

مبانی بیوانفورماتیک

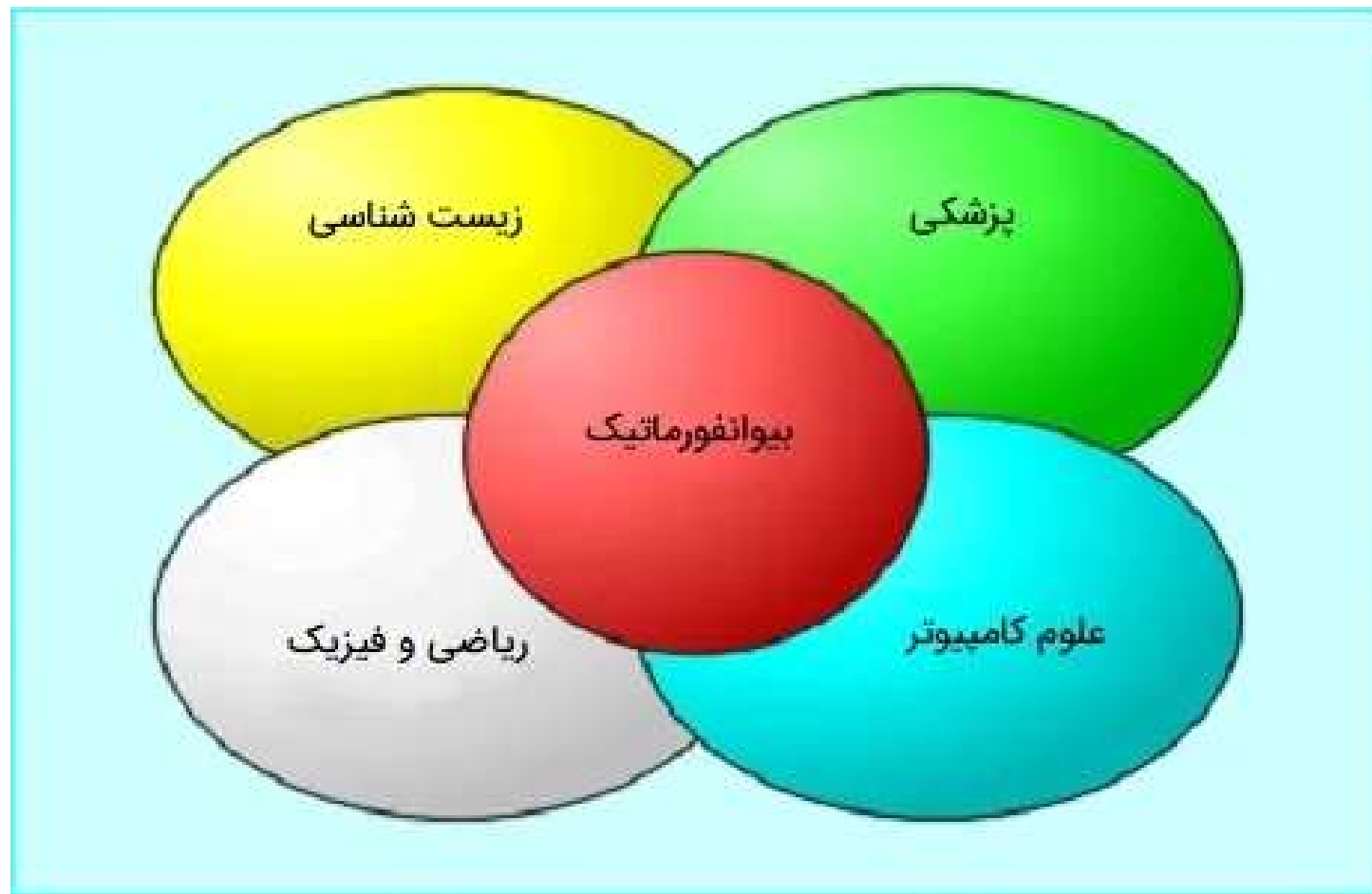
Fundamental Bioinformatics

بیوانفورماتیک چیست؟

- رشد سریع و توسعه فناوری رایانه‌ای، همراه با پیشرفت‌های چشمگیر در درک ما از زیست‌شناسی، رشد اخیر در بیوانفورماتیک را تسهیل کرده است.
- بیوانفورماتیک، علوم کامپیوتر را با زیست‌شناسی ادغام می‌کند تا داده‌های مربوط به سیستم‌های بیولوژیکی را ذخیره، تجزیه و تحلیل و به اشتراک بگذارد که اغلب این داده‌ها مربوط به DNA و توالی اسیدهای آمینه است.

بیوانفورماتیک چیست؟

- بیوانفورماتیک ترکیبی از زیست شناسی و فناوری اطلاعات است.
- اساساً، بیوانفورماتیک علمی است که اخیراً توسعه یافته و از فناوری اطلاعات برای درک پدیده بیولوژیکی استفاده می کند.
- بیوانفورماتیک به طور گسترده ای شامل ابزارها و روش های محاسباتی است که برای مدیریت، تجزیه و تحلیل و اداره کردن داده های حجیم بیولوژیکی استفاده می شود.
- بیوانفورماتیک همچنین ممکن است به عنوان بخشی از زیست شناسی محاسباتی در نظر گرفته شود.
- زیست شناسی محاسباتی با استفاده از تکنیک های تحلیلی کمی در مدل سازی و حل مشکلات در سیستم های بیولوژیکی درگیر است.
- بیوانفورماتیک یک رویکرد میان رشته ای است که به دانش پیشرفته علوم کامپیوتر، ریاضیات و روش های آماری برای درک پدیده های بیولوژیکی در سطح مولکولی نیاز دارد.



