ӘЛ-ФАРАБИ АТЫНДАҒЫ ҚАЗАҚ ҰЛТТЫҚ УНИВЕРСИТЕТІ
БИОЛОГИЯ ЖӘНЕ БИОТЕХНОЛОГИЯ ФАКУЛЬТЕТІ
МОЛЕКУЛАЛЫҚ БИОЛОГИЯ ЖӘНЕ ГЕНЕТИКА КАФЕДРАСЫ

Пән атауы: Қолданбалы генетика және геномика

Дәріс 12

Геномдық деректерді биоинформатикалық талдау. Big Data және жасанды интеллект қолданысы

Лектор: к.б.н., Алтыбаева Н.А.

Дәрістің мақсаты: Геномдық деректерді биоинформатикалық талдаудың негізгі әдістерін, үлкен деректерді (Big Data) өңдеу технологияларын және жасанды интеллекттің генетикалық, геномдық зерттеулердегі қолданылу мүмкіндіктерін қарастыру.

Дәрістің жоспары:

1.Биоинформатиканың негізгі ұғымдары және геномдық зерттеулердегі маңызы.
2.Геномдық деректердің түрлері және оларды өңдеу кезеңдері.
3.ДНҚ тізбектерін салыстыру және генетикалық вариацияларды анықтау әдістері.
4.Big Data технологияларының геномдық деректерді талдаудағы рөлі.
5.Машиналық оқыту және жасанды интеллект әдістерінің геномикада қолданылуы.
6.Биоинформатика мен жасанды интеллекттің медициналық диагностикадағы және дербестендірілген медицинадағы маңызы.

Оқыту нәтижелері

Дәрісті меңгергеннен кейін докторант:

1.Биоинформатиканың негізгі ұғымдарын, қалыптасу тарихын және геномдық зерттеулердегі маңызын түсіндіре алады.
2.Геномдық деректердің негізгі түрлерін және оларды алу технологияларының ерекшеліктерін сипаттай алады.
3.ДНҚ секвенирлеу, генотиптеу, микрочип және RNA-seq әдістерінің қолданылуын ажыратады.
4.Геномды жинақтау, тізбектерді салыстыру және генетикалық вариацияларды анықтаудың негізгі принциптерін түсіндіре алады.
5.Big Data ұғымын және үлкен көлемдегі геномдық деректердің ерекшеліктерін сипаттай алады.
6.Бұлттық есептеу мен деректерді сақтау технологияларының биоинформатикалық талдаудағы рөлін түсіндіре алады.
7.GenBank, NCBI және BLAST ресурстарының геномдық зерттеулердегі қолданылуын сипаттай алады.
8.Машиналық оқыту мен жасанды интеллект әдістерінің генетикалық деректерді талдаудағы мүмкіндіктерін салыстыра алады.
9.Жасанды интеллекттің ауруларға генетикалық бейімділікті бағалау және дербестендірілген медицинадағы маңызын талдай алады.

Геном және геномдық деректер ұғымы

Геном - тірі организмнің барлық генетикалық ақпаратын қамтитын толық тұқым қуалау материалы. Яғни, ол жасушаның организмнің ДНҚ (дезоксирибонуклеин қышқылы) немесе кейбір вирустарда РНҚ (рибонуклеин қышқылы) түрінде сақталатын барлық гендерінің жиынтығы болып табылады.



Геном құрамына тек ақуыз синтезін бақылайтын гендер ғана емес, сонымен қатар реттеуші аймақтар, интрондар, кодталмайтын РНҚ, және қайталанатын ДНҚ тізбектері де кіреді. Геном организмнің барлық белгілері мен қасиеттерін - сыртқы пішіннен бастап, зат алмасу ерекшеліктеріне дейін - анықтайды. Адам геномы шамамен:

- 3,2 миллиард нуклеотидтен
- 20–25 мың белсенді геннен тұрады.



Геном ұғымын алғаш рет **1920** жылы неміс ғалымы **Ганс Винклер** (Hans Winkler) енгізген. Ол «ген» және «хромосома» сөздерін біріктіріп, «геном» терминін ұсынды.

Геномдық деректер

Геномдық деректер – бұл ДНҚ мен РНҚ секвенирлеу (тізбектеу) нәтижесінде алынған сандық және мәтіндік ақпараттардың жиынтығы.Басқаша айтқанда, бұл – организмнің генетикалық кодын сипаттайтын биологиялық және сандық деректер.Геномдық деректер бірнеше түрге бөлінеді:

Геномдық деректер арнайы секвенирлеу технологиялары арқылы алынады:

1. Sanger sequencing – классикалық әдіс;
2. Next Generation Sequencing (NGS) – жоғары өнімді жаңа буын секвенирлеу;
3. Whole Genome Sequencing (WGS) – тұтас геномды анықтау;
4. RNA-Seq – транскриптомды анықтау;
5. Metagenomic sequencing – микроорганизмдер қауымдастығының геномдарын анықтау.





1 Секвенирлеу деректері (DNA/RNA sequencing data) – ДНҚ тізбектерінің нақты нуклеотидтік реттілігі.

2 Экспрессия деректері (gene expression data) – гендердің белсенділік деңгейі туралы мәлімет.

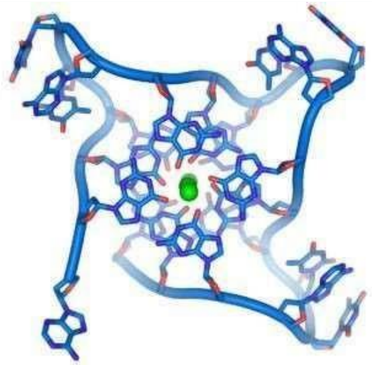
3 Эпигеномдық деректер – ДНҚ метилденуі мен гистон модификациясы сияқты реттеу тетіктері туралы ақпарат.

4 Варианттық деректер (variant data) – мутациялар мен генетикалық айырмашылықтарды көрсететін мәліметтер (мысалы, SNP, indel).

Геномдық зерттеу әдістері — тірі ағзалардың ДНҚ құрамындағы барлық генетикалық ақпаратты анықтап, талдауға мүмкіндік беретін заманауи технологиялар жиынтығы.
Бұл әдістер арқылы ғалымдар ген құрылымын, қызметін және тұқымқуалаушылық ерекшеліктерін зерттейді. Негізгі әдістер

Негізгі әдістер

1. ДНҚ секвенирлеу (Genome Sequencing)
   - Геномдағы нуклеотидтердің толық ретін анықтайды.
   - Қазіргі таңда Next Generation Sequencing (NGS) кең қолданылады

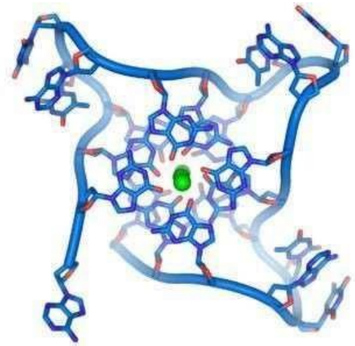


2. Генотиптеу (Genotyping)
   - Адам немесе ағзаның жеке генетикалық айырмашылықтарын табу.
   - Ауруларға бейімділікті анықтау үшін қолданылады.

