از:

AND : در صورت استفاده از این عملگر هر دو عبارت موجود در دو طرف AND، در نتایج جستجو وجود خواهد داشت. مثال: CIPK AND CDPK

OR : در صورت استفاده از این عملگر، یکی از دو عبارت سمت چپ یا راست OR در نتایج جستجو وجود خواهد داشت. مثال: CIPK OR CDPK

۳۴ National Library of Medicine

۳۵ Operators

۳۶ Qualifiers

۳۷ Boolean operator

NOT : با استفاده از این عملگر عبارت سمت راست NOT در نتایج جستجو وجود نخواهد داشت.

مثال: CIPK NOT CDPK

حتماً با استفاده از حروف بزرگ عملگرهای ذکر شده را در کادر جستجو تایپ کنید.

به عبارت هایی که درون کروشه قرار گرفته و جستجو را محدودتر و اختصاصی تر می نمایند، توصیف گر می‌گویند. در جدول زیر برخی از مهم ترین توصیف گرها همراه با مفهوم و علامت اختصاری آنها نشان داده شده است (جدول ۱).

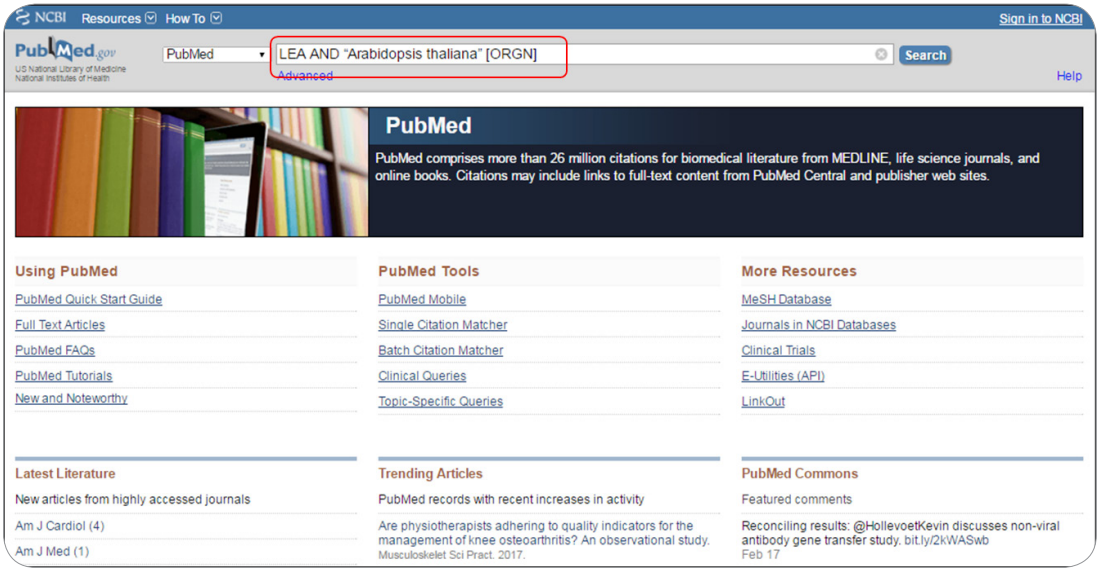
مفهوم	توصیف گر	علامت اختصاری
کد دسترسی به توالی نوکلئوتیدی	Accession	[ACCN]
موسسات و مراکز تحقیقاتی	Affiliation	[AD]
تمام فیلدها	All Fields	[ALL]
نام نویسنده	Author Name	[AU]
تاریخ ثبت مقاله	Entrez Date	[EDAT]
زبان	Language	[LA]
نام موجود زنده	Organism	[ORGN]
تاریخ چاپ مقاله	Publication Date	[PDAT]
نام پروتئین	Protein Name	[PRPT]

جدول۱. مهمترین توصیف گرهای مورد استفاده در جستجو

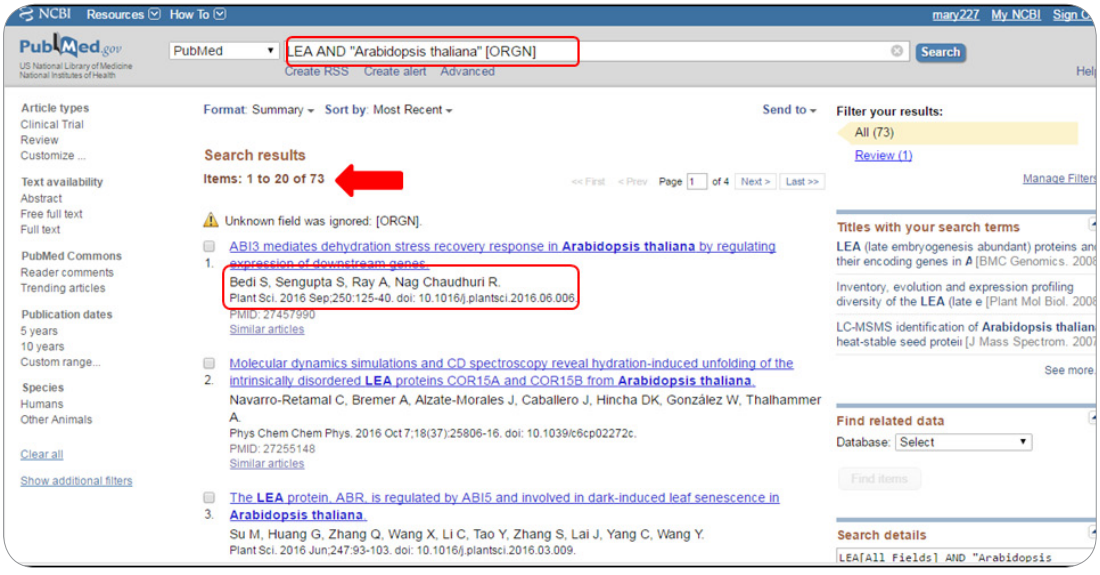
برای مثال فرض کنید که می خواهیم مقالاتی که در رابطه با پروتئین LEA[۳۸] در گیاه ارابیدوپسیس تالیانا منتشر شده است را در پایگاه PubMed جستجو کنیم. کافی است عبارت LEA AND "*Arabidopsis thaliana*"[ORGN] را در کادر مربوط به جستجو این پایگاه داده وارد کنیم (شکل ۸) در این عبارت، توصیف گر [ORGN] به موتور جستجو Entrez می‌گوید که *Arabidopsis thaliana* نام یک ارگانیسم هست. به همین ترتیب در صورتی که مقالات یک نویسنده خاص مد نظرتان باشد، کافی است که در کنار کلمه کلیدی خود نام نویسنده مورد نظر را همراه با توصیف گر [AU] که نشان دهنده ی Author Name است، استفاده کنید.

همانطور که در شکل ۹ مشاهده می کنید، این پایگاه داده ۷۳ مقاله را بعنوان نتیجه جستجو به ما ارائه داد، عنوان تمام مقالات به صورت هایپرلینک آبی رنگ هستند که با کلیک بر روی آنها می توانیم وارد صفحه مربوط به این مقاله در پایگاه داده PubMed شویم. در زیر عناوین مقالات، اطلاعاتی مانند نویسندگان مقاله، مجله ای که این مقاله در آن منتشر شده است، سال انتشار آن و همچنین کد DOI مقاله را ملاحظه می کنید.

۳۸ Late abundant embryogenesis

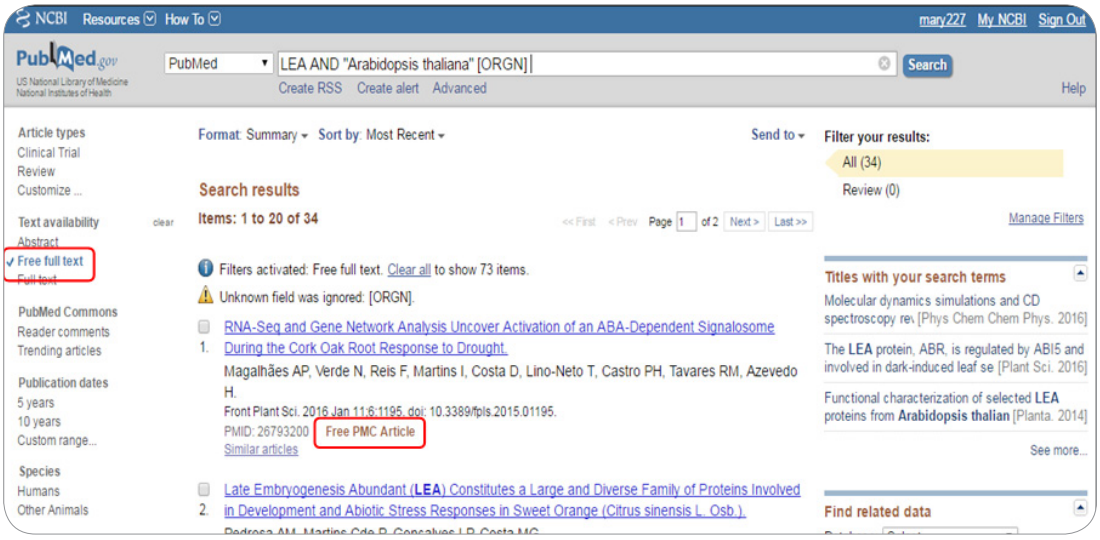


شکل ۸. جستجو مقاله در پایگاه داده PubMed



شکل ۹. صفحه نتایج حاصل از جستجو در پایگاه داده PubMed

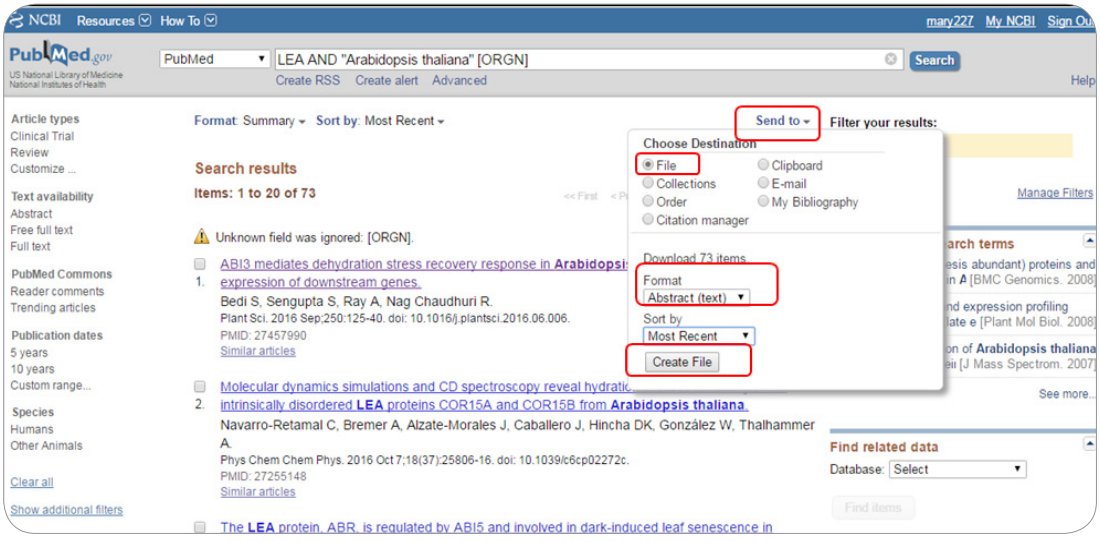
در صورتی که بخواهید تنها مقالاتی که متن کامل آنها به صورت رایگان در دسترس است، برای شما نمایش داده شود کافی است بر روی «Free full text» در سمت چپ صفحه نتایج کلیک نمایید. در این صورت تنها این مقالات برای شما نشان داده می شود که در ذیل عنوان هر یک نیز عبارت «Free PMC article» قابل مشاهده است (شکل ۱۰).



شکل ۱۰. انتخاب مقالات رایگان با متن کامل در میان نتایج حاصل از جستجو

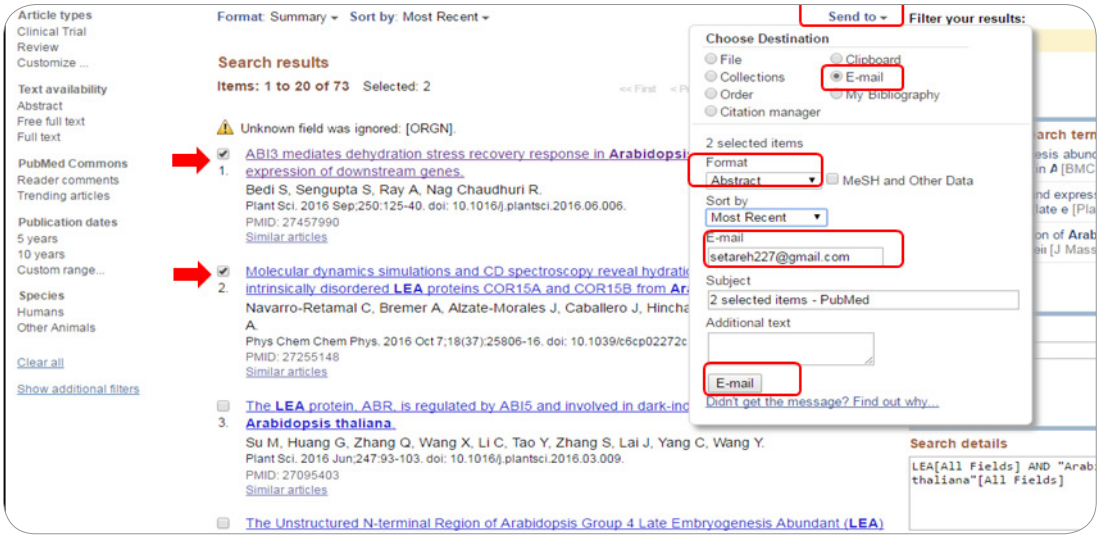
۱-۴. ذخیره نتایج حاصل از جستجو در پایگاه اطلاعاتی PubMed

در نتایج حاصل از جستجو در پایگاه داده PubMed، در سمت چپ عنوان هر مقاله یک مربع کوچک وجود دارد که از طریق آن می توانید مقالات مورد نظر خود را برای ذخیره انتخاب کنید، در صورتی که هیچ موردی انتخاب نشود، NCBI به صورت پیش فرض خلاصه تمام مقالات یافت شده را برای شما ذخیره خواهد کرد. برای ذخیره مقالات مورد نظر می توانید بر روی فلش رو به پایین در کنار Send to در سمت راست صفحه نتایج کلیک کرده، سپس گزینه File و در قسمت Format ، Abstract (txt) را انتخاب کرده و در نهایت بر روی گزینه Create File کلیک نمایید. به این ترتیب خلاصه مقالات برای شما در کامپیوترتان ذخیره خواهد شد. (شکل ۱۱).



شکل ۱۱. ذخیره نتایج حاصل از جستجو در پایگاه داده PubMed

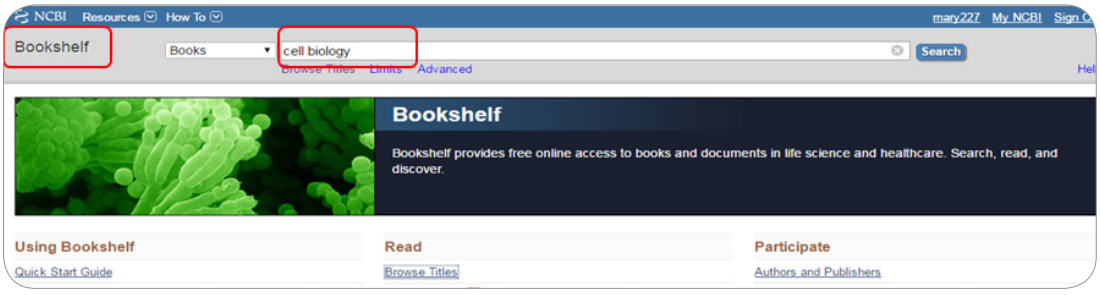
شما می توانید نتایج حاصل از جستجو را به ایمیل شخصی خود نیز ارسال نمایید. به این منظور بر روی فلش رو به پایین در کنار Send to کلیک کنید و گزینه E-mail را انتخاب و سپس در قسمت Format، مثلا Abstract (txt) را انتخاب کرده و در قسمت بعد نیز ایمیل خود را وارد کرده و در نهایت بر روی گزینه E-mail کلیک نمایید. برای مثال ما در شکل ۱۲ تنها ۲ مقاله را انتخاب کرده و خلاصه آنها را به ایمیل مورد نظر ارسال کردیم.



شکل ۱۲. ارسال خلاصه مقالات انتخاب شده (موارد ۱ و ۲) به ایمیل شخصی

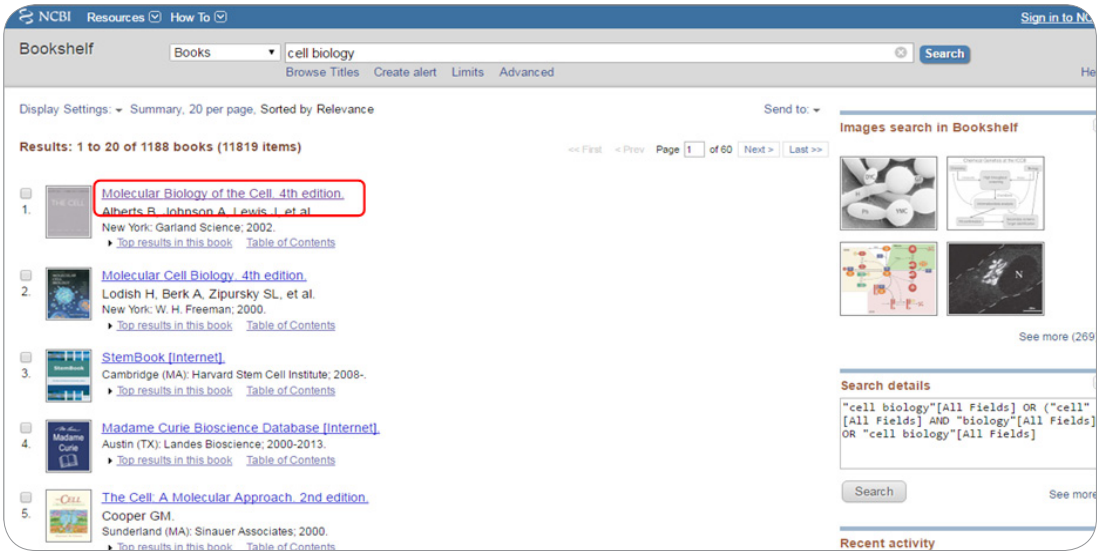
۵ آشنایی با پایگاه داده Bookshelf

پایگاه داده Bookshelf مرجع رایگان کتاب های زیستی است. برای دسترسی به صفحه اصلی این پایگاه داده، از فلش رو پایین در صفحه اصلی NCBI گزینه Books را انتخاب کنید و بدون وارد کردن هیچ کلمه ای بر روی گزینه Search کلیک نمایید. صفحه ای مشابه با شکل ۱۳ برای شما نمایش داده می شود که واژه Bookshelf در گوشه سمت چپ آن مشاهده می شود. اکنون در کادر مستطیلی شکل مربوطه نام کتاب مورد نظرخود را وارد کنید و بر روی Search کلیک نمایید. برای مثال ما در اینجا "cell biology" را وارد کردیم تا کتاب های مرتبط با زیست شناسی سلولی را جستجو کنیم (شکل ۱۳).

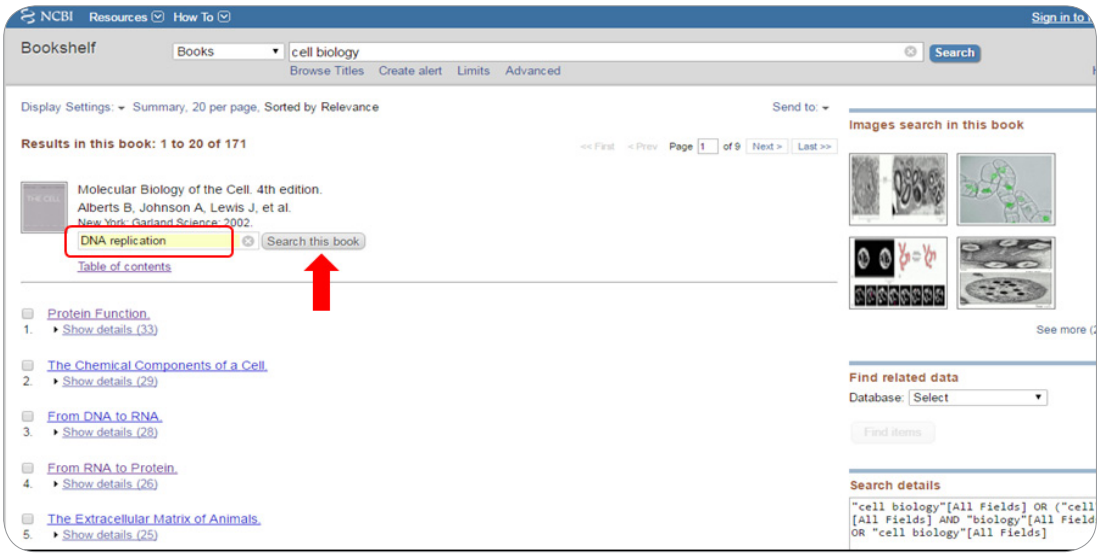


شکل ۱۳. صفحه اصلی پایگاه داده Bookshelf و جستجو کتاب در آن

در نتایج حاصل از جستجو در این پایگاه داده، لیستی از کتاب های مرتبط با زیست شناسی سلولی برای شما نمایش داده می شود (شکل ۱۴). در صورتی که مثلاً بر روی اولین گزینه که کتاب «Molecular Biology of the Cell» می باشد، کلیک نمایید وارد کتاب مورد نظر می شوید که در آن می توانید کلمات کلیدی مناسب را در کادر مشخص شده در شکل وارد کرده و با کلیک بر Search this book، آنها را در این کتاب جستجو نمایید. برای مثال ما در اینجا با وارد کردن کلمات DNA replication به دنبال مطالبی در مورد همانند سازی DNA در این کتاب هستیم (شکل ۱۵).



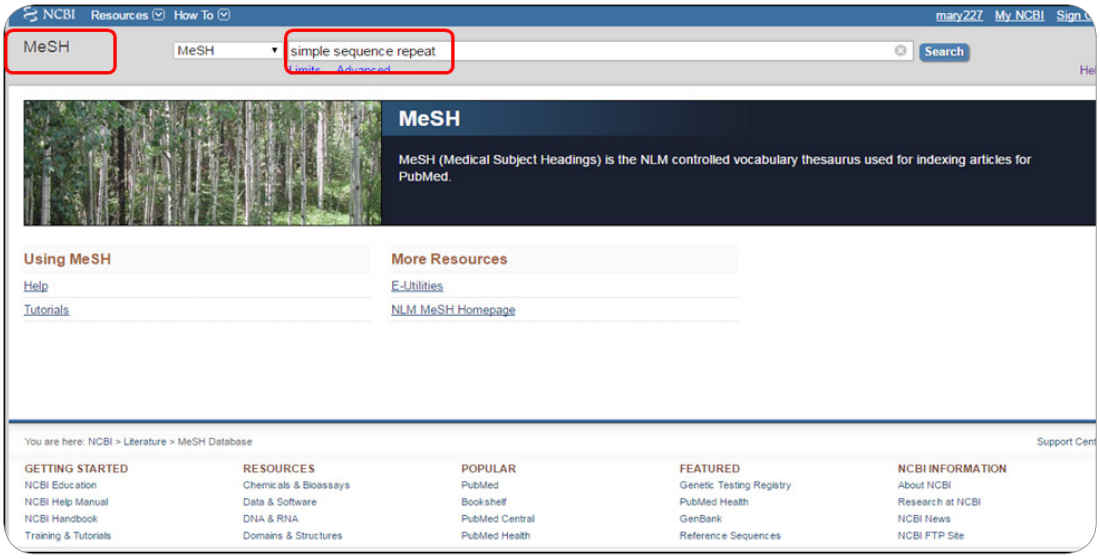
شکل ۱۴. نتایج حاصل از جستجو در پایگاه داده Bookshelf



شکل ۱۵ جستجو مطالب مورد نظر در کتاب انتخاب شده

۶ آشنایی با پایگاه اطلاعاتی MeSH

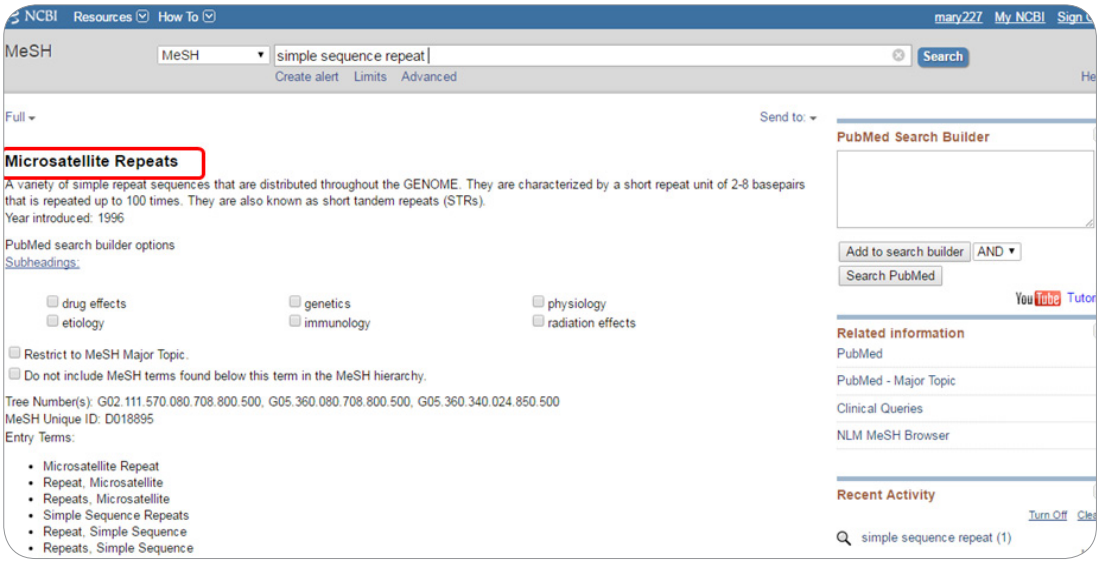
پایگاه اطلاعاتی MeSH، فرهنگ لغت NCBI است، اگر کلمه ای را در این پایگاه جست و جو کنید معنی و مفهوم آن برای شما نمایش داده می شود. برای دسترسی به صفحه اصلی این پایگاه داده مانند آنچه که قبلاً توضیح داده شد، بر روی فلش رو به پایین در کنار گزینه All database در صفحه اصلی NCBI کلیک کرده و از منو باز شده گزینه MeSH را انتخاب کنید و بدون وارد کردن هیچ کلمه ای بر روی گزینه Search کلیک نمایید. صفحه ای مشابه با شکل ۱۶ برای شما نشان داده می شود که در کادر مستطیلی شکل مشخص شده می توانید کلمات یا واژگان مورد نظر را وارد کنید. برای مثال، ما در اینجا عبارت «simple sequence repeat» را وارد کردیم.



شکل ۱۶. صفحه اصلی پایگاه داده MeSH و جستجو در آن

نتایج این جستجو در صفحه ای مشابه شکل ۱۷ برای شما ارائه می شود. همانطور که ملاحظه می کنید بر اساس توضیحات مندرج در این صفحه، simple sequence repeat همان تکرارهای ریزماهواره‌ای هستند که در سراسر ژنوم گسترده شدند. این تکرارها، تکرارهای کوتاه به طول ۸-۲ جفت باز هستند که تا ۱۰۰ بار در ژنوم تکرار شدند و همچنین به تکرارهای پشت سر هم کوتاه یا STR[۳۹] نیز معروف هستند.

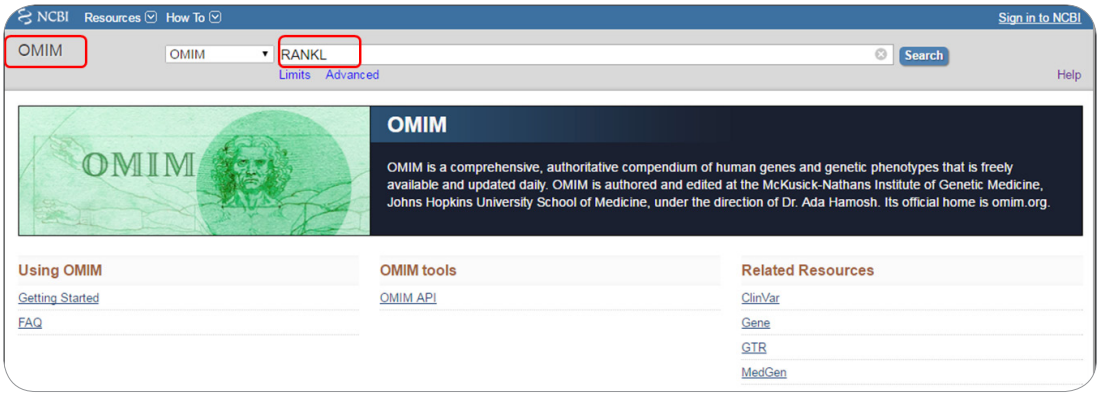
۳۹ Short tandem repeats



شکل ۱۷. نتایج حاصل از جستجو در پایگاه داده MeSH

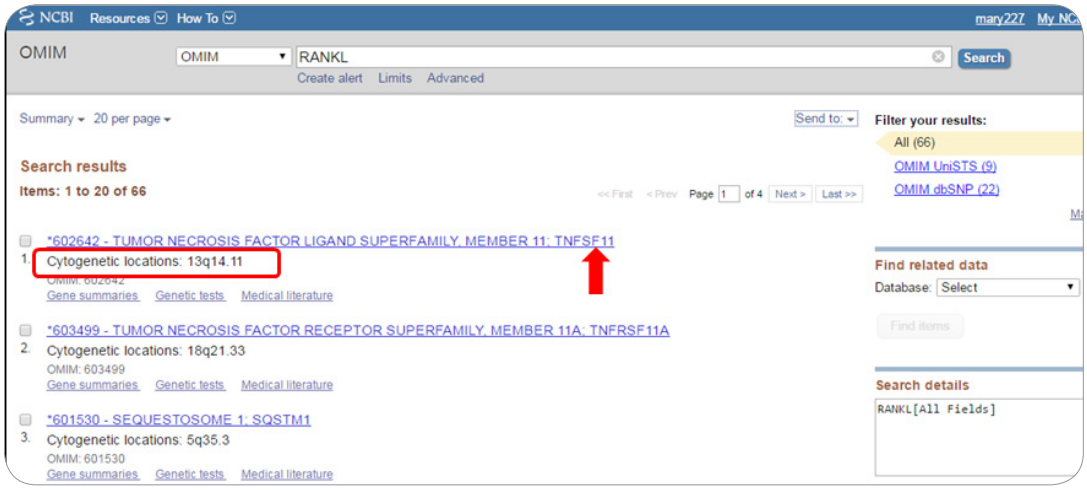
۷ آشنایی با پایگاه اطلاعاتی OMIM (Online Mendelian Inheritance in Man)

پایگاه داده OMIM مجموعه ای جامع، معتبر و به روز از ژن های انسانی و اختلالات ژنتیکی است. در این پایگاه داده می توان اطلاعاتی نظیر جایگاه کروموزومی ژن بیماری زا مورد نظر، عملکرد آن، نوع بیماری ایجاد شده توسط آن، نحوه توارث این بیماری و ... را دریافت کرد. برای دسترسی به صفحه اصلی این پایگاه داده مانند آنچه قبلا توضیح داده شد عمل کنید و از فلش رو به پایین در صفحه اصلی NCBI، گزینه OMIM را انتخاب کنید و بدون وارد کردن کلمه بر روی گزینه Search کلیک نمایید. صفحه ای مشابه با شکل زیر برایتان نمایش داده می شود که در کادر مستطیل شکل مربوطه می توانید نام ژن مورد نظر خود را وارد کنید. برای مثال، ما در اینجا می خواهیم بدانیم آیا ژن *RANKL* باعث ایجاد بیماری در انسان می شود یا خیر و در صورت بیماری زا بودن، نوع بیماری ایجاد شده توسط آن و نحوه توارث بیماری را دریابیم. بنابراین، نام این ژن (*RANKL*) را در کادر مربوطه وارد کردیم (شکل ۱۸).

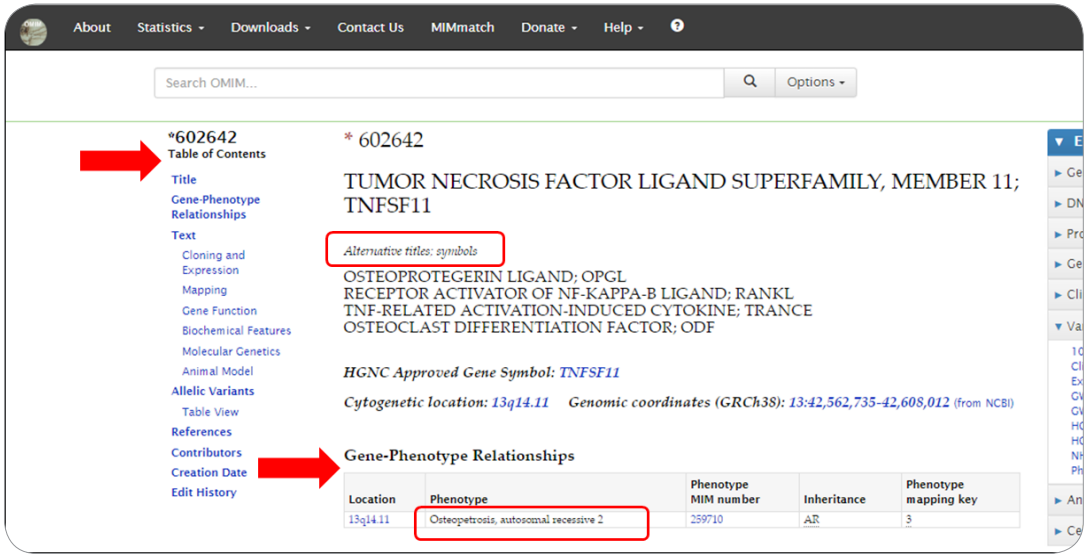


شکل ۱۸. جستجو در پایگاه داده OMIM

نتایج حاصل از جستجو در صفحه ای مانند شکل زیر نمایش داده می شود که نشان می دهد ژن *RANKL* ژنی است که می تواند باعث ایجاد بیماری در انسان شود (شکل ۱۹). در صورتی که به اولین خط در صفحه نتایج دقت نمایید، متوجه خواهید شد که ژن *RANKL* متعلق به ابرخانواده tumor necrosis factor ligand است و نام دیگر آن *TNFSF11* نیز می باشد. در زیر این خط جایگاه کروموزومی ژن مذکور نوشته شده است (در شکل ۱۹ مشخص شده)، این ژن بر روی بازوی بلند (q) کروموزوم شماره ۱۳ انسان قرار گرفته است. در صورتی که بر روی هایپرلینک آبی رنگ کلیک کنیم وارد صفحه ای می شویم که اطلاعات جزئی تر در مورد این ژن و محصول پروتئینی آن و همچنین بیماری ایجاد شونده توسط آن قید شده است (شکل ۲۰).



شکل ۱۹. نتایج حاصل از جستجو در پایگاه داده OMIM



شکل ۲۰. مشاهده اطلاعات جزیی تر و کاملتر در مورد ژن *RANKL* و بیماری ایجاد شونده توسط آن

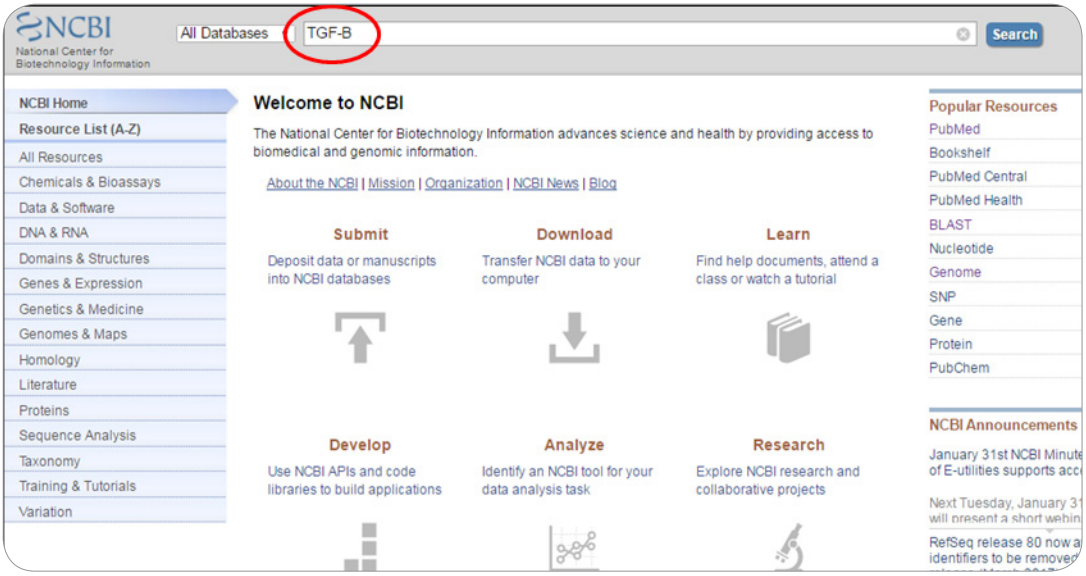
همانطور که در شکل ۲۰ مشاهده می کنید، در سمت چپ صفحه ستونی با عنوان Table of contents وجود دارد که در آن اطلاعات متنوعی در رابطه با ژن مورد نظر را می توان یافت، تمام موارد به صورت هایپرلینک هستند و در صورت کلیک بر آنها به توضیحات مربوط به هر کدام هدایت خواهید شد. برای مثال در صورتی که بر روی «Gene-Phenotype Relationship» کلیک کنیم، توضیحات این بخش برای ما نمایش داده می شود (با پیکان قرمز رنگ در شکل مشخص شده) که بیانگر جایگاه کروموزومی ژن مورد نظر است و نیز نشان می دهد که این ژن در ایجاد بیماری «Osteopetrosis» درگیر است و نحوه توارث بیماری مذکور به صورت اتوزومال مغلوب می باشد. همچنین سایر نام های این ژن در قسمت «Alternative titles; symbols» (با پیکان قرمز رنگ در شکل مشخص شده) نمایش داده شده است (شکل ۲۰).

۸ بازیابی اطلاعات مربوط به یک ژن معین در پایگاه داده NCBI

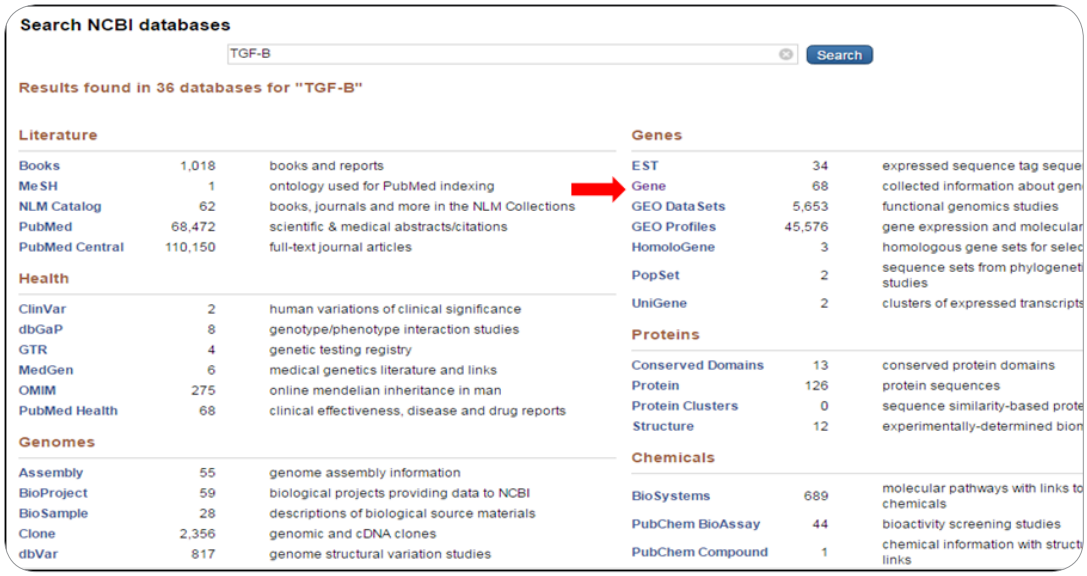
همانطور که در ابتدا ذکر شد اطلاعات بسیار متنوعی می توان در مورد یک ژن خاص از پایگاه داده NCBI دریافت کرد که بطور خلاصه شامل موارد ذیل هستند:

۱. شناسایی محل ژن بر روی کروموزوم و شناسایی ژن های اطراف آن
۲. شناسایی واریانت های بیانی ژن (نسخه های متفاوت mRNA مرتبط با یک ژن خاص)
۳. شناسایی جهش های موجود بر روی ژن (SNP)
۴. شناسایی پروتئین های کد شونده توسط ژن
۵. شناسایی ساختار پروتئین های مربوطه
۶. بررسی دومین های پروتئین های یک ژن
۷. مقالاتی که در مورد آن ژن منتشر شده است

اکنون برای بررسی موارد بالا، برای مثال ژن، *TGF-β* که فاکتور تغییر دهنده رشد را کد می کند در نظر می گیریم. برای یافتن اطلاعات موجود درباره این ژن، در صفحه اصلی NCBI، نام این ژن را وارد کنید و سپس بر روی Search کلیک نمایید (شکل ۲۱). موتور جستجوی NCBI به صورت پیش فرض در تمام پایگاه های داده (All databases) جستجو را انجام می دهد و نتایج را در صفحه ای مشابه با شکل ۲۲ نمایش می دهد (شکل ۲۲).