بیوانفورماتیک چیست؟

- استفاده از کامپیوتر برای حل مشکلات بسیار اختصاصی زیست شناسی
- قبل از عصر بیوانفورماتیک دو راه برای انجام آزمایشهای زیست شناسی
 - درون موجود زنده (*in vivo*)
 - در محیط مصنوعی (*in vitro*) یا درون شیشه
- بیوانفورماتیک=زیست شناسی *in silico*
- استفاده از تراشه های سیلیکونی تشکیل دهنده ریزپردازنده ها در زیست شناسی
- اولین قدم در درک بیوانفورماتیک=درک مفاهیم اصولی زیست شناسی مولکولی

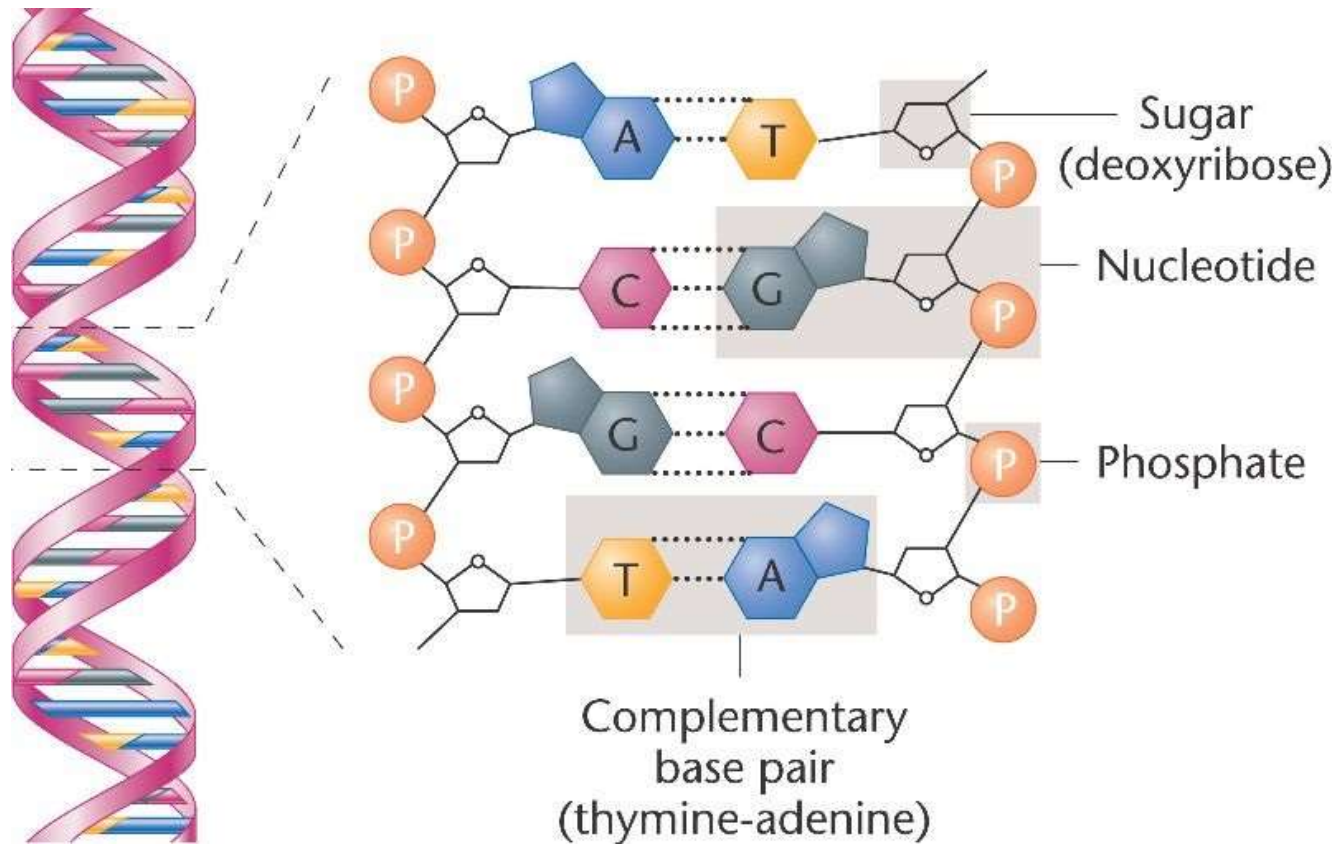
مروری بر زیست شناسی مولکولی



ساختار مولکولی DNA

- در ساختار مولکولی DNA، هر یک مولکول متشکل از دو زنجیره است که به دور هم پیچیده شده اند.
- دو رشته DNA به عنوان polynucleotide نیز شناخته می شوند. زیرا از واحدهای مونومری ساده تر به نام نوکلئوتید تشکیل شده اند.
- هر نوکلئوتید متشکل از یکی از چهار هسته بازی سیتوزین C، گوانین G، آدنین A یا تیمین T است. علاوه بر یک قند به نام دزوکسی ریبوز و یک گروه فسفات که شامل نوکلئوتیدها هستند.