3. Микрочип технологиясы (DNA Microarray)
   - Мыңдаған ген экспрессиясын бір уақытта талдау.
   - Дәрі-дәрмек әсерін және ген белсенділігін зерттеу.

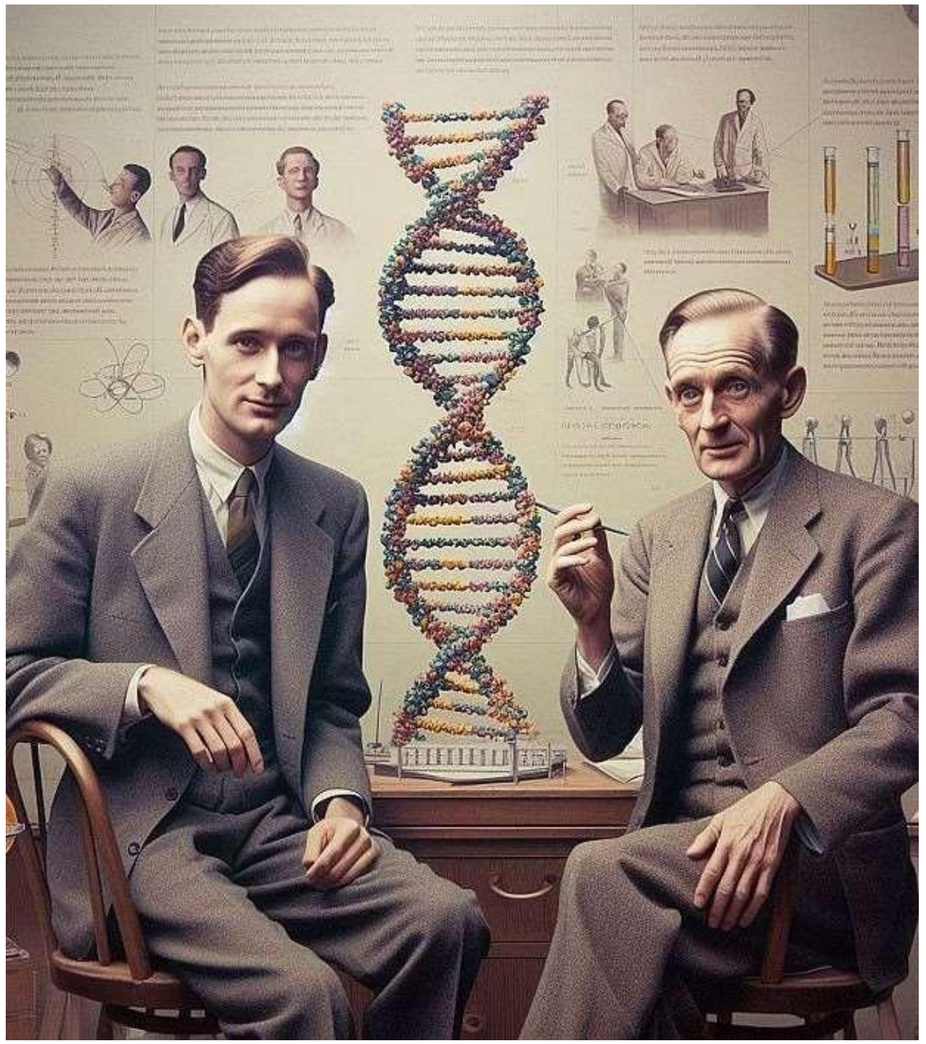
4. Транскриптомдық талдау (RNA-seq)
   - Белсенді жұмыс істейтін гендерді анықтайды.

5. Биоинформатикалық
   - талдау Алынған деректерді компьютерлік бағдарламалармен өңдеп, статистикалық талдау жасайды.



Биоинформатика ғылымының қалыптасуы және маңызы

Биоинформатика – биологиялық деректерді жинақтау, сақтау, талдау және түсіндірумен айналысатын ғылым саласы. Ол информатика, молекулалық биология, генетика, математика және статистика ғылымдарының тоғысында қалыптасқан көпсалалы бағыт болып табылады.Бұл ғылымның негізгі мақсаты – биологиялық молекулалар (ДНҚ, РНҚ, ақуыздар) туралы үлкен көлемдегі ақпаратты сандық әдістермен өңдеу және сол деректер негізінде биологиялық заңдылықтарды анықтау.Биоинформатика ғылымының бастауын XX ғасырдың ортасынан, яғни ДНҚ құрылымының ашылуымен (**1953** жылы **Джеймс Уотсон** мен **Фрэнсис Крик**) байланыстыруға болады. Сол кезеңнен бастап биологиялық ақпаратты есептеу арқылы талдауға деген қажеттілік арта түсті.



Қазіргі таңда биоинформатикасыз молекулалық биологияны елестету мүмкін емес. Ол:

Big Data талдау,жасанды интеллект және машиналық оқыту,бұлттық есептеу (cloud computing),математикалық модельдеу әдістерін кеңінен қолданады.

Биоинформатиканың даму кезеңдері

1950–1970 жж. – Негіз салу кезеңі:

1953 ж. ДНҚ-ның қос оралма құрылымы анықталды.1960 жылдары ақуыздардың аминқышқылдық тізбектерін анықтау басталды.1965 ж. алғашқы биологиялық мәліметтерді компьютерге енгізу әрекеттері жасалды.

1980–1990 жж. – Деректер базаларының қалыптасу кезеңі:

EMBL, GenBank және SwissProt сияқты алғашқы генетикалық мәліметтер банктері құрылды.1982 ж. АҚШ-та GenBank іске қосылды – онда ДНҚ тізбектерінің алғашқы электронды базасы жинақталды.1990 ж. Human Genome Project (Адам геном жобасы) басталып, биоинформатиканың стратегиялық рөлі артты.

2000 жылдан бастап – Жаңа буын секвенирлеу және Big Data дәуірі:

ДНҚ мен РНҚ тізбектерін жылдам анықтайтын Next Generation Sequencing (NGS) технологиялары пайда болды.Биоинформатикада үлкен деректер (Big Data) және жасанды интеллект (AI) әдістері қолданыла бастады.Мысалы, DeepMind компаниясының AlphaFold жүйесі ақуыз құрылымын болжауда үлкен жетістікке жетті.

