

Технологиите Big data и IomT в подкрепа на прецизната медицина

Прецизната медицина е една от най-горещите теми в днешно време и включва диагностика и лечение на заболявания, които отчитат индивидуалния генетичен профил, спецификата на околната среда и начина на живот на индивида. Откриването на знания посредством анализ на данни (KDD) е от решаващо значение за идентифицирането на сложни ДНК аномалии, свързани с генетични заболявания и рак. Капацитетът на ДНК базите данни нараства експоненциално през последното десетилетие. Прецизната медицина започва с геномика и разчита на платформите „omics“ за анализ и интерпретация на многомащабни данни.

Персонализирана медицина. Широко известно е, че лечение, което помага на някои пациенти, е неефективно за други и същото лекарство може да предизвика нежелани реакции само при определени пациенти. В рамките на повече от две хилядолетия, медицината запазва стремежа си да бъде персонализирана. В древни времена Хипократ съчетава оценката на четирите телесни течности - кръв, хрчка, жълта жлъчка и черна жлъчка - за да определи най-доброто лечение за всеки пациент. Понастоящем, анализът на последователността на четирите химически градивни блока, които изграждат ДНК, съчетана с наличието на известни протеини в кръвта, дава възможност за по-точни медицински прогнози.

Мисията на Геномно здраве (Genomic Health) е да трансформира решенията и резултатите от лечението на рак чрез предоставяне на клинично приложими диагностика и услуги.

Коалицията за персонализирана медицина (Personalized Medicine Coalition – PMC), обединява новатори, учени, пациенти, доставчици и потребители, насърчава разбирането и приемането на персонализирани медицински концепции, услуги и продукти в полза на пациентите и здравната система. Според PMC персонализираната медицина е развиваща се област, в която лекарите използват диагностични тестове, за да определят кои медицински процедури ще работят най-добре за всеки пациент. Чрез комбиниране на данните от тези тестове с медицинската история, обстоятелствата и ценностите на индивида, доставчиците на здравни услуги могат да разработят целеви планове за лечение и превенция.

Терминът "*Прецизна медицина*" се отнася до приспособяването на медицинското лечение към индивидуалните характеристики на всеки пациент. Това не означава

буквално създаването на лекарства или медицински устройства, които са уникални за пациента, а по-скоро способността да се класифицират индивидите в субпопулации, които се различават по чувствителността си към определена болест, в биологията и/или прогнозата на тези заболявания, които те могат да развият, или в отговор на специфично лечение. Превантивните или терапевтичните интервенции могат да бъдат съсредоточени върху онези групи пациенти, при които терапията ще има положителен ефект, ще бъдат намалени разходите, както и страничните ефекти за тези, при които те не са ефективни.

Персонализирана и прецизна медицина. Персонализираната и прецизна медицина естествено възникват от пресичането на технологиите "omics" и алгоритмите Big Data.

Персонализираната медицина стана възможна благодарение на възможностите да се секвенира човешкия геном и презумпцията, че е икономически приемливо да се секвенират геномите на цялата популация.

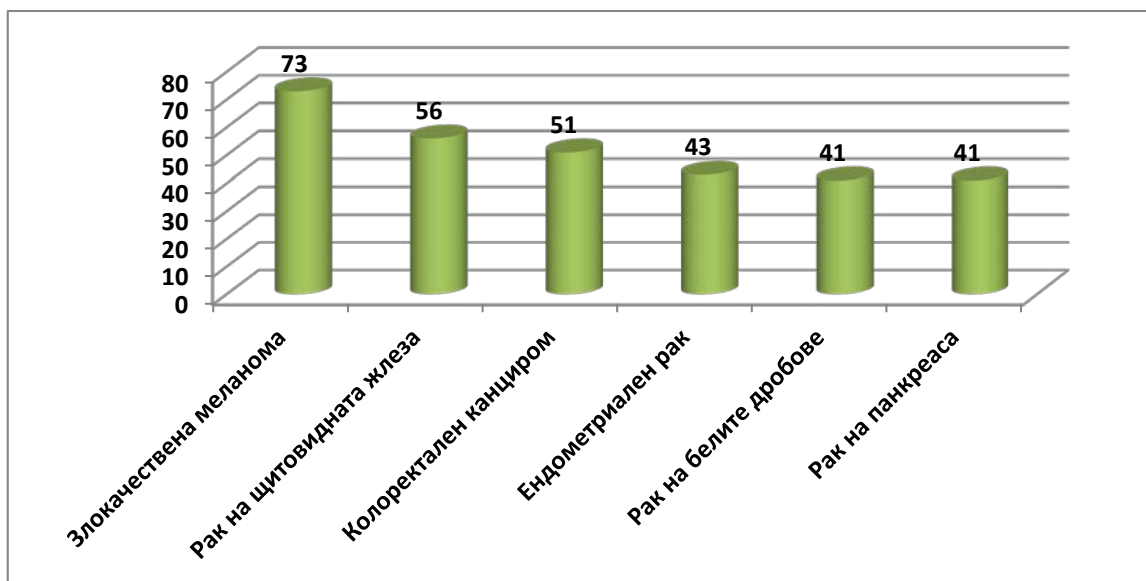
По този начин може да се обедини цялата тази информация в големи бази данни, обхващащи демографските аспекти, медицинската история, и ефекта на съществуващите лекарства, с цел след анализ да се определи кои терапии са били най-добри за отделните пациенти.