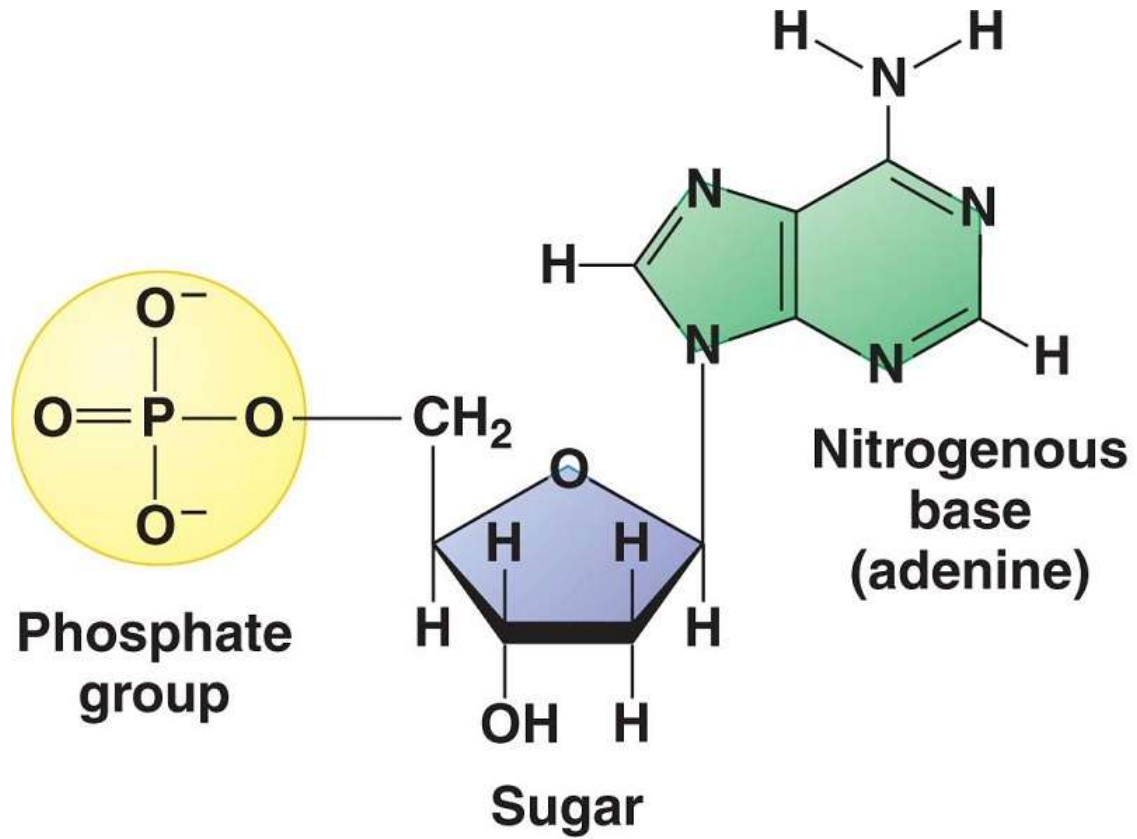
ساختار مولکولی DNA



نوکلئوتید

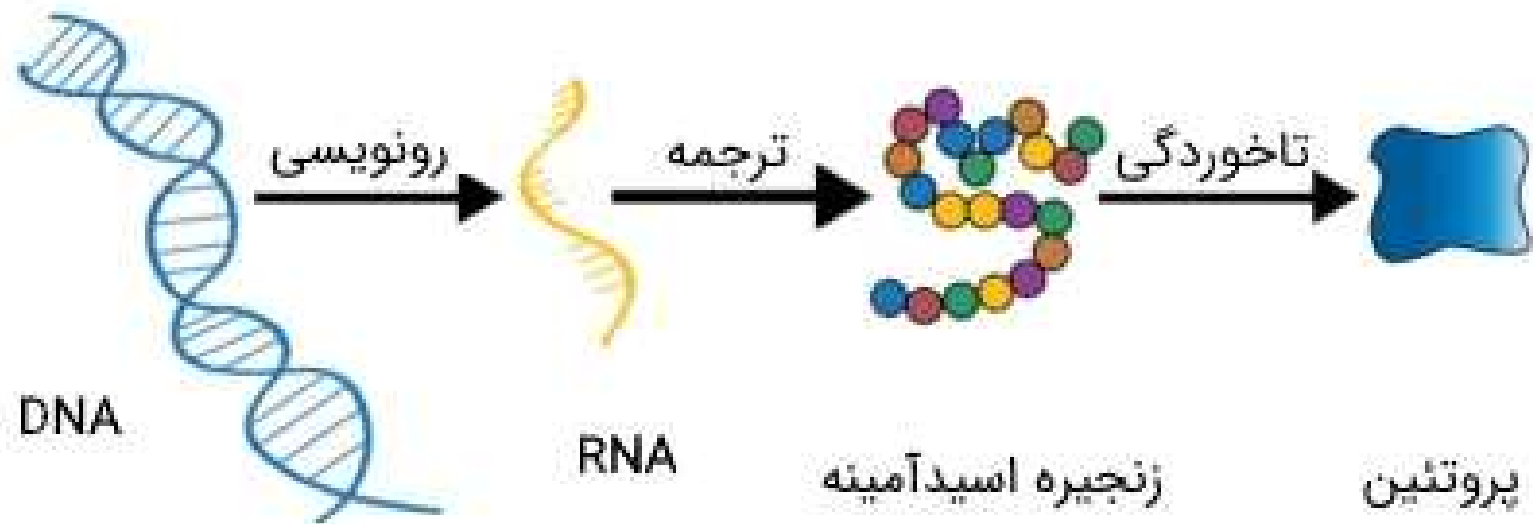
- در ساختار مولکولی DNA اسید نوکلئیک یکی از ماکرومولکولهای زیستی است که وظیفه ذخیره اطلاعات ژنتیکی را در سلول بر عهده دارد.
- جایگاه اسیدهای نوکلئیک در هسته و سیتوپلاسم سلول است که از واحدهایی به نام نوکلئوتید ساخته شده اند.
- نوکلئوتیدها اعمال متنوعی را در داخل سلول انجام می دهند.
- آنها به عنوان زیر واحدهای اسیدهای نوکلئیک حامل اطلاعات ژنتیکی به حساب می آیند.
- ساختمان هر پروتئین و نهایتا هر بیومولکول ، محصولی از اطلاعات موجود در توالی نوکلئوتیدی اسیدهای نوکلئیک سلول می باشد.
- توانایی ذخیره و انتقال اطلاعات ژنتیکی از نسلی به نسل بعد شرط اساسی زندگی است.
- در ساختار مولکولی DNA هر نوکلئوتید از قسمتهای زیر تشکیل شده است:
 - یک مولکول اسید فسفریک
 - مولکول قند ۵ کربنی
 - یک مولکول باز نیتروژن