Биоинформатика бірнеше зерттеу саласын қамтиды:



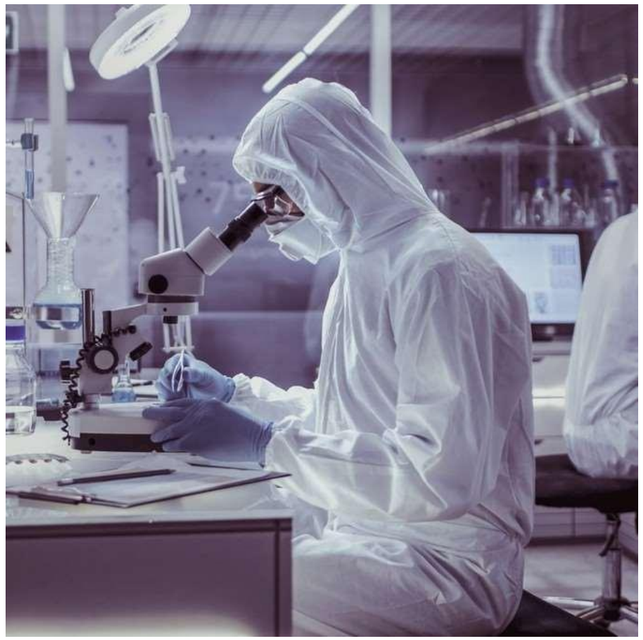
1.Медицина саласында: ауруларының генетикалық себептерін анықтауға, диагностика мен емдеуді дараландыруға (personalized medicine) дәрілік препарат-тарды дәл бағытта жасауға (targeted therapy) мүмкіндік береді.



2.Ауыл шаруашылығында: Жоғары өнімді және ауруға төзімді өсімдік пен жануар тұқымдарын шығару.Геномдық селекция және биотехно -логиялық өнімдер өндіру.



3.Экология мен микробиологияда: Қоршаған ортадағы микроорганизмдердің рөлін анықтау. Ластанған аймақтарды тазалауға арналған микробиологиялық шешімдер жасау.



4.Ғылыми зерттеулерде: Эволюциялық қатынастарды анықтау, филогенетикалық ағаштар құру;Геномдардың салыстырмалы талдауы арқылы жаңа биологиялық заңдылықтарды табу;

Геномдық деректерді талдаудың негізгі бағыттары

Қазіргі таңда геномдық зерттеулердің басты мақсаты – организмнің тұқым қуалау ақпаратын толық талдап, оның құрылымы мен қызметін түсіну. **Секвенирлеу** (тізбектеу) технологияларының дамуы нәтижесінде ғалымдар өте көп көлемде геномдық деректерге қол жеткізді. Бұл ақпаратты тиімді өңдеу мен түсіндіру үшін биоинформатикалық талдау бағыттары қалыптасты. **Геномдық деректерді талдау** – бұл ДНҚ немесе РНҚ тізбектері жөніндегі ақпаратты компьютерлік әдістермен өңдеу, салыстыру және биологиялық мәнін анықтау процесі. Геномды жинақтау – секвенирлеу нәтижесінде алынған қысқа ДНҚ фрагменттерін (ридтерді) біріктіріп, толық геном тізбегін қалпына келтіру процесі. Бұл екі тәсіл арқылы жүзеге асырылады:

De novo assembly – бұрын зерттелмеген организмнің геномын нөлден бастап құрастыру;





Reference-based assembly – белгілі референстік геномнен салыстыра отырып жинақтау. Жинақтау нәтижесінде алынған толық геном жаңа гендерді, мутацияларды және генетикалық ерекшеліктерді анықтауға мүмкіндік береді.

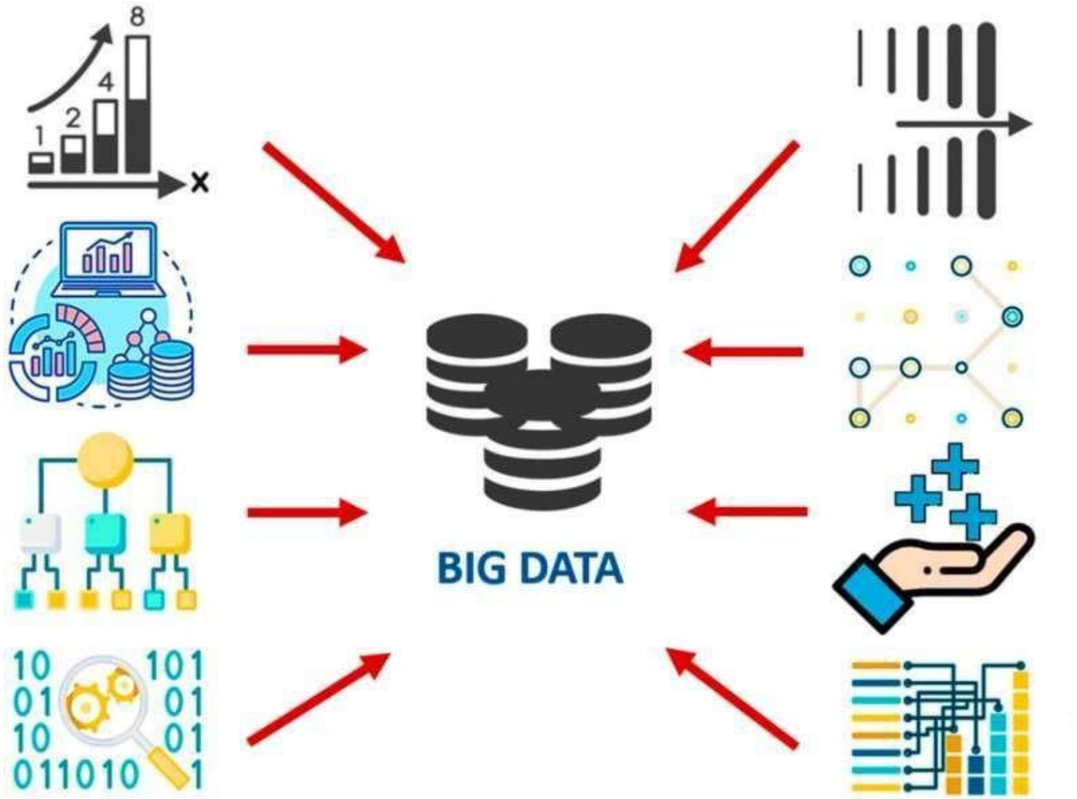
Big Data ұғымы және оның биологиядағы орны

Big Data (үлкен деректер) – бұл көлемі, жылдамдығы және алуан түрлілігі бойынша дәстүрлі әдістермен өңдеуге келмейтін аса ауқымды ақпараттар жиынтығы.Ол тек көлеммен ғана емес, сонымен қатар жинау, сақтау, талдау және визуализациялау тәсілдерінің күрделілігімен ерекшеленеді.XXI ғасырда биология ғылымы сандық (цифрлық) сипатқа ие бола бастады.Жаңа технологиялар арқылы зерттеушілер геном, транскриптом, протеом, метаболом, эпигеном сияқты биомолекулалық деңгейдегі деректерді жинай алады.