شکل ۲۱. جستجو ژن *TGF-B* در پایگاه داده GenBank



شکل ۲۲. نتایج حاصل از جستجو ژن *TGF-B* در پایگاه داده NCBI

همانطورکه شکل ۲۲ نشان می دهد، اطلاعات بسیاری در رابطه با ژن *TGF-β* در این پایگاه داده وجود دارد که در بخش های مختلف تحت عناوینی مانند «Proteins»، «Genes»، «Genomes»، «Health»، «Literature» و «Chemicals» طبقه بندی و سازماندهی شده اند. به قسمت Gene توجه کنید (با پیکان قرمز رنگ در شکل

مشخص شده)، ۶۸ ژن *TGF-β* در پایگاه داده NCBI موجود است. در صورتی که بر روی Gene کلیک کنید، وارد صفحه ای مشابه با شکل ۲۳ می شوید. به کمک موس به سمت پایین صفحه حرکت کنید و بر روی *TGFB1* انسانی (Homo sapiens یا human) کلیک کنید تا به اطلاعاتی از این ژن در انسان دست یابیم (شکل ۲۳).

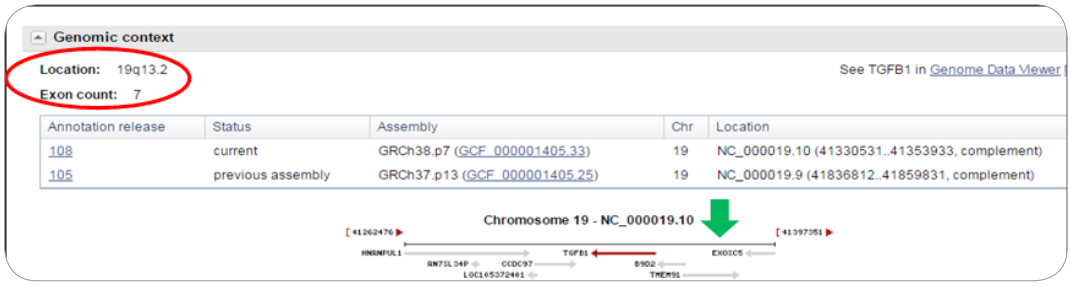
Acvrl1 ID: 11482	activin A receptor, type II-like 1 [*Mus musculus* (house mouse)]	Chromosome 15, NC_000081.6 (101128522..101145336)	AI115505, AI427544, Acvrlk1, Alk1
Acvrl1 ID: 25237	activin A receptor like type 1 [*Rattus norvegicus* (Norway rat)]	Chromosome 7, NC_005106.4 (142769942..142787336)	R-3, R3, SETHKIR
Acvr1 ID: 79558	activin A receptor type 1 [*Rattus norvegicus* (Norway rat)]	Chromosome 3, NC_005102.4 (44432476..44539680, complement)	
ATVR1 ID: 395246	serine/threonine-protein kinase receptor [*Gallus gallus* (chicken)]	Chromosome 7, NC_006094.4 (36448665..36493593, complement)	ACVR1, ALK2
DIGIT ID: 108868751	divergent to GSC, Induced by TGF-b family signaling [*Homo sapiens* (human)]		
HM90_gp084 ID: 19738087	TGF-B [*Penguinpox virus*]	NC_024446.1 (88363..89373, complement)	HM90_gp084, pepv_083
HM89_gp083 ID: 19737806	TGF-B [*Pigeonpox virus*]	NC_024447.1 (87247..88254, complement)	HM89_gp083, fep_081
TGFB1 ID: 7040	transforming growth factor beta 1 [*Homo sapiens* (human)]	Chromosome 19, NC_000019.10 (41330531..41353933, complement)	CED, DPD1, LAP, TGFB, TGFbeta
SMAD3 ID: 4088	SMAD family member 3 [*Homo sapiens* (human)]	Chromosome 15, NC_000015.10 (67065698..67195195)	HSPC193, HsT17436, JV15-2, LDS1C, LDS3, MADH3
Smad2 ID: 17126	SMAD family member 2 [*Mus musculus*	Chromosome 18, NC_000084.6	7120426M23Rik, Madh2, Madr2, Smad-2, mMad2

شکل ۲۳. انواع ژنهای *TGF-B* متعلق به موجودات مختلف در پایگاه داده Gene و انتخاب *TGFB1* انسانی

پس از کلیک بر روی *TGFB1* صفحه ای برای شما نمایش داده می شود که ابتدای آن مشابه شکل۲۴ است.
در این بخش که تحت عنوان «summary» نامگذاری شده است، بطور خلاصه اطلاعاتی مانند نام ژن، نوع ژن، وجود یا عدم وجود آن در پایگاه داده RefSeq (در ادامه توضیح می دهیم)، گونه مورد نظر، سایر اسامی این ژن و خلاصه ای از عملکرد این ژن ارائه شده است (شکل ۲۴).

شکل ۲۴. خلاصه (summary) اطلاعات در مورد ژن *TGFB1* انسانی

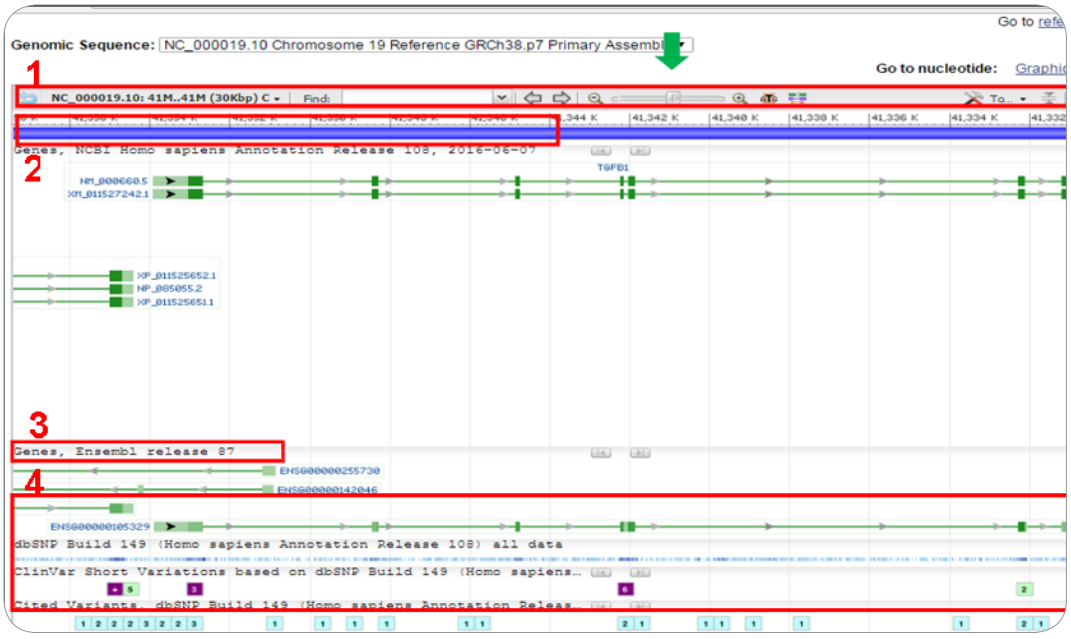
به قسمت بعدی اطلاعات مندرج در صفحه ای که پیش رو دارید یعنی «Genomic context» توجه نمایید (شکل ۲۵). در این بخش جایگاه کروموزومی ژن *TGFB1* و تعداد اگزون های آن نشان داده شده است، همانطور که ملاحظه می کنید، این ژن بر روی بازوی بلند (q) کروموزوم ۱۹ و در ناحیه نواربندی ۱۳٫۲ قرار گرفته است و واجد ۷ اگزون می باشد. در جدول نیز اطلاعات مربوط به محل قرارگیری ژن در کروموزوم در اسمبلی های مختلف نشان داده شده است. در این مثال مدل ۱۰۸ که مربوط به اسمبلی GRCh38.p7 است جدیدتر از مدل ۱۰۵ است. در قسمت آخر این بخش، خط سیاه رنگ (با پیکان سبز رنگ در شکل مشخص شده) معرف کروموزوم شماره ۱۹ است و پیکان قرمز رنگ، ژن *TGFB1* و جهت رونویسی آن بر روی این کروموزوم را نشان می دهد، سایر پیکان های خاکستری رنگ نیز نشان دهنده ژن های اطراف این ژن و جهت رونویسی آنها هستند (شکل ۲۵).



شکل ۲۵. اطلاعاتی در مورد جایگاه کروموزومی ژن مورد نظر، تعداد اگزون های آن و ژن های اطراف این ژن

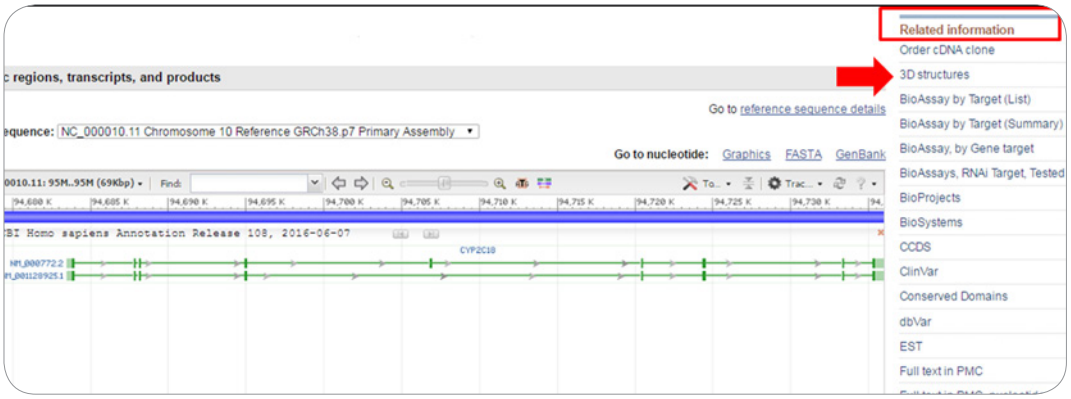
به قسمت بعدی همین صفحه یعنی «Genomic regions, transcript, and products» توجه کنید، شکلی مشابه با شکل ۲۹ مشاهده می نمایید که در آن کادر ۱ طول ژن و جفت نوکلئوتید های گسترده شده در این طول را

نشان می دهد. کادر ۲ و ۳ واریانت های بیانی این ژن را به ترتیب در پایگاه داده NCBI و Ensemebl نشان می دهند و کادر ۴ نشان دهنده جهش های موجود در این ژن است که در پایگاه داده SNP ثبت شده است. برای مشاهده نام جهش ها لازم است تا بزرگنمایی تصویر را افزایش دهید. برزگنمایی تصویر را از طریق مکانی که با پیکان سبز رنگ در شکل نشان داده شده است، تغییر دهید (شکل ۲۶).



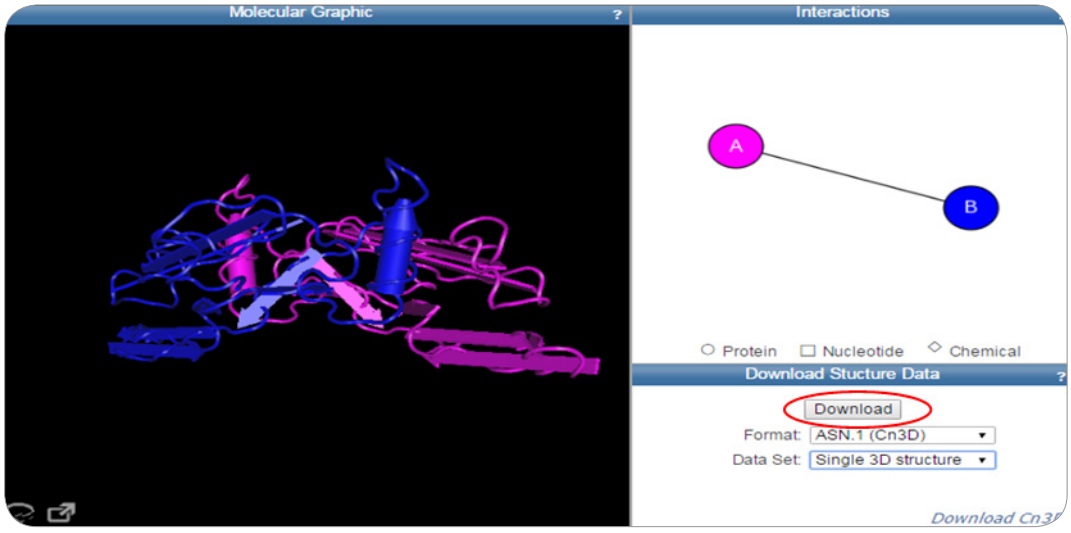
شکل ۲۶. اطلاعاتی در مورد واریانت های بیانی ژن مورد نظر و SNP موجود در این ژن

اکنون به سمت راست همین صفحه توجه نمایید، همانطور که ملاحظه می کنید ستونی به نام «Related information» وجود دارد که می توان اطلاعات بسیار مهم و کاملی از ژن و پروتئین های کد شونده توسط آن به دست آورد که برخی از مهم ترین آنها در ادامه شرح داده می شوند (شکل۲۷).



شکل ۲۷. ستون Related information و اطلاعات موجود در آن

برای مثال، در صورتی که در قسمت Related information بر روی 3D structure (با پیکان در شکل ۲۷ نشان داده شده) کلیک کنید، می توانید ساختمان سه بعدی پروتئین های کد شونده توسط این ژن را مشاهده و برای بررسی های بیشتر با سایر نرم افزارها آن را دانلود کرد. برای مثال، ساختار سه بعدی یکی از پروتئین های کد شونده توسط ژن *TGFB1* مشابه شکل زیر است که از محل مشخص شده قابل دانلود نیز می باشد (شکل ۲۸).



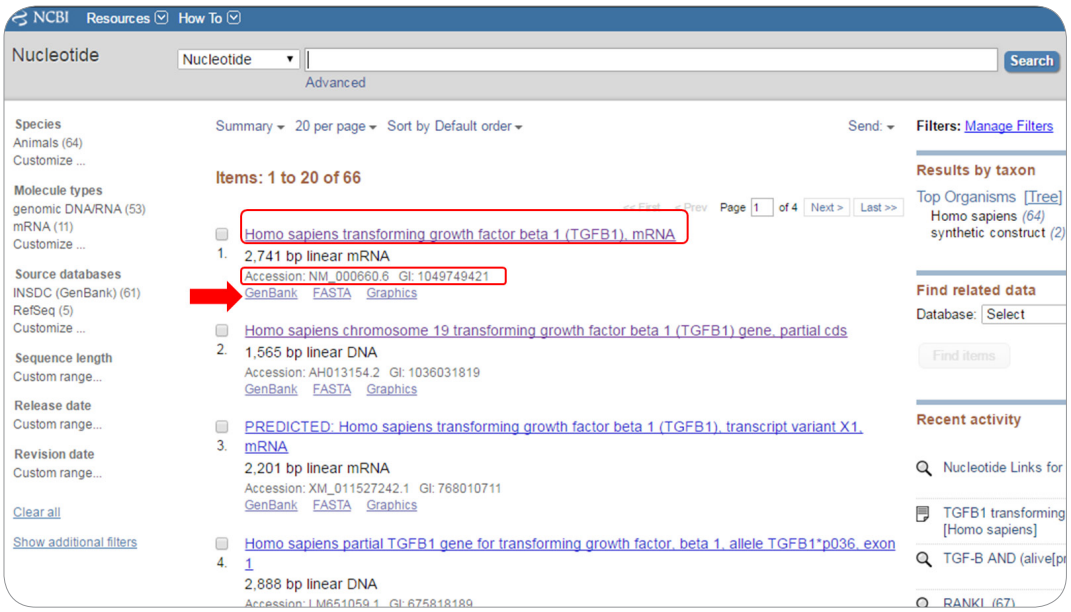
شکل ۲۸. مشاهده ساختار سه بعدی یکی از پروتئین های کد شونده توسط ژن *TGFB1*

همچنین با کلیک بر گزینه Nucleotide در بخش Related information می توانید اطلاعاتی در رابطه با توالی نوکلئوتیدی این ژن کسب کنید. در صورت کلیک بر Nucleotide وارد پایگاه داده Nucleoid که زیر مجموعه GenBank است، می شوید و صفحه ای مشابه با شکل ۲۹ مشاهده می کنید. در این بخش هر رکورد (توالی) با یک شماره دسترسی[۴۰] و یک شناسه ژنی[۴۱] یا GI مشخص شده است. شماره های دسترسی معمولا به صورت ۱+۵ یعنی ۱ حرف انگلیسی بزرگ و ۵ رقم به دنبال آن و یا ۲+۶ یعنی دو حرف انگلیسی بزرگ و ۶ رقم بعد از آن هستند. در صورتی که فرد دیگری همین توالی را با اندکی تفاوت (مثلا تغییر یک نوکلئوتید) در GenBank ثبت کند، شماره دسترسی به توالی تغییر نمی کند و تنها شماره نسخه (سطر بعد) تغییر می کند. همچنین شناسه ژنی یا GI در صورت وجود تغییر می کند. برخلاف عدد GI که فقط مخصوص GenBank است، شماره های دسترسی شماره های منحصر به فردی هستند که در بین پایگاه های اطلاعاتی مختلف مشترک می باشند.

برای مثال، کادر قرمز رنگ در شکل ۲۹ یکی از نسخه های بیانی ژن مورد نظر به طول ۲۷۴۱ جفت باز را نشان می دهد. با کلیک بر گزینه GenBank به صفحه مربوط به اطلاعات و توالی نوکلئوتیدی این نسخه هدایت می شوید. همچنین با کلیک بر گزینه FASTA (FASTA نوعی فرمت فایل است که در فصل ۴ توضیح داده می شود) می توانید مستقیماً توالی نوکلئوتیدی مورد نظر را مشاهده و ذخیره نمایید (شکل۲۹).

۴۰ Accession number

۴۱ Gene identifier



شکل ۲۹ . بررسی ژن *TGFB1* در پایگاه داده Nucleotide

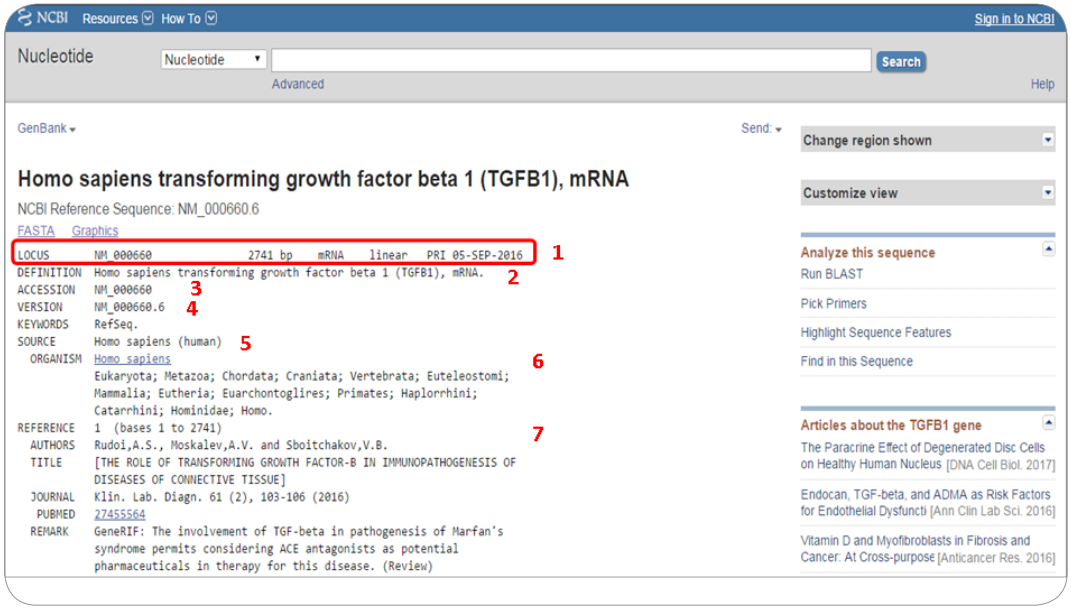
در صورت کلیک بر گزینه GenBank وارد صفحه ای می شوید که اطلاعات موجود در آن در سه بخش مرتب و طبقه بندی شده اند، این بخش ها عبارتند از:

۱. Header که حاوی اطلاعات کلی در مورد توالی است.

۲. Features که شامل توضیحات جزئی تر در مورد ویژگی های توالی است.

۳. Origin که واجد توالی نوکلئوتیدی ژن مورد نظر است.

شکل ۳۰ نشان دهنده بخش Header توالی است. اولین سطر Header درهمه رکوردهای (توالی ها) GenBank، سطر لوکوس (Locus) است که با کادر قرمز رنگ و شماره ۱ در شکل ۳۰ مشخص شده است. این بخش در پایگاه های اطلاعاتی مختلف ممکن است به طرق مختلفی نمایش داده شود. در GenBank، اطلاعاتی مانند نام رکورد، طول توالی، ماهیت آن (DNA یا RNA)، نوع آن (خطی یا حلقوی) و تاریخ اضافه شدن رکورد به GenBank و یا به روز شدن آن نشان داده شده است.



شکل ۳۰. اطلاعات بخش header ژن مورد نظر در پایگاه داده GenBank

در سطر بعد یعنی definition (شماره ۲)، نام ژن مورد نظر و گونه ای که این ژن در آن شناسایی شده است، نشان داده شده است. در سطر بعد شماره دسترسی (شماره ۳ شکل) و شماره نسخه[۴۲] (شماره ۴ شکل) این رکورد (توالی) قرار دارد. در بخش Header، اطلاعات تاکسونومیکی در سطرهایی با عنوان Source (شماره ۵ شکل) و Organism (شماره ۶ شکل) ارائه شده است. در سطر بعد یعنی Reference (شماره ۷ شکل)، مقالات که در آنها به معرفی این ژن پرداخته شده است همراه با نام نویسنده ها، نام مجله و سال انتشار آن ارائه شده است (شکل ۳۰).

بخش میانی رکورد، ویژگی یا features نامیده می شود که واجد اطلاعاتی در مورد ویژگی های توالی است. نواحی مختلف ژنی مانند اگزون، CDS، جایگاه شروع نسخه برداری، سیگنال پپتید[۴۳]، جایگاه دم پلی A همراه با موقعیت قرارگیری آنها در توالی در این محل درج شده اند.

۴۲ Version

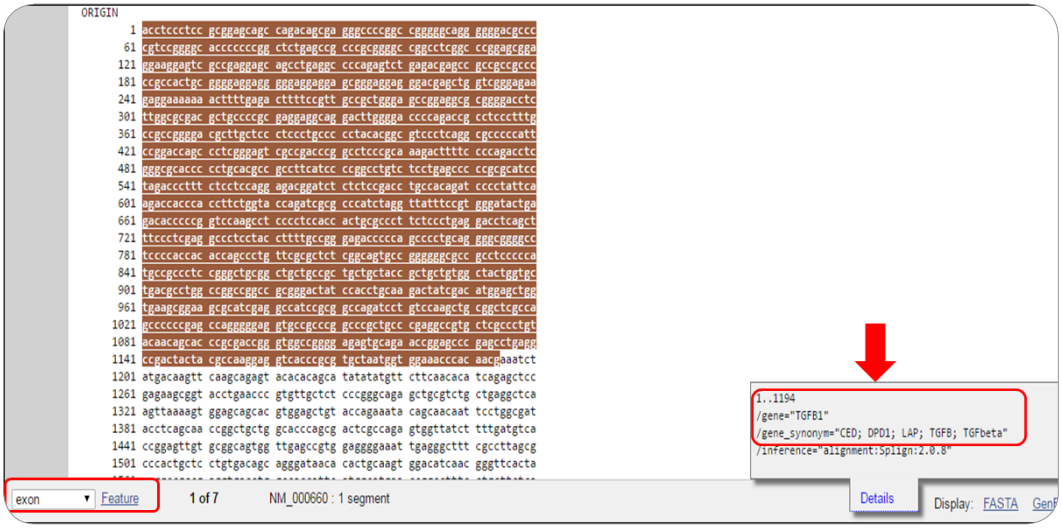
۴۳ Signal peptide

شکل ۳۱. بخشی از features مربوط به رکورد *TGFB1* در پایگاه داده GenBank

بخشی از features مربوط به رکورد *TGFB1* در شکل ۳۱ نشان داده است. در این مثال، این بخش شامل اطلاعاتی در مورد کل ژن، اگزون و جایگاه شروع نسخه برداری که با misc_feature مشخص شده است، می باشد (شکل ۳۱). در صورتی که به کمک موس به انتهای صفحه ای که پیش رو دارید، حرکت کنید آخرین بخش رکورد را مشاهده خواهید کرد. در این بخش توالی نوکلئوتیدی ژن مورد نظر ارائه شده است، برای مثال بخشی از توالی نوکلئوتیدی ژن *TGFB1* مشابه شکل۳۲ می باشد (شکل ۳۲).

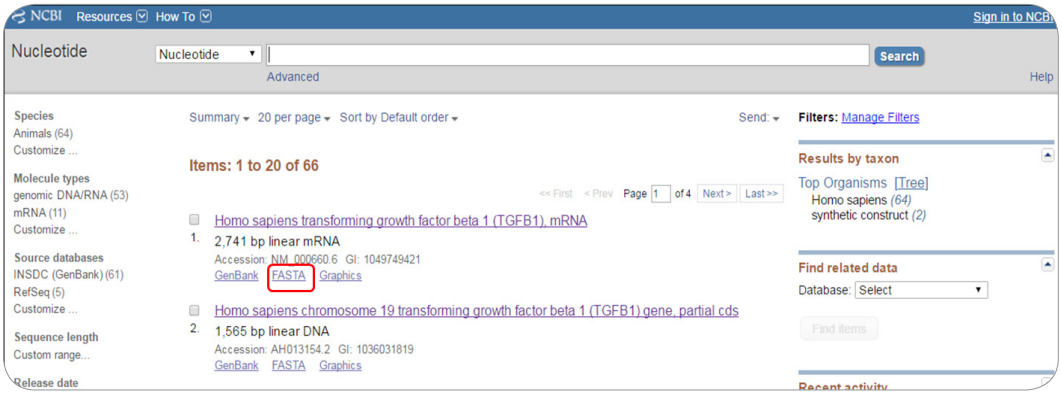
شکل ۳۲. بخشی از توالی ژن *TGFB1* در پایگاه داده GenBank

در صورت کلیک بر هر ویژگی یا features مربوط به یک ژن، محل آن در توالی ژن مربوطه هایلایت می شود. بعنوان مثال با کلیک بر اگزون در بخش features ژن *TGFB1*، این ناحیه در توالی نوکلئوتیدی هایلایت قهوه ای رنگ می شود و می توان موقعیت اگزون را در کل توالی مشاهده کرد، همچنین در پایین صفحه، اطلاعاتی در مورد ویژگی (features) هایلایت شده درج شده است. برای مثال، این اطلاعات در شکل ۳۳ شامل نوع ویژگی انتخاب شده که در اینجا اگزون می باشد (در سمت چپ صفحه با کادر قرمز رنگ در شکل مشخص شده) و موقعیت این ویژگی در توالی که در اینجا از نوکلئوتید شماره ۱ تا شماره ۱۱۹۴ است، نام ژن و سایر اسامی این ژن می باشد (در سمت راست صفحه با کادر قرمز رنگ در شکل مشخص شده) (شکل ۳۳).



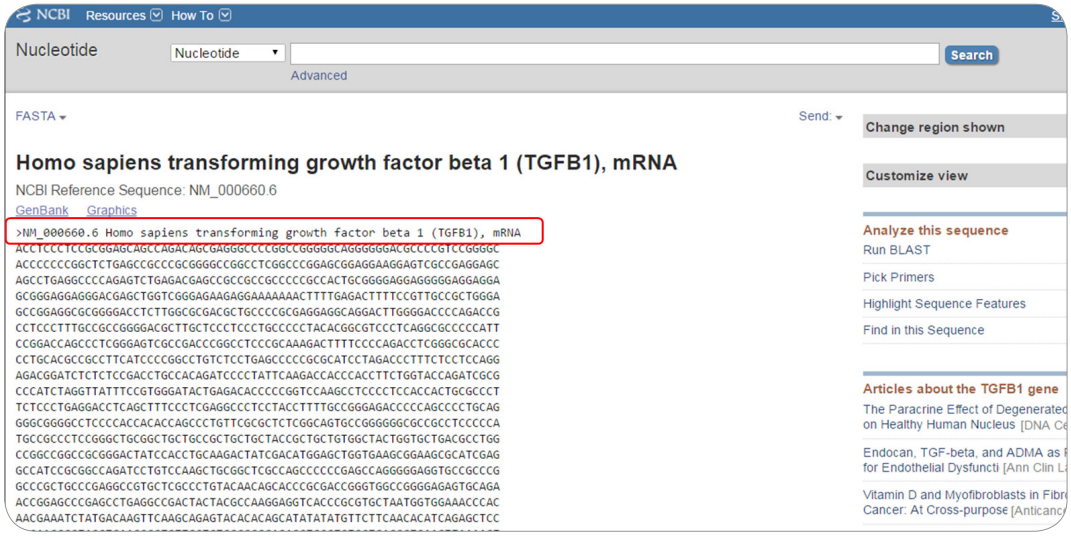
شکل ۳۳. بخش انتخاب شده features در توالی ژن *TGFB1*

همانطور که اشاره شد، زمانی که در پایگاه داده Nucleotide هستیم، در صورت کلیک برگزینه FASTA در ذیل رکورد مورد نظر می توان مستقیماً توالی نوکلئوتیدی ژن مورد نظر را مشاهده و ذخیره کرد (شکل ۳۴).



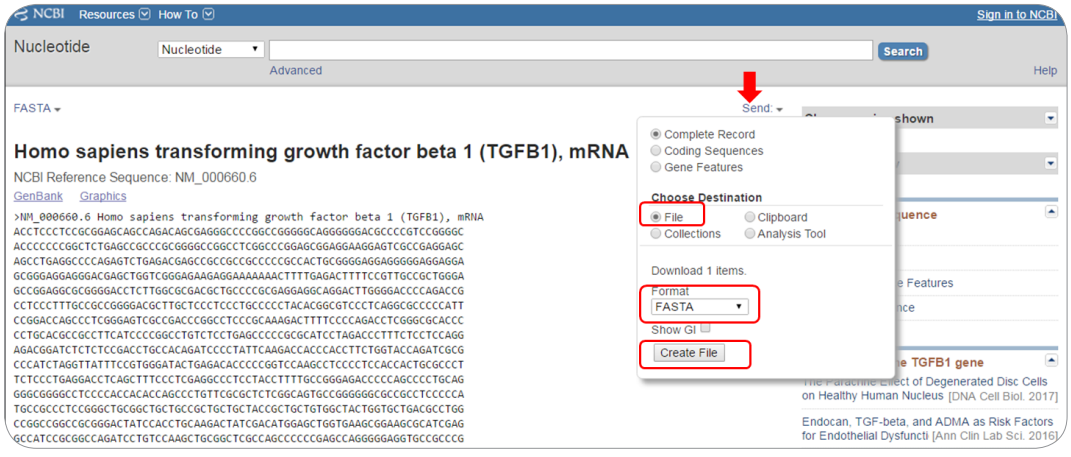
شکل ۳۴. انتخاب گزینه FASTA در رکورد *TGFB1* در پایگاه داده GenBank

در صورت کلیک بر گزینه FASTA وارد صفحه ای مشابه شکل زیر می شوید که در آن توالی نوکلئوتیدی ژن مورد نظر (در اینجا *TGFB1* انسان) را به فرمت FASTA مشاهده می کنید.



شکل ۳۵. مشاهده توالی نوکلئوتیدی ژن *TGFB1* به فرمت FASTA

رکوردهای موجود در GenBank واجد نام توالی و ماهیت مولکولی آن (DNA یا RNA) است، این ناحیه با کادر مستطیل شکل قرمز رنگ در شکل ۳۵ نشان داده شده است. برای ذخیره توالی نوکلئوتیدی مشاهده شده (شکل ۳۵) می توانید بر روی گزینه Send (با پیکان قرمز در شکل ۳۶ مشخص شده) کلیک کرده و پس از انتخاب گزینه File، و سپس Format که به صورت پیش فرض همان FASTA می باشد، بر روی Create File کلیک نمایید تا توالی نوکلئوتیدی به فرمت FASTA در کامپیوترتان ذخیره شود (شکل ۳۶). توالی ذخیره شده را می توانید با استفاده از برنامه Word یا Notepad باز کنید.



شکل ۳۶. ذخیره توالی نوکلئوتیدی به فرمت FASTA

۹ آشنایی با پایگاه داده RefSeq[۴۴]

در مواردی که چندین توالی نوکلئوتیدی یا پروتئینی برای یک ژن یا پروتئین خاص در یک موجود معین توسط محققان مختلف در GenBank وارد شود، کارمندان NCBI از بین آنها بهترین توالی که تا حد امکان فاقد جهش، خطای ناشی از کلونینگ و تعیین توالی می باشد را انتخاب کرده و در پایگاه داده ای به نام RefSeq قرار می دهند. بنابراین برخلاف GenBank، در پایگاه داده RefSeq، هر رکورد یا توالی مربوط به یک ژن یا فرم پیرایش شده از یک ژن است، از این رو این پایگاه داده واجد توالی های غیرتکراری[۴۵] می باشد. رکوردهای ثبت شده در RefSeq در سه سطح متفاوت پیش بینی شده[۴۶]، مشروط یا محتاطانه[۴۷] و بررسی شده[۴۸] قرار می‌گیرند، بدیهی است که رکوردهای بررسی شده قابل اعتماد تر از سایر رکوردها هستند.

اطلاعات مربوط به توالی های ژنی و پروتئینی در دو پایگاه اطلاعاتی RefSeq متفاوت قرار می گیرند. شماره دسترسی توالی های بیانی (mRNA) موجود در این پایگاه با NM و شماره دسترسی توالی های پروتئینی این پایگاه با NP شروع می شوند. در جدول ۲ فرمت شروع شماره های دسترسی رکوردهای ثبت شده در پایگاه داده Refseq بطور کامل نشان داده شده است.

فرمت	شرح
NC	توالی های ژنومی که بطور کامل توالی یابی شدند
NG	توالی های ژنومی که بطور کامل توالی یابی نشدند
NM	نسخه های بیانی کد کننده (mRNA)
NR	نسخه های بیانی غیر کد کننده (ncRNA)
NP	توالی های پروتئینی
XM	نسخه های پیش بینی شده mRNA
XR	نسخه های پیش بینی شده RNA های غیر کد کننده
XP	توالی های پیش بینی شده پروتئینی

جدول ۲. فرمت شماره های دسترسی انواع توالی های موجود در پایگاه داده Ref Seq

۴۴ Reference sequence

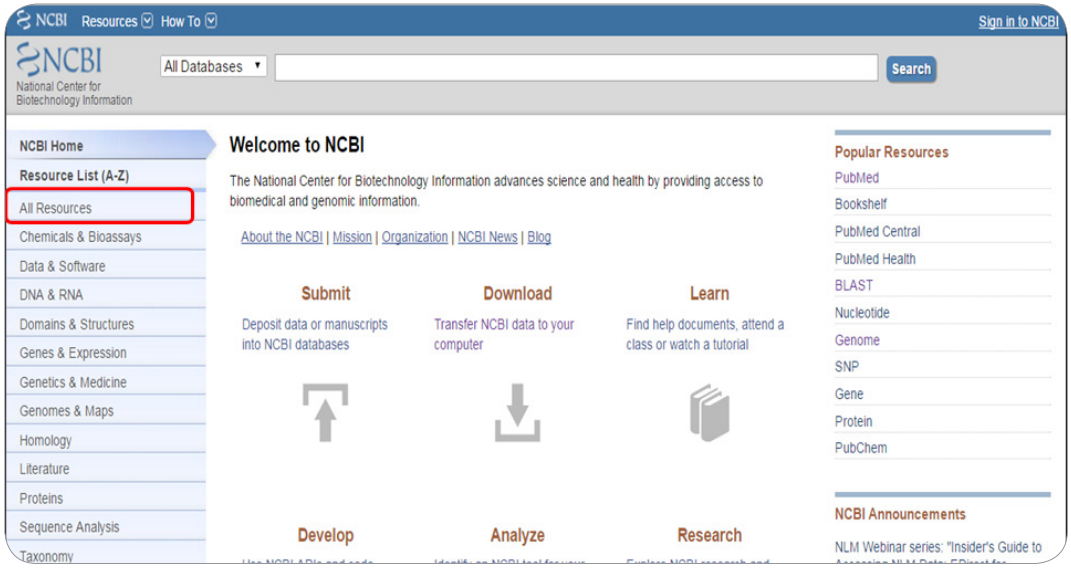
۴۵ Non-redundant

۴۶ Predicted

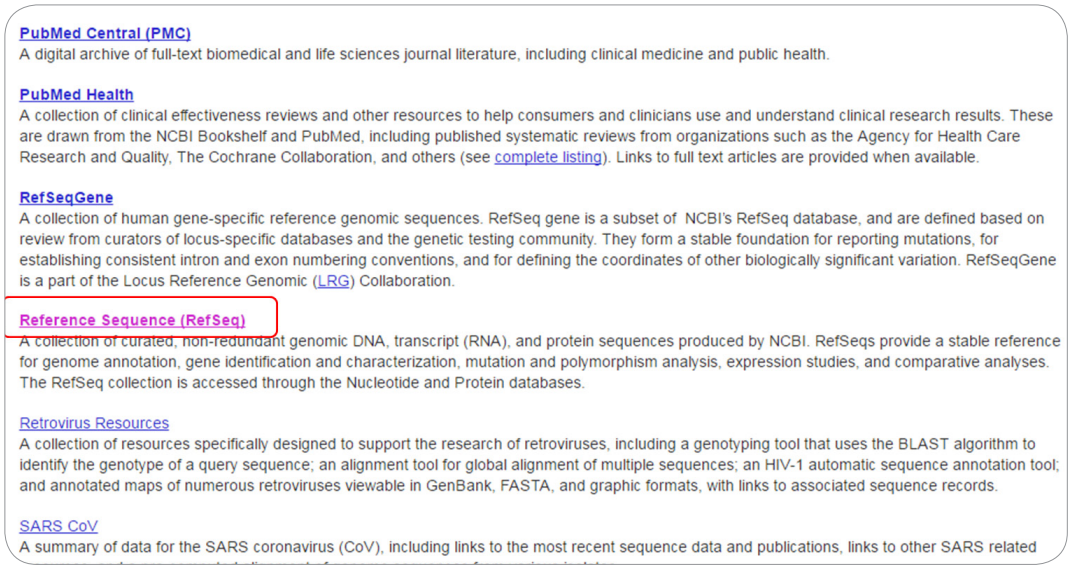
۴۷ Provisional

۴۸ Reviewed

در صورتی که مجدداً به شکل۳۰ برگردید، مشاهده خواهید کرد که شماره دسترسی ژن *TGFB1* به صورت NM_000660 می باشد که نشان دهنده نسخه بیانی موجود از این ژن در پایگاه داده RefsSeq است. برای دسترسی به پایگاه داده RefSeq، به صفحه اصلی پایگاه داده NCBI به آدرس https://www.ncbi.nlm.nih.gov بروید و بر روی All Resources کلیک نمایید (شکل ۳۷)، صفحه ای مشابه شکل ۳۸ برای شما نمایش داده می شود، در این صفحه بر روی (Ref Seq) Reference Sequences کلیک نمایید (شکل۳۸).