نوكلئوتيد



ساختار مولکولی DNA

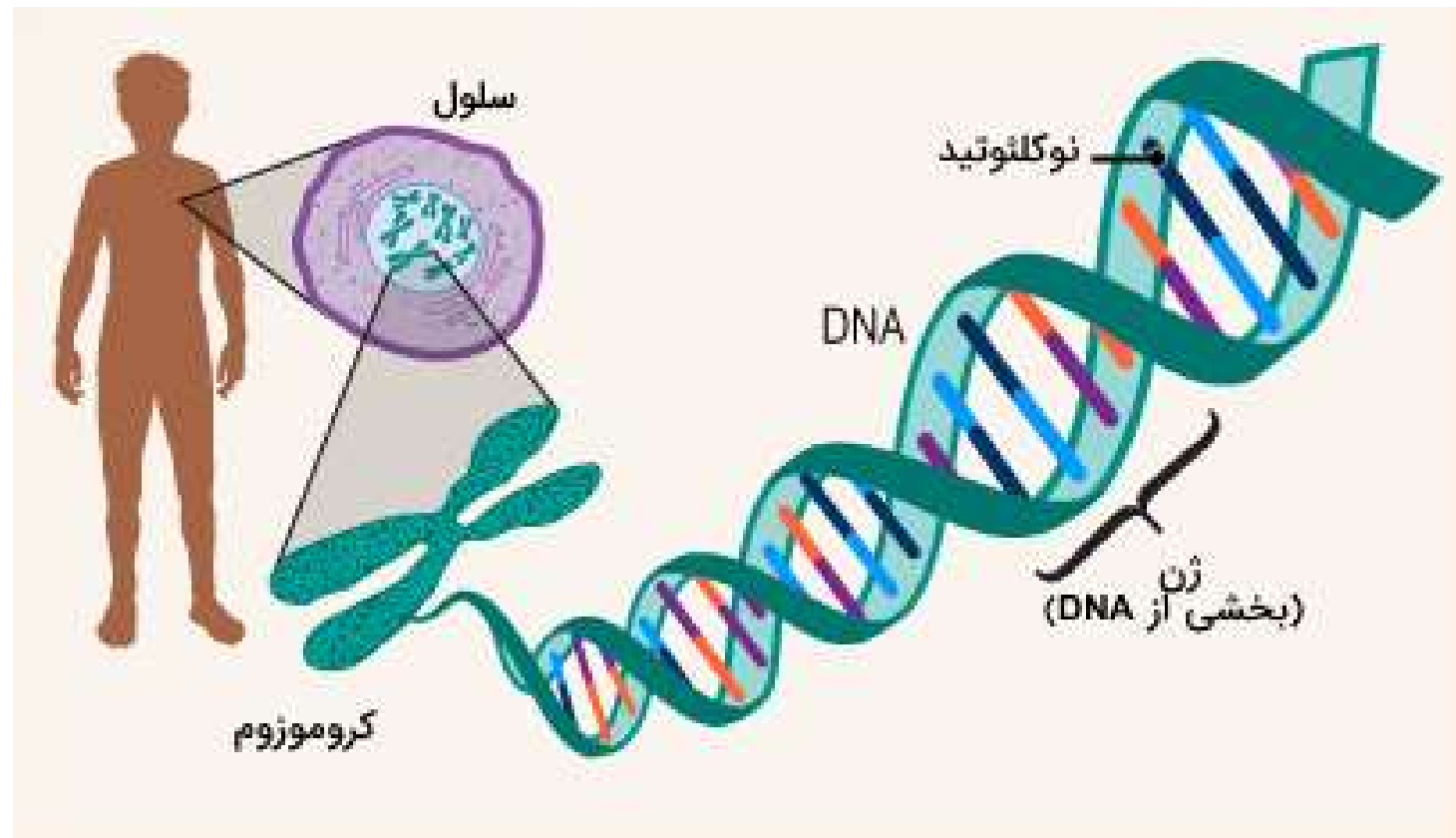
- در ساختار مولکولی DNA عملکردهای سلول از طریق پروتئین های مختلفی انجام می گیرد که از روی رمزهای موجود بر روی مولکول DNA کدگذاری می شوند. بدین صورت که مولکول های تک رشته ای به نام RNA از روی مولکول DNA الگوبرداری شده (فرایند رونویسی Transcription) و توسط دستگاه ترجمه سلولی به پروتئین ها ترجمه و تبدیل می شوند (فرایند ترجمه، Translation).
- بخش بزرگی از DNA یعنی بیش از ۹۸ درصد، برای انسانها رمزگذاری نمی شود. به این معنی که این بخش ها به عنوان الگویی برای ساخت پروتئین ها عمل نمی کنند.



با رونویسی از ژن، رشته‌های **RNA** ایجاد می‌شوند که پس از فرایند ترجمه زنجیره‌های اسید آمینه یا پپتیدها را می‌سازند. این زنجیره‌ها پس از ایجاد تاخوردگی، ساختارهای فضایی و عملکردی پروتئین‌ها را به وجود می‌آورند.