Volume (көлемі)	деректердің өте үлкен мөлшері (терабайттардан петабайттарға дейін);
Velocity (жылдамдығы)	деректердің үздіксіз және тез жиналуы;
Variety (әртүрлілігі)	деректердің түрлі форматта болуы (мәтін, сурет, геномдық тізбек, сигнал, т.б.)
Veracity (сенімділігі)	алынған ақпараттың дұрыстығы мен сапасын бақылау;
Value (құндылығы)	деректерден нақты пайда мен мағына алу қабілеті.

Big Data технологияларының биологиядағы қолданылуы

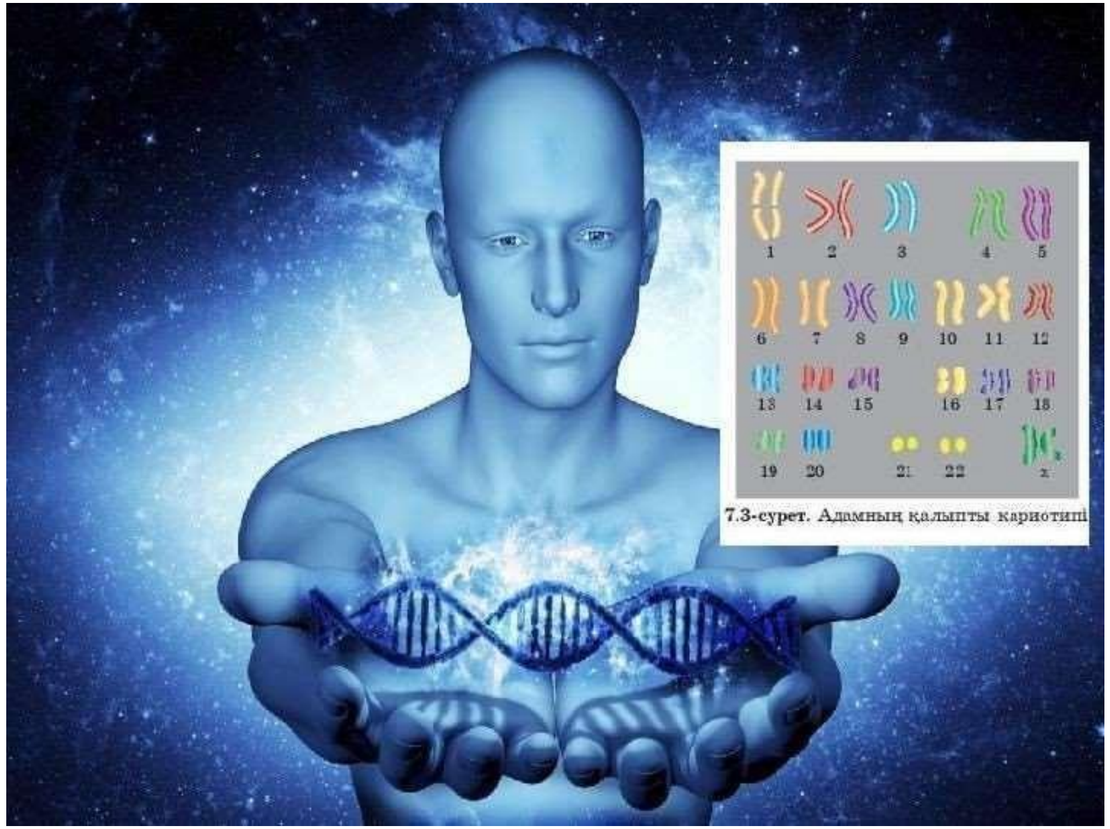


Big Data биоинформатиканың негізгі құрамдас бөлігіне айналды.Биоинформатика деректерді жинақтау, сақтау және талдау алгоритмдерін жасаса, Big Data осы процестерді жылдам әрі масштабты түрде жүзеге асыруға мүмкіндік береді.

Big Data технологиялары биологияда үлкен көлемді және күрделі биологиялық деректерді талдауға мүмкіндік береді. Олар геномдық зерттеулерде мутацияларды анықтау, ауруларға бейімділікті болжау үшін қолданылады. Медицинада Big Data клиникалық ақпаратты өңдеп, жеке (персональды) емдеу тәсілдерін жасауға көмектеседі. Микробиом және экожүйелерді зерттеуде түр құрамын, экологиялық өзгерістерді жылдам бағалауға мүмкіндік береді. Фармакологияда жаңа дәрілерді модельдеу және олардың уыттылығын алдын ала болжау жүзеге асады. Нейробиологияда ми сигналдарын талдап, нейрондық құруға қолданылады. Жалпы, Big Data биологиядағы зерттеулердің дәлдігін арттырып, жылдам әрі тиімді нәтижелер алуға жағдай жасайды.

Геномдық деректерді Big Data ауқымындағы ерекшеліктері

Геномдық деректер — бұл адамның немесе басқа организмдердің толық ДНҚ тізбектерінен алынған ақпарат. Оларды талдау кезінде көлемі өте үлкен деректермен жұмыс жасалады, сондықтан оларды Big Data ауқымында қарастыру қажет.1. Көлемі (Volume): Геномдық деректер өте үлкен. Мысалы, бір адамның толық геномы шамамен 3 миллиард нуклеотидтен тұрады, ал геномдарды мыңдаған адамнан жинаса, деректердің көлемі терабайттармен есептеледі.



7.3-сурет. Адамның қалыпты кариотипі

2. Әртүрлілігі (Variety): Геномдық деректер әртүрлі форматта болады: ДНҚ тізбектері (FASTA, FASTQ), гендердің экспрессиясы (RNA-seq), эпигенетикалық деректер, вариациялар (SNP, CNV) және т.б. 3.Жылдамдық (Velocity): Жаңа секвенирлеу технологиялары (мысалы, NGS — Next Generation Sequencing) деректерді өте жылдам өндіреді. Сондықтан оларды өңдеу және сақтау жылдамдығы маңызды. 4. Шынайылық пен дәлдік (Veracity): Геномдық деректерде қателер болуы мүмкін (секвенирлеу қателері, техникалық шудар). Сондықтан сапаны бақылау және дұрыстау қажет.5. Құнды ақпарат (Value): Бұл деректер ауруларды зерттеуде, дәрі-дәрмек жасауда, популяциялық генетика мен жеке медицинада маңызды.