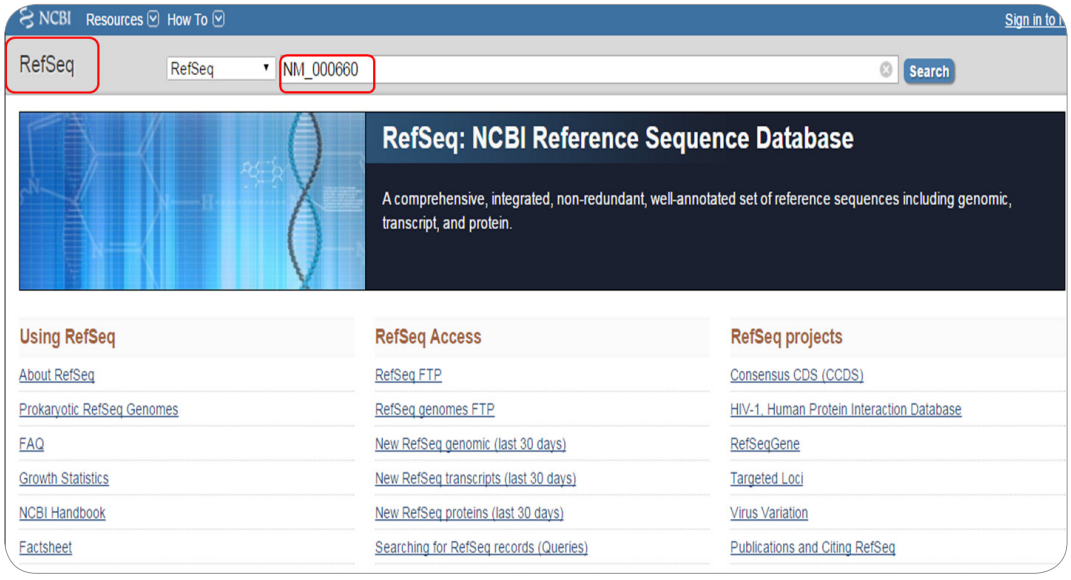


شکل ۳۷. انتخاب All Resource در صفحه اصلی پایگاه داده NCBI



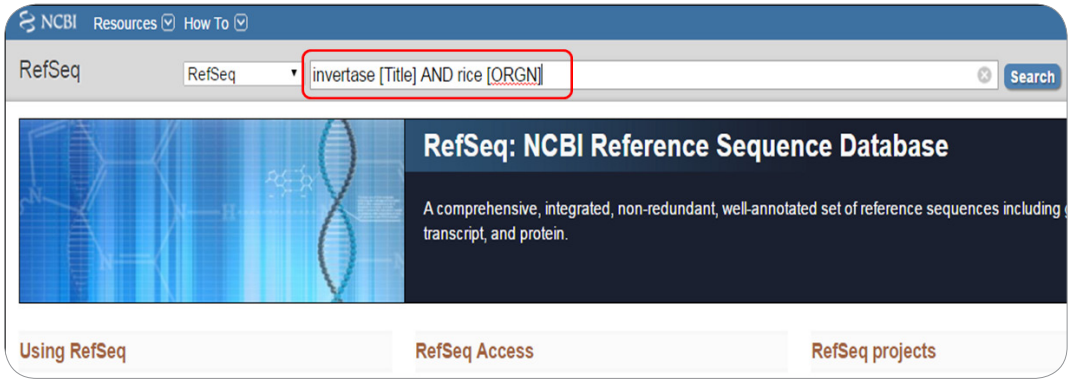
شکل ۳۸. انتخاب پایگاه داده Reference sequence (RefSeq)

اکنون شما در پایگاه داده RefSeq هستید، به لوگوی این پایگاه داده در سمت چپ تصویر دقت کنید. حال می توانید با وارد کردن کلمات مورد نظر در مستطیل مقابل Search در این پایگاه داده به جستجو بپردازید. برای مثال، با وارد کردن شماره دسترسی NM_000660 که مربوط به نسخه بیانی *TGFB1* است می توانیم مستقیماً به این نسخه در پایگاه داده RefSeq دست یابیم (شکل ۳۹).



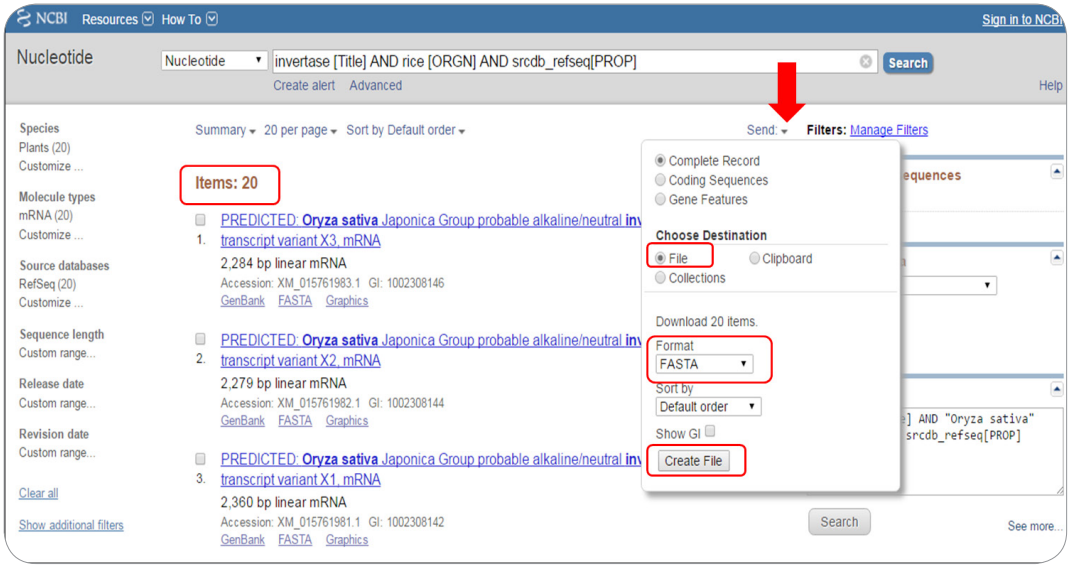
شکل ۳۹. صفحه اصلی پایگاه داده RefSeq و جستجو با استفاده از شماره دسترسی در آن

در صورتی که شماره دسترسی ژن مورد نظرتان را نداشته باشید، می توانید با استفاده از کلید واژه های مناسب در پایگاه های داده مختلف از جمله RefSeq جستجو کنید. برای مثال، فرض کنید که شما قصد جستجو توالی کد کننده اینورتازها درگیاه برنج را دارید، با وارد کردن عبارت[ORGN] invertase [Title] AND rice در پایگاه داده RefSeq می توانید توالی های تصحیح شده و کامل مربوط به اینورتازهای برنج را به دست آورید (شکل ۴۰).



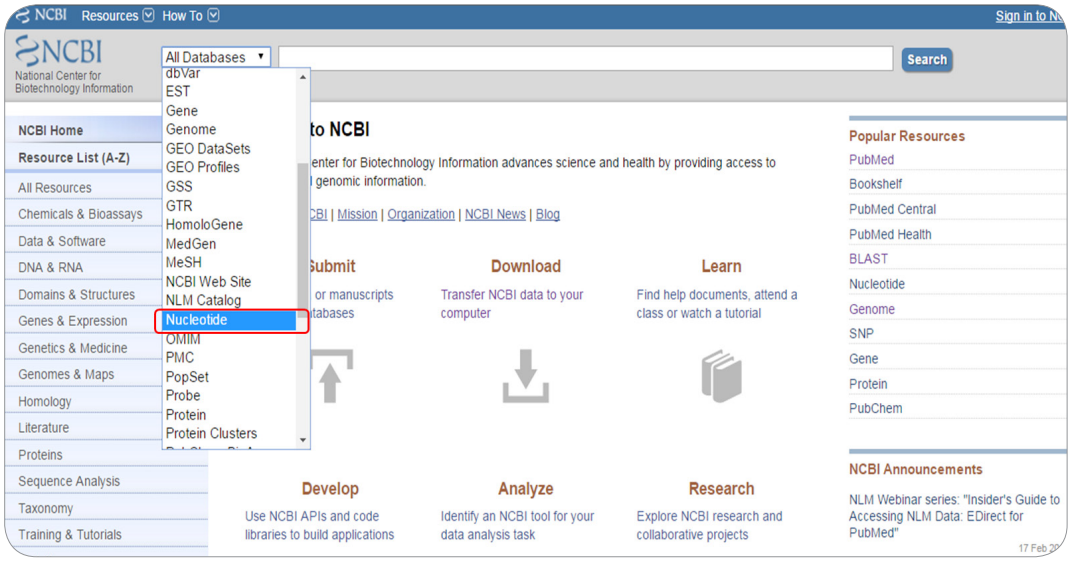
شکل ۴۰ . جستجو با استفاده از کلید واژه در پایگاه داده RefSeq

ملاحظه خواهید کرد که ۲۰ رکورد (توالی) بعنوان نتیجه این جستجو خواهید داشت (شکل ۴۱). با کلیک بر هر رکورد می‌توان به اطلاعات دقیق‌تری در مورد آن دست یافت. در صورتی که توالی نوکلئوتیدی همه رکوردها را بخواهید، کافی است بر روی فلش رو به پایین در کنار گزینه Send کلیک کرده و از منوی باز شده، گزینه File و FASTA را بعنوان فرمت توالی انتخاب کنید و در نهایت بر روی Create file کلیک کنید. به این ترتیب توالی نوکلئوتیدی تمام رکوردها در کامپیوتر شما ذخیره خواهد شد (شکل ۴۱). در صورتی که توالی نوکلئوتیدی تنها یک یا چند رکورد را بخواهید، کافی است تا از طریق مربع کوچک در کنار هر رکورد آنها را انتخاب نمایید و سپس مراحل بالا برای ذخیره توالی های انتخاب شده را انجام دهید.



شکل ۴۱. نتایج حاصل از جستجو اینورتازهای برنج در پایگاه داده RefSeq و ذخیره توالی های نوکلئوتیدی

البته باید توجه داشت که این پایگاه حاوی توالی های نوکلئوتیدی تمام موجوداتی که توالی آنها در GenBank است، نمی باشد. در صورتی که توالی مورد نظر خود را در این پایگاه نیافتید آن را در قسمت (Nucleotide) در GenBank جستجو نمایید. به این منظور کافی است به آدرس ncbi.nlm.nih.gov بروید و از فلش رو به پایین در کنار کادر All Database گزینه Nucleotide را انتخاب و سپس با وارد کردن شماره دسترسی مورد نظر یا کلمات کلیدی مناسب جستجو را انجام دهید (شکل ۴۲).



شکل ۴۲. انتخاب پایگاه داده Nucleotide در صفحه اصلی NCBI

۱۰ آشنایی با پایگاه داده UniProt [۴۹]

این پایگاه داده یکی از جامعترین و کاملترین منابع توالی های پروتئینی است که در سال ۲۰۰۲ ایجاد شد. UniProt از سه پایگاه داده کلیدی تشکیل شده است:

Swiss-Prot: یکی از کاملترین پایگاه های داده پروتئینی تفسیر شده[۵۰] است که شرح ساختار و عملکرد پروتئین های آن توسط متخصصین صورت گرفته است و از این رو اطلاعات موجود در این پایگاه از کیفیت بالایی برخودار می باشند.

TrEMBL[۵۱] : پروتئین های موجود در این پایگاه داده ترجمه توالی های نوکلئوتیدی در EMBL هستند و به صورت اتوماتیک (و نه توسط متخصص) تفسیر شده اند.

PIR[۵۲]: پایگاه داده پروتئینی دیگری است که توالی های آن تحت نظارت متخصصین بررسی و وارد شدند.

پایگاه های داده پروتئینی دیگری نیز از UniProt مشتق شده اند که مهمترین آنها عبارتند از :

UniParc[۵۳]: یک پایگاه داده غیر تکراری[۵۴] از توالی های پروتئینی است. در این پایگاه داده برای هر پروتئین تنها یک توالی وجود دارد که با شناسه منحصر بفرد و بدون تغییر موسوم به UPI مشخص می شود.

UniRef[۵۵]: توالی های پروتئینی موجود در UniProt و UniParc بر اساس میزان یکسانی[۵۶] با یکدیگر خوشه و گروه بندی شده و در پایگاه داده UniRef قرار می گیرند.

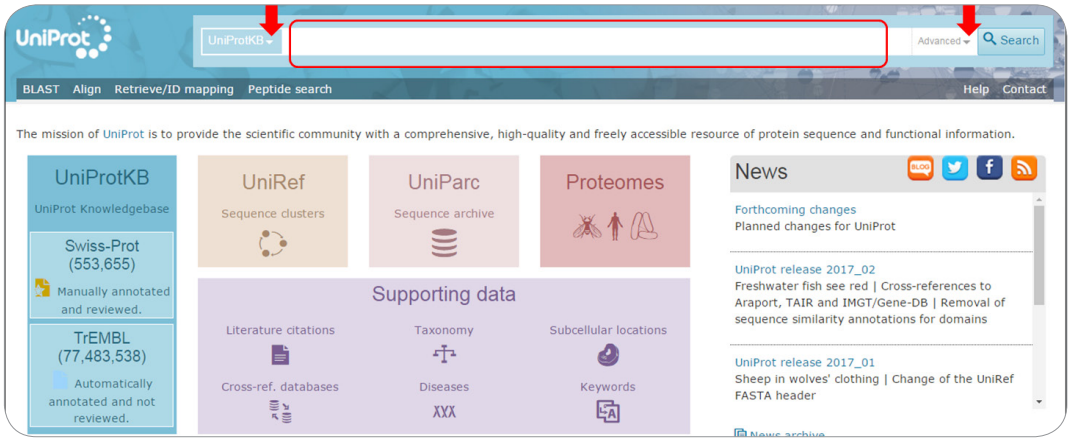
Proteomes: این پایگاه داده حاوی توالی های پروتئینی مربوط به ژنوم های تعیین توالی شده، می باشد.

برای دسترسی به پایگاه داده UniProt می توانید کلمه Uniprot را در موتور جستجوی گوگل وارد کنید و یا

---

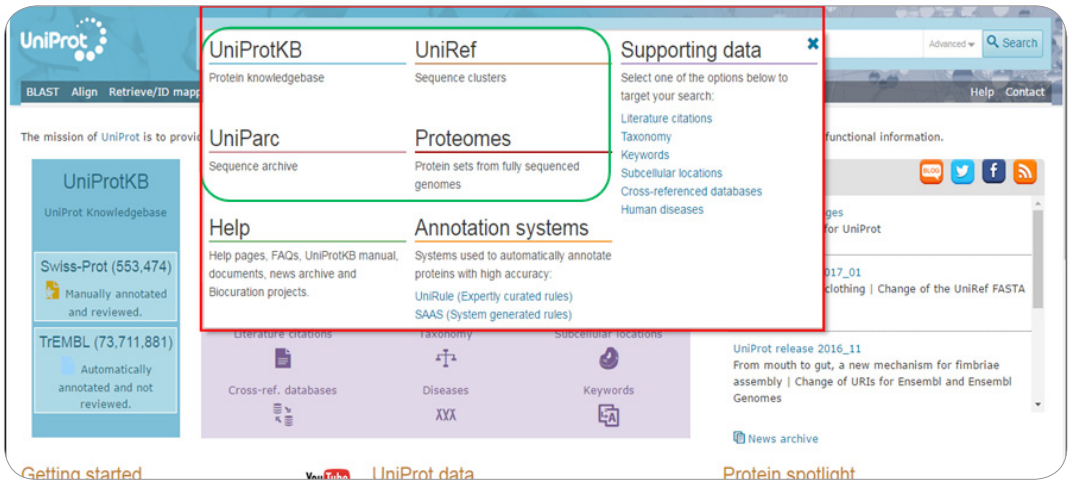
۴۹ Universal Protein Resource
۵۰ Annotated
۵۱ Translated EMBL
۵۲ Protein information resource
۵۳ UniProt Archive
۵۴ Non-redundant
۵۵ UniProt Reference Clusters
۵۶ Identity

به آدرس www.uniprot.org بروید تا به صفحه اصلی این پایگاه دست یابید. با وارد کردن کلمات کلیدی یا شماره دسترسی پروتئین مورد نظر در مستطیل مشخص شده می توانید جستجو خود را انجام دهید. همچنین با کلیک بر فلش رو به پایین در کنار کادر UniProtKB می توان یکی از پایگاه های داده پروتئینی مورد نظر را انتخاب کرد. با کلیک بر فلش رو به پایین در کنار کادر Advanced نیز می توانید جستجو خود را محدودتر و اختصاصی نمایید (شکل ۴۳).



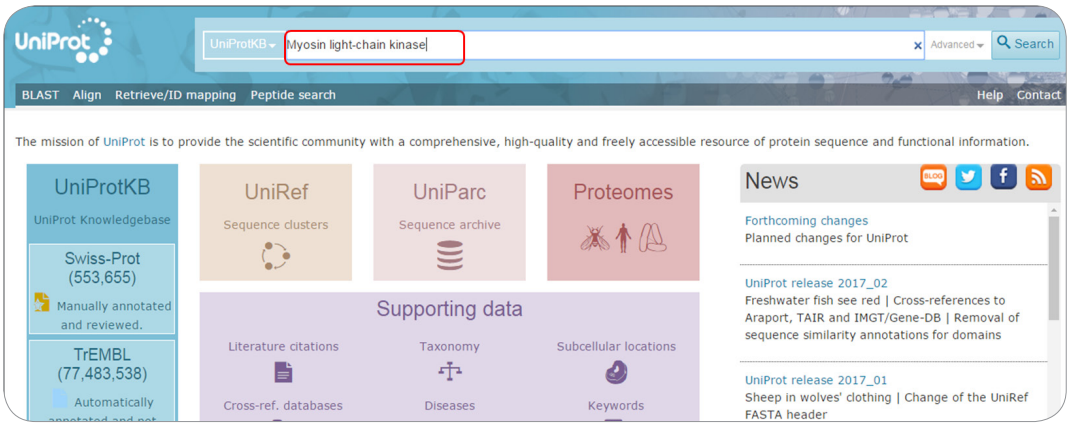
شکل ۴۳. صفحه اصلی پایگاه داده UniProt

در صورتی که بر فلش رو به پایین در کنار کادر UniProtKB کلیک نمایید، با کادر قرمز رنگ مشخص شده در شکل ۴۴ مواجه می شوید. در اینجا می توانید جستجو خود را در یکی از چهار پایگاه UniProtKB، UniRef، UniParc و Proteomes انجام دهید (شکل ۴۴). جستجو به صورت پیش فرض در پایگاه داده UniProtKB انجام می گیرد که کامل ترین پایگاه داده است و سایر پایگاه های داده دیگر را نیز پوشش می دهد.



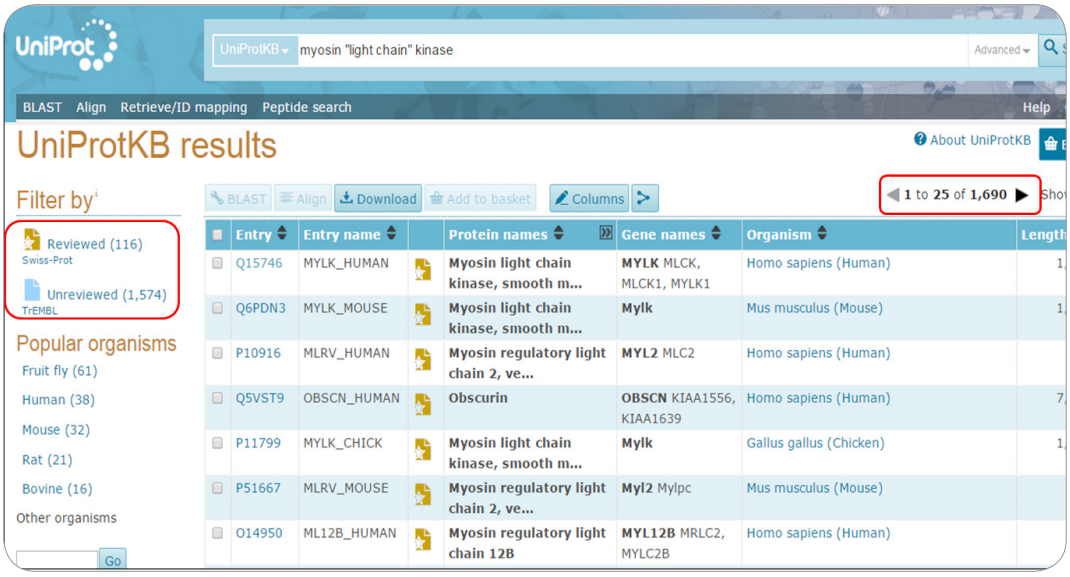
شکل ۴۴. انتخاب پایگاه داده پروتئینی به منظور جستجو در آن

برای مثال، فرض کنید که می خواهیم در پایگاه داده UniProt اطلاعاتی در مورد پروتئین کینازی که زنجیره سبک میوزین را فسفریله می کند، پیدا کنیم. با وارد کردن کلمات Myosin light-chain kinase در کادر مستطیل شکل مربوطه می توان این جستجو را انجام داد (شکل ۴۵).



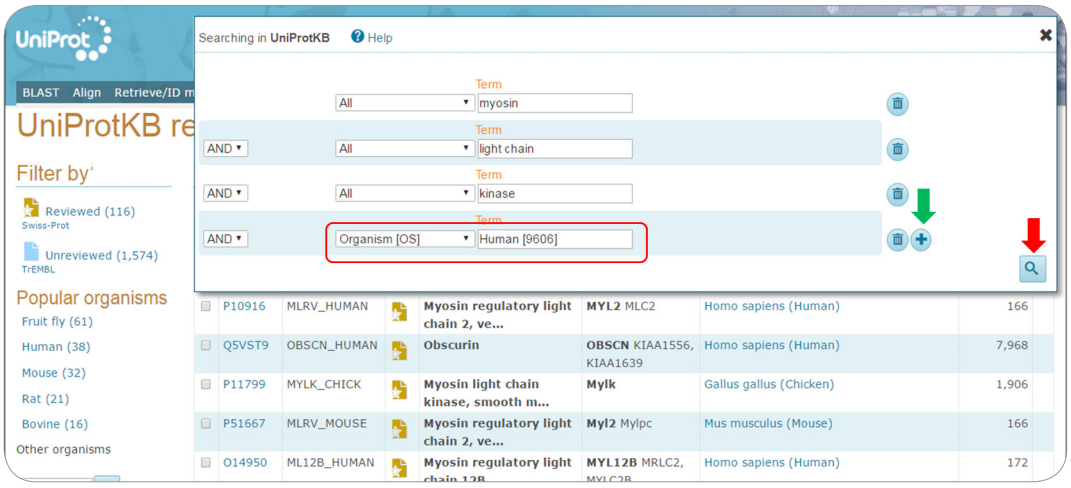
شکل ۴۵. جستجو با استفاده از کلید واژه در پایگاه داده UniProtKB

همانطورکه مشاهده می کنید ۱۶۹۰ رکورد بعنوان نتیجه جستجو حاصل شدند که از این تعداد ۱۱۶ رکورد متعلق به پایگاه داده Swiss-Prot و ۱۵۷۴ رکورد متعلق به پایگاه TrEMBL می باشند (شکل ۴۶). برای کاهش تعداد رکوردها می توانید جستجوی خود را به ارگانیسم خاصی محدود کنید. برای مثال ما این کار را برای انسان انجام می دهیم.



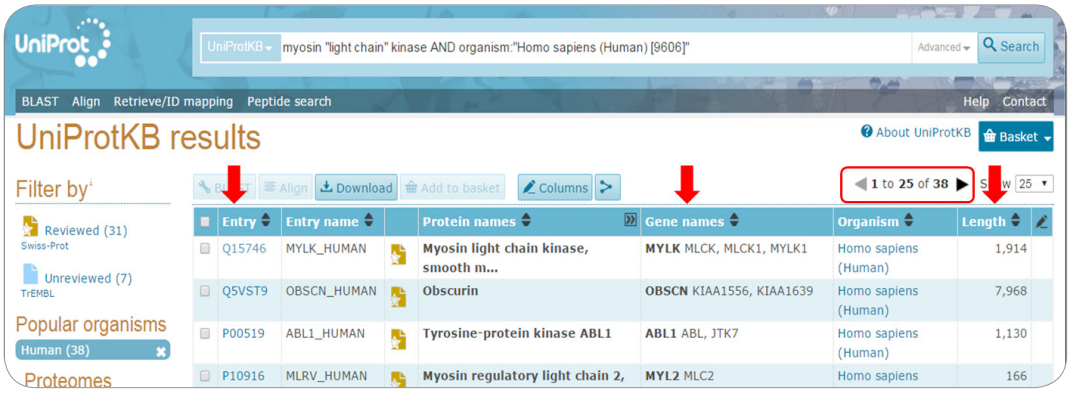
شکل ۴۶. نتایج حاصل از جستجو در پایگاه داده UniProt

به منظور اختصاصی کردن جستجو، بر روی فلش رو به پایین در کنار کادر Advanced کلیک کنید (شکل ۴۳) تا کادر شکل ۴۷ باز شود. در این کادر organism را انتخاب کرده و در مستطیل مقابل آن کلمه human را تایپ کنید و بر روی علامت ذره بین در پایین گوشه سمت راست کلیک کنید تا فقط پروتئین مورد نظر انسانی جستجو شود. در صورت نیاز به کادرهای بیشتر برای اختصاصی کردن بیشتر جستجو می توانید بر روی علامت + (با پیکان سبز رنگ در شکل مشخص شده) کلیک نمایید (شکل ۴۷).



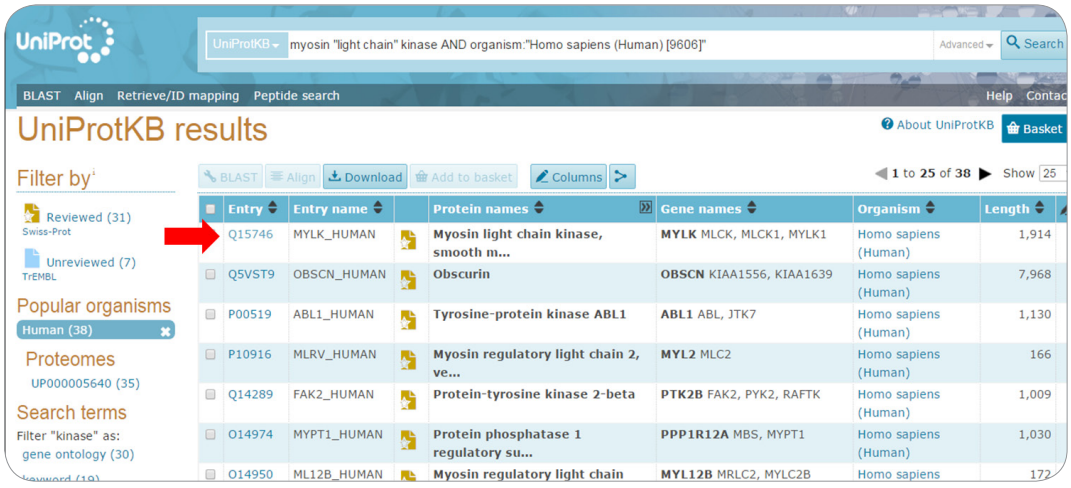
شکل ۴۷. محدود و اختصاصی کردن جستجو به انسان

همانطور که ملاحظه می کنید پس از اختصاصی کردن جستجو تنها ۳۸ رکورد از پروتئین کیناز مورد نظر یافت شد که تمام آنها متعلق به انسان می باشند. اولین ستون نتایج (Entry) دارای یک شناسه اختصاصی برای هر رکورد است که در بین پایگاه داده های مختلف مشترک است. بنابراین شما می توانید با استفاده از این شناسه، پروتئین مورد نظر خود را در پایگاه داده NCBI نیز جستجو کنید. در ستون چهارم (Gene name) اسامی مختلف این رکورد پروتئینی ارائه شده است و در ستون ششم (length) طول توالی آمینو اسیدی این پروتئین نشان داده شده است (شکل ۴۸).



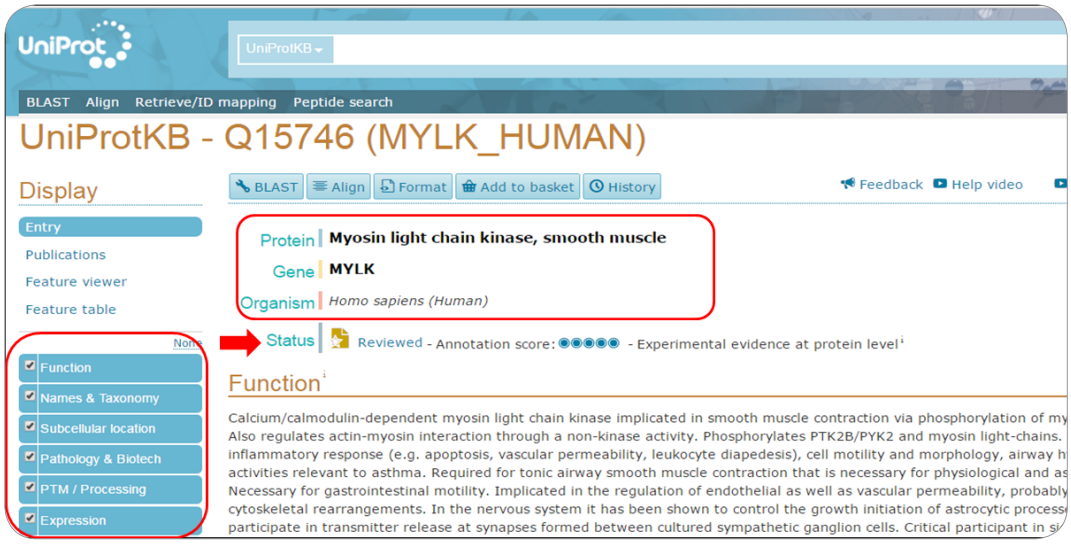
شکل ۴۸. نتایج حاصل از جستجو پس از محدود کردن جستجو به انسان

اکنون به بررسی یکی از رکوردها می پردازیم، در صورتی که بر روی شماره دسترسی هر رکورد که به صورت هایپرلینک می باشد، کلیک نمایید محتویات آن رکورد را می توانید مشاهده و مطالعه کنید. در این مثال بر روی Q15746 کلیک می کنیم (شکل ۴۹).



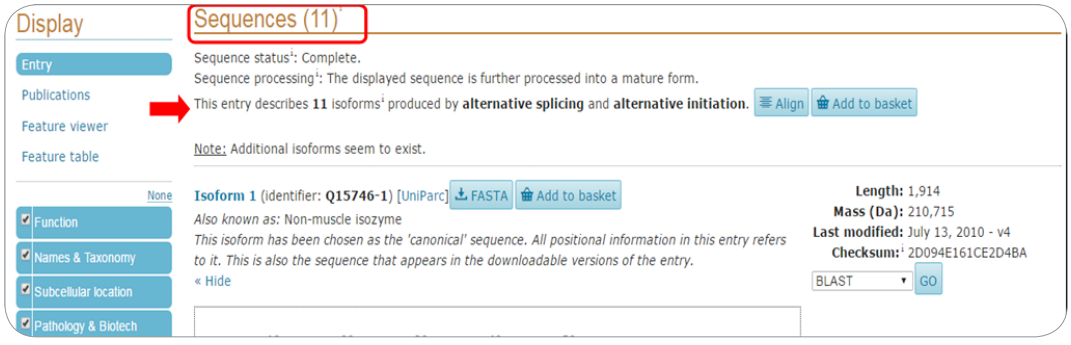
شکل ۴۹ . کلیک بر روی شماره دسترسی Q15746

همانطور که مشاهده می کنید صفحه ای مانند شکل ۵۰ باز می شود که در ابتدای آن نام پروتئین مورد نظر، ژن کد کننده آن و نوع ارگانیسم نوشته شده است. همچنین در قسمت Status (با پیکان قرمز رنگ در شکل مشخص شده) قید شده است که این رکورد توسط متخصصان بررسی شده و شواهد آزمایشگاهی برای وجود این پروتئین موجود است (یک پروتئین پیش بینی شده بر اساس آنالیز توالی DNA نیست)، لذا محتویات این رکورد بسیار قابل اطمینان هستند. در ستون سمت چپ صفحه، می توانید انواع اطلاعات مربوط به پروتئین مورد نظر که در این پایگاه داده (UniProt) موجود است را مشاهده نمایید. بعنوان مثال اطلاعاتی مانند عملکرد پروتئین، نام های دیگر آن، جایگاه درون سلولی پروتئین، بیماری زا بودن (مثلا در اثر جهش)، بیان آن در بافت های مختلف، برهمکنش آن با سایر پروتئین ها، توالی آمینواسیدی آن و ... در دسترس می باشد که با کلیک بر هر یک از آنها و یا کشیدن موس به سمت پایین صفحه می توانید آنها را مطالعه نمایید (شکل ۵۰).



شکل ۵۰. بررسی پروتئین یافت شده با شماره دسترسی Q15746 در نتایج حاصل از جستجو در پایگاه داده UniProt

برای مثال ما در اینجا بر روی sequence کلیک می‌کنیم، همانطور که در شکل۵۱ مشاهده می کنید بعلت فرآیندهای alternative initiation (استفاده از چندین کدون آغاز در یک نسخه mRNA و شروع ترجمه در این نقاط) و alternative splicing (اتصال اگزون‌های مختلف با ترتیب‌های متفاوت)، این پروتئین دارای ۱۱ ایزوفرم می باشد که توالی و طول تمام آنها در همین صفحه در دسترس است. با این وجود UniProt برای کاهش افزونگی[۵۷]، توالی یکی از ایزوفرم ها را بعنوان توالی استاندارد[۵۸] انتخاب می کند. معمولاً توالی ایزوفرمی که طول بلندتری دارد و در آن می توان به وضوح دومین ها، پلی مورفیسم ها، تغییرات پس ترجمه ای و... را تشریح و توصیف کرد تحت عنوان توالی استاندارد انتخاب می شود. در مواردی که ژنوم موجود توالی یابی شده باشد، توالی استاندارد معمولاً توالی حاصل از ترجمه توالی ژنومی است (شکل ۵۱).

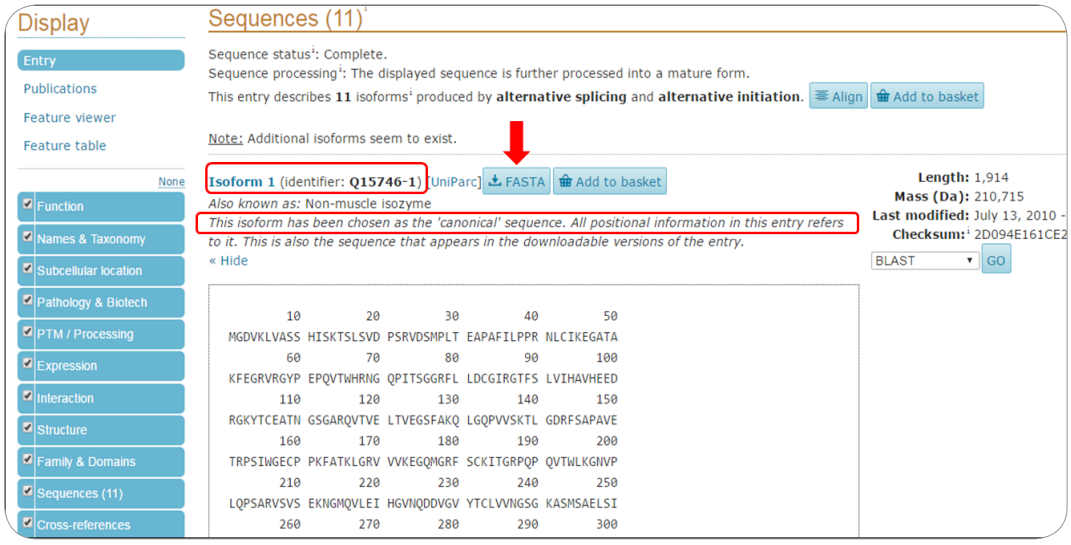


شکل ۵۱. بررسی توالی آمینو اسیدی پروتئین یافت شده
در نتایج با شماره دسترسی Q15746

۵۷ Redundancy

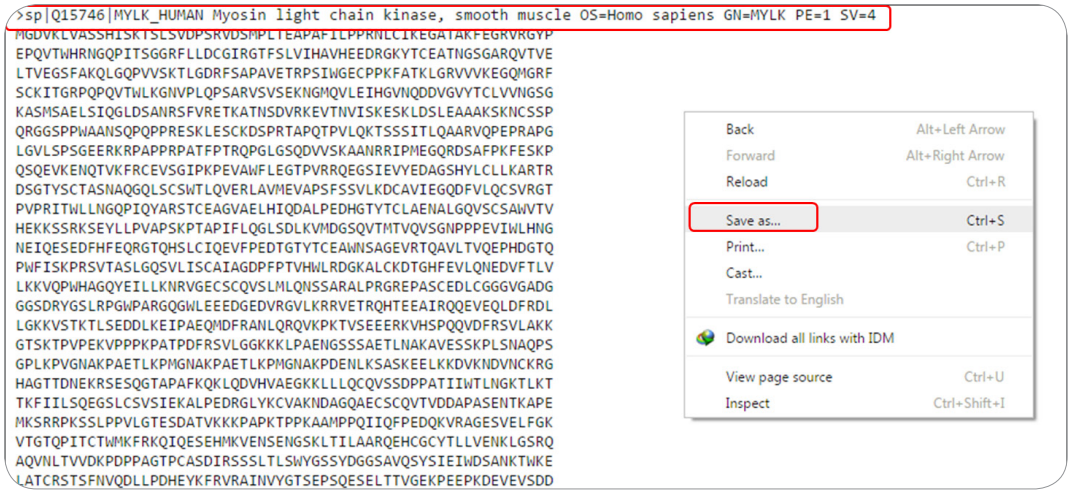
۵۸ Canonical sequence

در این مثال، توالی ایزفروم شماره ۱، توالی آمینواسیدی استاندارد است که در شکل ۵۲ نشان داده شده است. در صورت کلیک بر FASTA (با پیکان قرمز رنگ در شکل مشخص شده) می توانید توالی مربوطه را به فرمت FASTA مشاهده و ذخیره کنید (شکل ۵۲).



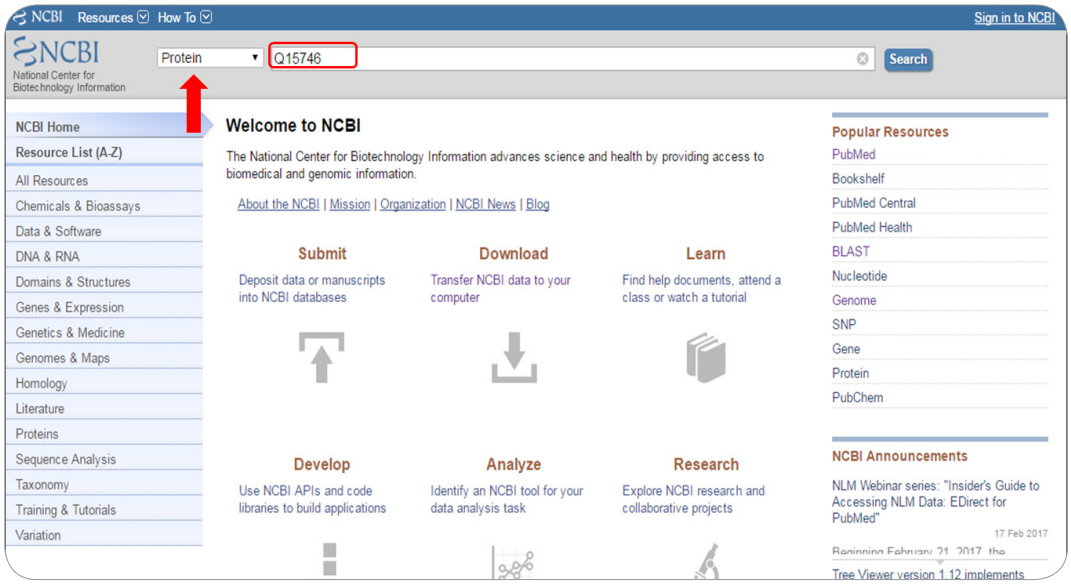
شکل ۵۲. مشاهده اطلاعاتی در مورد توالی آمینو اسیدی استاندارد پروتئین جستجو شده

با کلیک بر FASTA صفحه ای مشابه شکل ۵۳ را مشاهده خواهید کرد که در آن توالی آمینواسیدی پروتئین به فرمت FASTA نمایش داده شده است. با راست کلیک بر صفحه و انتخاب گزینه save as می توانید این توالی را در کامپیوتر خود ذخیره نمایید (شکل ۵۳). توالی ذخیره شده را می توانید با استفاده از برنامه Word یا Notepad باز کنید.