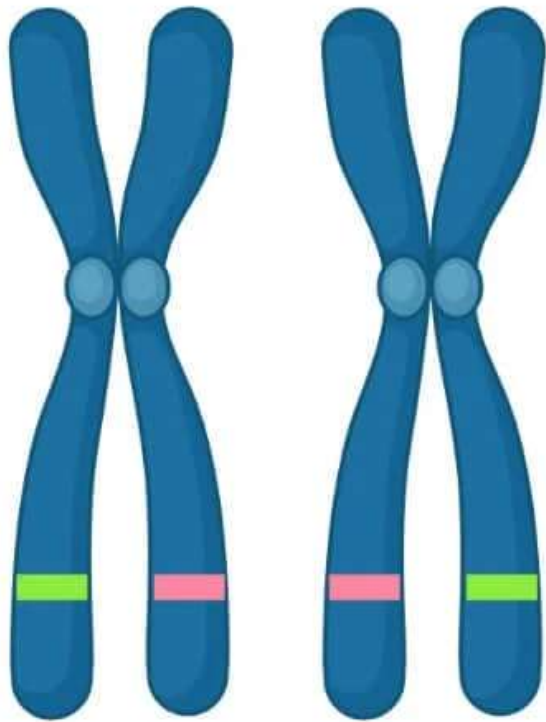
ژن

- در ساختار مولکولی DNA ژن کوچک ترین واحد شیمیایی اطلاعات ارثی است.
- یک ژن یک توالی از DNA و یا RNA است که به تعداد بی شماری درون میلیاردها سلول تشکیل دهنده بدن وجود دارد.
- چنین تخمین زده می شود که در ۲۳ جفت کروموزوم هسته سلول انسانی، بیش از صدهزار ژن اصلی می تواند وجود داشته باشد. شاید بیش از این مقدار ژن غیر اصلی هم وجود داشته باشد.
- مهمترین ویژگی در ژن ها، انتقال صفات والدین به فرزندان است.
- علاوه بر این، ژن ها همه اعمال سلول را کنترل می کنند.
- ژن ها که در هسته سلول قرار گرفته اند، همانند ستاد فرماندهی سلول عمل می کنند. همچنین همه اعمال درونی و بیرونی سلول، پیرو فرمان های صادر شده از سوی آنها است.
- یکی از مهم ترین این اعمال سلولی، عمل تولید و انتقال ساختار و صفات والدین به فرزندان می باشد.
- در واقع ژن واحد مولکولی وراثت یک موجود زنده است.



آلل

- در ساختار مولکولی DNA به انواع جایگزین اطلاعات ژنتیکی در یک جایگاه ژنی خاص یا به عبارت دیگر دو شکل مختلف یک ژن آلل می گویند.
- حداکثر تعداد آلل هایی که یک موجود دیپلوئید مانند انسان می تواند برای هر ژن داشته باشد ۲ آلل می باشد.
- هر کدام از این آلل ها روی یکی از کروموزوم های همولوگ قرار دارند.
- آلل ها هستند که تعیین می کنند ژن ها چگونه در یک فرد بروز پیدا کنند و آشکار شوند.
- همانطور که گفته شد موجودات زنده دو آلل برای هر صفت دارند.
- به وضعیتی که دو آلل متفاوت برای یک صفت وجود داشته دارد، هتروزیگوت گفته می شود.
- در نوعی از هتروزیگوت یکی از دو آلل غالب و دیگری مغلوب است. در نتیجه آلل غالب آشکار و آلل مغلوب پنهان می شود.
- به عنوان مثال در انسان آلل برای رنگ چشم قهوه ای غالب و آلل برای رنگ چشم آبی مغلوب است. غالب بودن رنگ قهوه ای بر رنگ آبی در چشم یک نمونه از تسلط کامل است.
- در روابط هتروزیگوتی که در آن هیچ کدام از آلل ها غالب نیستند اما هر دو به طور کامل آشکارند که آن هم بارز گفته می شود.
- نمونه وضعیت هم بارز، به ارث رسیدن گروه خونی AB است.