Прецизната медицина е възможна, защото могат да бъдат секвенирани геномите на повече хора и да бъдат събрани повече метаданни за множество съществуващи терапии, но се приема, че най-вероятният път към успеха идва от анализирането на точно дефинирани подгрупи пациенти.

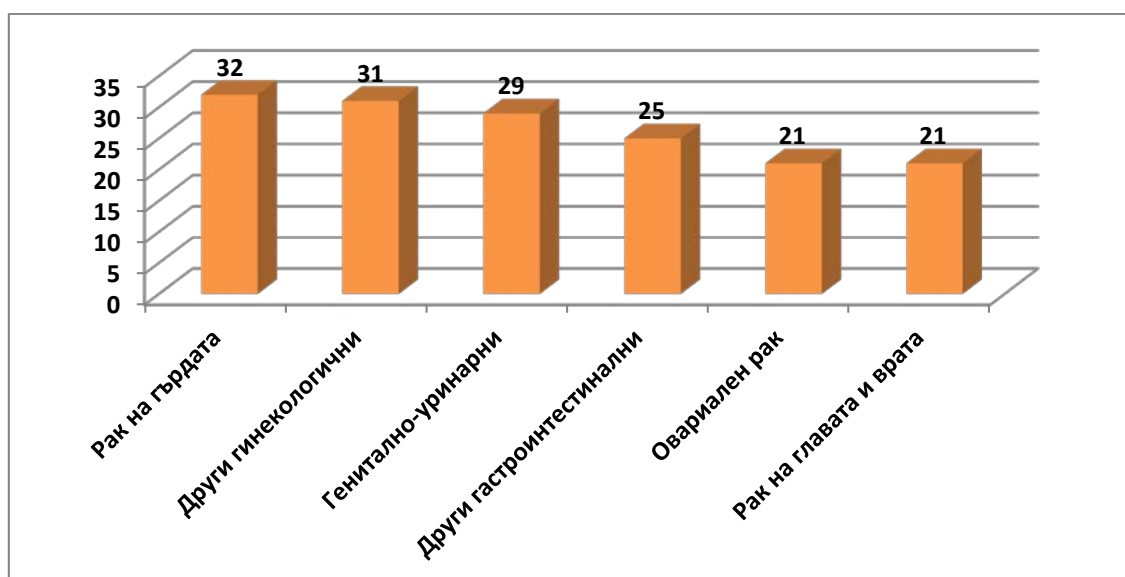
През 2015 г. президентът Обама (Белия дом), САЩ, стартира инициативата Precision Medicine Initiative с цел подобряване на здравето и лечението на болестите чрез приспособяване на терапевтичната и превантивна стратегия, така че да отговаря на спецификата на отделния пациент, вместо универсалния подход за средния пациент ("one-size-fits-all"). "Лекарите винаги са признавали, че всеки пациент е уникален, и те винаги са се опитвали да приспособят лечението си по най-добрия начин към отделните индивиди", заяви президентът Обама в обръщението си за състоянието на Съюза.

В контекста на лечението на рака, получаването на генетичния профил е изключително важно, защото много тежки заболявания, като рак на гърдата, се считат за генетично заболяване, дължащо се на мутации в специфични гени.

На фиг.1 и фиг.2 са показани данни, публикувани от РМС за % пациенти със специфични генетични мутации, които биха могли да бъдат третирани със специфични лекарства при различни ракови заболявания.