شکل ۵۳. مشاهده و ذخیره توالی آمینو اسیدی به فرمت FASTA در پایگاه داده UniProt

همانطور که اشاره شد شماره های دسترسی هر رکورد در بین پایگاه های داده مختلف مشترک و مختص همان رکورد هستند. برای مثال، ما با استفاده از شناسه Q15746 که شماره دسترسی اولین رکورد در نتایج مثال قبل است، می توانیم پروتئین مربوطه را در پایگاه داده NCBI جستجو نماییم. به این منظور پس از ورود به پایگاه داده NCBI، به کمک فلش رو به پایین در کنار کادر All databases گزینه protein را انتخاب کرده و در مستطیل کنار آن شماره دسترسی پروتئین مورد نظر یعنی Q15746 را وارد می کنیم (شکل ۵۴).

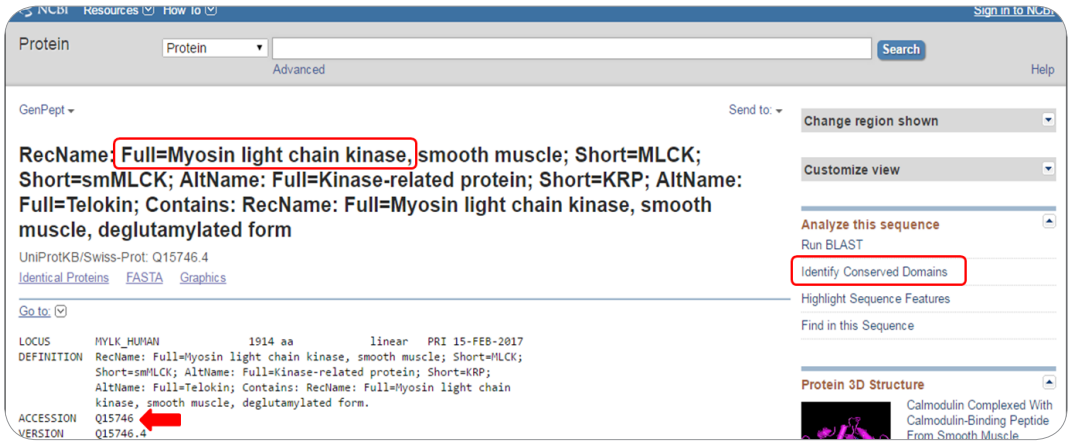


شکل ۵۴ . جستجو پروتئین با استفاده از شماره دسترسی در پایگاه داده Protein در NCBI

همانطور که ملاحظه می کنید صفحه ای مشابه با شکل ۵۵ بعنوان نتیجه این جستجو نمایش داده می شود که مشابه با صفحه مربوط به توالی های نوکلئوتیدی است که قبلاً بررسی کردیم. در ابتدای این صفحه، نام پروتئین کیناز مورد نظر نشان داده شده است. همچنین شماره دسترسی این پروتئین که همان Q15746 است در مقابل Accession درج شده است (با پیکان در شکل ۵۵ مشخص شده).

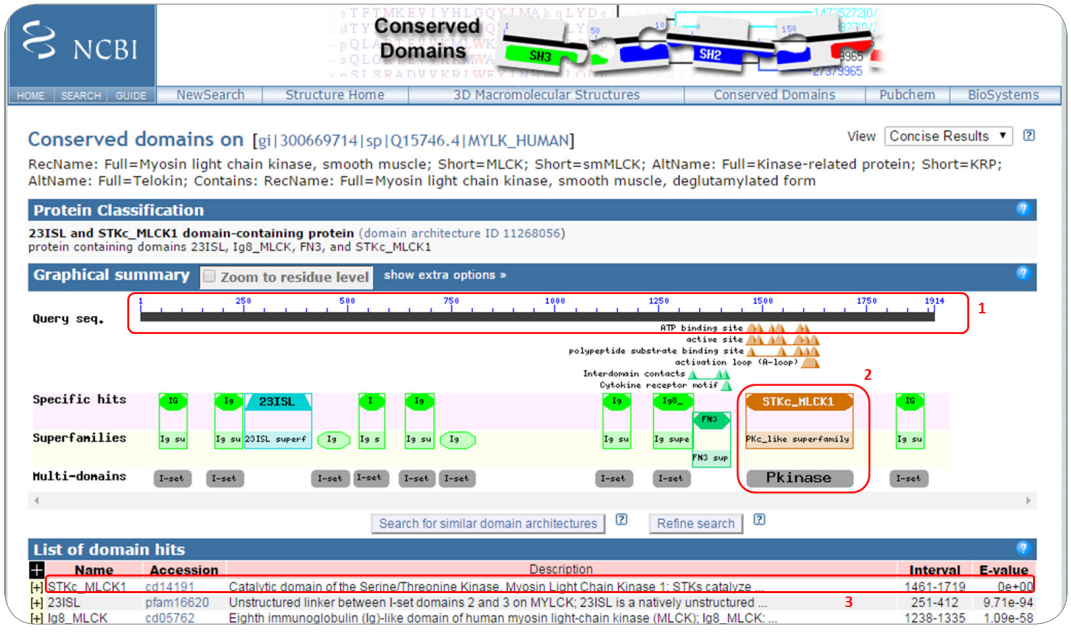
یکی از بخش‌های بسیار مهم که در طی تکامل در پروتئین ها حفاظت شده است، دومین[۵۹] می باشد، انواع مختلفی از دومین ها مانند دومین های کاتالیتیکی، تنظیمی و اتصالی در پروتئین ها وجود دارند. با کلیک بر روی identify conserved domain در سمت راست صفحه (در شکل ۵۵ مشخص شده) می توانید دومین های حفاظت شده در پروتئین کیناز مورد بررسی را مشاهده کنید.

۵۹ Domain



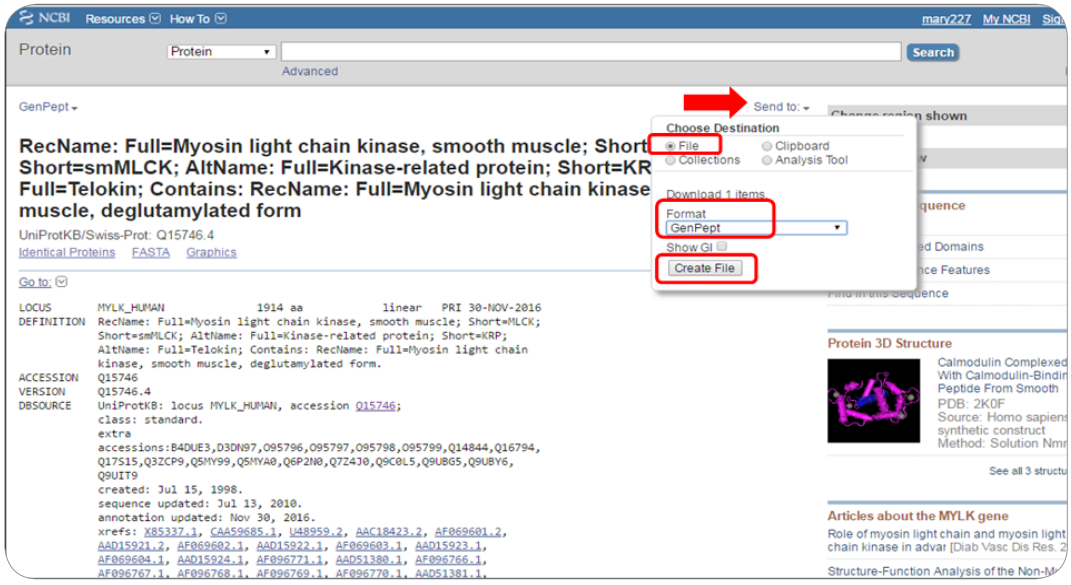
شکل ۵۵. صفحه مربوط به پروتئین مورد نظر با شماره دسترسی Q15746 در پایگاه داده NCBI

پس از کلیک بر روی identify conserved domain با صفحه ای مشابه شکل ۵۶ مواجه خواهید شد که در آن می توانید انواع دومین های حفاظت شده و موقعیت قرارگیری آنها را در این پروتئین کیناز مشاهده کنید (شکل۵۶). کادر ۱ در شکل ۵۹ نشان دهنده توالی آمینواسیدی پروتئین مورد نظر است، دومین کاتالیتیکی این پروتئین که یکی از مهم ترین دومین های حفاظت شده در این پروتئین و مسئول خاصیت کینازی آن است، در موقعیت 1461-1719 این توالی آمینو اسیدی قرار گرفته است که در کادر ۲ و ۳ شکل نشان داده شده است. از آنجایی که این پروتئین به واسطه داشتن این دومین خاصیت کینازی دارد، هر گونه جهش که باعث حذف یا آسیب در این ناحیه شود می تواند باعث کاهش یا عدم فعالیت کینازی پروتئین مذکور شود.



شکل ۵۶. انواع دومین های حفاظت شده در پروتین کیناز Q15746

حال مجددا به شکل ۵۵ بازگردید، در صورتی که بخواهید کل اطلاعات این صفحه را ذخیره کنید و بعد به صورت آفلاین بررسی نمایید، می توانید بر روی گزینه Send to کلیک کرده و فرمت GenPept را انتخاب نمایید. به این ترتیب فایلی با پسوند gp در کامپیوترتان ذخیره خواهد شد که می توانید با استفاده از برنامه word آن را باز و مشاهده کنید (شکل ۵۷).



شکل ۵۷. ذخیره اطلاعات پروتئین مورد نظر در پایگاه داده NCBI

۱۱ یافتن توالی‌های پروتئینی یا نوکلئوتیدی مشابه با توالی مورد نظر با استفاده از برنامه BLAST[۶۰]

نرم افزار بلاست یکی از پرکاربردترین نرم افزارها در علم بیوانفورماتیک می باشد که توسط انستیتو ملی سلامت امریکا (NIH) طراحی شده است. شما به کمک این نرم افزار می توانید توالی آمینواسیدی یا نوکلئوتیدی مورد نظر خود را با یک توالی دیگر یا با مجموعه ای از توالی های موجود در یک پایگاه داده مقایسه کنید و توالی های مشابه با توالی خود را بیابید. اما چرا تشابهات مهم هستند؟ معمولاً موجوداتی با نیا یا جد مشترک دارای توالی های مشابه با یکدیگر هستند، این توالی ها احتمالاً دارای ساختار و عملکرد زیست شناختی مشابه نیز هستند. بنابراین اگر اطلاعاتی مانند توالی، ساختار یا عملکرد در مورد یک پروتئین خاص در یک موجود معین داشته باشیم، می توانیم همین اطلاعات را به پرتئین مشابه با این پروتئین در موجودات دیگر تعمیم دهیم. برای مثال، تصور کنید دانشمندان زمان بسیاری صرف تحقیق در مورد یک توالی آمینواسیدی یا نوکلئوتیدی خاص کرده اند و تمام اطلاعات مربوط به آن را استخراج کرده و در پایگاه های داده ذخیره کرده اند. اتفاقاً این توالی به توالی مورد نظر شما شباهت بسیاری دارد، در این شرایط برای یافتن اطلاعات مورد نیاز در مورد توالی خودتان لازم نیست تمام مراحل قبلی را مجدداً تکرار کنید، تنها کافی است

۶۰ Basic Local Alignment Search Tool

یک جستجو مناسب در پایگاه های داده به کمک برنامه بلاست (Blast) انجام دهید و در مدت چند دقیقه اطلاعات مورد نیاز خود را استخراج کنید.

زیست شناسان به توالی هایی مشابه که دارای نیا مشترک و همچنین ساختار و عملکرد مشابهی هستند، هومولوگ[۶۱] می گویند. اما در چه صورت می توان دو توالی را مشابه در نظر گرفت؟ فرض کنید شما توالی با ۱۰۰ آمینو اسید دارید، در صورتی که شباهت این توالی با توالی دیگری بیش از ۲۵٪ باشند، این دو توالی با یکدیگر همولوگ هستند. در حالی که اگر توالی شما یک توالی نوکلئوتیدی باشد، تشابه بیشتر از ۷۰٪ به معنی هومولوگ بودن است. این مثال تنها یک مقیاس معمولی و ساده بود، واقعیت اندکی پیچیده تر است، در واقع برای این که از هومولوگ بودن دو توالی مطمئن شویم باید اطلاعات دیگری نیز مانند طول نواحی مشابه بین دو توالی، میزان حذف و اضافه شدگی های ناشی از جهش و عددی به نام E-value[۶۲] را نیز در نظر بگیریم. عدد E-value نشان می دهد که به چه میزان شباهت بین دو توالی تصادفی و شانسی است، بدیهی است که هر چه میزان این عدد کمتر باشد، بهتر است زیرا دو توالی واقعاً (نه از روی تصادف و شانس) با یکدیگر مشابهت دارند. در این زمینه به این نکته مهم توجه کنید که شباهت و هومولوگی معادل یکدیگر نیستند، شباهت یک صفت قابل اندازه گیری است ولی همولوگی یک صفت مطلق است. به عبارت دیگر دو توالی یا با یکدیگر همولوگ هستند یا نیستند، شما نمی توانید بگویید دو توالی ۴۰٪ با یکدیگر همولوگ هستند ولی می توانید بگویید که دو توالی ۴۰٪ با یکدیگر مشابه هستند.

همانطور که ذکر شد می توانیم دو توالی را با یکدیگر و یا یک توالی را با مجموعه ی توالی های موجود در یک پایگاه داده با استفاده از برنامه بلاست مقایسه کنیم و به میزان مشابهت احتمالی آنها پی ببریم. در برنامه بلاست، توالی مورد نظر شما که قصد مقایسه آن با سایر توالی ها را دارید، توالی مورد سوال یا query نامیده می شود و توالی های مشابه با توالی مورد سوال که حاصل نتیجه بلاست است، جواب های جور شده یا subject نامیده می شوند. بر اساس نوع توالی و پایگاه داده (آمینو اسیدی یا نوکلئوتیدی) چندین نوع بلاست وجود دارد که عبارتند از:

BLASTP: در این نوع بلاست، یک توالی پروتئینی با پایگاه داده ای که فقط واجد توالی های پروتئینی است، مقایسه می شود. بنابراین این نوع بلاست زمانی کاربرد دارد که بخواهید بدانید پروتئین مشابه با پروتئین مورد نظر شما چیست.

BLASTN: در این نوع بلاست، یک توالی نوکلئوتیدی با پایگاه داده ای که فقط واجد توالی های نوکلئوتیدی است مقایسه می شود. زمانی از این نوع بلاست استفاده می شود که توالی مورد نظر یک توالی نوکلئوتیدی باشد و می خواهید بدانید که این توالی با کدام توالی های نوکلئوتیدی دیگر شباهت دارد و یا زمانی که می خواهیم بدانیم یک توالی نوکلئوتیدی خاص با کدام ناحیه از ژنوم یک موجود مشابهت دارد.

BLASTX: در این بلاست، یک توالی نوکلئوتیدی با پایگاه داده ای که واجد توالی های پروتئینی است مقایسه می شود. زمانی که شما برای اولین بار یک توالی نوکلئوتیدی را در یک موجود معین شناسایی می کنید (موجودی که قبل از شما کسی بر روی آن کار نکرده است) و نمی دانید که این توالی چه پروتئینی را کد می کند، برای شناسایی پروتئین (های) احتمالی کد شونده توسط این توالی از این نوع بلاست استفاده می شود. بنابراین کاربرد مهم این نوع بلاست یافتن ژن های کد کننده پروتئینی در موجوداتی است که تا به حال بر روی آنها کار نشده است.

---

۶۱ Homologues

۶۲ Expected value

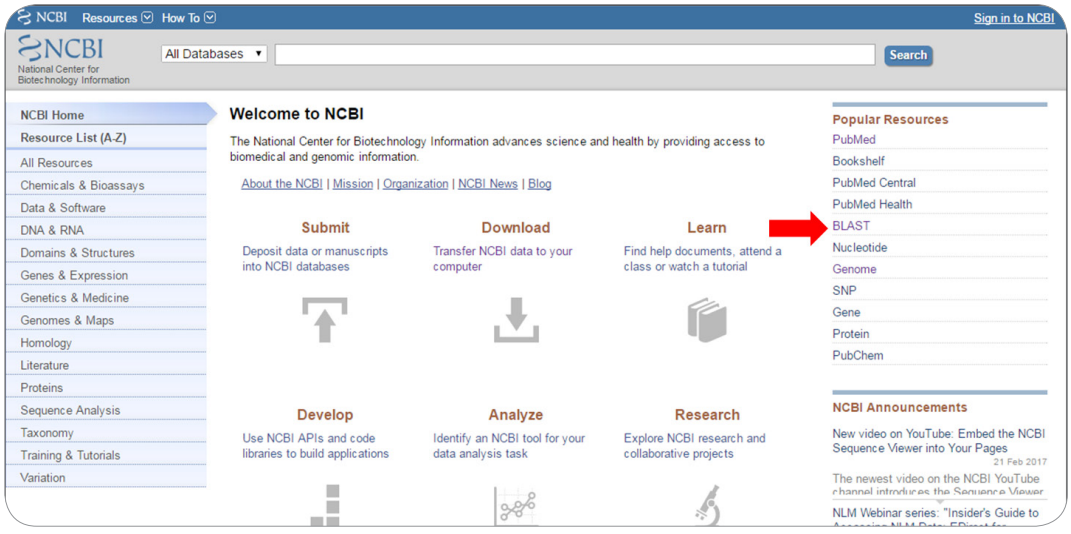
TBLASTN: در این نوع بلاست، توالی پروتئینی با پایگاه داده ای که واجد توالی های نوکلئوتیدی است که در تمامی قالب های خواندن[۶۳] ترجمه شده اند، مقایسه می شود. زمانی که می خواهید بدانید که یک پروتئین مشخص توسط چه ژن هایی کد می شود از این نوع بلاست استفاده می شود.

TBLASTX: در این نوع بلاست توالی نوکلئوتیدی که در تمام قالب های خواندن ترجمه می شود با پایگاه داده ای که واجد توالی های نوکلئوتیدی است که آنها هم در تمام قالب های خواندن ترجمه می شوند، مقایسه می شود. معمولاً این نوع بلاست برای مقایسه توالی های متعلق به گونه های دور از یکدیگر به کار می رود.

۱۲ اجرا BLASTP در NCBI

برنامه بلاست به صورت آنلاین در بسیاری از پایگاه های داده مانند NCBI و UniProt وجود دارد. در اینجا ما از برنامه موجود در پایگاه داده NCBI استفاده می کنیم. فرض کنید شما می خواهید بدانید که آیا پروتئینی مشابه با پروتئین نوکلئولین (نوعی پروتئین در ماده سازنده هستک) هامستر در پایگاه داده Prot-swiss وجود دارد یا خیر؟

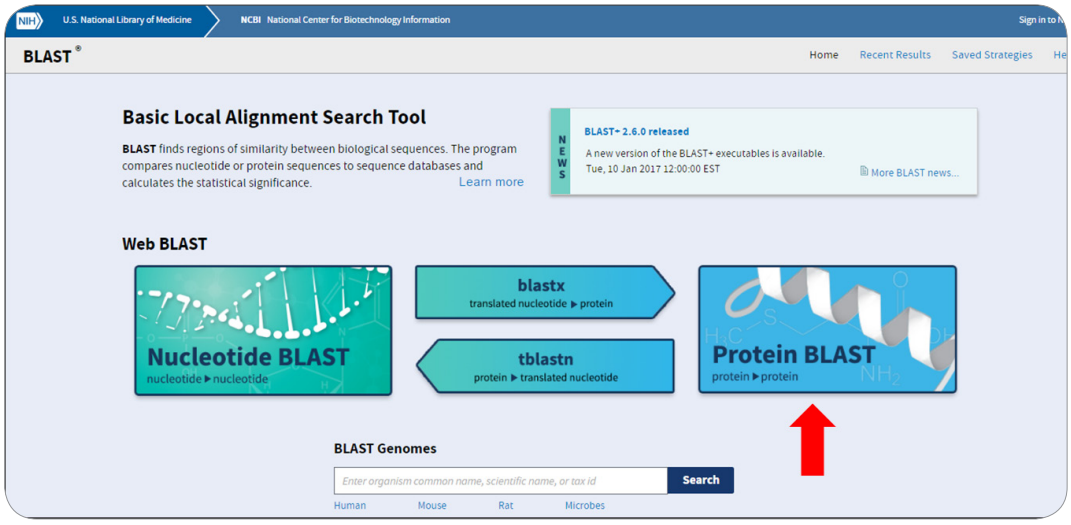
برای پاسخ به این سوال، وارد سایت (www.ncbi.nlm.nih.gov) NCBI شوید و بر روی گزینه BLAST در سمت راست صفحه کلیک نمایید (شکل ۵۸).



شکل ۵۸. دسترسی به برنامه BLAST از طریق پایگاه داده NCBI

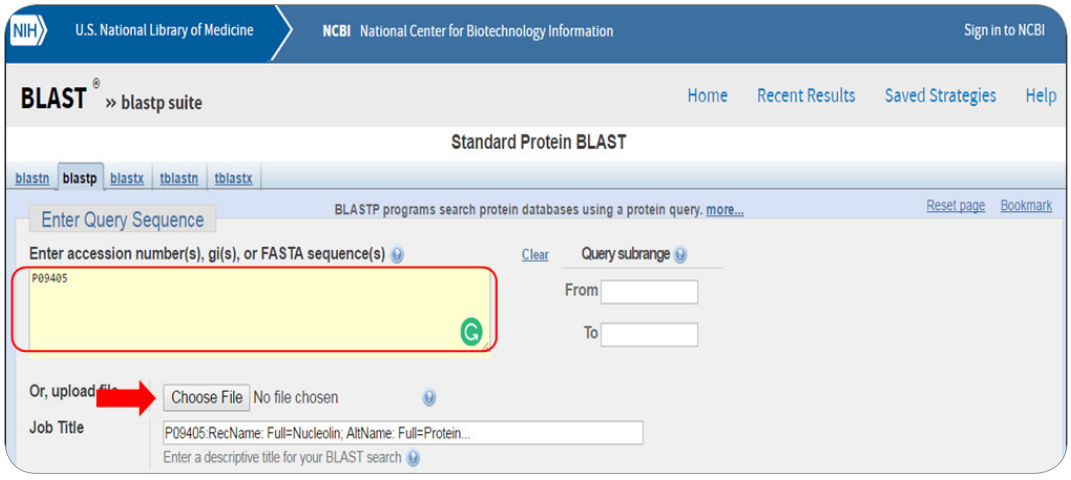
همانطور که ملاحظه می کنید، صفحه ای مشابه با شکل ۵۹ برای شما نمایش داده می شود. در این صفحه، بر روی Protein BLAST کلیک کنید (شکل ۵۹).

۶۳ Reading frame



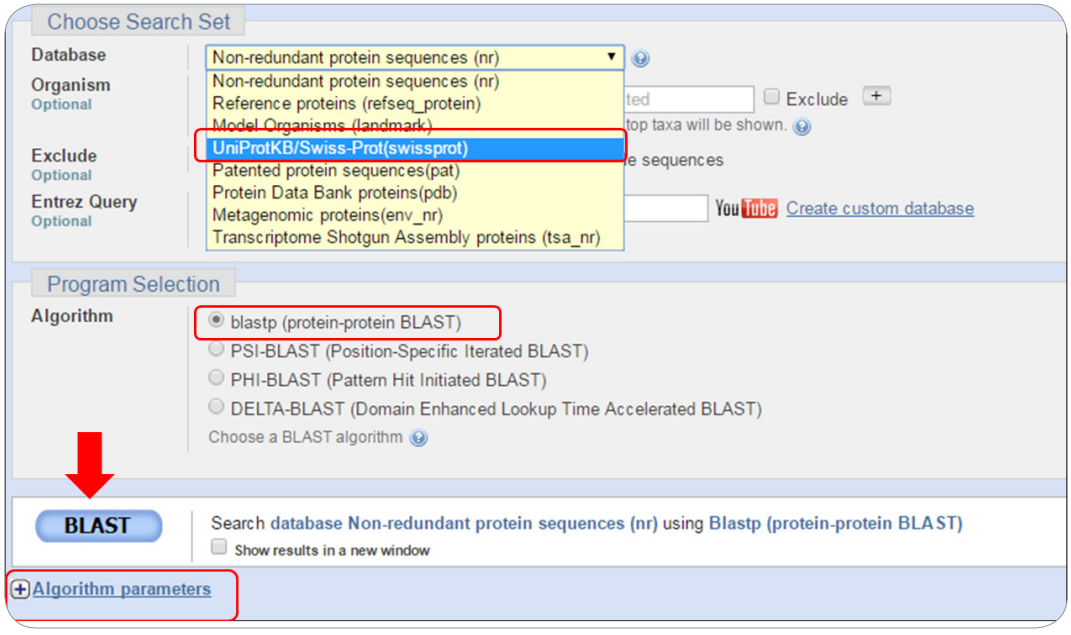
شکل ۵۹. صفحه اصلی BLAST در پایگاه داده NCBI

صفحه ای که پیش رو دارید، صفحه اصلی برنامه BLASTP در پایگاه داده NCBI است که بخش های مختلف آن شرح داده می شوند (شکل ۶۰). در ابتدا توالی مورد نظر خود را (در اینجا توالی پروتئین نوکلئولین هامستر) به فرمت FASTA در کادر مشخص شده در شکل ۶۰ وارد نمایید. در صورتی که این توالی در یک پایگاه اطلاعاتی وجود داشته باشد، می توانید تنها شماره دسترسی یا شناسه ژنی آن را در این کادر وارد نمایید. همچنین در صورتی که طول توالی شما بسیار بلند باشد و شما آن را در کامپیوتر خود ذخیره کرده اید، می توانید از قسمت Choose File (با پیکان در شکل مشخص شده) آن را انتخاب و به برنامه بلاست آپلود کنید. از آنجایی که پروتئین نوکلئولین یک پروتئین شناخته شده است که در پایگاه داده UniProt ثبت شده و دارای شماره دسترسی می باشد. ما شماره دسترسی آن یعنی P09405 را در کادر مشخص شده در شکل وارد کردیم (شکل ۶۰).



شکل ۶۰. وارد کردن توالی یا شماره دسترسی پروتئین مورد نظر به منظور انجام بلاست

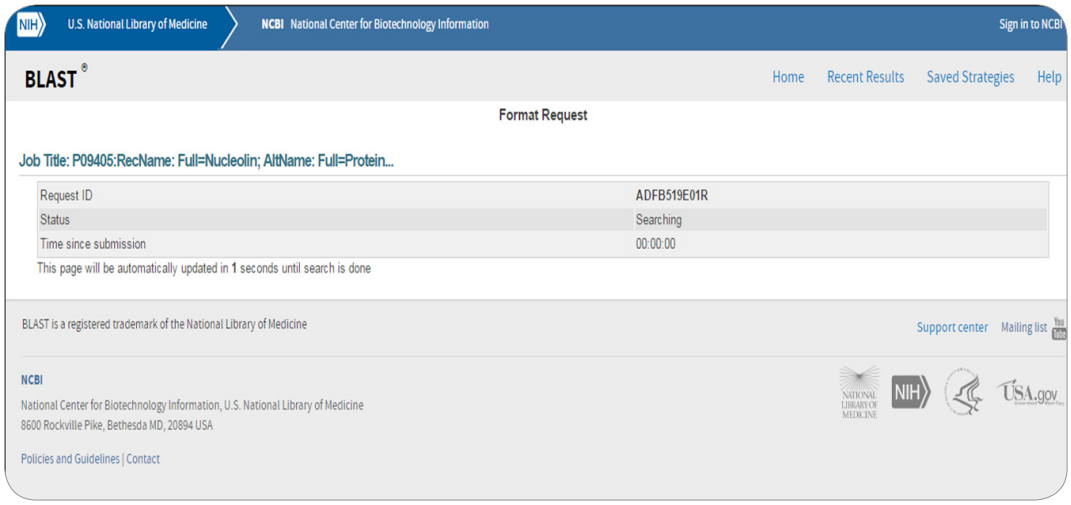
سپس باید پایگاه داده ای که قصد جستجو این پروتئین در آن را دارید، انتخاب کنید. برنامه بلاست بطور پیش فرض جستجو را در پایگاه داده NR یا Non-redundant protein sequences انجام می دهد. این پایگاه داده شامل تمام توالی های ثبت شده در NCBI است. همچنین در صورتی که بخواهید جستجو را در یک موجود خاصی انجام دهید، می توانید نام آن را درکادر مقابل Organism وارد کنید (شکل ۶۱). اگر نام هیچ موجودی وارد نشود، برنامه بلاست جستجو را در تمام موجودات انجام می دهد. در اینجا چون ما قصد جستجو در پایگاه داده Swiss-Prot را داریم، از فلش رو به پایین در کنار کادر Database، این پایگاه داده را انتخاب می کنیم (شکل ۶۱). در قسمت بعدی، انواع مختلف بلاست برای پروتئین ها نشان داده شده که در حالت پیش فرض BLASTP است، که آن را تغییر نمی دهیم. در نهایت بر روی BLAST کلیک کنید تا این برنامه اجرا شود (شکل ۶۱).



شکل ۶۱. انتخاب پایگاه داده مناسب و موجود مورد نظر به منظور انجام بلاست

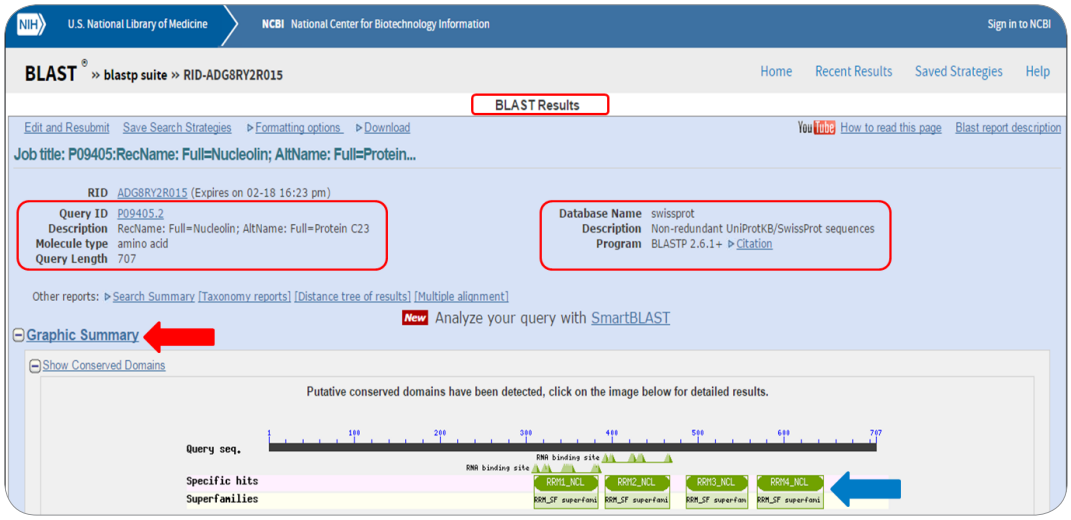
در صورتی که بر روی علامت + کنار Algorithm parameters کلیک کنید، می توانید پارامترهای مختلف برنامه بلاست را به دلخواه خود تغییر داده و تنظیم کنید. معمولاً پارامترهای پیش فرض بهترین پارامترها هستند که برای اجرای موفق یک برنامه تعیین شده اند، لذا ما این قسمت را تغییر نمی دهیم. زمانی که برنامه بلاست شروع به کار می کند، با صفحه ای مشابه شکل ۶۲ مواجه خواهید شدکه بطور اتوماتیک نوسازی[۶۴] می شود و پس از انجام عملیات صفحه نتایج نمایش داده خواهد شد (شکل ۶۲). توجه داشته باشید که یک جستجوی معمولی چند دقیقه طول می کشد، لذا صبور باشید و در هنگام انتظار، بر روی هیچ دکمه یا گزینه ای کلیک نکنید. در صورتی که برنامه بلاست بگوید جستجوی شما بیش از ۲۰ دقیقه طول خواهد کشید، بهتر است مجدداً از ابتدا این کار را انجام دهید.

۶۴ Update



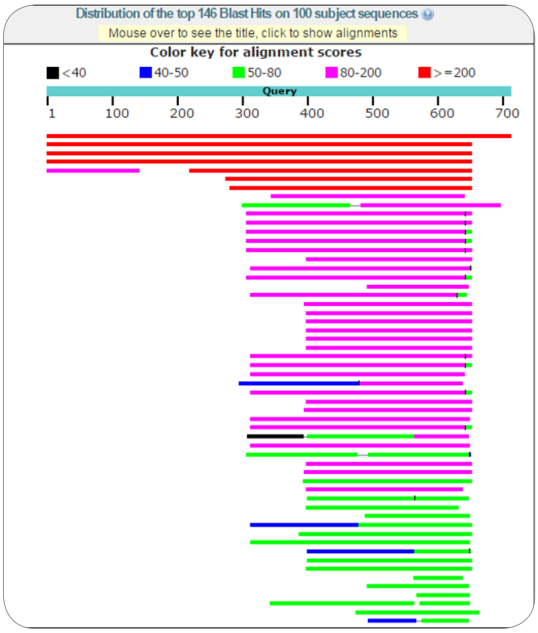
شکل ۶۲ . صفحه ای که نشان می دهد برنامه بلاست در حال جستجو است

پس از اتمام جستجو توسط برنامه بلاست، صفحه نتایج برای شما نمایش داده خواهد شد. این صفحه دارای چندین بخش است. در بخش اول، اطلاعات مختصری در مورد توالی مورد سوال شما (query) و پایگاه داده ای که جستجو در آن انجام شده همراه با نسخه نرم افزار بلاست مورد استفاده ارائه شده است (شکل ۶۳). برای مثال، در جستجویی که انجام دادیم، نام توالی همراه با شماره شناسه آن (query ID)، طول توالی که 707 آمینواسید است، پایگاه داده ای که این جستجو در آن انجام شده (Swiss-Prot) و نسخه ای از نرم افزار بلاست که این جستجو با آن انجام شده است و در اینجا 2.6.1+ است، نشان داده شده است. در بخش بعدی، نتایج حاصل از بلاست به صورت گرافیکی ارائه شده اند. در اولین کادر، موقعیت دومین های حفاظت شده در توالی آمینواسیدی پروتئین نمایش داده شده است که با پیکان آبی رنگ در شکل نشان داده شده اند (شکل ۶۳).



شکل ۶۳. نتایج حاصل از جستجو تشابه با استفاده از BLASTP

در قسمت بعدی نتایج، نتایج حاصل از جستجوی توالی ها به کمک بلاست به صورت گرافیکی نمایش داده شده است. در واقع ۱۰۰ توالی اول حاصل از این جستجو به صورت خطوط رنگی نشان داده می شوند که رنگ ها نشان دهنده میزان تشابه توالی های یافت شده با توالی مورد سوال یا query است. هر چقدر تشابه توالی های یافت شده با توالی مورد سوال بیشتر باشد با رنگ های صورتی و قرمز و هر چقدر این تشابه کمتر باشد با رنگ های رو به سیاه نشان داده می شود، بنابراین رنگ های سیاه پروتئین هایی هستند که نقاط اشتراک بسیار اندکی با توالی query دارند و بهتر است که از آنها صرفنظر شود (شکل ۶۴). به این ترتیب شما به کمک این تصویر می توانید در یک نگاه کلی میزان تشابه توالی پروتئین مورد نظر خود را با سایر پروتئین ها مشاهده نمایید. همچنین در صورتی که موس را بر روی هر یک از خطوط رنگی نگه دارید، نام توالی مربوطه نمایش داده خواهد شد.

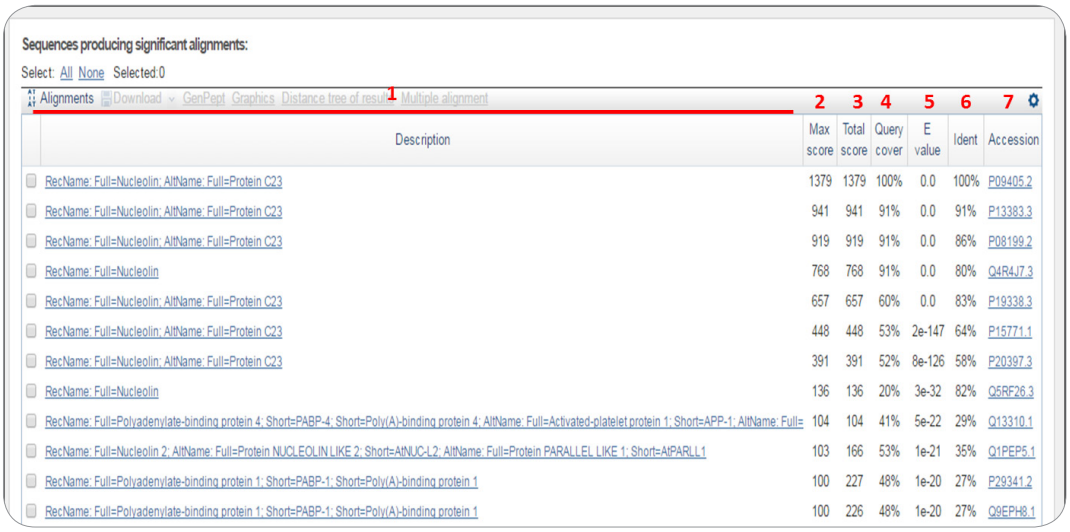


شکل ۶۴. نمایش گرافیکی نتایج حاصل از بلاست

در بخش بعد، نام توالی های یافت شده (شماره ۱) و پارامترهایی (شماره های ۲ تا ۶) که بر اساس آنها توالی مربوطه بعنوان توالی مشابه با توالی مورد سوال توسط بلاست مشخص شده است، در جدولی مشابه با شکل ۶۵ ارائه شده اند، همچنین شماره دسترسی هر توالی (شماره ۷) در این جدول نشان داده شده است (شکل ۶۵). شماره های ۲ و ۳ نشان دهنده امتیاز[۶۵] جور شدگی یا همردیفی هر توالی با توالی مورد سوال است، هر چه میزان این امتیاز بیشتر باشد، شباهت بین دو توالی بیشتر است. معمولاً توالی هایی با امتیاز کمتر از عدد ۵۰ ارزش چندانی ندارند و نمی توانید بگویید که آنها مشابه با توالی پروتئین مورد نظر شما هستند. شماره ۴ در شکل

۶۵ Score

۶۵ نشان می دهد که چند درصد از توالی های یافت شده توسط بلاست با توالی مورد سوال همپوشانی دارند. شماره ۵ یا E-value یکی از مهم ترین معیارهایی است که در این قسمت ارائه شده و نشان می دهد که به چه میزان شباهت توالی های یافت شده با توالی مورد سوال تصادفی و شانسی بوده است. هر چه عدد E-value کمتر باشد شما با اطمینان بیشتری می توانید بگویید که توالی مربوطه با توالی پروتئین مورد نظر شما مشابهت دارد. بطوری که این عدد برای توالی های کاملاً یکسان با توالی مورد نظر شما (هومولوگ) برابر با صفر است. معمولاً در این جستجو، توالی هایی با عدد E-value بیشتر از ۰/۰۱ بعنوان توالی های مشابه با توالی مورد سوال در نظر گرفته نمی شوند. شماره ۷ شماره دسترسی هر توالی است که شما به کمک آن می توانید اطلاعات بیشتری از آن توالی به دست آورید (شکل ۶۵).