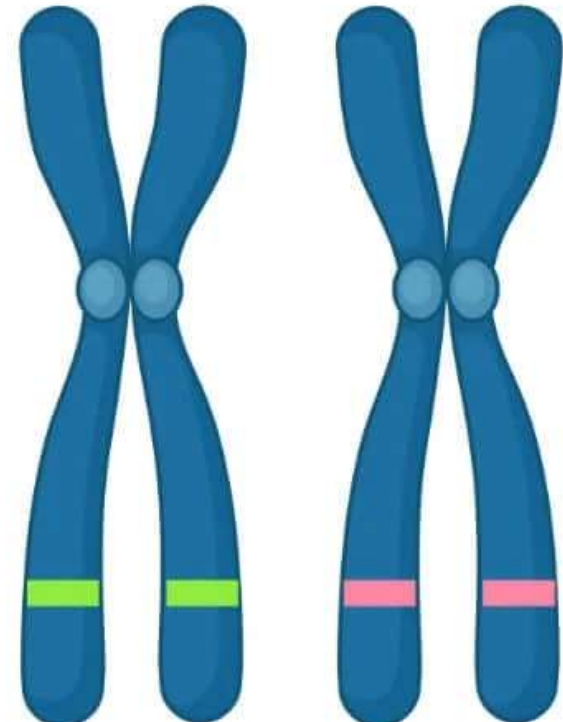


Aa

aA

Heterozygous Alleles

Alleles



AA

aa

Homozygous Alleles

موتاسیون یا جهش

- در ساختار مولکولی DNA جهش یا موتاسیون یک تغییر ژنتیکی است که صفات زیستی بعضی از افراد را تغییر می دهد.
- به عبارت دقیق تر جهش ها تغییراتی در توالی DNA هستند که می توانند در هر ناحیه ای از DNA رخ دهند.
- در هر یک از فعالیت های سلولی نظیر فرایندهای همانندسازی، رونویسی، ترجمه، ترکیب مجدد یا نوترکیبی کروموزوم ها، بروز و ظهور اطلاعات ژنتیکی احتمال خطا و اشتباه و ایجاد موتاسیون وجود دارد. در موارد نادر ممکن است تغییر خود به خود در قسمتی از DNA رخ دهد.
- این تغییرات که جهش نامیده می شوند، ممکن است تغییر در رمز ایجاد نموده و به تولید یک پروتئین ناقص منجر شوند.
- گاهی نتیجه خالص به صورت تغییری در ظاهر فرد و یا تغییری در یک شاخص قابل اندازه گیری موجود زنده، به نام ویژگی یا صفت مشاهده می شود.
- طی فرایند جهش، یک ژن ممکن است به دو یا چند شکل متفاوت به نام آلل تغییر یابد.

انواع جهش ژنی

جایگزینی

TGG **C** AG



TGG **T** AG

درج

TGGCAG

TGG **TAT** CAG

حذف

TGG ~~C~~ AG

TGGG



تاریخچه بیوانفورماتیک

- پایه‌های بیوانفورماتیک در اوایل دهه ۱۹۶۰ با استفاده از روش‌های محاسباتی برای تجزیه و تحلیل توالی پروتئین به ویژه، مونتاژ توالی‌های جدید، پایگاه داده‌های توالی بیولوژیکی و مدل‌های جایگزینی گذاشته شد.
- بعداً از آن، تجزیه و تحلیل DNA نیز به دلیل پیشرفت موازی روش‌های زیست‌شناسی مولکولی (که دستکاری آسان‌تر DNA و همچنین توالی آن را فراهم می‌کند) و همچنین پیشرفت در علوم کامپیوتر و ظهور رایانه‌های کوچک (که قدرتمندتر بوده و دارای نرم افزارهای جدیدتر و مناسب‌تر برای انجام کارهای بیوانفورماتیک بود) ساده‌تر شد.
- اصطلاح بیوانفورماتیک، برای اولین بار در دهه ۱۹۹۰ مطرح شد. در اصل، این کار با مدیریت و تجزیه و تحلیل داده‌های مربوط به توالی DNA، RNA و پروتئین بود.¹⁹

تاریخچه بیوانفورماتیک

- اولین توالی آمینواسیدی مربوط به پروتئین انسولین شامل ۱۱۰ اسید آمینه در سال ۱۹۵۱ تعیین گردید.
- اولین توالی پروتئین گزارش شده مربوط به انسولین گاوی در سال ۱۹۵۶ بود که از ۵۱ اسید آمینه تشکیل شده بود.
- واحد ساختمانی پروتئین=اسید آمینه یا آمینواسید
- آمینواسید=مولکولهای آلی پیچیده شامل اتمهای کربن، هیدروژن، اکسیژن، نیتروژن و گوگرد
- زنجیره پروتئین متشکل از ترکیب و ردیف صحیح آمینواسیدهای تشکیل دهنده

تاریخچه بیوانفورماتیک

- تقریباً یک دهه بعد، اولین توالی یابی اسید نوکلئیکی گزارش شد که مربوط به tRNA آلانین مخمر با ۷۷ باز بود.
- در دهه ۱۹۵۰ DNA به عنوان عنصر سازنده ژنها یک ماکرومولکول بزرگ به صورت مولکول زنجیره ای بلند و مارپیچ دوتایی پیچ خورده است.
- این زنجیره با جفت شدن دو جزء از چهار جزء سازنده به نام نوکلئوتید به هم متصل می شوند.
- در دهه ۱۹۷۰ تعیین توالی مولکولهای DNA=انقلاب بزرگ
- سانگر جایزه نوبل دریافت کرد، زیرا الفبای کوچک توالی DNA (۴ نوکلئوتید در مقایسه با ۲۰ آمینواسید) امکان خواندن سریعتر و راحت تر را فراهم نمود.
- ¹نرخ جهانی تعیین توالی DNA سریعتر از نرخ تعیین توالی پروتئین است.

تاریخچه بیوانفورماتیک

- از آنجایی که داده‌های بیولوژیکی با سرعت بی سابقه‌ای تولید می‌شوند، مدیریت و تفسیر آن‌ها به بیوانفورماتیک نیازمند است.
- بنابراین، بیوانفورماتیک اکنون انواع مختلفی از داده‌های بیولوژیکی را نیز شامل می‌شود.
- اولین پایگاه اطلاعاتی بیوانفورماتیک/بیولوژیکی چند سال پس از در دسترس بودن اولین توالی‌های پروتئینی ساخته شد.
- مارگارت دیلوف (۱۹۲۵-۱۹۸۳) که یک شیمی-فیزیک دان آمریکایی بود تمام داده‌های توالی‌های موجود را برای ایجاد اولین پایگاه داده بیوانفورماتیک جمع آوری کرد.
- پروتئین Data Bank در سال ۱۹۷۲ با جمع آوری ده پروتئین کریستالوگرافی شده با اشعه ایکس توالی‌یابی شد و در سال ۱۹۸۷ تأسیس اولین پایگاه داده توالی پروتئین به نام The SWISS PROT آغاز شد.



در این تصویر COMPROTEIN، اولین نرم افزار بیوانفورماتیک نشان داده شده است. (الف) یک فریم مرکزی IBM 7090 که برای اجرای آن COMPROTEIN ساخته شده است. (ب) کارت پانچ حاوی یک خط کد FORTRAN زبان COMPROTEIN با آن نوشته شده است. (ج) کد منبع کامل یک برنامه در کارت‌های پانچ. (د) یک نمای کلی از ورودی COMPROTEIN یعنی توالی پپتید Edman و خروجی (یک توالی پروتئین کلی).