Фиг.1. Процент на пациентите със специфични генетични мутации, които биха могли да бъдат третирани със специфични лекарства



Фиг.2. Процент на пациентите със специфични генетични мутации, които биха могли да бъдат третирани с персонална терапия

Генетичните фактори са известни като основни фактори, причиняващи болести, така че наследственият рисков фактор е отговорен за 3% до 10% от всички случаи на рак на гърдата и до 30% от всички ранни ракови заболявания. Като цяло мутациите в гените BRCA1 и BRCA2, както и в гените, които са необходими за запазване на генома интактен

(непокътнат), като гени STK11, ATM, CHEK2, P53 и CDH1, водят до 50% от всички случаи на фамилен рак на гърдата и една трета от всички случаи на рак на гърдата

Относно генетичните фактори при рака на гърдата двата гена BRCA1 и BRCA2 действат потискащо на клетъчния растеж. Протеините, които тези два гена кодират, се наричат анти-онкогени и те помагат на клетките да „коригират“ повредените участъци на ДНК с цел запазване на генетичния материал. Следователно, ако един от двата гена е повреден, променените ДНК участъци не могат да възстановени, като в резултат могат да възникнат допълнително още промени и мутации в клетъчната ДНК, клетката ще бъде предразположена към развитие на тумор и евентуално развитие раково заболяване.

Генът BRCA2 е открит като втори предразполагащ ген за рак на гърдата. Този ген е важен при поправката на прекъсването на двойно-верижната структура на ДНК и регулацията на транскрипцията. Генът на BRCA2 в здравите клетки би осигурил стабилност на ДНК клетките и контролиран растеж на клетките. Генната мутация на BRCA1 причинява вероятност 60% до 80% за рак на гърдата при жените, както и повишена възможност за развитие на рак на яйчниците при жените и рак на простатата при мъжете. Мутации на BRCA2 се наблюдават при приблизително 35% от семействата с ранен рак на гърдата при жени и също предизвикват повишен риск от рак на яйчниците при жени и рак на гърдата при мъжете.

Персонализирането на лечението на пациентите е трудна и скъпа задача и вместо това, в случай на прецизна медицина, пациентите се групират според спецификата си и им се дава оптимално лечение за целевата група.

Основната цел на инициативата Precision Medicine е индивидуализираното здравеопазване. Разгръщането на прецизната медицина в грижата за пациента включва 3 решаващи фактора: генетични особености, фактори на околната среда и индивидуален начин на живот.

За да се улови генетичния профил на широк кръг от индивиди, се използват специални инструменти - автоматизирани ДНК секвенатори. Технологиите за секвениране от следващо поколение (NGS), също наречена масово паралелно или дълбоко секвениране, дава възможност да се подреди цял човешки геном за един ден, но цената за това е все още доста висока. Всеки нуклеотид от генома (символът, който го представлява) се прочита няколко пъти, за да се намали вероятността от грешка. Размерът на личния секвениран геномен файл е около 200 Gbytes. Подходът за

секвениране на целия геном печели популярност, защото повече от 90% от ДНК областите с клинично значение всъщност са некодирани региони.

Първият човешки геном е секвениран преди повече от 15 години на цена от 3 милиарда долара. Днес наградата за секвениране на човешки геном е по-малко от 1000 долара. Въпреки това, все пак тази награда се смята за твърде висока и здравната общност разчита на технологични иновации, които ще намалят тази цена под \$ 100.

Според *Nebula Genomics*, освен цялото геномно секвениране, два други фактора са от решаващо значение за стимулиране на прецизната медицина и развитие на успешен геномен пазар: (1) лесно споделяне на придобитите лични геномни данни и (2) „засилена защита на данните“. За да се гарантира необходимата защита на личните данни, мъглявината *Genomics* възнамерява да внедри технология *blockchain*.

Глобалната мрежа от лични геномни проекти (стартирала през 2005 г.) е коалиция от проекти в световен мащаб, насочени към установяване на данни за обществения геном, здравето и характеристиките. Обхватът на участващите партньори включва: Харвардското медицинско училище, САЩ (пилотното място на мрежата), Канада, Великобритания, Австрия и Китай.