Sequences producing significant alignments:

Select: All None Selected:0

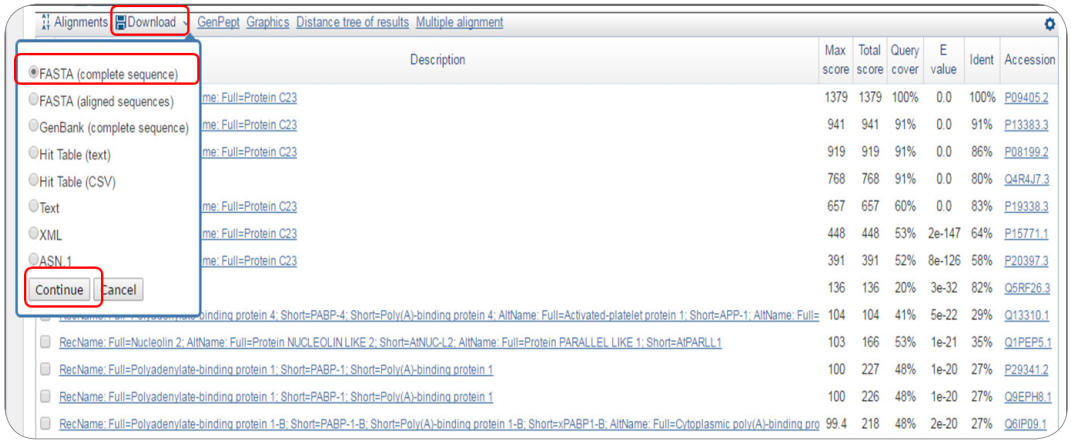
Alignments Download GenPept Graphics Distance tree of result Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
RecName: Full=Nucleolin; AltName: Full=Protein C23	1379	1379	100%	0.0	100%	P09405.2
RecName: Full=Nucleolin; AltName: Full=Protein C23	941	941	91%	0.0	91%	P13383.3
RecName: Full=Nucleolin; AltName: Full=Protein C23	919	919	91%	0.0	86%	P08199.2
RecName: Full=Nucleolin	768	768	91%	0.0	80%	Q4R4J7.3
RecName: Full=Nucleolin; AltName: Full=Protein C23	657	657	60%	0.0	83%	P19338.3
RecName: Full=Nucleolin; AltName: Full=Protein C23	448	448	53%	2e-147	64%	P15771.1
RecName: Full=Nucleolin; AltName: Full=Protein C23	391	391	52%	8e-126	58%	P20397.3
RecName: Full=Nucleolin	136	136	20%	3e-32	82%	Q5RF26.3
RecName: Full=Polyadenylate-binding protein 4; Short=PABP-4; Short=Poly(A)-binding protein 4; AltName: Full=Activated-platelet protein 1; Short=APP-1; AltName: Full=	104	104	41%	5e-22	29%	Q13310.1
RecName: Full=Nucleolin 2; AltName: Full=Protein NUCLEOLIN LIKE 2; Short=AtNUC-L2; AltName: Full=Protein PARALLEL LIKE 1; Short=AtPARLL1	103	166	53%	1e-21	35%	Q1PEP5.1
RecName: Full=Polyadenylate-binding protein 1; Short=PABP-1; Short=Poly(A)-binding protein 1	100	227	48%	1e-20	27%	P29341.2
RecName: Full=Polyadenylate-binding protein 1; Short=PABP-1; Short=Poly(A)-binding protein 1	100	226	48%	1e-20	27%	Q9EPH8.1

شکل ۶۵. نتایج حاصل از بلاست

معمولاً بهترین نتایج حاصل از بلاست یا به عبارت دیگر توالی هایی با بیشترین تشابه به توالی مورد نظر شما مواردی هستند که در ابتدای جدول در شکل ۶۵ لیست شده اند. این موارد در نمایش گرافیکی نتایج بلاست با خطوط قرمز رنگ نشان داده شده اند. برای دانلود توالی های دلخواه، کافی است با زدن تیک درکنار مربع هر توالی آنها را مشخص کرده (شکل ۶۶.) و سپس بر روی گزینه download کلیک کرده و از منوی باز شده اولین گزینه یعنی FASTA (complete sequence) انتخاب و بر روی continue کلیک نمایید تا توالی های مورد نظر خود را به فرمت FASTA دانلود و در کامپیوتر خود ذخیره نمایید. فایل ذخیره شده را می توانید با استفاده از برنامه word یا notepad باز کنید (شکل ۶۷).

شکل ۶۶. انتخاب توالی های مورد نظر جهت دانلود و ذخیره



شکل ۶۷. دانلود و ذخیره توالی های موردنظر در نتایج حاصل از بلاست

فصل چهارم

اصول آماده سازی، ویرایش و هم ردیف سازی توالی

۱ آماده سازی توالی

از آنجا که فرمت فایل های ورودی نرم افزارهای مختلف متفاوت است، بنابراین اولین قدم برای کار با توالی نوکلئوتیدی یا پروتئینی این است که فرمت فایل توالی مورد نظر به فرمت فایل ورودی نرم افزارهای مورد استفاده تبدیل شود. فرمت FASTA فرم رایج برای نمایش توالی های نوکلئوتیدی و آمینواسیدی است و همچنین فرمت استاندارد ورودی بسیاری از نرم افزارها می باشد. در صورتی که فایل حاوی توالی مورد نظر به این فرمت نباشد می توان با کمک بسیاری از نرم افزارها آن را به فرمت FASTA تبدیل کرد. این فرمت بصورت شکل ۱ می باشد.

در این فرمت هر نوکلئوتید یا آمینواسید با استفاده از کدهای تک حرفی نشان داده می شود. همانطور که در شکل ۱ نشان داده شده است، برای تهیه فرمت FASTA، باید Notepad را باز نموده و قبل از وارد نمودن توالی در اولین سطر علامت «>» را قرار داد. بعد از این علامت می توان نام دلخواه را به توالی داد که می تواند بر اساس نام ارگانیسم یا ژن مورد نظر باشد. این سطر Header نامیده می شود. سپس می توان توالی ها را وارد نمود. باید دقت شود که برای ایجاد فاصله در نام فایل باید از علامت «_» استفاده شود. در پایان پس از ذخیره ی فایل باید پسوند نام فایل را از « txt*. » به «fas *. » تغییر داد (شکل ۱).

Untitled - Notepad

File Edit Format View Help

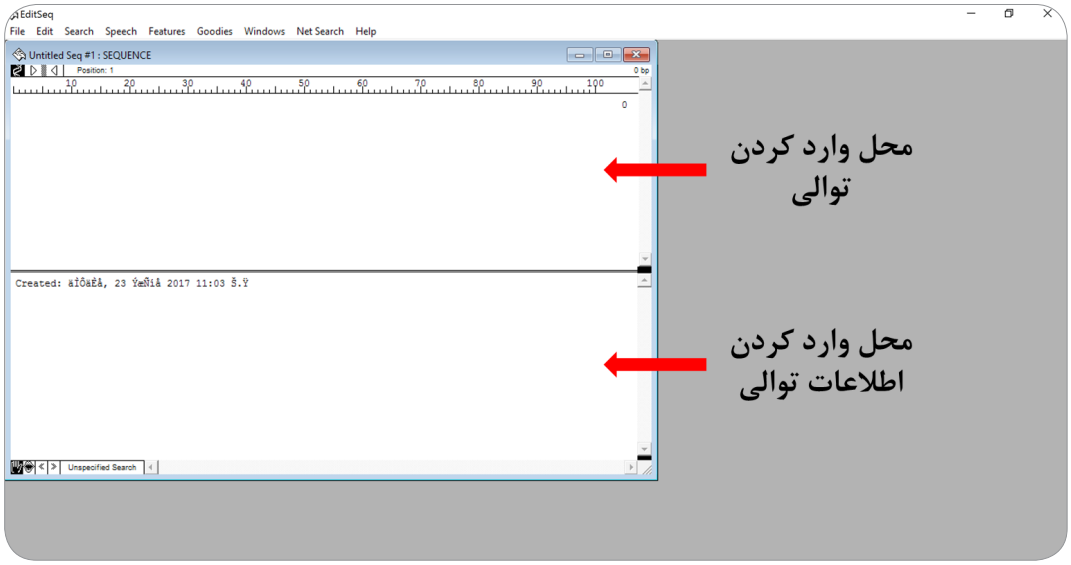
```
>c50249_g1_i2
CTTTATAAGTACAAGTCCATTACCAAAGGAGCTCATGAAATACTTGGAAGAATATAGTGATATAATTCAGTGGTCAGCAATCATTATGCGATTTGCAAACGACATGGGGACGTCATCTTACGAGATAGCAAGAGGGGATAATCCTAAATCTATC
AACTGTTACATGCATGAAACTGGAGTGTCTGAAGAAGAGGCTCGTGAACACATACAGAAAATGATTGCTGCAACGTGGATGAAGATGAACAGAGATCGGCTTGGAAACCCTCATTTTCCAGAAACTTTTGTTACAGTTGCGATGAACCTAGCCA
GGGTGGCTCAGTGTGTATACCAATATGGAGATGGGCATGGTATTCAAGACACGAGTAAAGATTGCATAATGTCATTATTTGTTGATCCCATTCCATTAATTAATGGAGATGGTTACAAGAAATGATGAAAGTTGAGATAACTTTCTTCGTCTTT
TATTTTTAACG

>c50354_g2_i1
ATTAATCTATTTATTATTAAGATTTATTTGTTTATCATATATCTTTTTTGTAACTGTCAATCATAAATTTTGTCCCTCTAGGTAAAACCTCAGATCTAAAGAGGAGTTCATAAACTCTAATTGTTTATAAAATATTTTTACCACTTCTTGATTA
TTTGGTAAAATATTTGCTTCTATGTTTTTCTTTTCAAATTTAGCCCGAGACAGCAAGAGGAGATGCTCTTAAATCTATCCACTGTTACATGCATGAAACTGGAGCTTCTGAAGAACAAGCTCATGAACACATACAAGAAATGATTGTTGTAGCG
TGGATGAAGATGAACAAAGATCGACTTGGAAACCCTCATTTTGCAGAATTTGTTACAACTGCGATGAGCATAAGCAGGATGT

>c47886_g1_i2
CTAATGTTCTGCTTTTGCGTTATGAAGGCGTCCGGTTGAAAGACCTCCACGTGATCATGGAACATTTCACGTGCACCGAACTATTAACTCGCGATTTTGTTCACGAATTGGATTCCATCAACTACTCCCTTCCCTCTCTTTTTCTATAAAATCA
AACCACAGGACAAAACTCAACCTCAAACTCAATTCGAACATTTTATTCTTGTTTTTTATCGTTTTAACATTGTCCGTTAAAAATGAATCCCTCACCCTCCGTCTCCGATGGCCAGGTCGAACACGTCACCTTCTCCTCCACTGTTAAACCTCCT
GGCTCCTCTAAAACTCATTTCCTCGGTGGCTCAGGGGTGAGAGGTTTGGAAATTCAAGGCAAGTTCGTGAAGTTCACGGCGATCGGAGTGTACGTTGAGGATAGCGCCGTGTCGTCACTCGCCAGCAAGTGGAAGGGCAAGAGTGCTGATGAGT
TATTTGAATCCGTTCCGTTCTTTAGAGATGTCGTTACAGGTCCCTTTGAGAAATTCATTAGAGTGACAATGATCTTGCCATTAACTGGTCAACAATACTCAGAGAAAGTTGCAGAAAATTGCGTCGCCATTTGGAAATCACTTGGACTTTACAC
TGACGCTGAATCCAAAGCCATTGAGCAATTTCTTGAGGCCTTCAAGGGCGAAAACTTTCCACCTGGCGCTTCAATTCTTTTCACACTATCACCCTTCGGATCATTGACAATTGGCTTTTCGAAGGACGAATCGATAGGGGAATCTGGGAAAGTA
GTGATAGATAATAAACTACTTGCAGAGTCAGTTTTGGAGTCAATTATTGGGAAGAATGGAGTTTCTCCTGCGGCAAGGAAGAACTTGGCTGAAAGATTTTCAAAATTGTTGAATGGAAAAGAGAAATCAGAAGCCGATAATTGTGCTAAAGAAT
TCAAAGGAGTGGACACGAATATCATCCAAATTGAAACCGGAAAAGTGTGAAGAGACCGGCGTAGATGTCTGGTTCATAATAATGTTAATGTATGGTTTTCAACTCTTGTGATGGCTTTACCTGAGAATGAACTTTTATGTTTAGAGATTTGAGA
TCCCTTAACTTGAATTTTGATTACGTATACAAATTGTTCTCGTGATATGATAAAAAACAAAAAGGTTAGGCCTTGTTTGTATTTCCAATCGAATCAATAATTAAAATAAAATACACGTGGAGATAGGCCATTACCTGATTAATGGGGTAAAGAA
ATACCTTCTTAAAAGTTACAATACATCCAGGGCAGTGTCAATGTAAAGG

>c47886_g1_i1
CCCTTTGAGAAATTCATTAGAGTGACAATGATCTTGCCATTAACTGGTCAACAATACTCAGAGAAAGTTGCAGAAAATTGCGTCGCCATTTGGAAATCACTTGGACTTTACACTGACGCTGAATCCAAAGCCATTGAGCAATTTCTTGAGGCCT
TCAAGGGCGAAAACTTTCCACCTGGCGCTTCAATTCTTTTCACACTATCACCCTTCGGATCATTGACAGTAAGTAAAAGTTTACATCAAAATAAAAATATAAATGGAAGTCTGAAAACAGAGAAGAGGCCTAACCCTTGCCTTCGTATTATTCT
GCATGAAAATTGCAGTTGACTTTTGTATACGAAGACGTATCTAAATGAAAAATGTATAAAAAAGGTTATTTAAGGGGTCACATTTATATATGTTTTTTATTGTTCTGCAGATTGGCTTTTCGAAGGACGAATCGATAGGGGAATCTGGGAAAGT
AGTGATAGATAATAAACTACTTGCAGAGTCAGTTTTGGAGTCAATTATTGGGAAGAATGGAGTTTCTCCTGCGGCAAGGAAGAACTTGGCTGAAAGATTTTCAAAATTGTTGAATGGAAAAGAGAAATCAGAAGCCGATAATTGTGCTAAAGAA
TTCAAAGGAGTGGACACGAATATCATCCAAATTGAAACCGGAAAAGTGTGAAGAGACCGGCGTAGA

>c47798_g2_i1
TCTCGCAGATTCTAAGTAAAACGGTAATATATATATTCTTATTTTGTATATAATGTTCTATAGTTTATACTTTTCTCTAAAAGTATCTTTTTCTAACTACATTTATCCTATGCACATTAGTGGGCAGGTTCGTTCCAAGCCTACATGAAGGAGG
CACAGTGGTACCACAGTCAATATAAACCAACGTTAGAAGAATATATGAAAAACGCATGGTTATCGATTACTGGCCCGACGATGACACTTGTATCATACCTTGCTTCAAAATATCCAATAACCGAAAAGGAACTAGAGTTTTTTGAAAGCGGACC
GGATTTAGTATACTGGGCATCCATGGTTTTCCGGCTTCAAGATGATTTGGGAACTTCATCGGATGAGGTACAAAGAGGGGATGTTCCTAAATCAGTGCAGTGTTACATGAATGACACTAATTGCTCTGAGGAAGTTTCTCGAGAACACATGAAG
CATCTAATGAGGCAGTTGTGGAAGAAAATAAATTTTTATAGAGCTGATACAAATTGTCCCCTACCTAAAACTATTAAAGAAATGATATTGAACCTGGTGAGAACGTCCCATTGCGTGTATCTACATGGAGATGGGCATGGTATTGAAGGCCAAG
AGGTTACGGATGTTCTTTCGTCATTATTCTTTCAGTCTATTCCACTGTAGGAGAAAAACATGTATTTGCATTATCTTCTGATTGCAAGACAAAGAATATCCATGATGTATTTCAAACAAATATCGGAAATTTTAATCATCTCTTTGAAGAACAC
ATAAAATTCAGAGTAATTAAGAAAGGCATTAGACATTCTGGTTCCAATTTCTTGGTCTATTTGTCACCACAAAATTGTTATTTGTTGATGCTCGATTCCAGGTGAAGGATCCTTTCCCAACTTGGAATCCATTGACACTAGCCTCAACACCCGT
```

شکل ۱. نمایی از فرمت FASTA در پنجره ی Notepad

علاوه بر نرم افزارهای ویرایشگری مانند Notepad، نرم افزارهای تخصصی مانند EditSeq و Chromas هم به راحتی قادر به ساخت فرمت FASTA می باشند. نرم افزار EditSeq را می توان در مجموعه ی Laser gene تهیه نمود. آخرین نسخه ی این نرم افزار از نشانی زیر قابل تهیه می باشد.

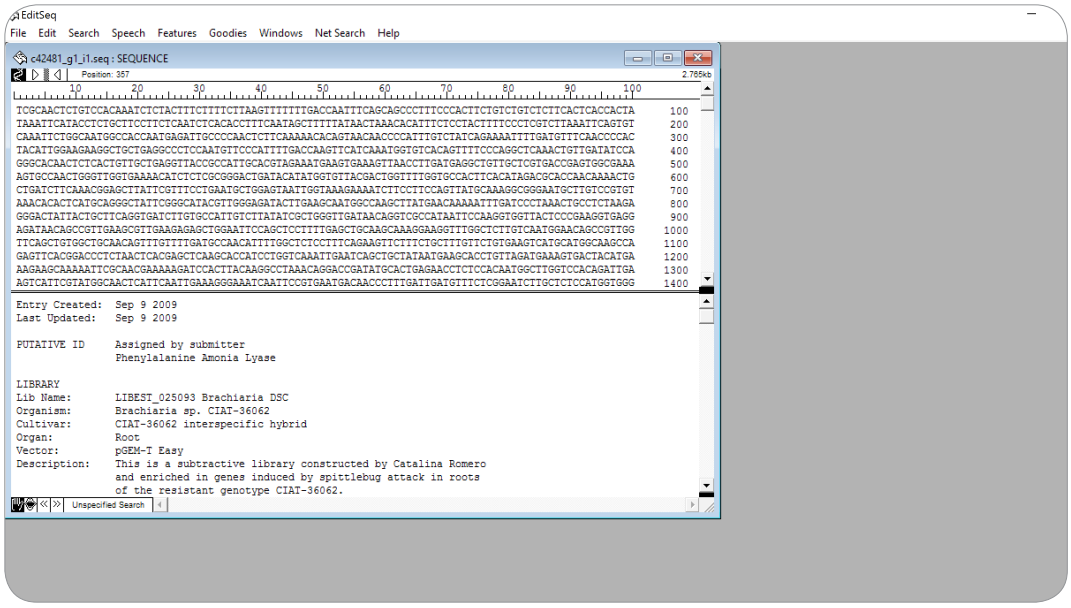
https://www.dnastar.com/t-editseq.aspx

پس از اجرای نرم افزار از منوی File و سپس New و در آخر انتخاب گزینه ی New DNA پنجره ی زیر باز می شود (شکل ۲).



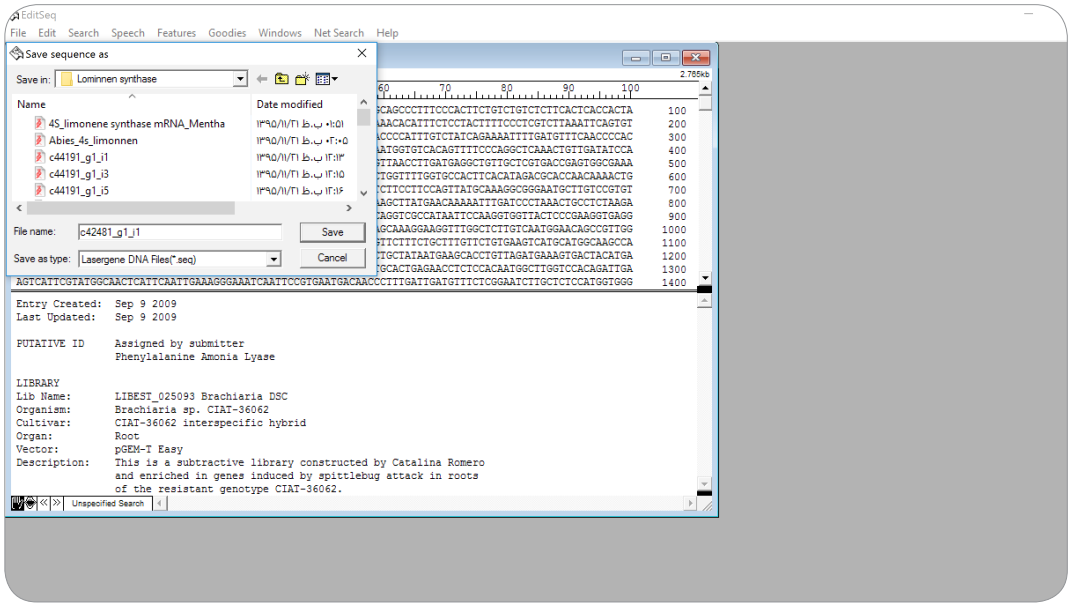
شکل ۲. محل وارد کردن توالی و اطلاعات آن در نرم افزار EditSeq

پس از وارد کردن توالی در قسمت بالای پنجره و اطلاعات مربوط به آن که می تواند اطلاعات بانک ژن باشد در قسمت پایین پنجره، صفحه ی زیر نمایان می شود (شکل ۳).



شکل ۳. وارد کردن توالی و اطلاعات بانک ژن در نرم افزار EditSeq

از طریق منوی File و انتخاب گزینه ی Save as می توان توالی مورد نظر را بصورت فرمت FASTA ذخیره نمود (شکل ۴).



شکل ۴. ذخیره سازی اطلاعات به فرمت FASTA بوسیله ی نرم افزار EditSeq

۲ ویرایش[۱] و هم ردیف سازی[۲] توالی

ویرایش توالی به معنی هر نوع تغییری است که روی توالی اولیه مانند ترجمه ی توالی DNA به پروتئین، حذف بخشی از توالی با کیفیت پایین مثل ابتدا و انتهای توالی، ساخت توالی مکمل معکوس و ... اعمال می شود. ویرایش توالی مورد نظر می تواند یک پیش نیاز برای انجام آنالیزها و مطالعات بعدی باشد.

هم ردیف سازی یعنی جابجا نمودن دو یا چند توالی در کنار هم تا آن حد که بیشترین جزء مشابه (نوکلئوتید یا اسید آمینه) روبروی هم قرارگیرند. هم ردیف سازی از طریق الگوریتم ها و نرم افزارهای مختلفی انجام می گیرد. بطور کلی دو نوع هم ردیف سازی وجود دارد. نوع اول هم ردیف سازی دوگانه نام دارد که توالی مورد نظر ما بصورت تک تک با توالی های پایگاه داده هم ردیف می شود. انواع بلاست در بانک ژن از این نوع هم ردیف سازی استفاده می کنند. نوع دوم هم ردیف سازی، هم ردیف سازی چند گانه است که توالی های مورد نظر ما با یکدیگر با استفاده از الگوریتم های پیچیده تری هم ردیف می شوند.

هم ردیف سازی می تواند با اهداف متنوعی انجام گیرد ولی از پرکاربردترین این اهداف، تعین عملکرد توالی مورد نظر از طریق بلاست، پیدا کردن جهش ها و همچنین یافتن نقاط حفاظت شده در توالی می باشد. در برخی موارد برای طراحی پرایمر لازم است نقاط حفاظت شده در توالی ژن مورد نظر هدف طراحی پرایمر قرار گیرند. در ادامه به آموزش کاربردی برخی از نرم افزارهای قابل استفاده برای ویرایش و هم ردیف سازی توالی می پردازیم.

۳ آموزش کاربردی ویرایش توالی با نرم افزار Chromas 2.6.2

نرم افزار Chromas، نرم افزاری رایگان بوده که بعنوان نرم افزار نمایش دهنده ی کروماتوگرام توالی DNA، می تواند آن را در هر دو جهت Forward و Reverse complement رسم نماید. با استفاده از این نرم افزار می توان کیفیت توالی را بررسی و ویرایش توالی را انجام داد. علاوه براین این نرم افزار با کمک نرم افزارهای هم ردیف ساز در تشخیص موقعیت جهش و همچنین سنجش دقت خوانش توالی ها در محل رخداد جهش کاربرد دارد. داده های حاصل از توالی یابی به فرمت هایی نظیر seq و abi هستند. این نرم افزار قادر به خوانش فایل هایی با پسوندهای abi، scf و ztr می باشد. آخرین نسخه ی این نرم افزار از طریق نشانی زیر قابل دانلود می باشد.

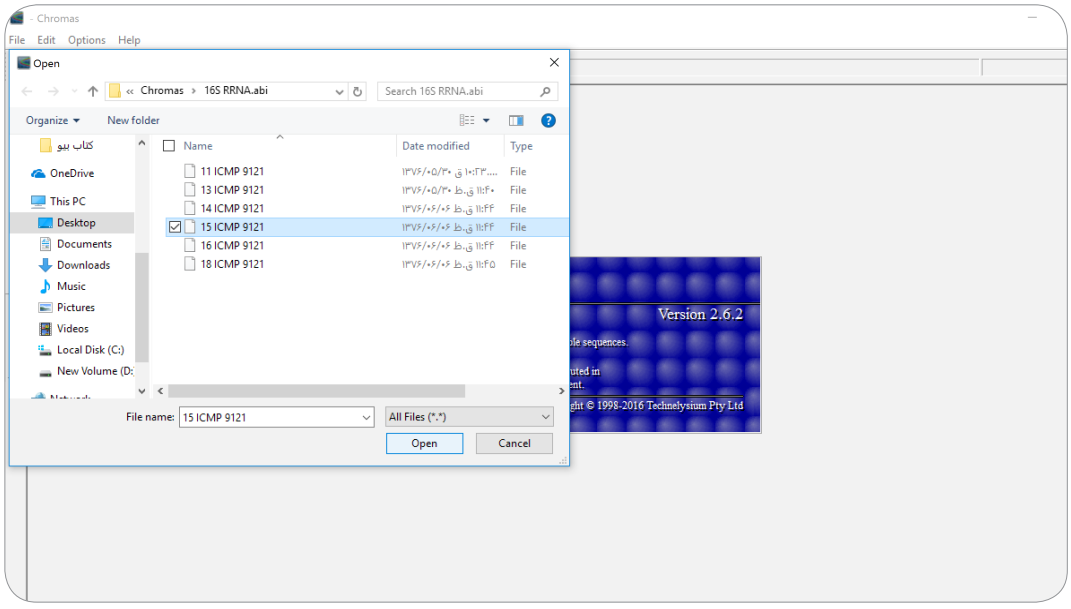
http://technelysium.com.au/wp/chromas/

پس از اجرای نرم افزار برای باز کردن فایل مورد نظر محتوی توالی DNA باید از منوی File گزینه ی Open، فایل مورد نظر را از محل ذخیره شده باز نمود (شکل ۵).

---

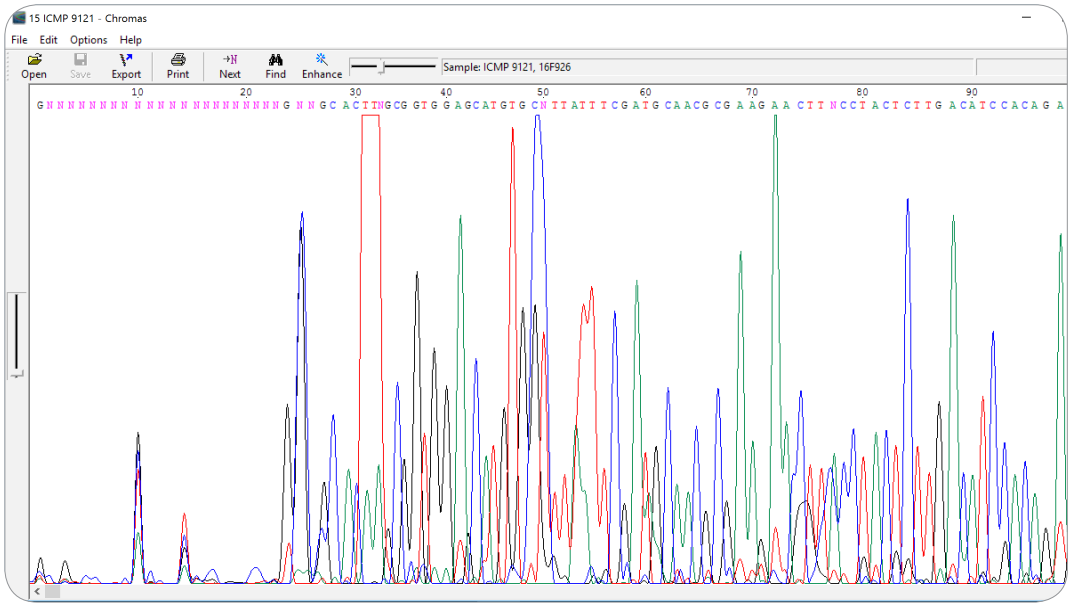
۱ Edit

۲ Alignment



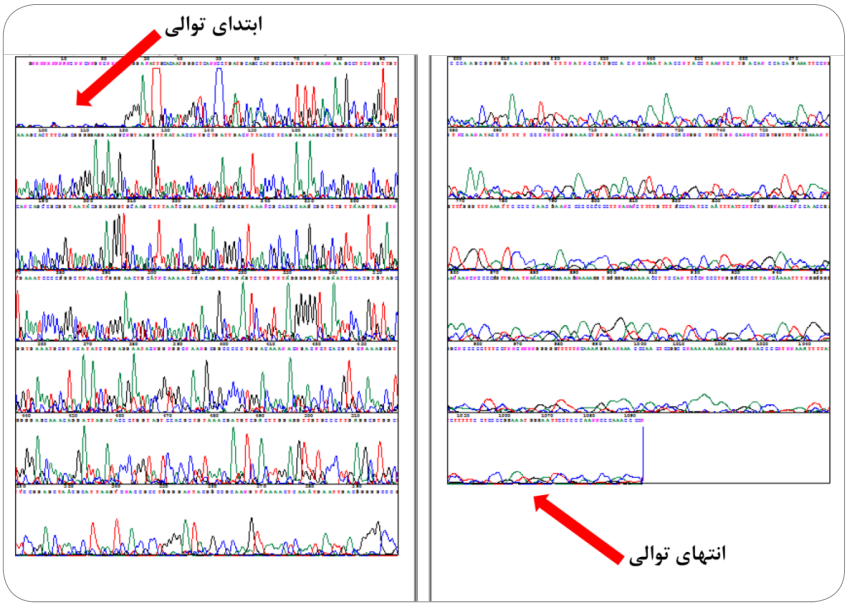
شکل ۵. صفحه ی اول نرم افزار Chromas

پس از باز کردن فایل مورد نظر پنجره ی زیر نمایان می شود. در این نرم افزار مانند سایر نرم افزار ها هر کدام از نوکلئوتیدهای G ، A ، T و C با رنگی خاص مشخص هستند (شکل ۶).



شکل ۶. نمایش کروماتوگرام در نرم افزار Chromas

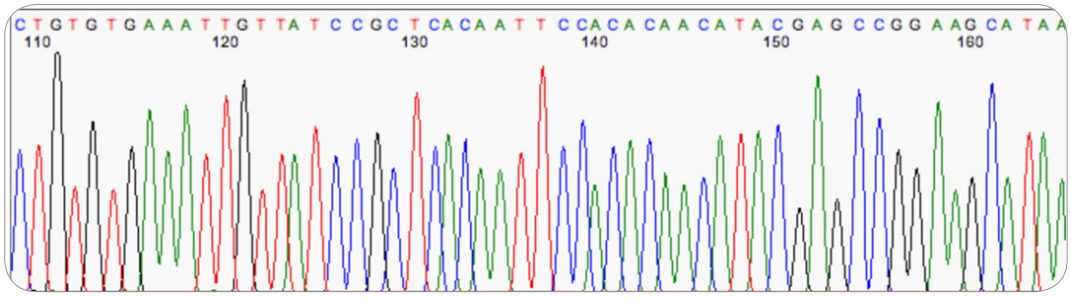
پس از باز شدن کروماتوگرام باید کیفیت توالی را مورد بررسی قرار داد و کیفیت ابتدا و انتهای توالی را با ناحیه ی میانی آن مقایسه نمود. معمولاً کیفیت ابتدا و انتهای توالی پایین تر از ناحیه ی میانی آن است (شکل ۷).



شکل ۷. مقایسه ی ابتدا و انتهای توالی با ناحیه ی میانی آن در کروماتوگرام

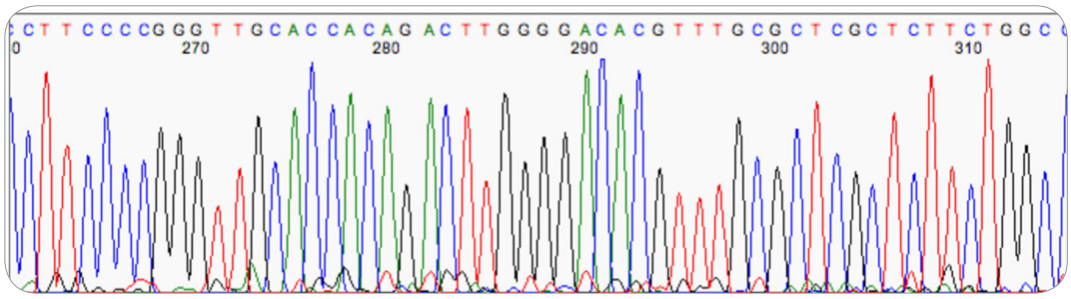
در کروماتوگرام هر نوکلئوتید حداقل بصورت یک منحنی نمایش داده می شود هر چقدر این منحنی ها تفکیک یافته تر و دارای نقطه ی اوج بالاتری باشند، کیفیت توالی یابی بهتر است. Noise در Baseline ممکن است وجود داشته باشد ولی با وجود یک DNA الگو و پرایمر مناسب باید حداقل باشد.

شکل زیر یک نمونه از توالی با کیفیت عالی را نشان می دهد که در آن فاصله ی پیک ها مساوی است و Noise در Baseline وجود ندارد (شکل ۸).



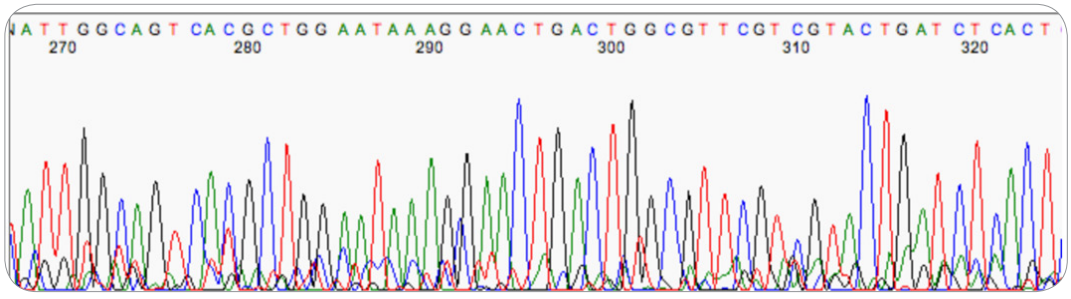
شکل ۸- نمونه ای از توالی بدون Noise در Baseline

شکل زیر یک نمونه از توالی است که مقدار کمی Noise در Baseline دارد، ولی به دلیل اینکه هنوز قله های حقیقی در پیک ها قابل تشخیص هستند، این توالی از لحاظ کیفیت قابل قبول است (شکل ۹).



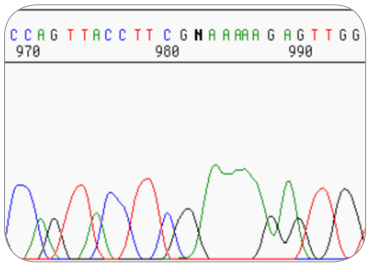
شکل ۹. نمونه ای از توالی با مقدار کمی Noise در Baseline

این نمونه از توالی، مثالی از توالی با کیفیت پایین است که مقدار زیادی Noise در Baseline دارد. در نواحی ۲۷۱ و ۲۷۳ قله های چند رنگ و قله های بینابینی نزدیک به نواحی ۲۹۱ و ۳۰۱ وجود دارد و همچنین تعیین نوکلئوتید واقعی در ناحیه ی ۳۱۰ غیر ممکن است (شکل ۱۰).



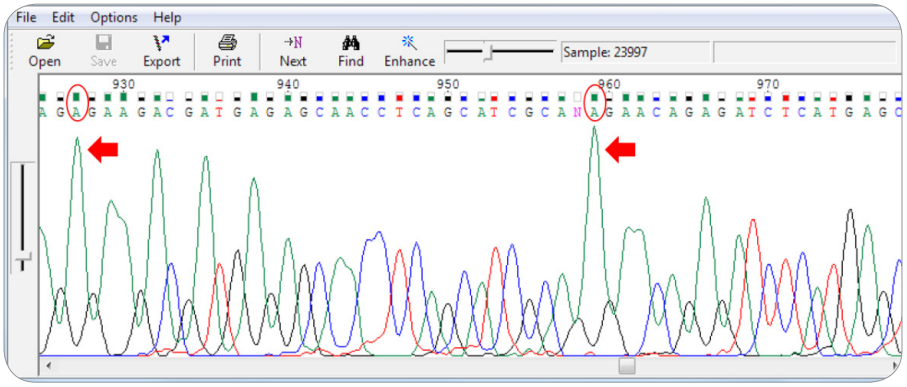
شکل ۱۰. نمونه ای از توالی با مقدار زیادی Noise در Baseline

شکل زیر نمونه ای از توالی با کیفیت پایین را نشان می دهد که فاصله ی پیک ها یکسان نبوده و منحنی ها تفکیک شده نیستند و نقطه ی اوج بالا و قابل توجهی را ندارند. فقط ماهیت تعداد کمی از نوکلئوتیدها را بصورت قابل اطمینان می توان تعیین کرد (شکل ۱۱).



شکل ۱۱. نمونه ای از توالی با کیفیت پایین

مربع رنگی بالای توالی نوکلئوتیدی (Quality value) نشان دهنده ی دقت و تمایز خوانش نوکلئوتیدها است. هر چقدر فضای بیشتری از این مربع پر باشد دقت خوانش بیشتر است. همانطور که در شکل زیر مشاهده می کنید بلندترین و تفکیک یافته ترین پیک ها دارای پرترین مربع ها هستند (شکل ۱۲).

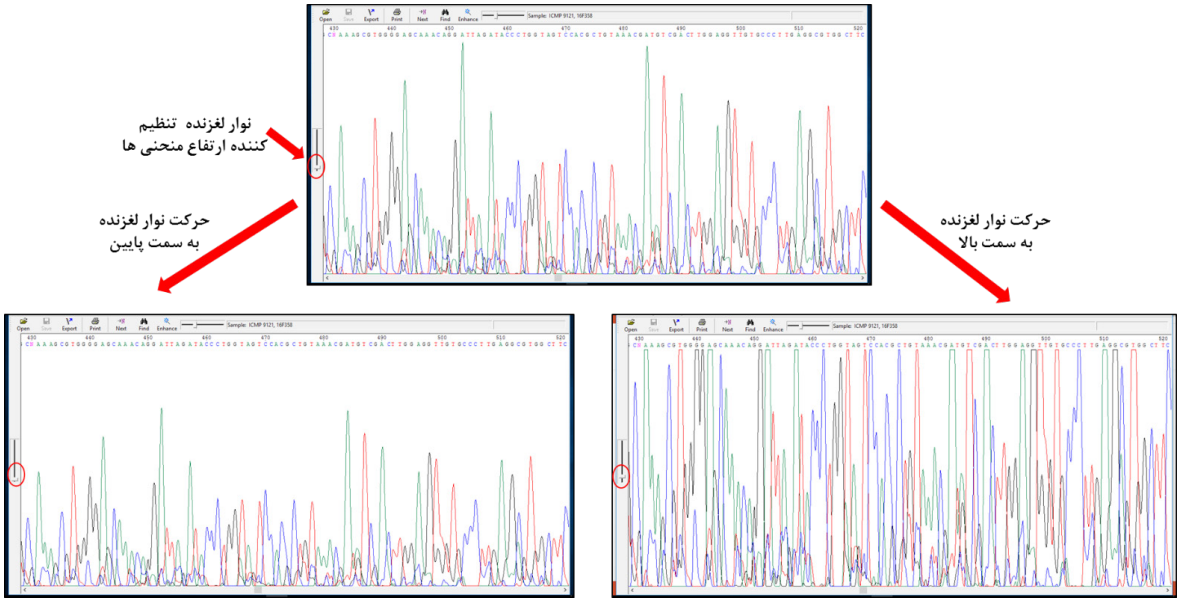


شکل ۱۲. مقایسه ی Quality value خوانش نوکلئوتیدها با یکدیگر در نرم افزار Chromas

برای بررسی بهتر کروماتوگرام با کلیک کردن بر روی نوار ابزار (شکل ۱۳) و گردش آن به سمت راست و چپ می توان منحنی ها را از یکدیگر مجزاتر یا متراکم تر نمود (شکل ۱۳).