Деректер орталықтарын сақтау және басқарудағы бұлттық технологиялар

1. Бұлттық технологиялардың рөлі

Бұлттық сақтау (Cloud Storage) — деректерді физикалық серверлерге емес, интернет арқылы сақтау;

Бұлттық басқару (Cloud Management) — деректер орталықтарын қашықтан бақылау, ресурстарды тиімді бөлу және автоматтандыру;Танымал платформалар: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform.



2. Деректер орталықтарын сақтау ерекшеліктері

Физикалық шектеу жоқ: Дәстүрлі серверлермен салыстырғанда масштабталу мүмкіндігі шексіз;

Қауіпсіздік: Деректер шифрланады, резервтік көшіру жасалады және рұқсаттар бақыланады;

Орташа мәлімет: Орташа деректер орталығы 1–10 ПБ деректерді өңдей алады (2023 жылы).3. Бұлттық технологияларды пайдалану артықшылықтары Масштабталу: Қосымша серверсіз сақтау көлемін көбейту.Икемділік: Қызметтерді нақты қажеттілікке қарай өзгерту. Шығынды азайту: Серверді сатып алу мен күтіп ұстаудан тиімді.Қолжетімділік: Әлемнің кез келген нүктесінен деректерге қол жеткізу.Автоматтандыру: Ресурстарды басқару, резервтік көшіру, жаңарту процестері автоматты түрде жүзеге асады.

Биоинформатикалық бағдарламалар мен платформалар (GenBank, NCBI, BLAST және т.б.)

1. GenBank

• Сипаттама: Дүниежүзілік ДНҚ және РНҚ тізбектерінің ашық мәліметтер базасы.Мақсаты: Ғалымдарға генетикалық деректерді сақтау, бөлісу және зерттеу мүмкіндігін беру.Мәліметтер көлемі: 2025 жылы GenBank-та 250 миллионнан астам тіркелген ДНҚ/RNA тізбектері бар. Қолданылуы: Гендерді сәйкестендіру, эволюциялық зерттеулер, молекулалық биологиядағы тәжірибелер.

2. NCBI (National Center for Biotechnology Information)

Сипаттама: АҚШ-тағы биоинформатика орталығы, көптеген мәліметтер базаларын біріктіреді (GenBank, PubMed, Protein Database).Мақсаты: Геномдық, протеомдық, әдеби және клиникалық деректерді біріктіру.Қолданылуы: Деректерді іздеу, талдау және визуализация жасау, ғылыми зерттеулерді қолдау.

3. BLAST (Basic Local Alignment Search Tool)

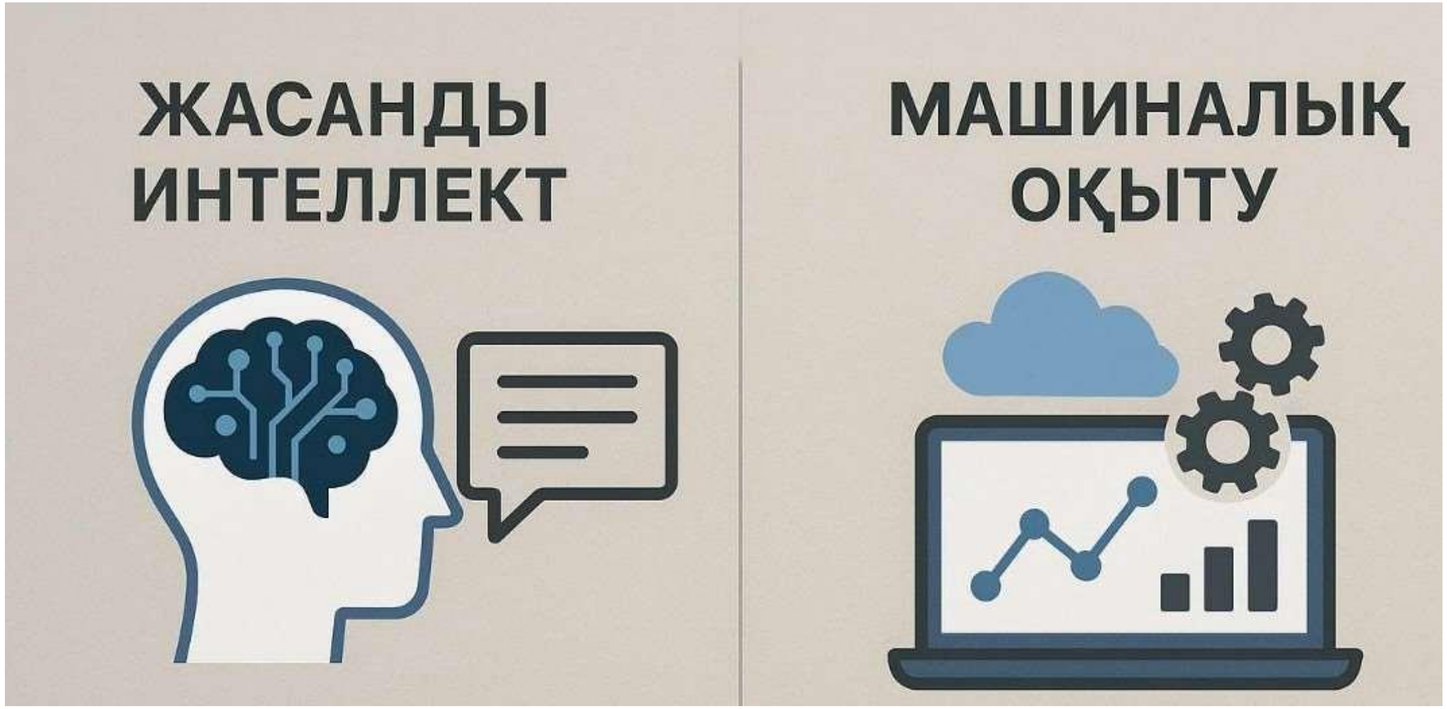
Сипаттама: Генетикалық тізбектерді салыстыруға арналған бағдарлама.Мақсаты: Белгілі бір ДНҚ/РНҚ немесе ақуыз тізбегін мәліметтер базасында іздеп, ұқсас тізбектерді анықтау.Қолданылуы: Гендердің функциясын болжау, эволюциялық байланыстарды зерттеу, жаңа организмдердің геномын талдау.