داده های بیوانفورماتیکی

- داده های کلاسیک بیوانفورماتیک شامل توالی DNA ژن ها یا ژنوم های کامل، توالی اسیدهای آمینه پروتئین ها، ساختارهای سه بعدی پروتئین ها و اسیدهای نوکلئیک و مجموعه های اسید نوکلئیک-پروتئین است.
- جریان داده های فرعی یا omics عبارتند از:
- Transcriptomics که به معنی الگوی سنتز RNA از DNA است،
- Proteomics که توزیع پروتئین در سلول ها بوده،
- Interactomics که الگوهای فعل و انفعالات پروتئین-پروتئین و پروتئین-اسید نوکلئیک
- Metabolomics که طبیعت و الگوهای ترافیکی تبدیل مولکول های کوچک توسط مسیرهای بیوشیمیایی فعال در سلول ها تعریف شده است.
- در هر حالت علاقه به دستیابی به داده های دقیق و جامع برای انواع خاصی از سلول ها و شناسایی الگوهای تغییر در داده ها وجود دارد.

داده های بیوانفورماتیکی

- به عنوان مثال، بسته به نوع سلول، زمان جمع آوری اطلاعات (در طول چرخه سلولی یا تغییرات روزانه، فصلی یا سالانه)، مرحله رشد و شرایط مختلف محیطی ممکن است داده‌ها در نوسان باشند.
- متاژنومیکس و متاپروتومیکس این اندازه گیری‌ها را به شرح جامعی از ارگانیسم‌های موجود در یک نمونه گرفته شده از محیط مانند یک سطل آب اقیانوس یا یک نمونه خاک گسترش می‌دهد.
- بیوانفورماتیک با شتاب زیاد فرآیندهای تولید داده در زیست شناسی را به جلو می‌برد.
- روش‌های تعیین توالی ژنوم شاید بارزترین اثرات آن را نشان دهند.
- در سال ۱۹۹۹ بایگانی توالی اسید نوکلئیک حاوی ۳/۵ میلیارد نوکلئوتید بود که کمی بیشتر از طول یک ژنوم انسانی است.
- یک دهه بعد این اطلاعات حاوی بیش از ۲۸۳ میلیارد نوکلئوتید یعنی طولی به اندازه ۹۵ عدد ژنوم انسانی بودند.

ذخیره و بازیابی داده ها

- در بیوانفورماتیک، از بانک‌های داده برای ذخیره و سازماندهی داده‌ها استفاده می‌شوند
- محققان توالی DNA و RNA بسیاری از این موجودات را از مقالات علمی و پروژه‌های ژنوم جمع‌آوری می‌کنند.
- بسیاری از پایگاه‌های اطلاعاتی در اختیار کنسرسیوم‌های بین‌المللی است.
- به عنوان مثال، یک کمیته مشورتی متشکل از اعضای بانک اطلاعاتی توالی نوکلئوتید آزمایشگاه زیست‌شناسی مولکولی اروپا (EMBL-Bank) در انگلستان، بانک داده DNA ژاپن (DDBJ) و GenBank از مرکز ملی اطلاعات بیوتکنولوژی در ایالات متحده (NCBI) بر همکاری بین‌المللی بانک اطلاعات توالی نوکلئوتیدی (INSDC) نظارت می‌کنند.

اهداف بیوانفورماتیک

- توسعه الگوریتم‌های کارآمد برای اندازه‌گیری تشابه توالی،
- گسترش داده‌های تجربی توسط پیش‌بینی‌ها