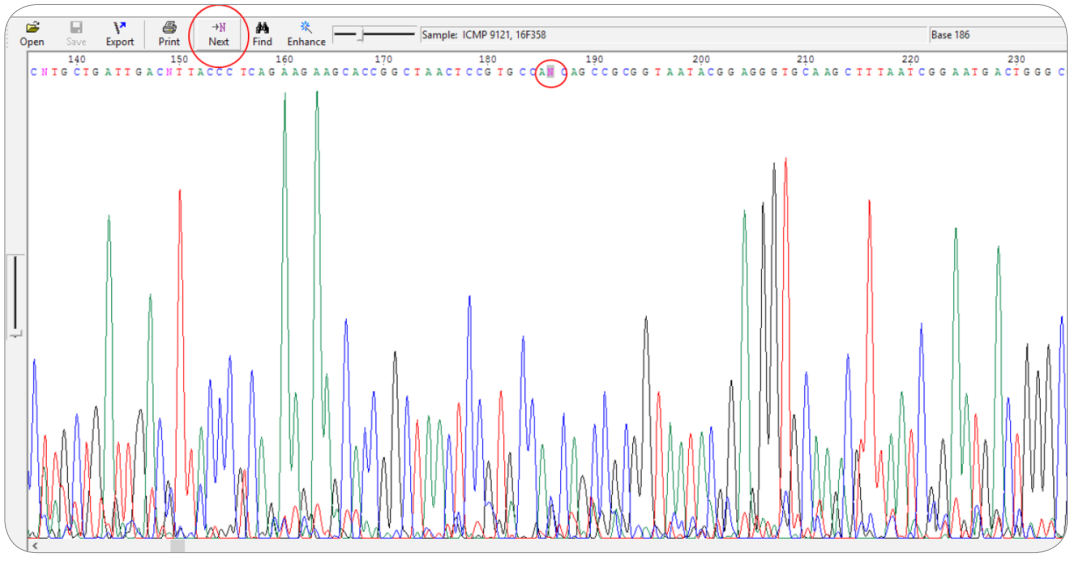
شکل ۱۳. تنظیم تراکم منحنی های کروماتوگرام در نرم افزار Chromas

با کلیک روی نوار لغزنده (شکل ۱۴) و حرکت آن به سمت بالا و پایین می توان ارتفاع منحنی را تغییر داد (شکل ۱۴).



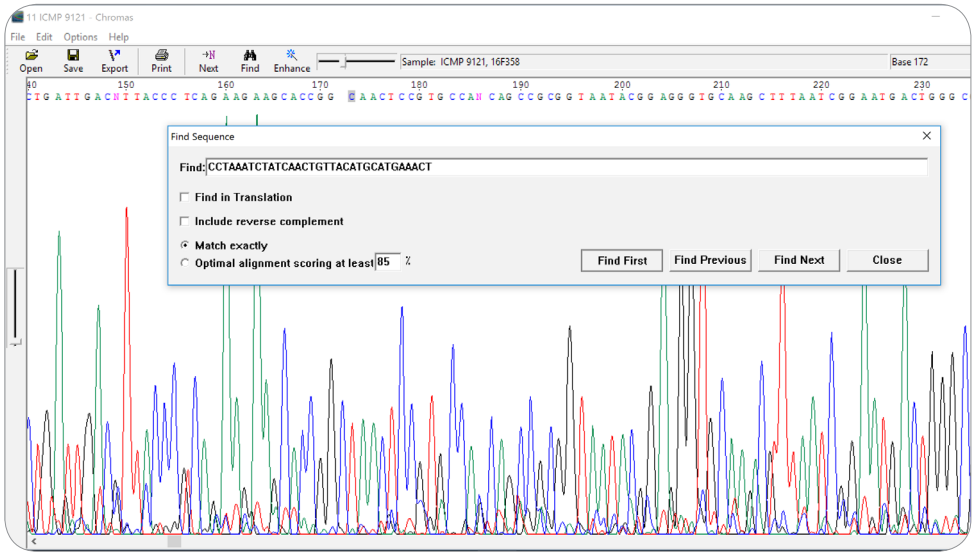
شکل ۱۴. تنظیم ارتفاع منحنی های کروماتوگرام در نرم افزار Chromas

گزینه ی نشان داده شده در (شکل ۱۵) جهت مشخص نمودن نوکلئوتیدهایی با هویت نا معلوم «N» به کار می رود (شکل ۱۵).



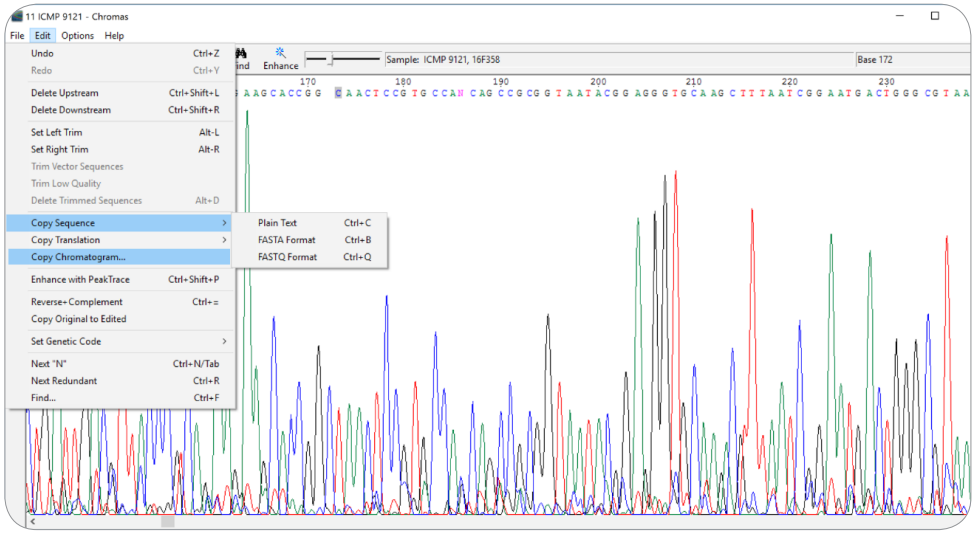
شکل ۱۵. مشخص نمودن نوکلئوتیدهای ناشناخته «N» در نرم افزار Chromas

در صورتی که جستجوی توالی نوکلئوتیدی خاصی مانند توالی پرایمر مدنظر باشد باید از منوی Edit گزینه‌ی Find را انتخاب و سپس در پنجره ی باز شده توالی مورد جستجو را قرار داد و کلید Find First را فشار داد (شکل ۱۶).



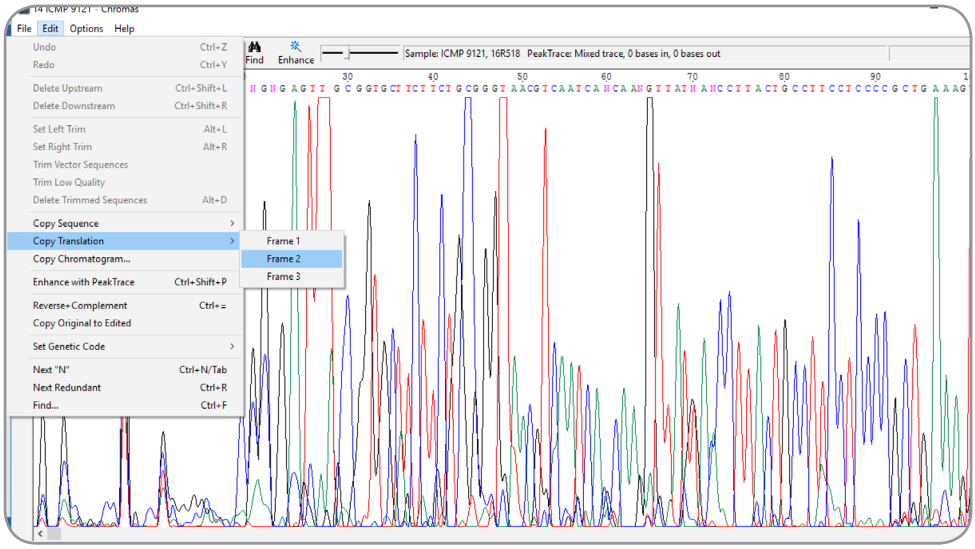
شکل ۱۶. جستجوی توالی نوکلئوتیدی خاصی با استفاده از پنجره ی Find Sequence در نرم افزار Chromas

می توان داده های مربوط به توالی را به سه صورت Plain Text، FASTQ Format و FASTA Format از طریق منوی Edit و سپس Copy Sequence کپی و ذخیره نمود. برای کپی و ذخیره ی گراف باید از منوی Edit و سپس گزینه ی Copy chromatogram استفاده نمود (شکل ۱۷).



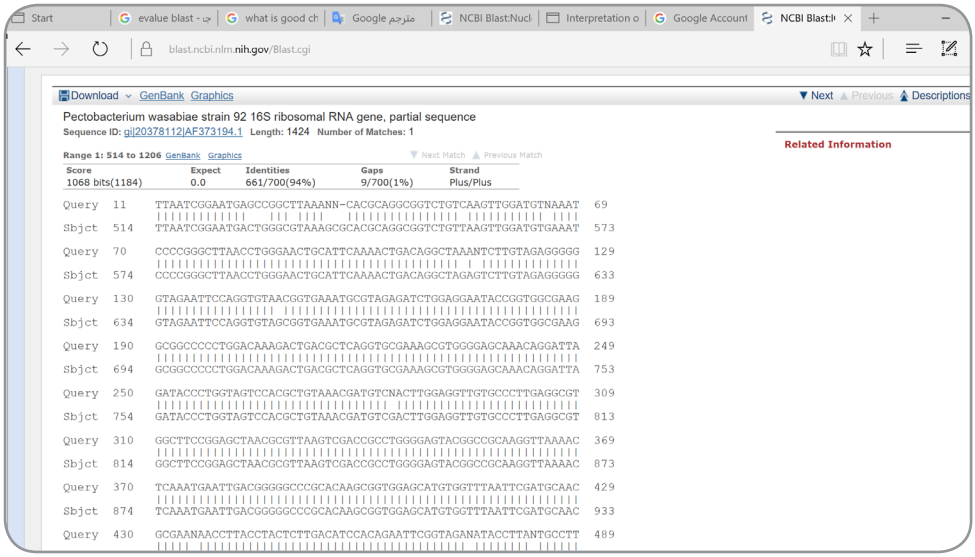
شکل ۱۷. ذخیره ی داده ها و گراف خروجی نرم افزار Chromas

برای ترجمه توالی DNA به پروتئین باید از منوی Edit گزینه ی Copy Translation انتخاب گردد. این گزینه دارای ۳ نوع قالب خواندن است که تفاوت آنها براساس انتخاب نقطه ی شروع ترجمه می باشد (شکل ۱۸).



شکل ۱۸. ترجمه توالی DNA به پروتئین از طریق نرم افزار Chromas

این نرم افزار قابلیت بلاست نمودن توالی ژن مورد نظر را از طرق منوی File گزینه ی BLAST Search دارا می‌باشد. اگر بلاست را با گزینه ی پیش فرض نرم افزار و بدون تغییر آن انجام دهیم، بلاست در پایگاه داده ی NCBI انجام شده و نتیجه بصورت شکل زیر نمایان می شود (شکل ۱۹).



شکل ۱۹. بلاست توالی DNA مورد نظر در نرم افزار Chromas

۴ آموزش کاربردی ویرایش و هم ردیف سازی توالی با نرم افزار BioEdit 7.2.5

نرم افزار BioEdit، یک نرم افزار رایگان ویراش کننده و همچنین هم ردیف ساز توالی ها می باشد. این نرم افزار علاوه بر ویرایش یک توالی ساده قادر به ویرایش توالی های هم ردیف شده نیز می باشد. نرم افزار BioEdit علاوه بر اعمال گفته شده می تواند با استفاده از نرم افزارهای دیگری که روی آن سوار شده است آنالیزهای دیگری را نیز انجام دهد. برخی از این آنالیزها و کابردهای این نرم افزار شامل:

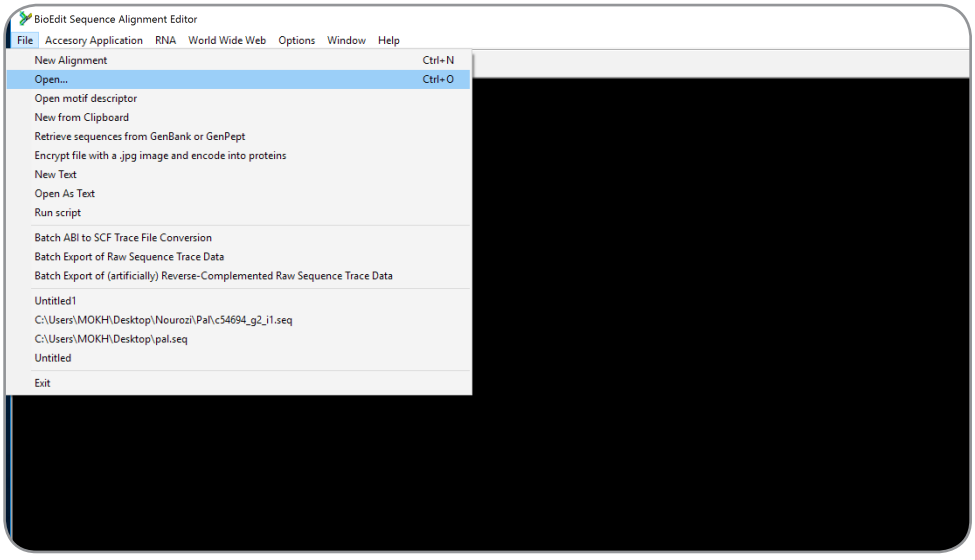
- ویرایش یک توالی و اطلاعات مربوط به آن از بانک ژن
- رسم پلات های Hydrophobicity
- ترسیم پلاسمید و وارد کردن مشخصات آن
- آزمون های گرافیکی داده ها
- نمایش پیک های ABI و SCF، ویرایش و تبدیل آنها
- جستجوی موتیف ها
- لینک خودکار به ClustalW با امکان استفاده از خط فرمان

این نرم افزار ابزارهای زیادی دارد. در اینجا ما تنها روی ابزارهای ویرایشگر و هم ردیف ساز آن که در این مبحث مورد نیاز ما می باشند تکیه می‌کنیم. آخرین نسخه ی این نرم افزار از طریق نشانی زیر قابل دانلود می باشد.

http://www.mbio.ncsu.edu/bioedit/bioedit.html

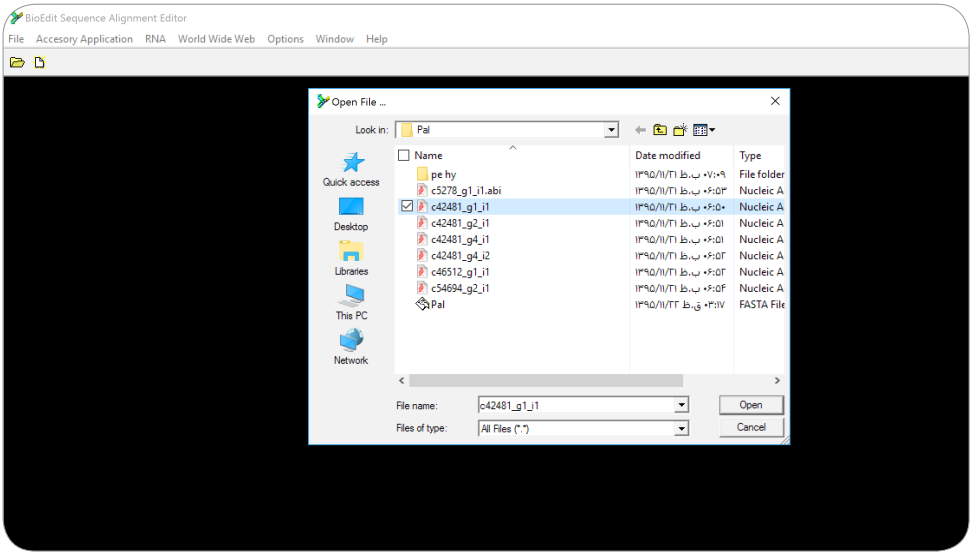
۴-۱. ویرایش توالی با نرم افزار BioEdit 7.2.5

این نرم افزار فرمت های بسیاری از جمله FASTA را می پذیرد. پس از اجرای نرم افزار برای وارد کردن فایل ورودی از منوی File گزینه ی Open انتخاب می شود (شکل ۲۰).



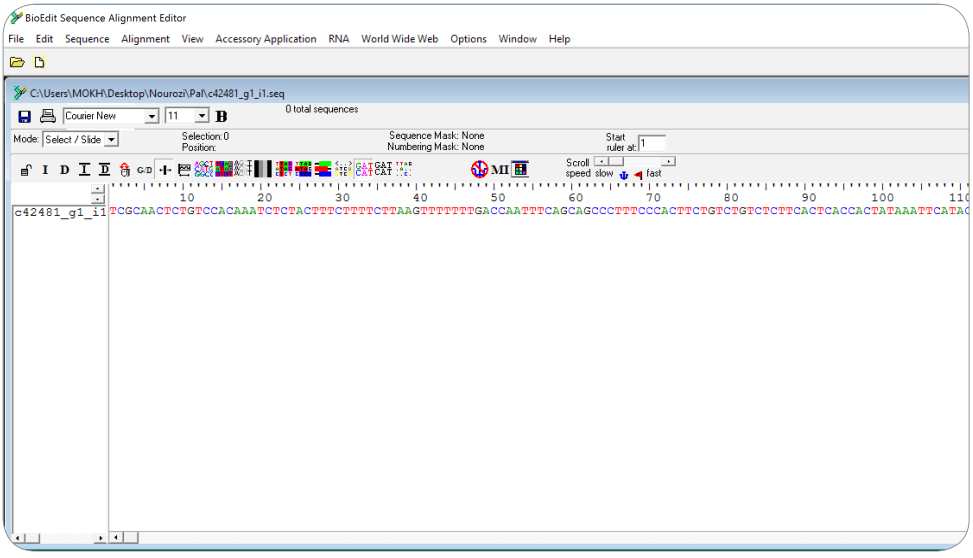
شکل ۲۰. صفحه ی اول نرم افزار BioEdit

سپس فایل مورد نظر را انتخاب و روی گزینه ی Open کلیک می شود (شکل ۲۱).



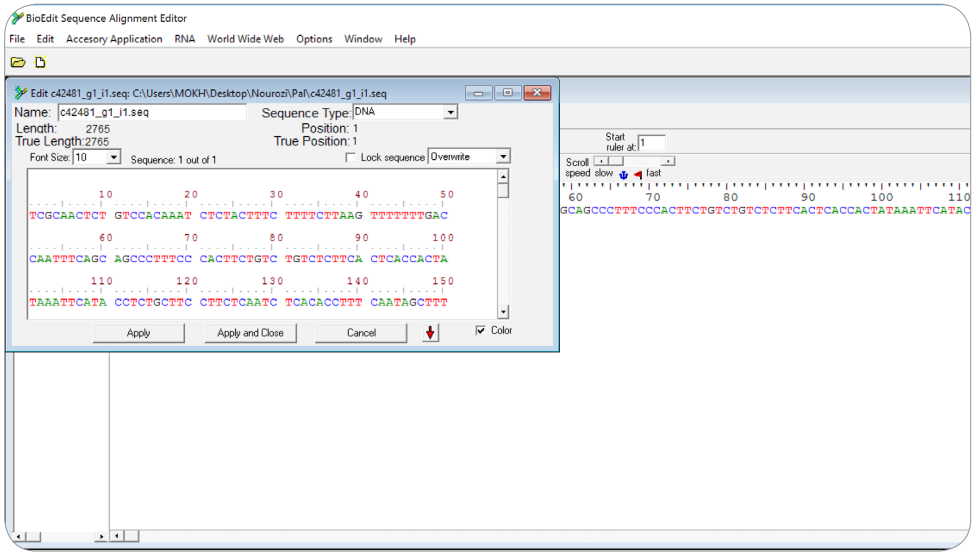
شکل ۲۱. باز کردن فایل مورد در نرم افزار BioEdit

پس از باز شدن فایل مورد نظر، در نرم افزار، صفحه ی زیر نمایش داده می شود (شکل ۲۲).



شکل ۲۲. پنجره ی حاصل از باز کردن فایل مورد نظر در نرم افزار BioEdit

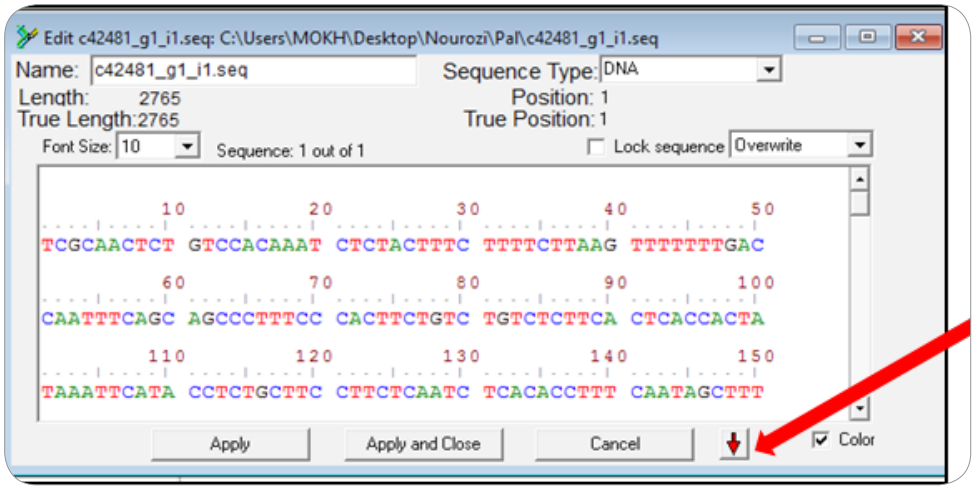
برای باز کردن پنجره ی ویرایش توالی، باید روی نام توالی که در سمت چپ پنجره ی باز شده قرار دارد دو بار کلیک کرد یا اینکه پس از انتخاب کل توالی از منوی Sequence گزینه ی Sequence Edit را انتخاب کرد. در این صورت می توان پنجره ویرایش را بصورت نشان داده شده در شکل ۲۳ باز نمود (شکل ۲۳).



شکل ۲۳. نمایش پنجره ی ویرایش در نرم افزار BioEdit

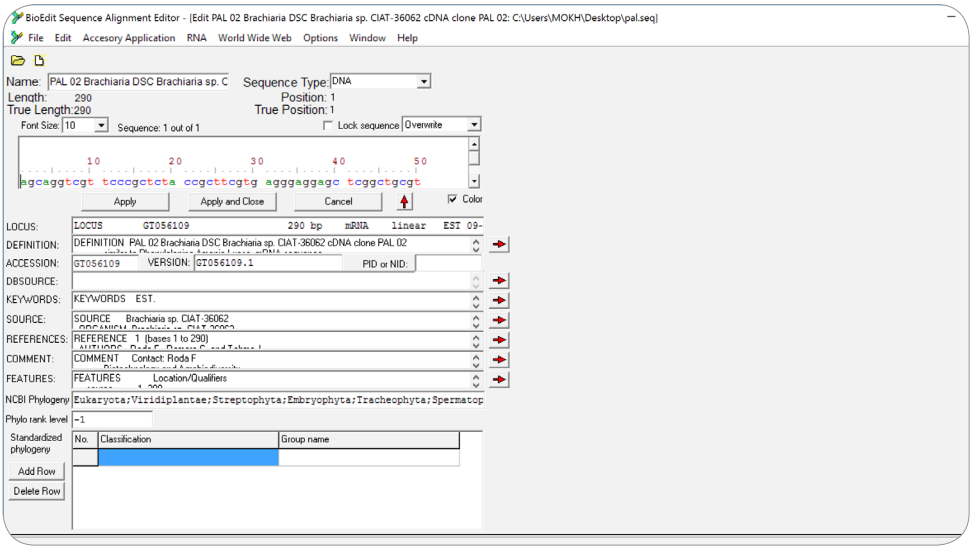
در پنجره ی باز شده از طریق گزینه ی Sequnce Type می توان نوع داده را مشخص کرد و همچنین در این پنجره می توان بصورت دستی توالی را ویرایش و نوکلئوتیدها را حذف یا اضافه کرد. در صورتی که مایل به ذخیره ی تغییرات باشیم باید در انتهای ویرایش گزینه ی Apply یا Apply and Close را انتخاب نمود (شکل ۲۳).

در صورتی که اطلاعات توالی در فرمت GenBank و یا BioEdit ذخیره شده باشد می توان با فشار دادن دکمه زیر پنجره ویرایش را گسترش داد (شکل ۲۴).



شکل ۲۴. گسترش پنجره ی ویرایش در نرم افزار BioEdit

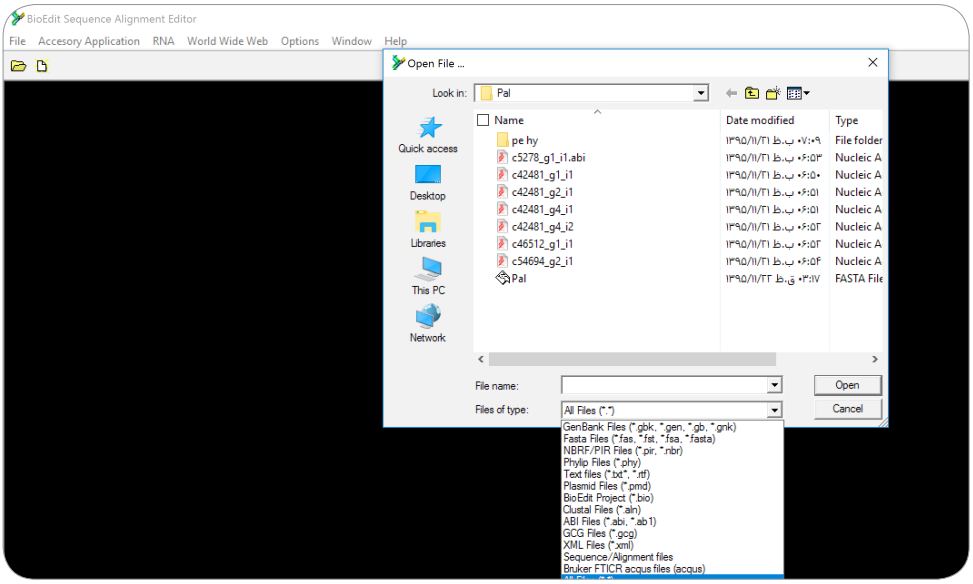
پنجره ی گسترش یافته حاوی اطلاعات مربوط به بانک ژن است (شکل ۲۵).



شکل ۲۵. نمایش پنجره ی ویرایش گسترش یافته در نرم افزار BioEdit

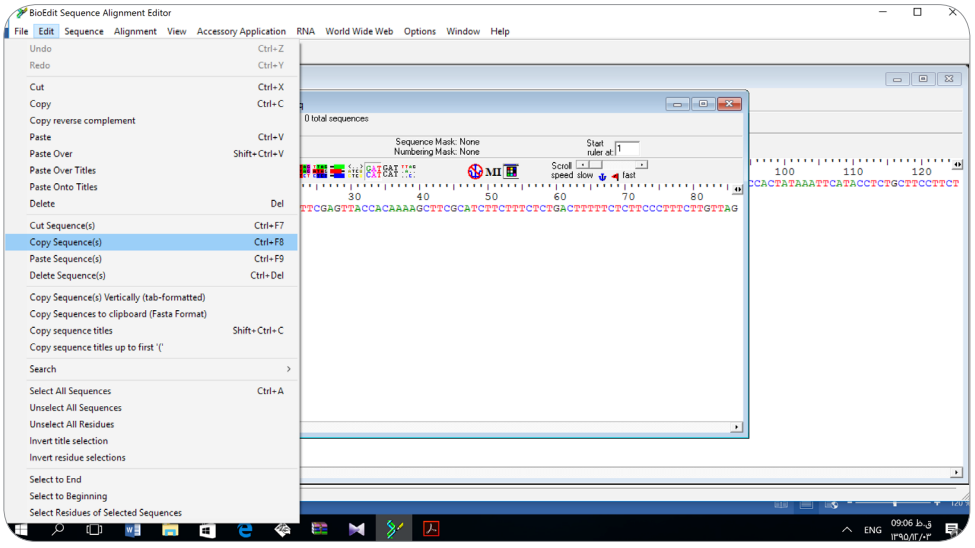
۲-۴. هم ردیف سازی توالی ها با نرم افزار BioEdit 7.2.5

برای وارد کردن فایل ورودی از منوی File، گزینه ی Open را کلیک نمایید. در پنجره ی باز شده در قسمت نوع فایل، روی All files کلیک نموده سپس فایل مورد نظر را انتخاب و باز کنید. همانطور که در شکل ۲۶ مشاهده می‌کنید این نرم افزار فرمت های زیادی از جمله FASTA را می پذیرد (شکل ۲۶).



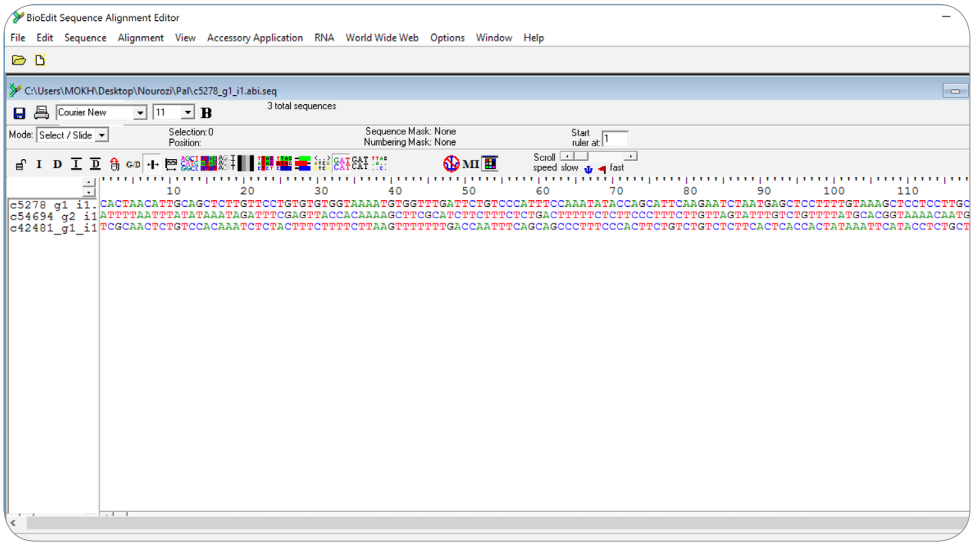
شکل ۲۶. انواع فرمت های ورودی به نرم افزار BioEdit

لازم است تمام توالی هایی که قصد هم ردیف سازی آنها را داریم در یک صفحه قرار داشته باشند. به این منظور باید پس از باز کردن فایل ها ابتدا نام توالی مورد نظر که در سمت چپ پنجره ی توالی باز شده قرار دارد را انتخاب کرده و سپس از منوی Edit گزینه ی Copy Sequence (s) را کلیک نمایید (شکل ۲۷).



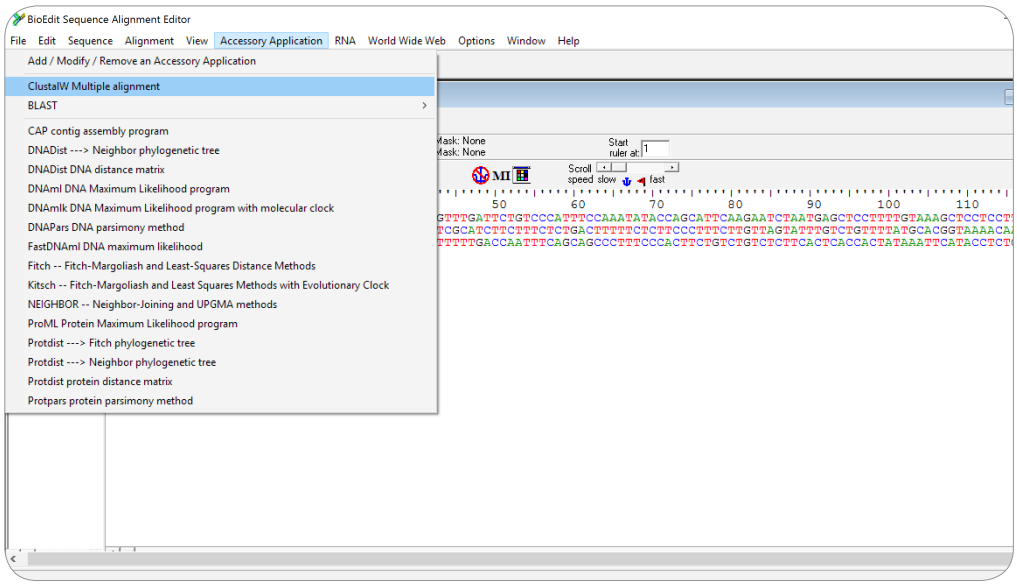
شکل ۲۷. کپی نمودن توالی مورد نظر از پنجره ی مربوطه

سپس در پنجره ای که در نظر داریم همه ی توالی ها به منظور هم ردیف سازی در آنجا حضور داشته باشند از منوی Edit باید گزینه ی Paste Sequence (s) را کلیک کنیم. در نهایت پنجره ای مطابق شکل زیر خواهیم داشت (شکل ۲۸).



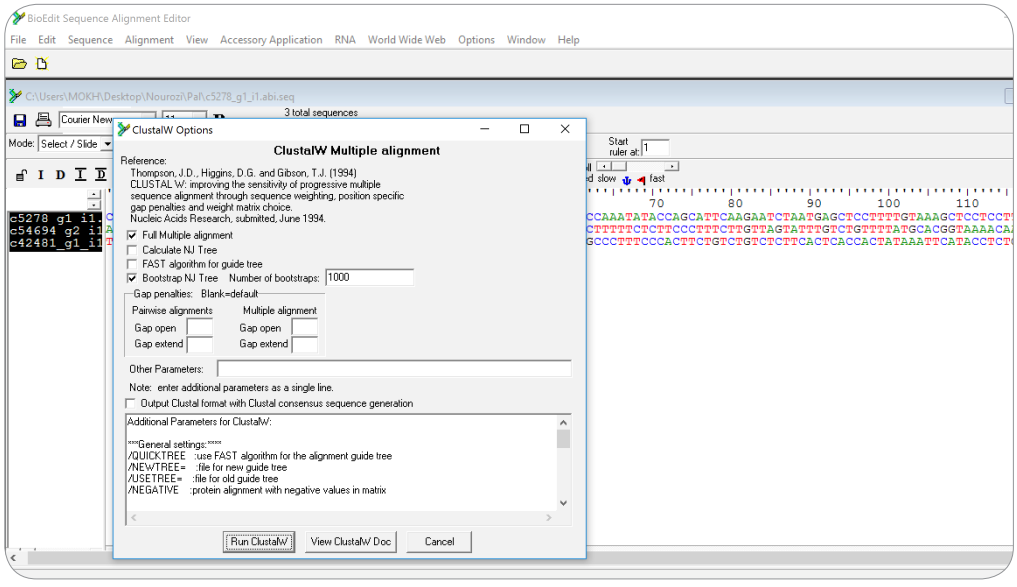
شکل ۲۸. قرار دادن همه ی توالی های مورد نظر به منظور هم ردیف سازی در یک صفحه

برای انجام هم ردیف سازی چندگانه روی Accessory Application کلیک کرده و سپس گزینه ی ClustalW Multiple Alignment را انتخاب نمایید (شکل ۲۹).



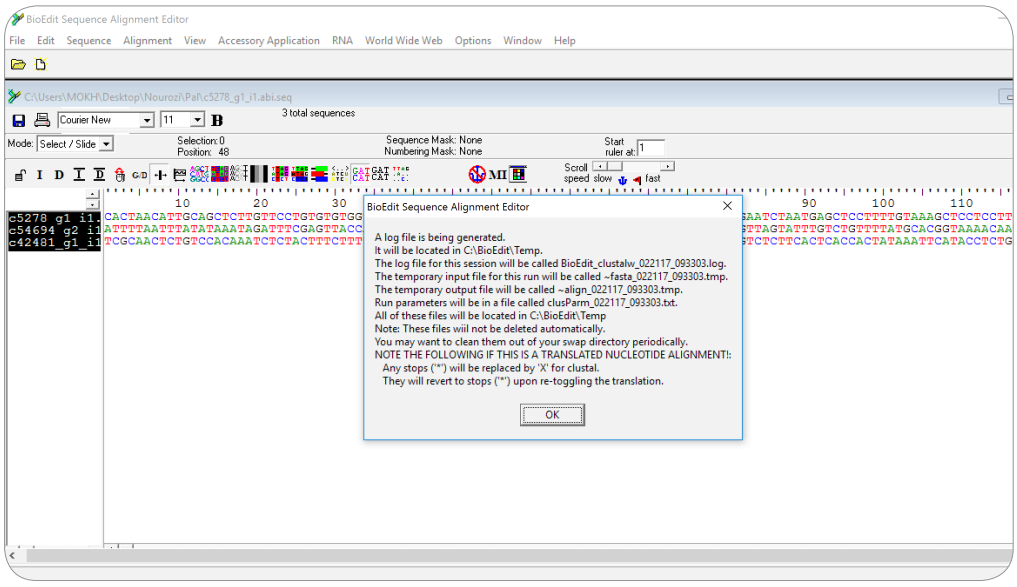
شکل ۲۹. لیست نرم افزارهای جنبی و انتخاب گزینه ی هم ردیف سازی چند گانه با ClustalW در نرم افزار BioEdit

سپس روی گزینه ی Run ClustalW کلیک نمایید (شکل ۳۰).



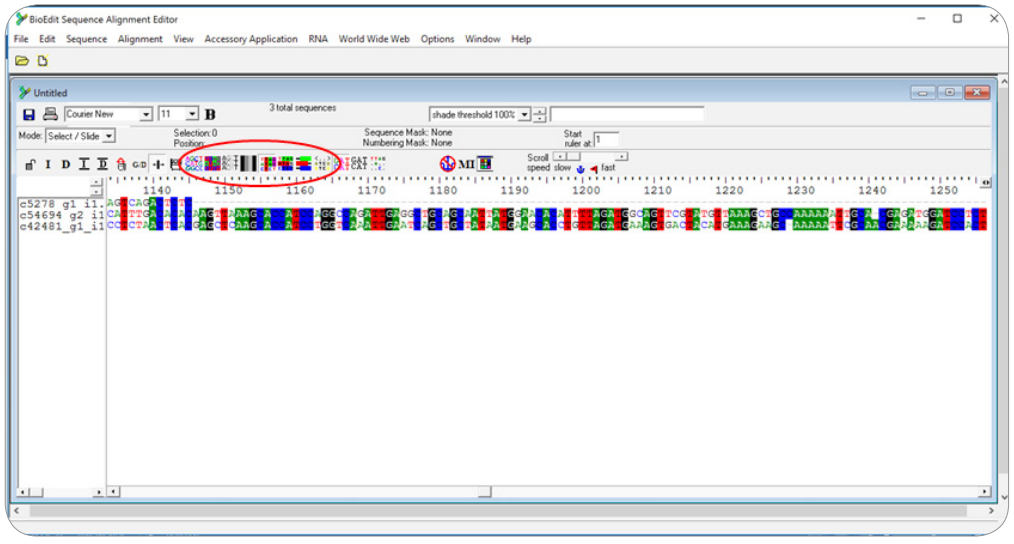
شکل ۳۰. اجرای هم ردیف سازی چند گانه در نرم افزار BioEdit

در این حالت پیام زیر نمایان می‌شود که گویای محل و مشخصات ذخیره سازی نتایج هم ردیفی می باشد (شکل ۳۱).



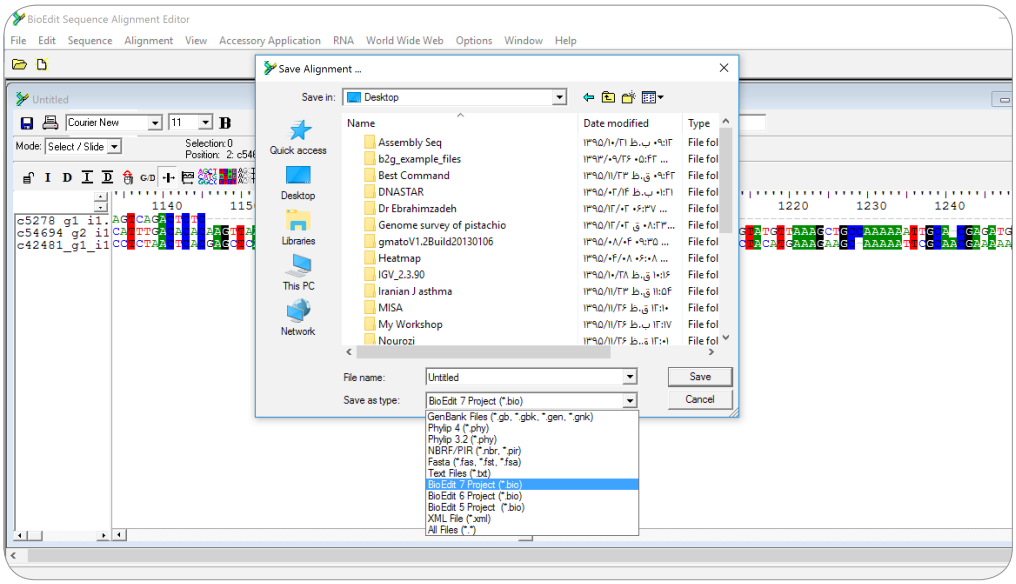
شکل ۳۱. گزارش فایل های خروجی نرم افزار BioEdit

نتیجه ی هم ردیف سازی در پنجره ای جدید ظاهر می شود که می توان دکوراسیون رنگی آن را با استفاده از بخش مشخص شده در نوار ابزار بالای پنجره تغییر داد (شکل ۳۲).