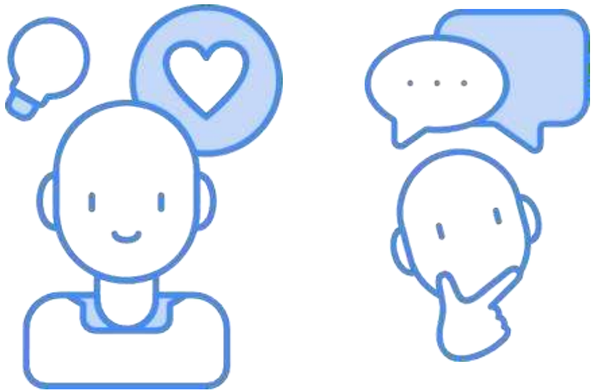
4. Басқа платформалар

Ensembl: Геномдық деректерді интеграциялайды, гендердің құрылымын және олардың варианттарын көрсетеді.UniProt: Ақуыз тізбектері мен олардың функцияларын сипаттайтын база;.KEGG: Метаболизм жолдары мен биологиялық процестерді визуалдау платформасы.

Жасанды интеллект және машиналық оқыту туралы түсінік

Жасанды интеллект (ЖИ) – бұл компьютерлік жүйелердің адам санасына тән ойлау, үйрену, шешім қабылдау, сөйлеу және қорытынды жасау қабілеттерін модельдеу технологиясы. Яғни, ЖИ адамның интеллектуалды әрекеттерін орындай алатын бағдарламалар мен алгоритмдерді қамтиды. Мысалы, сөйлеуді тану, мәтінді аудару, суретті талдау, деректер негізінде болжау сияқты қызметтерді атқарады.

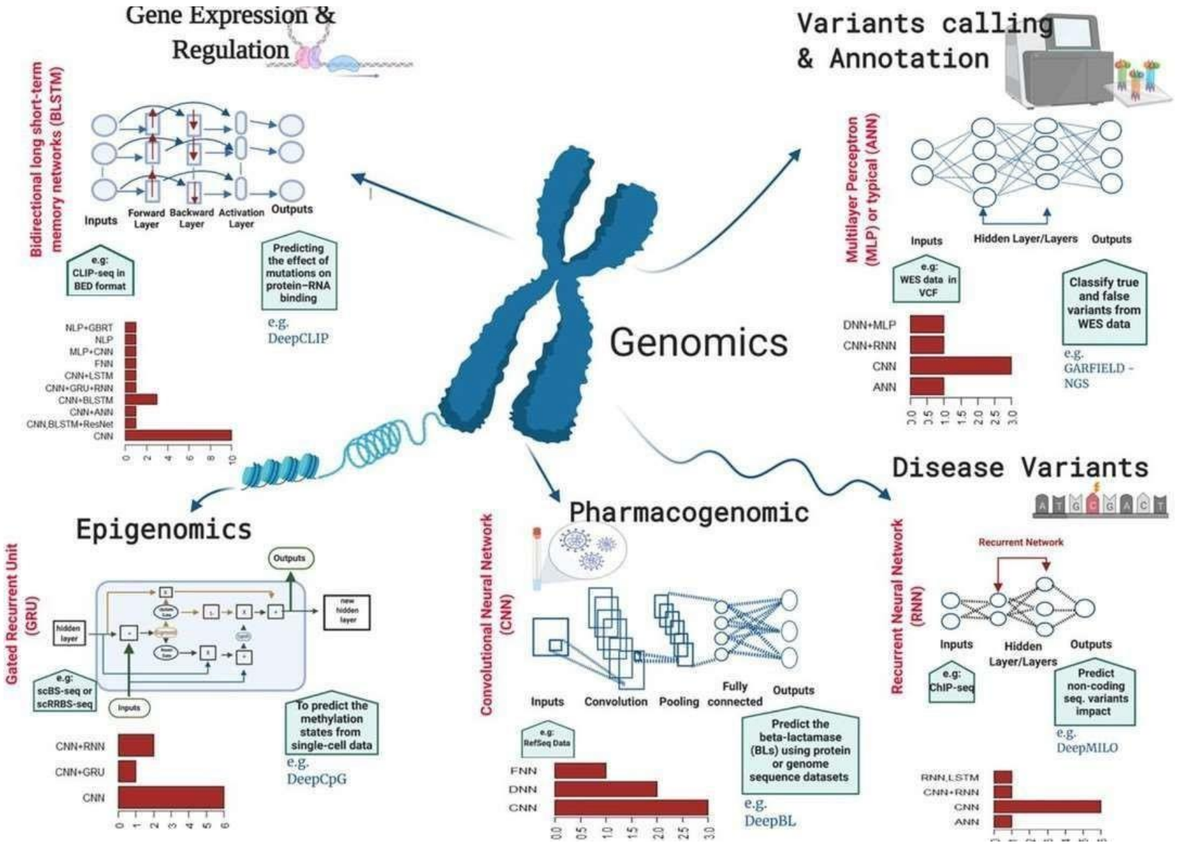




Машиналық оқыту (МО) – жасанды интеллектінің бір бағыты. Ол компьютерді нақты бағдарламаламай-ақ, деректер арқылы өздігінен үйренуге және тәжірибеден қорытынды шығаруға үйретеді. Машиналық оқытуда жүйе үлкен көлемдегі деректерді талдап, белгілі бір заңдылықтарды табады және сол заңдылықтарға сүйеніп жаңа жағдайларда шешім қабылдай алады.

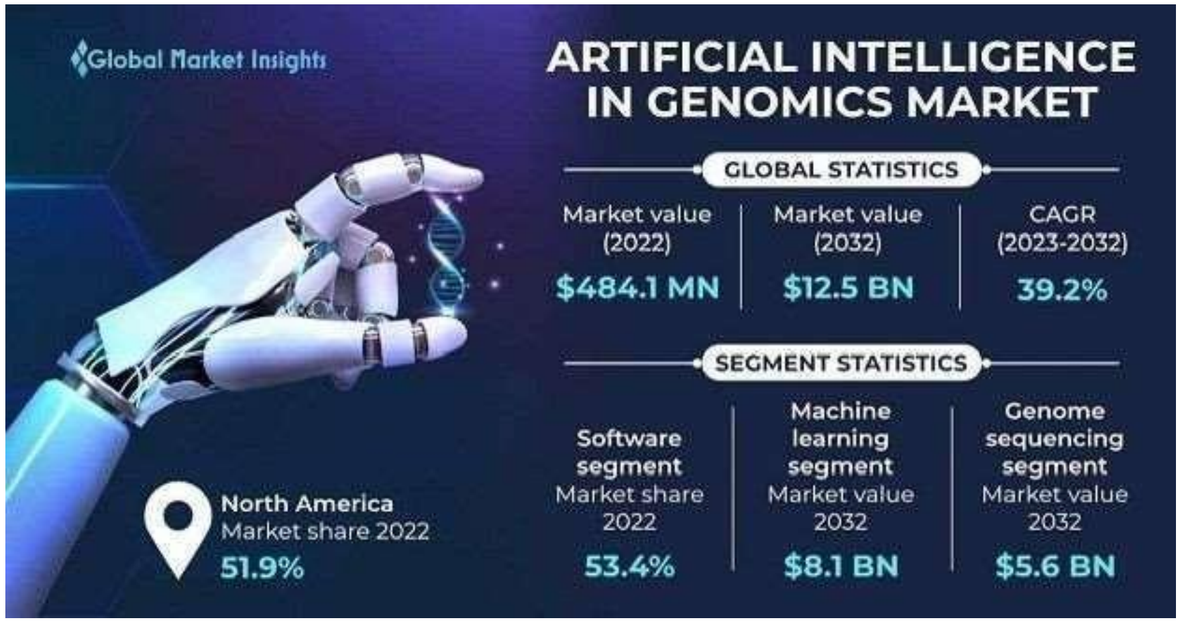
Жасанды интеллект (ЖИ) геномдық деректерді талдауда маңызды құрал ретінде қолданылады. Қазіргі заманғы биология мен медицинада геномнан алынған ақпараттың көлемі өте үлкен болғандықтан, оны қолмен талдау мүмкін емес. Сондықтан ЖИ мен машиналық оқыту әдістері үлкен деректерді өңдеп, заңдылықтарды анықтауға мүмкіндік береді.