اهداف بیوانفورماتیک

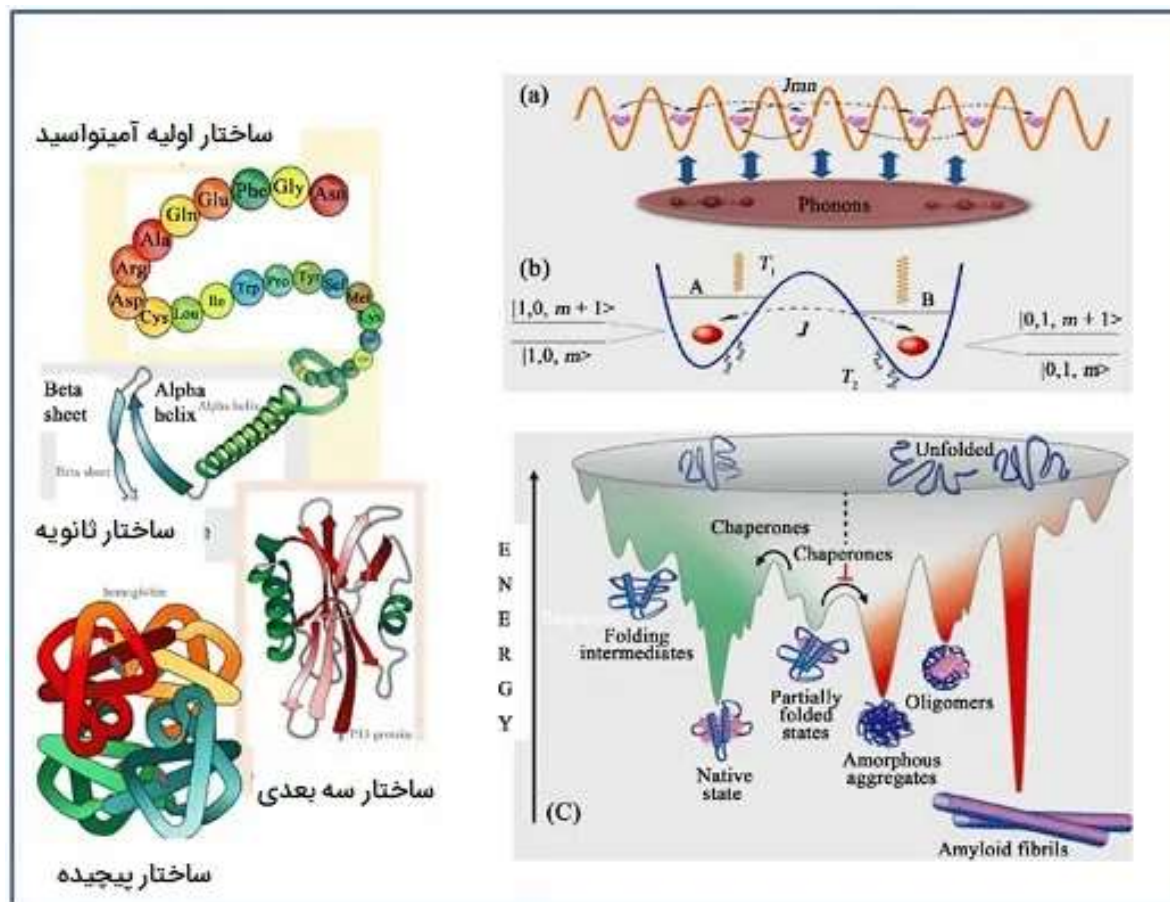
- توسعه الگوریتم‌های کارآمد برای اندازه گیری تشابه توالی، هدف مهم بیوانفورماتیک است.
- الگوریتم Needleman - Wunsch، که مبتنی بر برنامه نویسی پویا است، یافتن چینش بهینه جفت توالی‌ها را تضمین می‌کند.
- این الگوریتم اساساً یک مسئله بزرگ (دنباله کامل) را به مجموعه‌ای از مسائل کوچک‌تر (بخش‌های توالی کوتاه) تقسیم می‌کند و از راه حل‌های مسئله‌های کوچک‌تر برای راه حل مسئله بزرگ استفاده می‌کند.
- شباهت در توالی‌ها در یک ماتریکس امتیاز بندی می‌شوند و الگوریتم امکان تشخیص شکاف‌ها در «تراز بندی توالی‌ها» (Sequence Alignment) را فراهم می‌کند.
- اگرچه الگوریتم Needleman - Wunsch موثر است، اما به عنوان کاوش یک پایگاه داده توالی‌یابی بزرگ بسیار کند است.

اهداف بیوانفورماتیک

- بنابراین، توجه زیادی به یافتن الگوریتم‌های بازیابی سریع اطلاعات شده است که می‌توانند با مقادیر زیادی از داده‌ها در بایگانی‌ها سر و کار داشته باشند.
- برنامه «بلاست» (BLAST یا Basic Local Alignment Search Tool یکی از آن‌ها است.
- یک بخش پیشرفته از BLAST، شناخته شده به عنوان موقعیت خاص تکرار شونده یا PSIBLAST که برگرفته از position-specific iterated BLAST است، باعث استفاده از الگوهای حفاظت‌شده در توالی‌های مرتبط می‌شود و ترکیبی از سرعت بالای BLAST همراه حساسیت بسیار بالا برای یافتن توالی‌های مرتبط است.

اهداف بیوانفورماتیک

- هدف دیگر بیوانفورماتیک، گسترش داده‌های تجربی توسط پیش‌بینی‌ها است.
- یک هدف اساسی از زیست‌شناسی محاسباتی، پیش‌بینی ساختار پروتئین از یک توالی اسید آمینه است.
- پیشرفت در توسعه روش‌های پیش‌بینی «حالات قرارگیری فضایی» (Folding) پروتئین توسط برنامه‌های بیوانفورماتیک انجام شده است.
- از بیوانفورماتیک برای پیش‌بینی تعاملات بین پروتئین‌ها با توجه به ساختارهای فردی آمینواسیدها نیز استفاده می‌شود. این نوع پژوهش به عنوان داکینگ پروتئین‌ها شناخته می‌شود. مجموعه‌های پروتئین-پروتئین مکمل بودن خوبی با هم در شکل سطحی و قطبیت دارند و عمدتاً توسط فعل و انفعالات ضعیف مانند پیوندهای سطح آبگریز، پیوندهای هیدروژنی و نیروهای وان در والس تثبیت می‌شوند.
- برنامه‌های نرم افزاری رایانه‌ای این تعاملات را برای پیش‌بینی رابطه فضایی مطلوب بین زیرواحدهای متصل به هم شبیه‌سازی می‌کند.
- یک چالشی که می‌تواند کاربردهای درمانی مهمی داشته باشد، طراحی آنتی‌بادی‌هایی است که با میل زیاد به پروتئین هدف متصل شوند.
- در ابتدا، بسیاری از تحقیقات بیوانفورماتیک، تمرکز نسبتاً کمی بر الگوریتم‌هایی برای تجزیه و تحلیل انواع خاصی از داده‌ها، مانند توالی ژن‌ها یا ساختارهای پروتئینی داشته است. با این حال، اکنون، اهداف بیوانفورماتیکی یکپارچه هستند و هدف آن‌ها این است که دریابند چگونه می‌توان از ترکیب انواع مختلف داده‌ها برای درک بهتر پدیده‌های طبیعی، از جمله ارگاناسم‌ها و بیماری‌ها استفاده کرد.



در این تصویر شبیه سازی ساختار پروتئین هموگلوبین و بررسی سطح انرژی های آن با استفاده از نرم افزارهای بیوانفورماتیکی نشان داده شده است.

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G