شکل ۳۲. نتیجه هم ردیف سازی در نرم افزار BioEdit

برای ذخیره سازی فایل نتیجه باید از منوی File گزینه ی Save as را کلیک نمایید. از امتیازات این نرم افزار همانطور که در شکل زیر نشان داده شده است، قابلیت ذخیره ی فایل خروجی به فرمت های مختلف می باشد (شکل ۳۳).



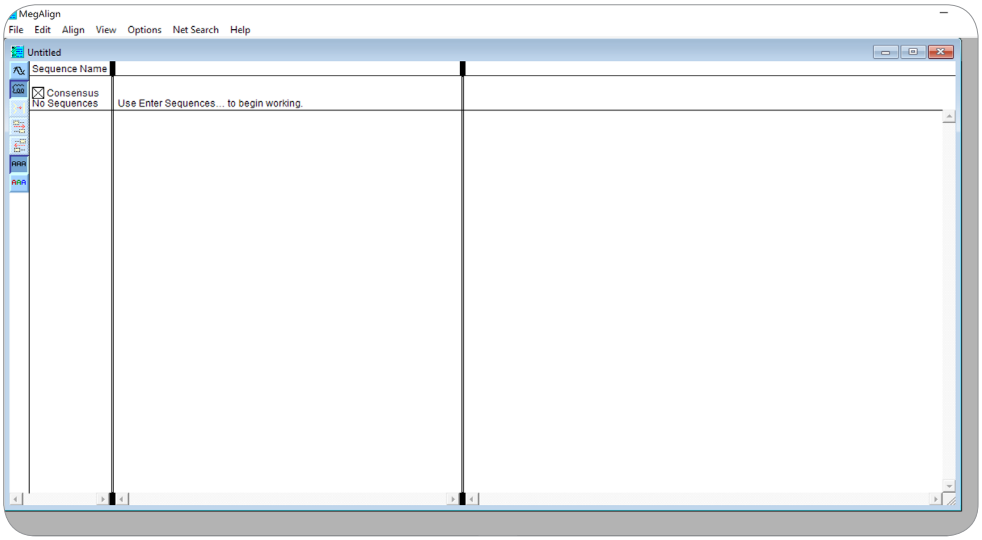
شکل ۳۳. ذخیره ی نتیجه ی هم ردیف سازی به فرمت های مختلف در نرم افزار BioEdit

۵ آموزش کاربردی هم ردیف سازی با نرم افزار MegAlign

نرم افزار MegAlign ، نرم افزاری است که برای انجام هم ردیف سازی به کار می رود. این نرم افزار را می توان در مجموعه ی Laser gene تهیه نمود. آخرین نسخه ی این نرم افزار از نشانی زیر قابل تهیه می باشد.

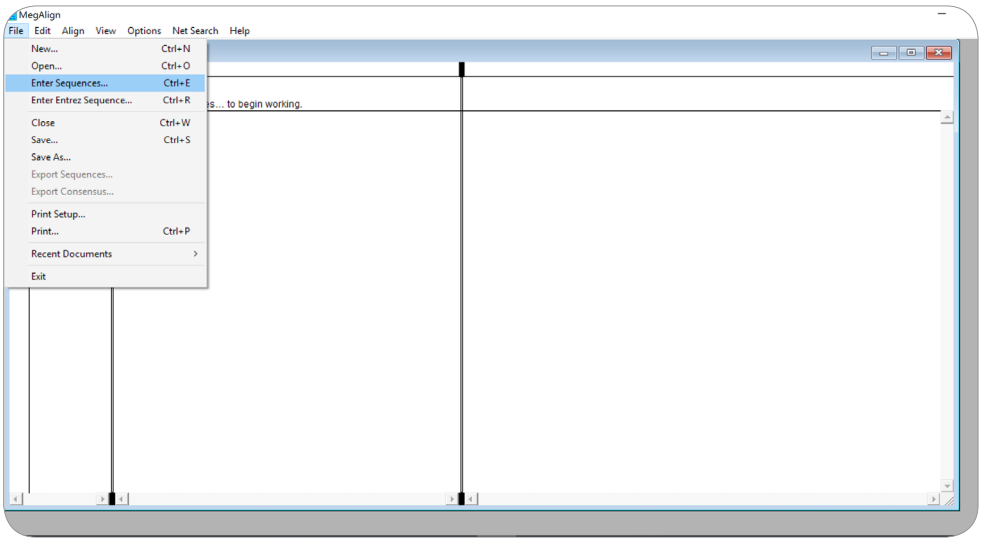
https://www.dnastar.com/t-megalign-pro.aspx

پس از اجرای نرم افزار پنجره ی زیر قابل مشاهده می باشد (شکل ۳۴).



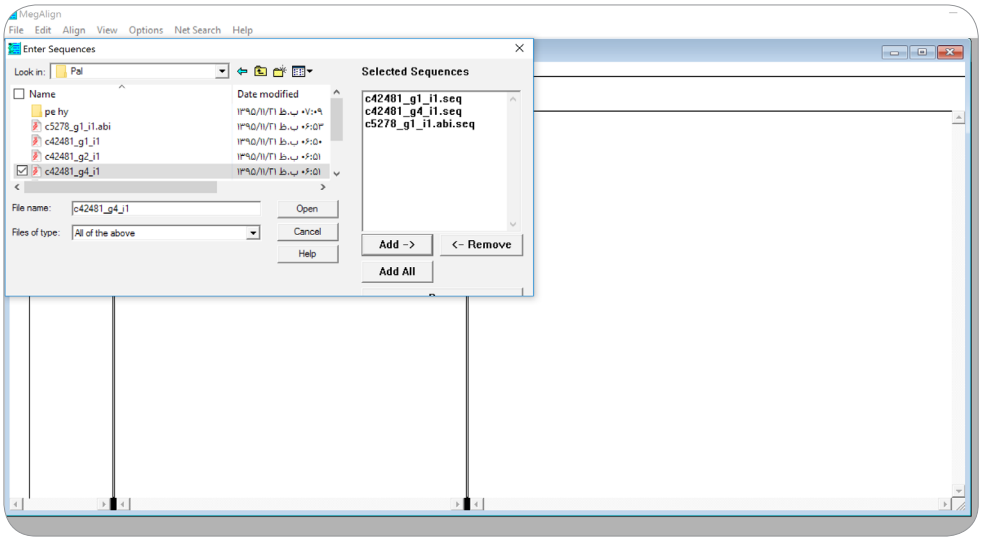
شکل ۳۴. صفحه اول نرم افزار MegAlign

برای وارد کردن توالی باید از منوی File گزینه ی Enter Sequence را انتخاب نمایید (شکل ۳۵).



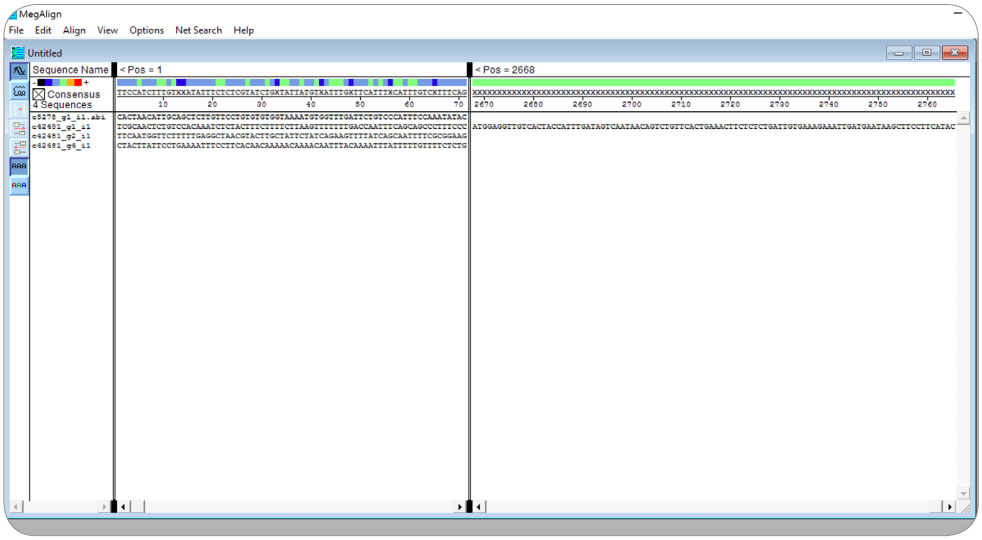
شکل ۳۵. وارد نمودن توالی در نرم افزار MegAlign

در پنجره باز شده به محل ذخیره ی فایل ها رفته و تمام توالی های مورد نظر را انتخاب و برای اضافه شدن آنها به لیست Selected Sequence، گزینه ی Add و نهایتا هم گزینه ی Done را کلیک نمایید (شکل ۳۶).



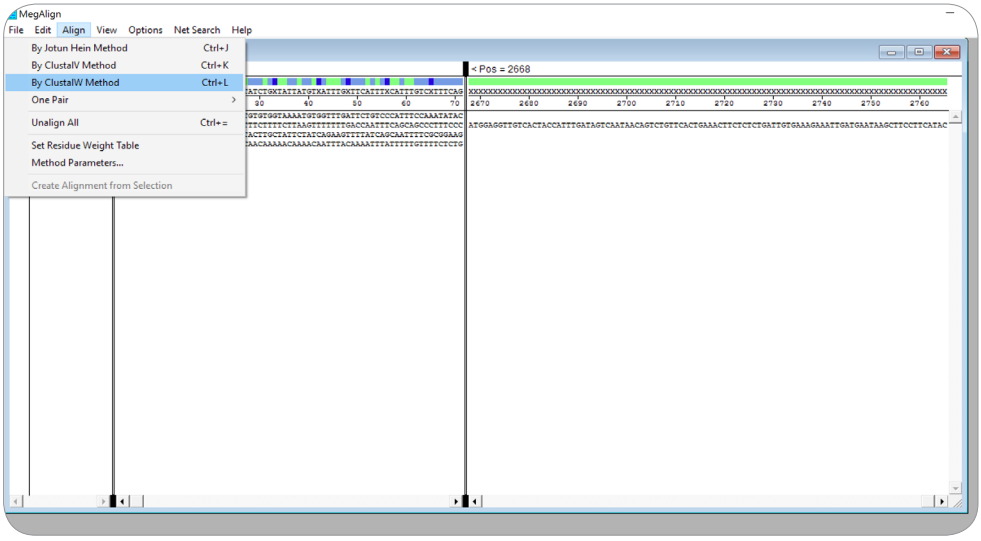
شکل ۳۶. وارد نمودن توالی‌های مورد نظر در نرم افزار MegAlign

سپس پنجره‌ی زیر باز می شود که نشان دهنده‌ی نام و توالی فایل های انتخاب شده برای هم ردیف سازی می باشد (شکل ۳۷).



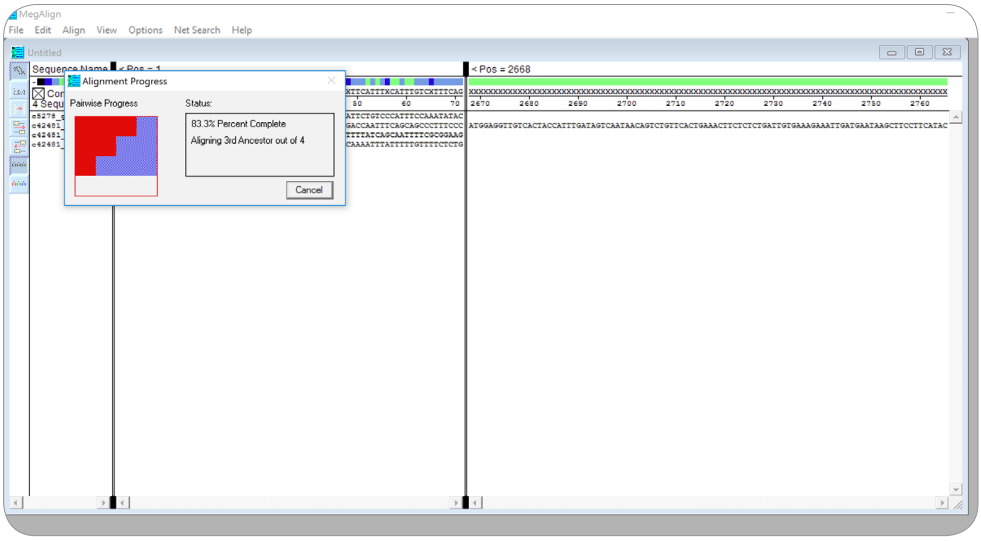
شکل ۳۷. توالی‌های انتخاب شده برای هم ردیف سازی در نرم افزار MegAlign

برای انجام هم ردیف سازی روی منوی Align کلیک نموده و مطابق شکل زیر روش هم ردیف سازی را انتخاب کنید (شکل ۳۸).



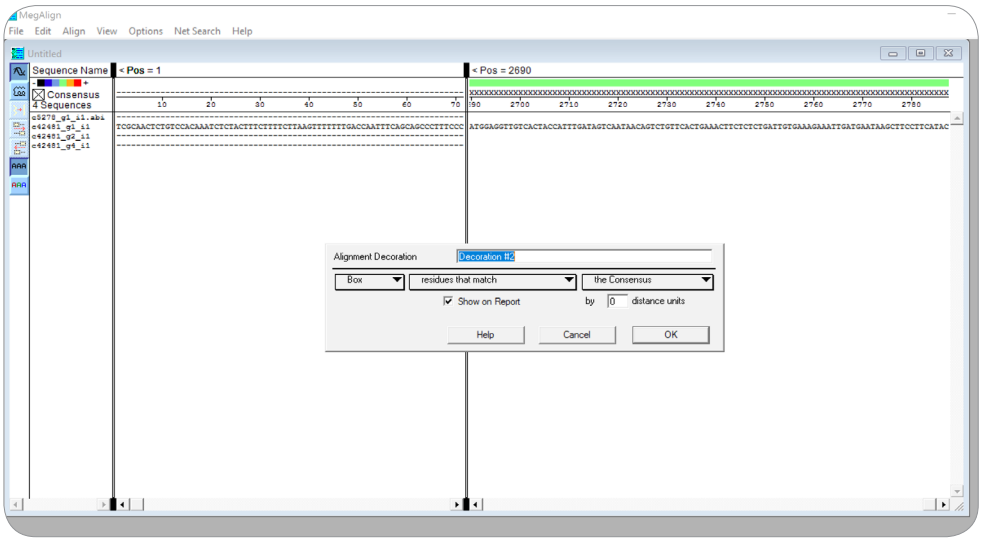
شکل ۳۸. انتخاب روش هم ردیف سازی در نرم افزار MegAlign

مطابق شکل زیر بر حسب میزان شباهت یا تفاوت و همچنین حجم اطلاعات در حال آنالیز هم ردیفی بین چند ثانیه تا چند دقیقه به اتمام می رسد (شکل ۳۹).



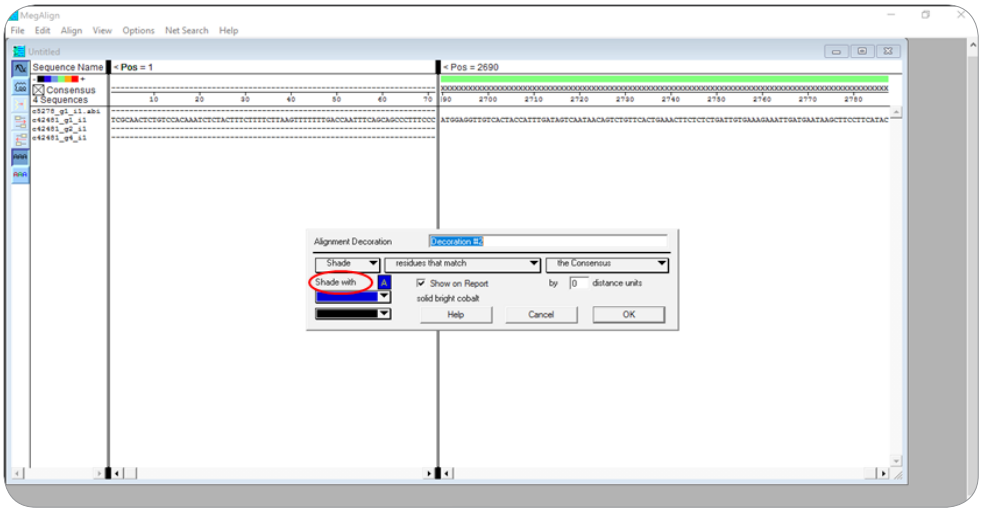
شکل ۳۹. انجام هم ردیف سازی در نرم افزار MegAlign

برای مشاهده ی بهتر نتایج هم ردیفی از منوی Options گزینه ی New Decoration را کلیک نمایید. پنجره ی زیر باز می شود (شکل ۴۰).



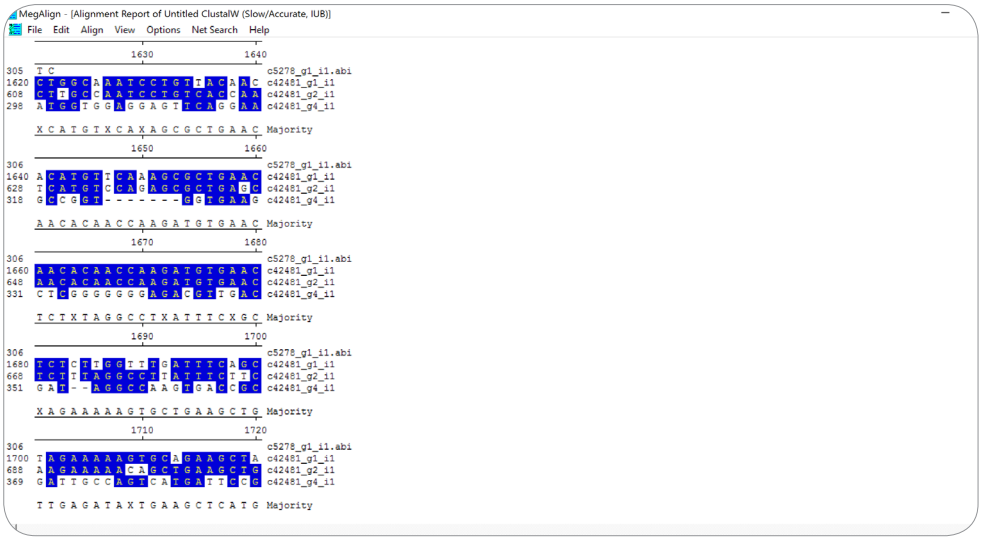
شکل ۴۰. انتخاب دکوراسیون مناسب برای مشاهده ی بهتر نتایج هم ردیفی در نرم افزار MegAlign

سپس گزینه ی Box و پس از آن Shade را انتخاب نمایید (شکل ۴۰). در بخش Shade with مطابق سلیقه می توان برای بهتر دیده شدن شباهت ها و تفاوت ها رنگ های متفاوتی را انتخاب کرده و در نهایت دکمه Ok را کلیک کنید (شکل ۴۱).



شکل ۴۱. انتخاب دکوراسیون رنگی مناسب برای مشاهده‌ی بهتر نتایج هم ردیفی در نرم افزار MegAlign

از منوی View گزینه ی Alignment Report، نتیجه هم ردیفی را مطابق شکل زیر در پنجره ای مجزا نشان می دهد (شکل ۴۲).



شکل ۴۲. نمایش نتایج هم ردیفی در نرم افزار MegAlign

پس از اتمام هم ردیف سازی می توان نتایج را از طریق منوی File گزینه ی Save As ذخیره نمود.

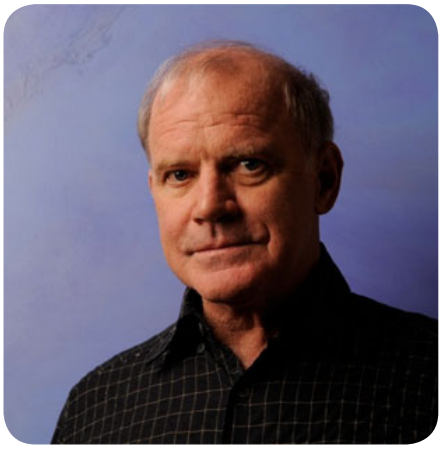
اصول طراحی پرایمر

فصل پنجم

۱ مقدمه ای بر واکنش زنجیره ای پلیمراز یا PCR

PCR به روش ازدیاد مقادیر جزئی DNA یا RNA (ازدیاد RNA با روش RT-PCR امکان پذیر است) تا حد فراهم شدن امکان مشاهده آنها توسط روش های ساده و رایج آزمایشگاهی اطلاق می شود. این روش جهت تقویت ناحیه و یا قطعه خاصی از DNA (DNA هدف) که می تواند یک ژن واحد، قسمتی از یک ژن یا توالی غیر کد کننده باشد، به کار می رود. در فرایند PCR، آنزیم DNA پلیمراز از یك مولکول یا تعداد بسیار کمی مولکول DNA به عنوان الگو استفاده کرده و میلیون ها نسخه از همان مولکول را تولید می کند. محصول نهایی PCR را آمپلیکون (Amplicon) می نامند که به معنی ماده تقویت شده یا آمپلی فاید شده است.

از زمان معرفی PCR در سال ۱۹۸۳ توسط کری مولیس[۱]، به کارگیری آن در تحقیقات زیست شناختی پایه و کاربردی موجب ایجاد تحولات اساسی در حوزه ی زیست شناسی گردیده است (شکل ۱). به دلیل اهمیت این اختراع، کاربردهای فراوان و نقش ارزنده آن در پیشرفت علم ژنتیک و زیست شناسی مولکولی، وی جایزه نوبل شیمی را در سال ۱۹۹۳ دریافت کرد.



شکل ۱. تصویر کری مولیس بیوشیمیست، نویسنده و سخنران آمریکایی که در سال ۱۹۹۳ برای ابداع روش واکنش زنجیره پلیمراز (PCR) موفق به کسب جایزه نوبل شیمی شد.

مواد لازم برای واکنش PCR شامل DNA الگو، آنزیم Taq polymerase، پرایمرها، بافر، یون منیزیم، نوکلئوتیدها و آب می باشند. عملکرد اختصاصی PCR به پرایمر بستگی دارد. PCR به کمک چرخه های حرارتی (Thermal Cycling) با دماهای مختلف انجام می گیرد، هر چرخه، جهت انجام فرایندهای خاصی که در ادامه اشاره خواهد شد، تکرار می شود. بطور کلی هر چرخه کامل حرارتی که شامل سه مرحله می باشد را یک سیکل (Cycle) گویند. معمولا هر واکنش PCR حدوداً به ۴۰ -۲۰ سیکل نیاز دارد. در ابتدا PCR بطور دستی و با قرار دادن و انتقال لوله های آزمایش بین حمام های آب دارای دمای مورد نیاز انجام می شد. امروزه PCR با

۱ Kary Mullis

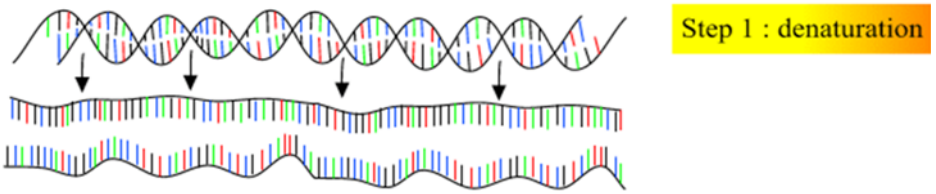
استفاده از دستگاه هایی به نام ترموسایکلر یا دستگاه PCR که در آنها جایگاه های لوله ای با بلوک فلزی حرارت پذیر تعبیه شده است و قابل برنامه ریزی برای تغییر سریع بین دماهای لازم می باشد، انجام می شود (شکل۲).

شکل ۲. دستگاه ترموسایکلر

هر چرخه PCR شامل سه مرحله، جدا شدن دو رشته DNA از همدیگر (واسرشتی اولیه)، اتصال پرایمرها و در نهایت تکثیر می باشد. در ادامه سه مرحله ی واکنش PCR توضیح داده می شود.

۱. مرحله واسرشتی اولیه[۲]

در این مرحله مولکول های دو رشته ای DNA بوسیله حرارت بالا (حدود ۹۴ درجه سانتیگراد) از همدیگر جدا گردیده و به مولکول های تك رشته ای تبدیل می شوند (شکل۳).

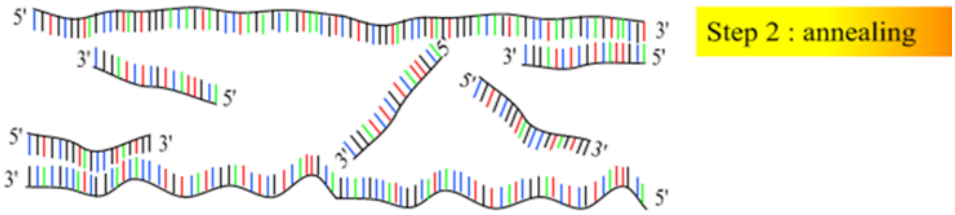


شکل ۳. مرحله ی واسرشتی واکنش زنجیره ای پلیمراز

۲ Denaturation Step

۲. مرحله اتصال[۳]

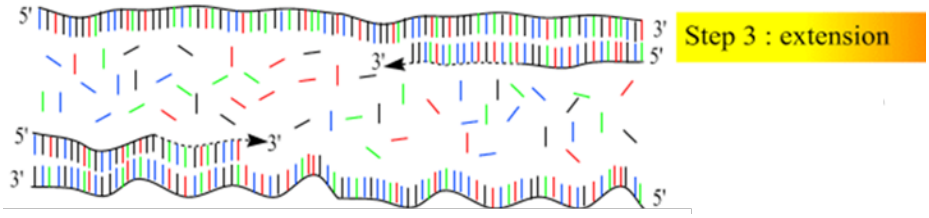
در این مرحله کاهش دمای واکنش صورت می گیرد تا اینکه پرایمرها بتوانند به مولکول های DNA تك رشته ای الگو پیوند زده شوند. پرایمرها قطعات کوچکی از DNA و محل آغاز تکثیر DNA توسط آنزیم DNA پلیمراز می باشند. درجه حرارت مورد استفاده برای این مرحله می تواند از ۳۰ تا ۷۲ درجه سانتیگراد متغیر باشد (شکل۴).



شکل ۴. مرحله ی اتصال واکنش زنجیره ای پلیمراز

۳. مرحله توسعه[۴]

در این مرحله که آخرین مرحله یك چرخه PCR می باشد، آنزیم Taq پلیمراز (DNA پلیمرازی است مقاوم به حرارت که از یك باکتری ترموفیل به نام ترموس آکواتیکوس Thermus aquaticus استخراج می شود) دئوکسی نوکلئوتیدهای تری فسفاته مختلف و مکمل با تک رشته ی الگو را به پرایمرها اضافه کرده تا تکثیر قطعه مورد نظر آغاز شود. درجه حرارت لازم برای انجام این مرحله ۷۲ درجه سانتیگراد است. این درجه حرارت دمای بهینه ی فعالیت آنزیم های DNA پلیمراز مقاوم به حرارت است که بطور عادی مورد استفاده قرار می گیرند. هر چرخه PCR نهایتاً با دو برابر شدن تعداد مولکول های DNA هدف همراه است. تکرار چرخه PCR بدفعات زیاد (۲۰ تا ۴۰ بار) منجر به افزایش در تعداد قطعه ی مورد نظر و در نتیجه ازدیاد DNA هدف به میزان یك میلیون برابر یا بیشتر خواهد شد (شکل ۵).



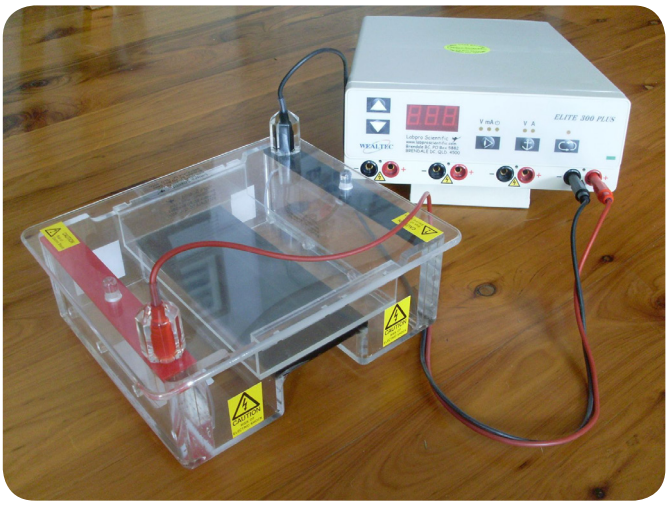
شکل ۵. مرحله ی توسعه ی واکنش زنجیره ای پلیمراز

۳ Annealing Step

۴ Extension Step

۱-۱. ارزیابی محصولات PCR

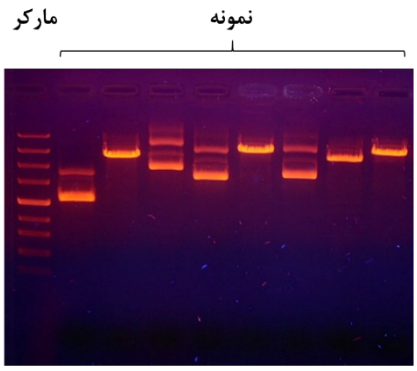
یکی از ساده ترین و مرسوم ترین روش های مورد استفاده در ارزیابی PCR، الکتروفورز است (شکل ۶). الکتروفورز در واقع حرکت مولکول ها در یك میدان الکتریکی بر اساس وزن مولکولی آنها است، یعنی ذره ی کوچکتر دارای حرکت سریعتر و ذره ی بزرگتر دارای حرکت کندتر است.



شکل ۶. دستگاه الکتروفورز

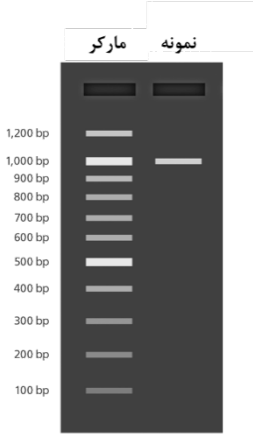
قسمت های اصلی یک دستگاه الکتروفورز همانطور که در شکل ۶ مشاهده می کنید شامل تانک الکتروفورز، منبع تغذیه الکتریکی و محیط نیمه جامد (ژل) به عنوان فاز ثابت می باشد. الکتروفورز بسته به اینکه در سطح افقی یا عمودی انجام گیرد به ترتیب الکتروفورز افقی و عمودی نامیده می شود. از طرف دیگر الکتروفورز را می توان برحسب نوع ژل به کار گرفته شده به دو نوع الکتروفورز ژل پلی اکریل آمید (PAGE) و الکتروفورز ژل آگارز تقسیم کرد. برای تفکیک اسیدهای نوکلئیک معمولا از ژل آگارز استفاده می شود. تهیه ژل مزبور به مراتب سریع تر و آسان تر از ژل پلی اکریل آمید بوده و هزینه کمتری را در بر می گیرد. این ژل یك شبکه ای دارای خلل و فرج را فراهم می آورد که دارای چاهك هایی برای نمونه گذاری می باشد. نمونه های حاوی DNA (محصول PCR) را در این چاهك ها ریخته، الکترود منفی نزدیك محل نمونه گذاری و الکترود مثبت دور از منطقه نمونه گذاری نصب می شود. چون نمونه دارای بار منفی است، پس از برقراری میدان الکتریکی با اعمال اختلاف پتانسیل در دو سوی این شبکه مولکول های DNA بر اساس وزن مولکولی و بار منفی خود با سرعت های متفاوتی درون شبکه متخلخل شروع به حرکت به سوی الکترود مثبت می کنند. برای ارزیابی وزن مولکولی (اندازه) محصولات PCR، از مارکر یا Ladder استفاده می شود. Ladder مخلوطی از قطعات DNA با اندازه های متفاوت و مشخص است که بعد از الکتروفورز از هم جدا می شوند. در پایان کار، ژل را معمولاً با رنگ هایی مثل اتیدیوم بروماید رنگ آمیزی نموده و در معرض لامپ UV قرار داده تا باند های DNA قابل مشاهده شود زیرا اتیدیوم بروماید در معرض نور UV دارای فلورسانس به رنگ نارنجی است (شکل ۷)، برای محافظت خود بویژه چشم ها از اشعه UV بایستی از

عینک های خاص استفاده کرد. امروزه از دستگاه ژل داک[۵] استفاده می شود که مجهز به ترانس ایلومیناتور، هود، فیلتر، یك محفظه ی جدا شده دارای لامپ UV و دوربین است.



شکل ۷. قطعات DNA در ژل آگارز رنگ آمیزی شده با رنگ اتیدیوم بروماید

مشاهده ی تك باند با سایز مورد انتظار پس از الکتروفورز نشان دهنده ی کیفیت خوب پرایمرها و اجرای موفق PCR می باشد (شکل ۸).



شکل ۸. ژل آگارز مربوط به نمونه ای که تک باند با سایز مورد نظر ایجاد کرده است

۱-۲. برخی از کاربردهای مهم PCR

- جداسازی DNA ژنومی
- تهیه نسخه های متعدد از ژن مورد نظر

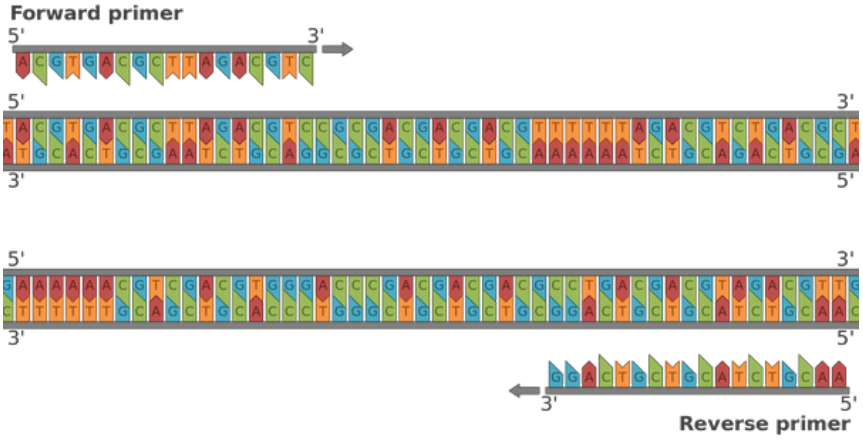
۵ Gel Doc

- بررسی حضور یا عدم حضور یک ژن
- تقویت و بررسی های کمّی DNA
- تشخیص بیماری ها
- تعیین جنسیت جنین

۲ پرایمر

پرایمر قطعه ی کوچکی از DNA تک رشته ای است که بعنوان نقطه شروع سنتز DNA در واکنش زنجیره ای PCR مورد استفاده قرار می گیرد. رشته های DNA مکمل یکدیگر هستند و در هنگام همانند سازی DNA این رشته ها از یکدیگر جدا می شوند. برای تکثیر قطعه هدف برای ابتدا و انتهای آن، پرایمرهای Forward و Reverse طراحی می شوند. پرایمرهای Forward معمولا به یکی از رشته های الگو متصل می شوند و سنتز DNA جدید را به سمت پرایمر Reverse پیش می برند در حالیکه پرایمرهای Reverse به نحوی طراحی می شوند که به رشته مکمل رشته ی الگوی اول متصل شده و سنتز DNA را در جهت معکوس به سمت پرایمر Forward پیش می برند. در واقع قطعه ای از DNA الگو که حد واسط دو پرایمر Forward و Reverse قرار دارد از طریق فعالیت آنزیم پلیمراز تکثیر می یابد (شکل ۹). توجه به اصول طراحی پرایمر مهمترین معیار در موفقیت واکنش PCR است زیرا یک پرایمر نامناسب منجر به تکثیر ناچیز قطعه مورد نظر و یا عدم انجام واکنش PCR به دلیل تکثیر غیر اختصاصی یا تشکیل پرایمر دایمر می شود. در صورتی که شرایط PCR بهینه نباشد، ممکن است پرایمرها به یکدیگر (نه به رشته الگو) متصل شوند و قطعاتی را به نام پرایمر دایمر ایجاد کنند.

طی ۱۵ سال اخیر، کامپیوتر تبدیل به یک ابزار مهم و ضروری برای زیست شناسان سلولی و مولکولی جهت طراحی پرایمر شده است که پیوند آن با علم زیست شناسی، شاخه ای از بیوانفورماتیک را تشکیل می دهد. پیش از طراحی پرایمر پرسیدن این سوال که هدف از طراحی پرایمر چیست بسیار ضروری و اساسی می باشد زیرا بر اساس اهداف مختلف مثلا تکثیر معمولی، مطالعه متیلاسیون و یا Real-time PCR برخی از خصوصیات پرایمرها متفاوت و اختصاصی است.



شکل ۹. پرایمرهای Forward و Reverse در حال تکثیر توالی هدف

۳ فاکتورهای مهم در طراحی پرایمر

فاکتورهایی وجود دارند که موفقیت در طراحی یک پرایمر مناسب وابسته به آنها هستند، این فاکتورها شامل موارد زیر می باشند:

۱. محتوای GC

۲. طول پرایمر

۳. ساختارهای ثانوی

۴. اختصاصی بودن

۵. دمای ذوب (Tm)

در ادامه توضیح مختصری در رابطه با هر یک از این فاکتورها ارائه می شود.

۳-۱. طول پرایمر

طول پرایمر یک فاکتور مهم در طراحی پرایمر می باشد. این فاکتور بطور معنی داری بر روی دمای اتصال پرایمر موثر است. طول پرایمر همراه با دمای اتصال پرایمر به DNA هدف تعیین کننده اختصاصی بودن واکنش PCR می باشد. طول مناسب برای پرایمر معمولاً بین ۱۸ تا ۳۰ جفت باز در نظر گرفته می شود. پرایمرهای کوتاهتر اتصال غیر اختصاصی را افزایش می دهند و پرایمرهای بلندتر سرعت هیبریداسیون را به دلیل ایجاد ساختارهای ثانوی کاهش می دهند.

۳-۲. دمای ذوب (Tm)

توجه به دمای ذوب (Tm) جفت پرایمرها فاکتور بسیار مهمی در طراحی پرایمرهای کارآمد می باشد. دمای ذوب، درجه حرارتی است که در آن نیمی از مولکول DNA از فرم دو رشته ای به تک رشته ای تبدیل می شود. دمای ذوب پرایمرها از طریق طول و غلظت پرایمرها و ترکیب بازها در آنها تعیین می شود. بهتر است Tm پرایمرهای Forward و Reverse مشابه هم باشد یا حداکثر تفاوت Tm آنها ۲ درجه سانتیگراد باشد. در ساده ترین روش برای محاسبه دمای Tm پرایمرهایی با طول کمتر از ۲۰ نوکلئوتید، از رابطه والاس استفاده می شود:

$$Tm\ (C^{\circ}) = 2\ (A+T) + 4\ (G+C)$$

نرم افزارهای طراحی پرایمر آنلاین[۶] مانند Primer3 و آفلاین[۷] مانند OLIGO 7 بصورت خودکار محاسبه Tm پرایمرها را با روش های مختلف انجام می دهند. دمای اتصال پرایمرها به DNA الگو بر اساس دمای ذوب محاسبه می شود. دمای اتصال پرایمرها به رشته الگو باید بقدر کافی پایین باشد تا پرایمرها قادر به اتصال به DNA الگو باشند و از سوی دیگر باید به اندازه ی مناسب بالا باشد تا از تشکیل اتصالات غیراختصاصی جلوگیری کند. بطورکلی دمای مناسب برای اتصال پرایمرها معمولاً بین ۵۵ تا ۶۵ درجه سانتیگراد است.

---

۶ Online

۷ Offline

۳-۳. محتوای GC

یکی از مهمترین مشخصات Tm وابستگی آن به ترکیب بازی DNA است. محتوای GC بر روی Tm پرایمرها تاثیر گذار است. درصد GC پرایمرها باید بین ۳۵ تا ۵۵ درصد باشد. علاوه بر این درصد GC پرایمرهای Forward و Reverse باید تقریبا مشابه بوده تا دمای Tm پرایمرها نزدیک هم باشد.