ЖИ жүйелері ДНҚ тізбегін талдап, мутацияларды, ауру тудыратын гендерді, және тұқым қуалайтын ауруларға бейімділікті анықтай алады. Мысалы, нейрондық желілер мен алгоритмдер генетикалық ақпарат негізінде аурулардың даму қаупін болжайды немесе жаңа дәрілік нысандарды табуға көмектеседі.Сондай-ақ, жасанды интеллект жеке медицина (персонализирленген медицина) саласында да кең қолданылады — әр адамның генетикалық ерекшелігіне қарай ең тиімді ем немесе алдын алу тәсілін таңдауға мүмкіндік береді.

Машиналық оқыту әдістері мен олардың генетикалық зерттеулердегі рөлі



Машиналық оқыту (МО) – бұл компьютердің алдын ала нақты бағдарламаланбай, берілген деректерден үйреніп, өздігінен қорытынды шығару қабілетін дамытатын жасанды интеллект әдісі. Генетикада бұл тәсілдер үлкен көлемдегі геномдық және биомедициналық деректерді талдау үшін кеңінен қолданылады.Негізгі машиналық оқыту әдістері: 1. **Бақыланатын оқыту** (Supervised learning) – алдын ала белгілі нәтижелері бар деректер негізінде модельді үйрету. Мысалы, ауру немесе сау адам генотиптерін салыстырып, ауруға бейімділікті болжау.2.**Бақыланбайтын оқыту** (Unsupervised learning) – белгісіз үлгілер арасындағы ұқсастықтарды анықтау. Гендердің кластерленуін немесе жаңа мутация түрлерін табу үшін қолданылады. 3. Нейрондық желілер мен терең оқыту (Deep learning) – күрделі геномдық деректерді өңдеп, жасырын заңдылықтарды анықтайды. Бұл әдіс ДНҚ тізбегін талдауда өте тиімді.

Генетикалық зерттеулердегі рөлі:
Генетикалық ауруларды ерте анықтау және болжау;
Генетикалық мутацияларды автоматты түрде тану;
Жеке медицинада (personalized medicine) әр адамның генетикалық ерекшелігіне сай ем тәсілдерін таңдау;
Эволюциялық байланыстар мен тұқым қуалаушылық заңдылықтарын талдау.

Аурулардың генетикалық бейімділігін анықтау – қазіргі медициналық генетика мен персоналдық медицинадағы басты мәселелердің бірі.

Жасанды интеллект (AI) үлгілері бұл бағытта үлкен рөл атқарады, өйткені олар үлкен көлемдегі геномдық деректерді тез әрі тиімді талдауға, гендер арасындағы күрделі өзара әрекеттерді анықтауға мүмкіндік береді.





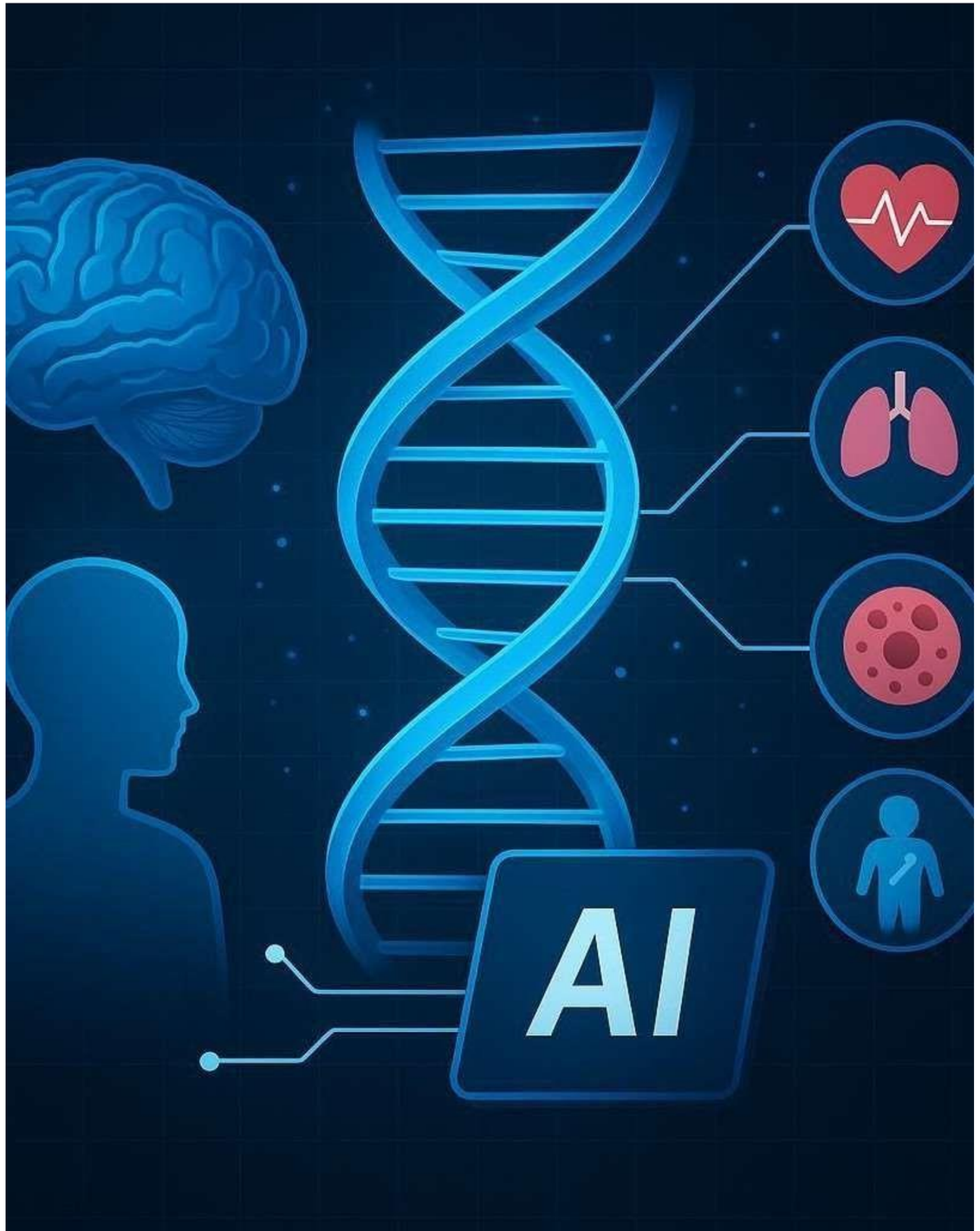
1. Машиналық оқыту (МО) үлгілері

МО алгоритмдері ауру қаупін бағалауда статистикалық және үлгілерді тану әдістерін біріктіреді. Негізгі әдістер:

Логистикалық регрессия – белгілі бір SNP немесе генотиптің ауру қаупіне қаншалықты әсер ететінін бағалайды.

2. Нейрондық желілер

Күрделі өзара әрекеттер мен эпистаздық байланыстарды анықтауға арналған.



3. Генетикалық алгоритмдер

Табиғи іріктеу және эволюциялық принциптерге негізделген.Гендер арасындағы өзара әрекеттерді модельдеп, ауру қаупін бағалауға көмектеседі.

Қолданылуы: Көптеген SNP комбинацияларын және олардың ауру қаупіне әсерін модельдеуде тиімді.

4. Гибридтік үлгілер

Бірнеше AI әдісін біріктіріп, геномдық және клиникалық деректерді салыстырады.Мысалы, Random Forest + Neural Network комбинациясы ауру қаупін дәл әрі толық бағалауға мүмкіндік береді.Артықшылығы: Әр әдістің күшті жақтарын біріктіреді, жоғары дәлдік береді.

5. Қолдану мысалдары

BRCA1/BRCA2 мутациялары – сүт безі рагының генетикалық бейімділігін анықтау.APOE ε4 аллелі –Альцгеймер ауруының генетикалық тәуекелін бағалау.Polygenic Risk Scores (PRS) – бірнеше геннің ықпалын біріктіріп, ауру қаупін сандық түрде көрсету.Жүрек-қан тамыр аурулары – түрлі SNP және ген комбинацияларын талдау арқылы жеке тәуекелді бағалау

Қорытынды

Биоинформатика - биологиялық деректерді жинақтау, сақтау, өңдеу және талдау арқылы олардың биологиялық маңызын анықтайтын пәнаралық ғылым саласы. Ол генетика, молекулалық биология, информатика, математика және статистика ғылымдарын біріктіреді.

Геномдық деректерге ДНҚ және РНҚ тізбектері, гендердің экспрессиясы, эпигеномдық өзгерістер және генетикалық вариациялар туралы ақпарат жатады. Бұл деректер Sanger секвенирлеу, NGS, WGS, RNA-seq, генотиптеу және микрочип технологиялары арқылы алынады.

Геномдық деректерді биоинформатикалық талдау геномды жинақтау, тізбектерді салыстыру, генетикалық вариацияларды анықтау және олардың биологиялық қызметін түсіндіру кезеңдерін қамтиды. De novo және референстік геномға негізделген жинақтау тәсілдері геномдық ақпаратты зерттеуде маңызды орын алады.

Big Data технологиялары үлкен көлемдегі, әртүрлі құрылымдағы және жылдам жинақталатын геномдық деректерді тиімді өңдеуге мүмкіндік береді. Бұлттық есептеу жүйелері деректерді сақтау, есептеу ресурстарын пайдалану және зерттеушілер арасындағы ақпарат алмасу мүмкіндіктерін кеңейтеді.

GenBank, NCBI және BLAST сияқты биоинформатикалық ресурстар генетикалық тізбектерді сақтау, іздеу, салыстыру және ғылыми талдау үшін қолданылады. Олардың көмегімен гендердің ерекшеліктерін, тізбектер арасындағы ұқсастықтарды және геномдық өзгерістерді зерттеуге болады.

Биоинформатика, Big Data және жасанды интеллект технологиялары геномдық ақпаратты жан-жақты зерттеудің маңызды құралдары болып табылады және қазіргі генетика мен медицинаның жаңа мүмкіндіктерін қалыптастырады.

Өзін-өзі тексеруге арналған сұрақтар:

1.Биоинформатика дегеніміз не және оның геномдық зерттеулердегі негізгі міндеттері қандай?
2.Геномдық деректердің қандай түрлері бар және олар қандай технологиялар арқылы алынады?
3.Sanger секвенирлеу, NGS, WGS және RNA-seq әдістерінің негізгі ерекшеліктері қандай?
4.Геномды de novo жинақтау мен референстік геномға негізделген жинақтаудың айырмашылығы қандай?
5.Геномдық деректерді биоинформатикалық талдаудың негізгі кезеңдері қандай?
6.Big Data дегеніміз не және геномдық деректердің қандай ерекшеліктері үлкен деректер технологияларын қолдануды қажет етеді?
7.Бұлттық есептеу технологиялары геномдық деректерді сақтау және өңдеуде қандай қызмет атқарады?
8.GenBank, NCBI және BLAST ресурстарының негізгі қызметтері қандай?
9.Машиналық оқыту мен жасанды интеллекттің айырмашылығы қандай және олар геномдық зерттеулерде қалай қолданылады?
10.Жасанды интеллект ауруларға генетикалық бейімділікті бағалау мен дербестендірілген медицинада қандай мүмкіндіктер береді?

Пайдаланылған әдебиеттер

1.Mount, D. W. (2004). *Bioinformatics: Sequence and genome analysis* (2nd ed.). Cold Spring Harbor Laboratory Press.

2.Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

3.Marx, V. (2013). Biology: The big challenges of big data. *Nature, 498*, 255–260. https://doi.org/10.1038/498255a

4.Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics, 16*, 321–332. https://doi.org/10.1038/nrg3920

5.Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: New computational modelling techniques for genomics. *Nature Reviews Genetics, 20*, 389–403. https://doi.org/10.1038/s41576-019-0122-6