۴-۳. پایداری نوکلئوتیدهای انتهای ´۳

ترکیب بازی در انتهای ´۳ پرایمرها برای اتصال اختصاصی به رشته الگو مهم است، زیرا این ناحیه، ناحیه ی شروع فعالیت DNA پلیمراز می باشد. حداقل یک G یا C باید در انتهای ´۳ پرایمر باشد ولی این انتها نباید غنی از G و C باشد. پایداری انتهای ´۳ پرایمر به ΔG پنج باز انتهایی آن وابسته است و باید کمتر از kcal/mol ۹- باشد. پایداری بالا در این انتها می تواند احتمال تولید محصول غیر اختصاصی را افزایش دهد.

۵-۳. اجتناب از توالی های تکراری

از تکرارهای تک و دو نوکلئوتیدی باید در طول پرایمر اجتناب کرد. حداکثر تکرارهای قابل قبول تک و دو نوکلئوتیدی ۴ باز می باشد. وجود تکرارها در طول پرایمر، اختصاصی بودن اتصال پرایمر به رشته الگو را کاهش می دهد.

۶-۳. اجتناب از Cross homology

منظور از اجتناب از Cross homology این است که سعی شود پرایمر از نواحی حفاظت شده توالی نوکلئوتیدی DNA مورد نظر طراحی شود، در این صورت احتمال اتصال اختصاصی پرایمر به ناحیه مورد نظر افزایش یافته و پرایمر مکان دیگری، غیر از محل مورد نظر را تکثیر نمی کند. برای رسیدن به این هدف پس از طراحی پرایمر بلاست کردن آن در پایگاه های داده مثل NCBI پیشنهاد می شود.

۷-۳. اجتناب از تشکیل ساختارهای ثانوی

تشکیل ساختارهای ثانوی در یک پرایمر یا بین پرایمرها می تواند دسترسی پرایمرها برای اتصال به رشته الگو و بنابراین تکثیر موفق مولکول های DNA مورد نظر را کاهش دهد و موجب عدم تشکیل محصول در PCR گردد.

۱-۷-۳. پرایمر دایمر[۸]

ایجاد پرایمر دایمر هم می تواند ناشی از اتصال دو پرایمر مشابه[۹] با هم و هم می تواند ناشی از اتصال پرایمرهای Forward و Reverse[۱۰] باشد (شکل ۱۰). ایجاد دایمر پرایمر می تواند میزان تولید محصول در PCR را کاهش دهد.

---

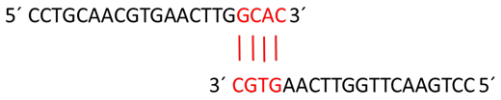
۸ Primer dimer

۹ Self dimer

۱۰ Cross dimer

در واقع یک حالت رقابتی بین محصول پرایمر دایمر و قطعه مورد نظر جهت تکثیر به وجود می آید. تشکیل پرایمر دایمر در انتهای ´۳ مشروط بر اینکه ΔG اتصال، حداکثر kcal/mol ۵- و برای پرایمر دایمرهای درونی تا ΔG برابر با kcal/mol ۶- باشد، قابل قبول است.

الف)

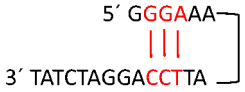


ب)

شکل ۱۰. دایمر پرایمر الف) Self dimer ب) Cross dimer

۳-۷-۲. ساختار سنجاق سری

وجود توالی پالیندرومی در پرایمر، منجر به تا شدن پرایمر و تشکیل ساختار سنجاق سری (Hairpin) می شود که یک ساختار درون مولکولی است (شکل ۱۱). این ساختار جفت شدن پرایمرها را با الگو تحت تاثیر قرار می دهد و پیامد این حالت نیز کاهش راندمان PCR است. تشکیل ساختار سنجاق سری تا ΔG برابر با kcal/mol ۲- در انتهای ´۳ و ΔG برابر با kcal/mol ۳- در نواحی درونی تر پرایمر قابل قبول است.



شکل ۱۱. ساختار سنجاق سری

۴ نرم افزارهای طراحی پرایمر

برای طراحی پرایمر، بیشتر محققان از توالی DNA برای یافتن پرایمر با خصوصیاتی که قبلا ذکر شد استفاده می کنند. تعداد زیادی نرم افزار برای طراحی پرایمر وجود دارند. در مجموع می توان نرم افزارهای طراحی پرایمر را به دو دسته آنلاین و آفلاین تقسیم کرد. در جدول ۱ تعدادی از نرم افزارهای آنلاین و آفلاین طراحی پرایمر لیست شده است.

نام نرم افزار	سیستم اجرای نرم افزار	آدرس سایت
Oligos	Offline	http://www.biocenter.- helsinki.fi/bi/bare-1_ html/oligos.htm
PCR Help!	Offline	http://www.techneuk. co.uk/CatMol/ pcrhelp.htm
Oligo	Offline	http://www.oligo.net/
iOligo	Offline	http://www.caesar software.com/pages/ products/ioligo/ ioligo.shtml
Primer3	Online	http://www-genome. wi.mit.edu/cgi-bin/ primer/primer3_ www.cgi
Web Primer	Online	http://genome- www2.stanford.edu/ cgi-bin/SGD/web primer
xprimer	Online	http://alces.med.umn. edu/webprimers.html
The Primer Generator	Online	http://www.med.jhu. edu/medcenter/ primer/primer.cgi

جدول ۱. نرم افزارهای آنلاین و افلاین طراحی پرایمر

اگر شما خودتان مایل به طراحی پرایمر نباشید و تمایل داشته باشید از توالی پرایمرهای موجود در مقالات استفاده کنید لازم است که Tm، احتمال تشکیل ساختارهای ثانوی و سنجاق سری و سایر خصوصیات پرایمرها را از طریق نرم افزارهای ویژه ای که در این زمینه وجود دارند مورد ارزیابی قرار گیرند. در جدول۲ نرم افزارهای آنلاین آنالیز پرایمرها لیست شده است.

نام نرم افزار	سیستم اجرای نرم افزار	آدرس سایت
Oligo Analyzer	Online	http://playground.idtdna.com/ program/oligocalc/oligocalc.asp
NetPrimer	Online	http://www.premierbiosoft. com/netprimer/netprimer.html

جدول ۲. نرم افزارهای آنلاین آنالیز کننده پرایمر

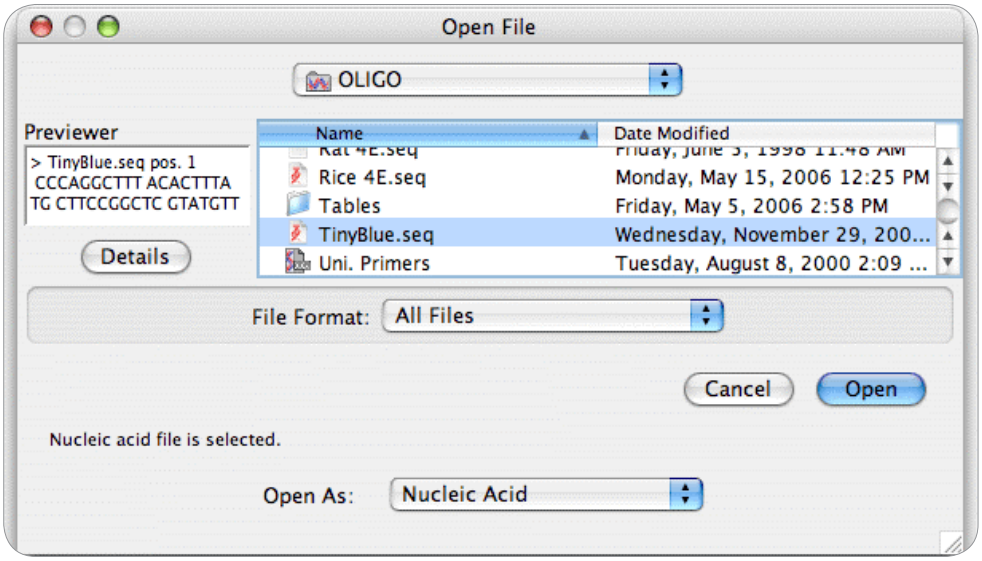
۵ آموزش کاربردی طراحی پرایمر با نرم افزار OLIGO 7

انواع نرم افزارهای آفلاین جهت طراحی پرایمر موجود هستند که قابل نصب در رایانه می باشند. از جمله این نرم افزارها، OLIGO است که نسخه اصلی[۱۱] آن به صورت تجاری به فروش می رسد. در این قسمت نحوه طراحی پرایمر با نسخه ی ۷ (OLIGO7) این نرم افزار آموزش داده می شود. آخرین نسخه ی این نرم افزار از نشانی زیر قابل تهیه می باشد.

http://www.oligo.net/

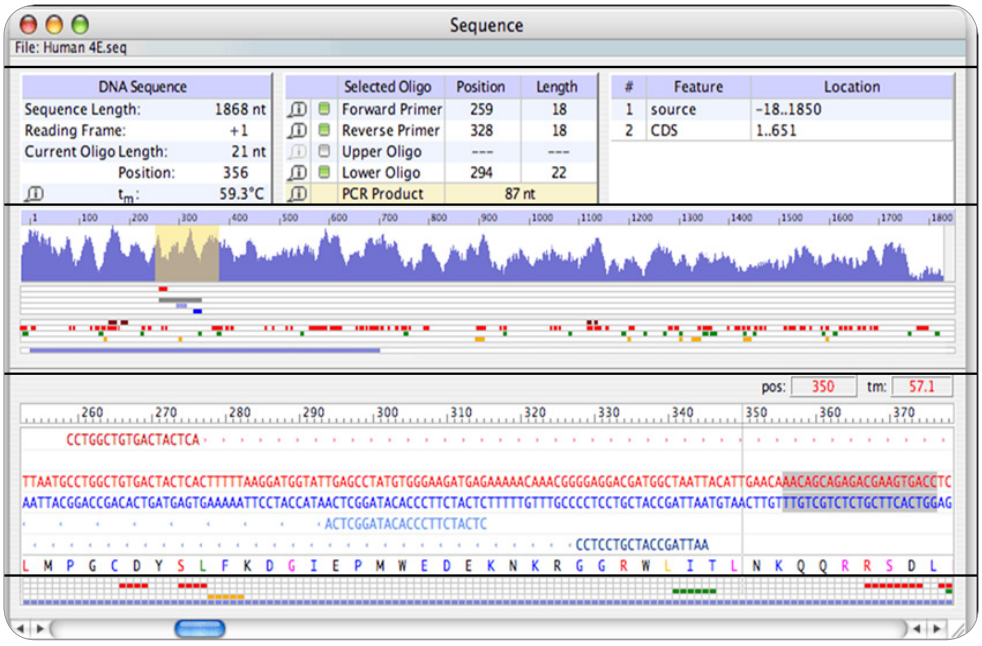
۱۱ Original

در ابتدا لازم است که توالی DNA مورد نظر به فرمت FASTA ذخیره شود. برای باز کردن توالی در نرم افزار باید از منوی File گزینه ی Open استفاده کرد. پنجره ی باز شده همانطور که در شکل ۱۲ مشاهده می کنید اجازه انتخاب و باز کردن فایل توالی مورد نظر را فراهم می کند (شکل ۱۲).



شکل ۱۲. باز کردن فایل مورد نظر در نرم افزار OLIGO

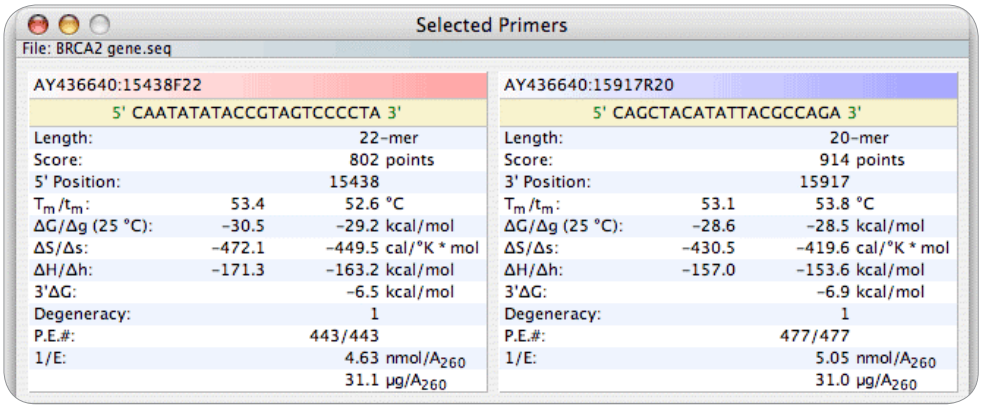
پس از باز کردن فایل توالی مورد نظر باید با استفاده از منوی Select گزینه ی Forward Primers یا Reverse Primers، نوع پرایمرها را انتخاب و موقعیت آنها را مشخص نمود. در ضمن از طریق منوی Change، گزینه ی Current Oligo Length می توان طول مورد نظر برای پرایمرها را انتخاب کرد. پس از انتخاب موقعیت و طول پرایمرها پنجره مشابه با شکل ۱۳ باز می شود که اطلاعات زیادی درباره توالی الگو و پرایمرها از جمله نمودار Tm، توالی های مستعد برای تشکیل ساختار سنجاق سری، موقعیت و طول پرایمرها و طول محصول PCR را نشان می دهد (شکل ۱۳).



شکل ۱۳. نمایش پنجره ی توالی باز شده در نرم افزار OLIGO

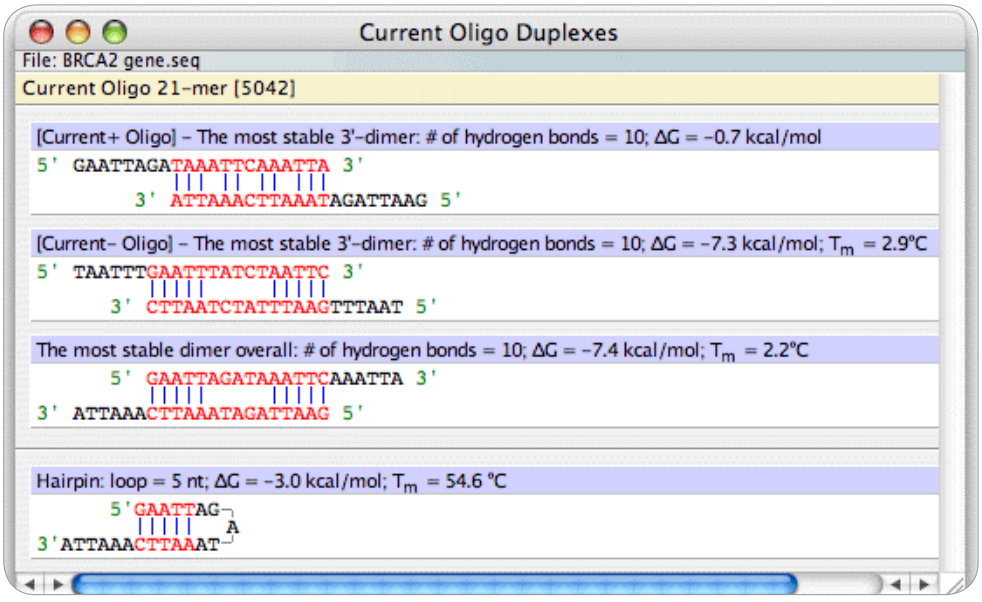
پس از انتخاب موقعیت و طول پرایمرها، در این مرحله برای آنالیز پرایمرها باید گزینه Analyze را کلیک نمایید. این گزینه پارامترهای مختلفی را مورد بررسی قرار می دهد که مهمترین آنها در ادامه بررسی می شود.

منوی Analyze و سپس Key Info گزینه هایی را دارا می باشد که اطلاعات مفیدی از جمله توالی، طول، ΔG و Tm پرایمرها را نمایش می دهد. در شکل زیر پنجره مربوط به گزینه ی Selected Primers نشان داده شده است (شکل ۱۴).



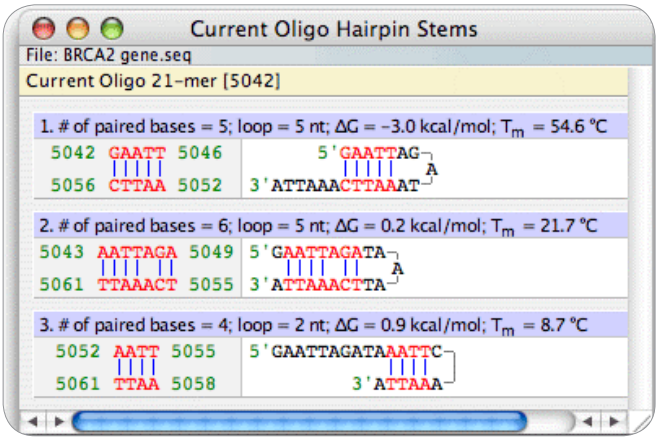
شکل ۱۴. نمایش پنجره ی Key Info>Selected Primers از منوی Analyze نرم افزار OLIGO

منوی Analyze و سپس Duplex Formation دارای گزینه هایی جهت نشان دادن دایمرهای بالقوه در یک پرایمر یا بین پرایمرهای مختلف است. ΔG و Tm مربوط به این ساختارهای ثانوی نیز در کنار هر ساختار نشان داده شده است. در شکل ۱۵ پنجره ی باز شده مربوط به گزینه ی Current Oligo Duplexes نمایش داده شده است (شکل ۱۵).



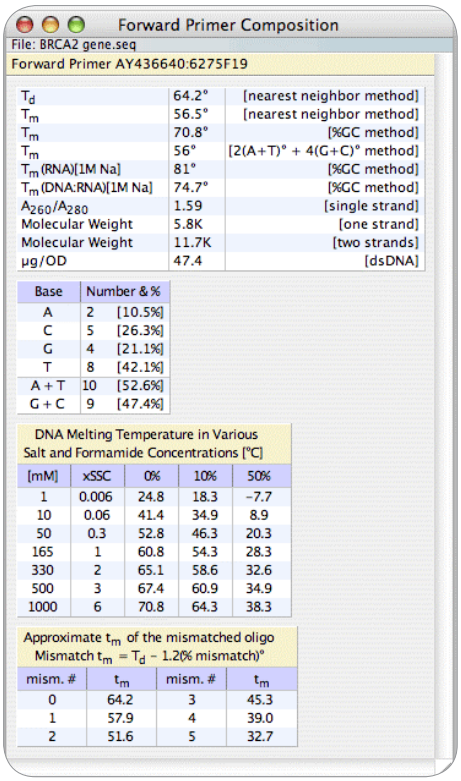
شکل ۱۵. نمایش پنجره ی Duplex Formation>Current Oligo Duplexes از منوی Analyze نرم افزار OLIGO

منوی Analyze و سپس Hairpin Formation دارای گزینه هایی برای نشان دادن ساختارهای سنجاق سری بالقوه در پرایمرهای طراحی شده است. ΔG و Tm مربوط به این ساختارهای ثانوی نیز در کنار هر ساختار نشان داده شده است. شکل ۱۶ پنجره ی باز شده مربوط به گزینه ی Current Oligo Hairpin Stems است که ساختارهای سنجاق سری بالقوه را در موقعیت انتخاب شده نمایش می دهد (شکل ۱۶).



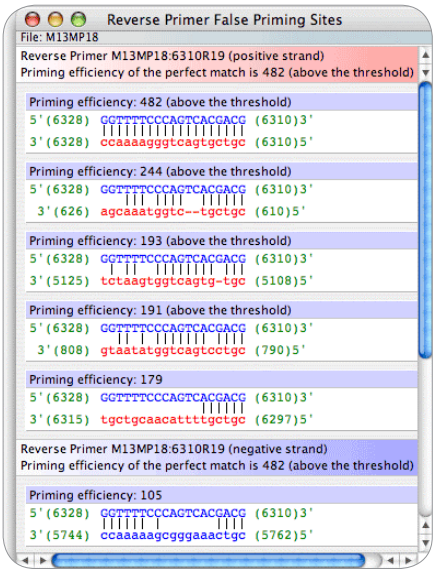
شکل ۱۶. نمایش پنجره ی Hairpin Formation>Current Oligo Hairpin Stems از منوی Analyze نرم افزار OLIGO

منوی Analyze و سپس Composition & Tm، دارای گزینه های است که ترکیب بازی شامل محتوای GC و Tm پرایمرهای انتخاب شده، توالی ورودی و محصول PCR را نمایش می دهد. شکل ۱۷ پنجره ی باز شده مربوط به گزینه ی Forward Primer Composition است که اطلاعات مربوط به GC و Tm پرایمر Forward را نمایش می دهد (شکل ۱۷).



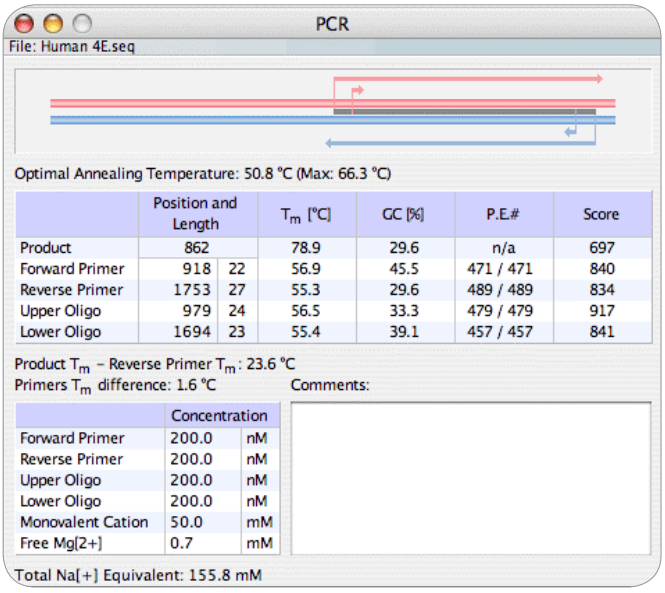
شکل ۱۷. نمایش پنجره ی Composition & Tm>Forward Primer Composition از منوی Analyze نرم افزار OLIGO

انتخاب گزینه False Priming Site از منوی Analyze، امکان بررسی اتصال غیر اختصاصی پرایمر به مکان هایی غیر از محل مورد نظر که تحت عنوان False Priming Site شناخته می شوند، را فراهم می کند. به کمك این گزینه، امکان آنالیز نواحی false priming برای پرایمرهای Forward و Reverse هم در رشته ی sense و هم antisense وجود دارد. بر اساس پارامترهای پیش فرض نرم افزار، اگر عبارت « above the threshold » در مقابل شاخص priming efficiency (بیانگر کارایی اتصال پرایمر به رشته الگو) نوشته شود، چون احتمال اتصال این پرایمر به مکانی غیر از محلی که برای آن طراحی شده در چنین شرایطی بالا خواهد بود از این پرایمر باید صرف نظر شود. شکل ۱۸ پنجره ی باز شده مربوط به گزینه ی Reverse Primer False Priming Sites است که اطلاعات مربوط به آنالیز دقیق نواحی false priming را برای پرایمر Reverse نمایش می دهد (شکل ۱۸).



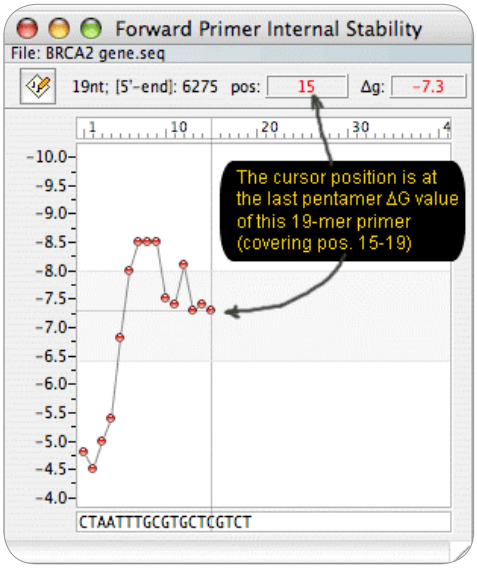
شکل ۱۸. نمایش پنجره ی False Priming Sites>Reverse Primer False Priming Sites از منوی Analyze نرم افزار OLIGO

منوی Analyze و سپس گزینه ی PCR اطلاعات متنوعی درباره محصول PCR مربوط به پرایمرهای انتخابی را نمایش می دهد. در این پنجره دمای اتصال بهینه (T_aOPT) و حداکثر دمای اتصال ($T_a$max) مشخص شده است. در ابتدا بهتر است که از T_aOPT برای انجام واکنش PCR استفاده شود، اگر در واکنش PCR باند غیر اختصاصی مشاهده شد می توان برای اختصاصی تر شدن واکنش از دماهای بالاتر و حتی $T_a$max نیز استفاده کرد (شکل ۱۹).



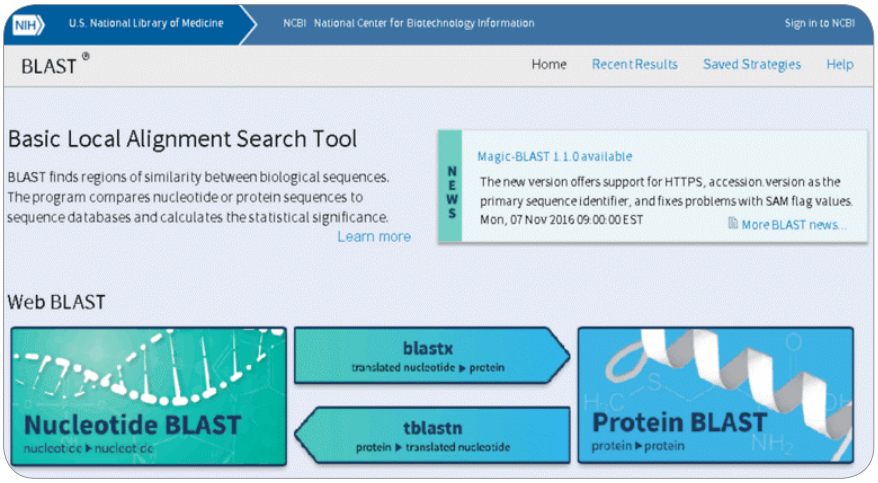
شکل ۱۹. نمایش پنجره ی PCR از منوی Analyze نرم افزار الیگو

گزینه Internal Stability از منوی Analyze نشان دهنده نمودار پایداری داخلی پرایمر طراحی شده است. میزان این پایداری در نواحی مختلف پرایمر وابسته به میزان CG آن ناحیه است. هر چه میزان CG در ناحیه‌ای از پرایمر بالاتر باشد، پایداری داخلی پرایمر در آن ناحیه بیشتر است. بهتر است که پایداری داخلی پرایمر در انتهای ´۳ کمتر از نواحی میانی پرایمر باشد (شکل ۲۰).



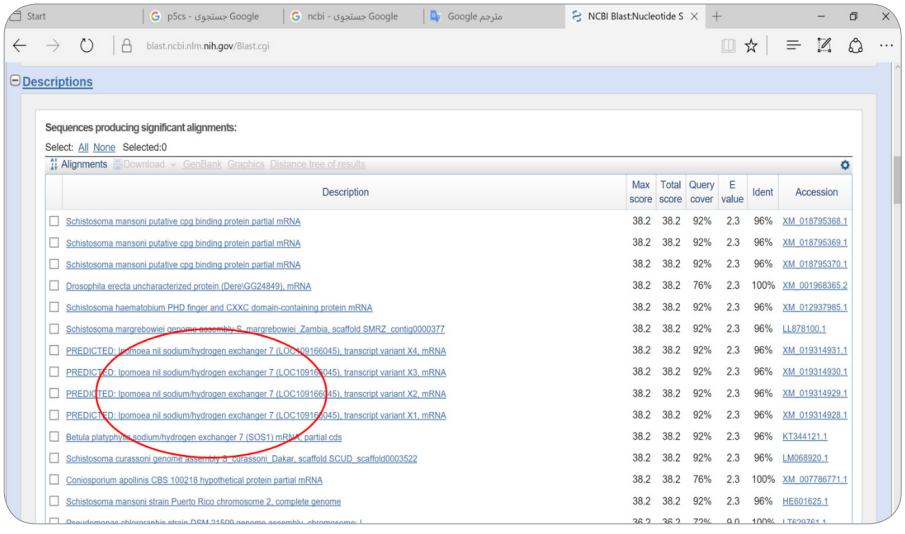
شکل ۲۰. نمایش پنجره ی Internal Stability از منوی Analyze نرم افزار الیگو

در انتها برای اینکه مطمئن شویم پرایمرهای Forward و Reverse انتخاب شده به توالی دیگری غیر از توالی که بطور اختصاصی برای آن طراحی شده اند اتصال نمی یابند، باید پرایمرها را در پایگاه داده ای مانند NCBI بلاست کرد که در این صورت می توان از Nucleotide Blast و یا Primer-Blast برای بلاست نمودن استفاده نمود (شکل ۲۱).



شکل ۲۱. صفحه ی بلاست پایگاه داده NCBI

بعنوان مثال اگر طراحی پرایمر برای ژن (Sodium Hydrogen Exchanger 7) *NHX7* انجام شده باشد، پس از Nucleotide Blast، نتایج بدست آمده در شکل ۲۲ نشان می دهد که پرایمر از ناحیه ی حفاظت شده ژن مورد نظر طراحی شده و پس از بلاست نیز به ژن مورد نظر چسبیده است، بنابراین پرایمر مناسبی می باشد. در صورت مطلوب بودن نتیجه بلاست می توان پرایمرها را به شرکت مربوطه برای ساخت سفارش داد (شکل ۲۲).



شکل ۲۲. نمایش نتیجه ی بلاست در پایگاه داده NCBI

منابع

1. Abd-Elsalam KA. Bioinformatic tools and guideline for PCR primer design. African Journal of Biotechnology. 2003. 2(5):91-5.

2. Applications of bioinformatics-biotech articles. http://www.biotecharticles.com

3. Basic local alignment search tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi

4. Pevsner J. Bioinformatics and Functional Genomics. 2015. by John Wiley & Sons, Inc.

5. Claverie JM and Notredame C. Bioinformatics for Dummies. 2007. Published by Wiley publishing Inc .

6. Bioinformatics web-comprehensive educational resource on bioinformatics. www.geocities.com/bioinformaticsweb

7. Breman AM, Chow JC, Uren L, Normand EA, Qdaisat S, Zhao L, et al. Evidence for feasibility of fetal trophoblastic cell-based noninvasive prenatal testing. Prenatal Diagnosis. 2016. 36(11): 1009–19.

8. Dennis A, Benson KM, David JL, James O and Eric W. Sayers. GenBank. Nucleic Acids Research. 2011. 32–7.

9. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Science Translational Medicine .2012. 4(136): 136-68.

10. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nature Reviews Genetics. 2012. 13(12): 840-52.

11. Genome news network. http://www.genomenewsnetwork.org/articles/07_02/tepidum.shtml

12. Bioinformatics: open access education resource. http://bioinformaticsweb.net/applications.html

13. Lewitter F and Bourne PE. Teaching bioinformatics at the secondary school level. PLoS Computational Biology. 2011. 7(10):e1002242.

14. Luscombe NM, Greenbaum D and Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods of Information in Medicine. 2001. 40(4): 346-58.

15. National center for biotechnology information: http://www.ncbi.nlm.nih.gov

16. Pease J and Sooknanan R. A rapid, directional RNA-seq library preparation workflow for Illumina sequencing. Nature Methods. 2012. 9(3).

17. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012; 13: 341-54.

18. Role of bioinformatics in agriculture and sustainable development. http://www.academia.edu/6808947/Role_of_Bioinformatics_In_Agriculture_and_Sustainable_Development

19. Salipante SJ, Scroggins SM, Hampel HL, Turner EH and Pritchard CC. Microsatellite instability detection by next generation sequencing. Clinical Chemistry. 2014. 60(9): 1192-9.

20. Sharma R, Khajuria R, Sharma C, Kapoor B, Goswami K and Kohli K. Gene therapy: current concepts. JK Science. 2004. 6:62-6.

21. Thampi SM. Introduction to bioinformatics. arXiv preprint arXiv:09114230. 2009.

22. Universal protein resource: http://www.uniprot.org/

23. Wallace RB, Shaffer J, Murphy R, Bonner J, Hirose T and Itakura K. Hybridization of synthetic oligodeoxyribonucleotides to ΦX 174 DNA: the effect of single base pair mismatch. Nucleic Acids Research. 1979. 6(11):3543-58.

24. Xu XP, Gan HY, Li FX, Tian Q, Zhang J, Liang RL, et al. A Method to quantify cell-free fetal DNA fraction in maternal plasma using next generation sequencing: its application in non-invasive prenatal chromosomal aneuploidy detection. PLoS ONE. 2016. 11(1):e0146997.

25. Xue J, Zhao S, Liang Y, Hou C and Wang J, editors. Bioinformatics and its applications in agriculture. International conference on computer and computing technologies in agriculture. Springer. 2007.

26. Zhou x, Ren L, Meng Q, Li Y, Yu Y and Yu J. The next generation sequencing technology and application. Protein Cell .2010. 520–36.

27. Zimmer C. The microbe factor and its role in our climate future. Yale Environment. 2015. 360:1355-59.

واژه نامه

DNA پلیمراز: آنزیمی که با افزودن نوکلئوتیدهای مکمل به یک رشته DNA در دو رشته ای شدن و در واقع همانند سازی DNA نقش دارد.

Gigabase (Gb): طولی از DNA که معادل با ۱۰۹ نوکلئوتید است.

امیکس: واژه کلی است که به مطالعات جامع و گسترده در سیستم های زیستی در سطوح مختلف ژنوم، ترانسکریپتوم، پروتئوم، متابولوم و سایر «اوم»ها اطلاق می شود.

پروتئوم: کل پروتئین های یک موجود، بافت یا سلول است که علم پروتئومیکس به مطالعه آن می پردازد.

چند شکلی طولی قطعات تکثیر شده (AFLP): نوعی مارکر مولکولی مبتنی بر PCR است که برای بررسی تنوع و تغییرات اندک توالی ژنومی در گونه های نزدیک به کار می رود. به این منظور قطعات DNA ژنومی نمونه های مورد نظر با استفاده از آنزیم های محدودالاثر اختصاصی در جایگاه های انتخابی برش داده می شوند و سپس این قطعات برش یافته از طریق PCR تکثیر می شوند. این تکثیر موجب ایجاد محصولات متنوعی می شود که به کمک الکتروفورز از یکدیگر قابل تفکیک هستند، وجود نوارهای DNA با طول های متفاوت در بین نمونه ها در سطح ژل نشان دهنده پلی مورفیسم (چند شکلی) در جایگاه شناسایی آنزیم های محدودالاثر است.

چند شکلی طولی قطعات ایجاد شده توسط آنزیم های محدودالاثر (RFLP): نوعی مارکر مولکولی مبتنی بر هیبریداسیون است که برای بررسی تنوع و تغییرات جزیی توالی ژنومی در گونه های نزدیک به کار می رود. قطعات DNA نمونه های مورد بررسی با استفاده از آنزیم های محدودالاثر اختصاصی برش داده شده و به این ترتیب قطعاتی با طول های مختلف ایجاد می شود که با استفاده از الکتروفورز از یکدیگر تفکیک می شوند. در این نوع مارکرها، یک قطعه DNA تک رشته ای نشاندار (پروب یا کاوشگر) برای هیبریداسیون با قطعات تفکیک شده بر روی ژل استفاده می شود. هیبریداسیون یا عدم هیبریداسیون پروب با قطعات تفکیک شده (متعلق به نمونه های مختلف) به ترتیب نشان دهنده وجود یا عدم وجود جایگاه شناسایی اختصاصی آنزیم های محدود کننده و در واقع پلی مورفیسم (چند شکلی) در این جایگاه در بین نمونه های مورد بررسی است.

پیرایش متناوب: در طی این فرایند توالی مربوط به اینترون ها از مولکول RNA اولیه حذف شده و توالی مربوط

به اگزون های مختلف با ترتیب های متفاوتی به یکدیگر متصل می شوند و نسخه های متفاوت mRNA را ایجاد می‌کنند.

ترانسکریپتوم: تمام نسخه های RNA در یک نوع موجود، بافت یا یک نوع سلول است که علم ترانسکریپتومیکس به مطالعه آن می‌پردازد.

توالی پالیندرومی: نوعی از توالی های نوکلئوتیدی که از روی هر یک از دو رشته DNA که خوانده شوند (در جهت ´۵ به ´۳) یکسان هستند.

دالتون: واحد اندازه‌گیری جرم در مقیاس اتمی و مولکولی است و با علامت اختصاری u یا Da نمایش داده می شود. هر دالتون معادل $۱/۶۶۰۵۳۹۰۴۰ (۲۰) \times ۱۰^{-۲۷}$kg است.

دومین: واحدهایی با تا خوردگی مستقل و پایدار در ساختار سوم پروتئین ها هستند که می توانند به صورت مستقل از بقیه پروتئین وجود داشته باشند و عمل کنند. در پروتئین های مختلف، انواع متفاوتی از دومین ها مانند دومین های ساختاری، عملکردی یا تنظیمی وجود دارند.

رسوب دهی ایمونوکروماتین: نوعی روش رسوب دهی پروتئین است که برای بررسی برهمکنش پروتئین و DNA در داخل سلول استفاده می شود.

رشته Antisense: رشته non-coding یا غیر کدکننده است که توالی آن مکمل توالی mRNA است.

رشته Sense: رشته coding یا کدکننده DNA است که توالی آن مشابه با توالی mRNA است.

ریز آرایه یا میکروواری: روشی برای بررسی بیان ده ها، صدها و هزاران ژن در یک سطح کوچک است. در این روش، قطعات کوتاهی از DNA در سطح یک اسلاید میکروسکوپی به صورت لکه های متعدد ثابت شده‌اند، هر لکه نماینده یک ژن است. نمونه mRNA مورد بررسی به cDNA تبدیل شده و با رنگ های فلورسانت نشاندار می گردد و به سطح این اسلاید افزوده می شود. مولکول های cDNA که مکمل قطعات متصل به سطح اسلاید هستند، به

این سطح متصل شده و سایر مولکول ها در طی شستشو در چند مرحله حذف می‌گردند. از آنجایی که cDNAها فلورسانت هستند، مکان اتصال آنها در اسلاید به صورت نقاط رنگی مشخص می شود که این رنگ ها توسط دستگاهی اسکن می شوند. شدت نور فلورسانس نیز نشان دهنده میزان بیان ژن های مورد نظر در نمونه mRNA اولیه است.

ژنوم: مجموع تمام ژن های موجود در یک موجود، بافت یا سلول است که علم ژنومیکس به مطالعه آن می پردازد.

ژن های هومولوگ: هومولوگ بودن به معنای داشتن جد یا نیا مشترک است. ژن های هومولوگ توالی هایی مشابه و همچنین ساختار و عملکرد مشابهی دارند.

کاسمید: نوعی وکتور کلونینگ که قادر است قطعات DNA با طول حداقل ۳۸ کیلو جفت باز (۳۸kb) را در خود جای دهد و به صورت پلاسمید در باکتری باقی بماند. کاسمید ها ویژگی های مشترکی از وکتورهای پلاسمیدی و فاژ لامبدا دارند.

کروموزوم مصنوعی باکتری یا BAC: وکتور باکتریایی بر پایه فاکتور F باکتری E.coli که برای کلون کردن قطعات بسیار بلند (با طول بیش از ۳۰۰ کیلو جفت باز) DNA در باکتری استفاده می شود.

کروموزوم مصنوعی مخمر یا YAC: کروموزوم مصنوعی متشکل از سانترومر مخمر، دو توالی تلومری و یک مارکر که برای کلون کردن قطعات بسیار بزرگ DNA در مخمر استفاده می شود.

کلون کردن (کلونینگ): روشی برای تولید نسخه های متعدد از یک مولکولDNA. در این روش قطعه DNA واجد ژن یا توالی نوکلئوتیدی مورد نظر به پلاسمید (وکتور) وارد می شود و به سلول میزبان (باکتری) منتقل می گردد، به این ترتیب همراه با همانند سازی پلاسمید در سلول میزبان، قطعه DNA مورد نظر نیز به مقدار زیاد تولید می شود.
متابولوم: کل متابولیت های یک موجود، بافت یا سلول است که علم متابولومیکس به مطالعه آن می پردازد.

وکتور: یک قطعه DNA بیگانه است که قادر به همانند سازی مستقل در سلول میزبان (معمولاً باکتری) می باشد و برای کلون کردن ژن های مورد نظر استفاده می‌گردد، پلاسمیدها متداولترین نوع وکتورها هستند.