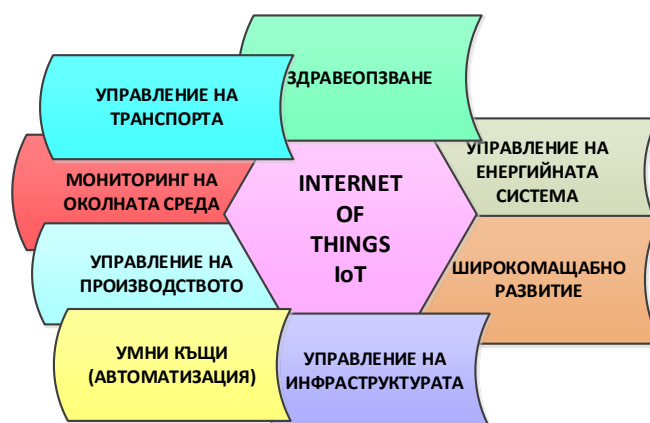
Очевидно персонализираната медицина и прецизна медицина се появяват в резултат на пресичането на *omics* технологиите и *Big Data*.

Ролята на Интернет на медицинските неща (Internet of medical Things) за развитието на прецизната медицина.

Интернет на медицинските неща е взаимосвързана инфраструктура на мобилни устройства “вградени с електроника, софтуер, сензори и мрежова свързаност, която позволява на тези обекти да събират и обменят данни”. Интернет на нещата позволява на обектите да бъдат разпознати или управлявани дистанционно в съществуващата мрежова инфраструктура, като създават възможности за по-директна интеграция на физическия свят в компютърни системи, което води до подобрена ефективност, точност и рентабилност в допълнение към намалената човешка намеса.

В случай, че интернет на нещата се допълва от сензори и задвижващи механизми, технологията става пример за по-общ клас кибер-физически системи, включително технологии като интелигентни мрежи, виртуални електроцентрали, интелигентни домове, интелигентен транспорт и интелигентни градове.

Всяко "нещо" е уникално идентифицирано от вградената компютърна система, но е в състояние да работи в рамките на съществуващата интернет инфраструктура.



Фиг.3. Internet of Things – спектър.

Европейската комисия признава необходимостта от холистичен подход към развитието на цифровото общество и цифровата икономика, включително развитието на бъдещия интернет. Развитието на единния европейски дигитален пазар е определено като един от основните приоритети на ЕС. Това може да се направи въз основа на високоскоростен интернет, като се създадат интелигентни инфраструктури и услуги, основани на ИКТ, като: интелигентни мрежи, интелигентни системи за енергийна ефективност на сградите, интелигентни системи за наблюдение и мониторинг на околната среда, интелигентни системи за транспорт и мобилност, интелигентни системи за здравеопазване и т.н.

Визията на бъдещия интернет е модификация или интеграция на някои от следните модели, подкрепени от обща мрежова платформа: интернет от и за хората, интернет на съдържанието и знанията, интернет на услугите и интернет на нещата.

Европейският клъстер за Интернет на нещата е динамична глобална мрежова инфраструктура със самоконфигуриращи се възможности, базирани на стандартни и оперативно съвместими комуникационни протоколи, където физическите и виртуалните "неща", които имат идентичност, физически качества и виртуални личности използват интелигентни интерфейси и са безпроблемно интегрирани в информационна мрежа.

Паневропейските центрове за върхови постижения, изпълнявани в съответствие с пътната карта на Европейския стратегически форум за научноизследователски инфраструктури (ESFRI), са важни стълбове на европейските изследвания и иновации.

Едно от най-успешните решения за здравната индустрия в инфраструктурата, поддържаща IomT, са мобилните приложения за дистанционно наблюдение на здравето, т.е. Digital Health Advisors, които улесняват комуникацията между пациентите и лекарите чрез сигурна връзка.

Безжичната телесна мрежа (WBAN), наречена също телесна сензорна мрежа (BSN), всъщност дава възможност да се придобият и натрупат лични медицински данни от носими устройства с широк спектър (физиологични биосензори), вградени в или на повърхността на тялото, или подходящи за носене в дрехи, чанти и др. (Фиг.2.4). Данните, получени на място за физиологичния статус на дадено лице, се предават по интернет и по този начин се правят достъпни в разумен срок и по сигурен начин за лекарите, независимо от разпределението на пациента.

Концепцията за свързване и дистанционно наблюдение на медицинско оборудване за изобразяване през Интернет датира преди около 20 години. Големите три компании за медицински изображения Siemens Healthcare, Philips Healthcare и GE Healthcare заедно изградиха стратегия за установяване на „All-in-One Health Cloud“ през 2015 г. Основната цел е да се премести интензивната обработка на изображения в компютърната екосистема.

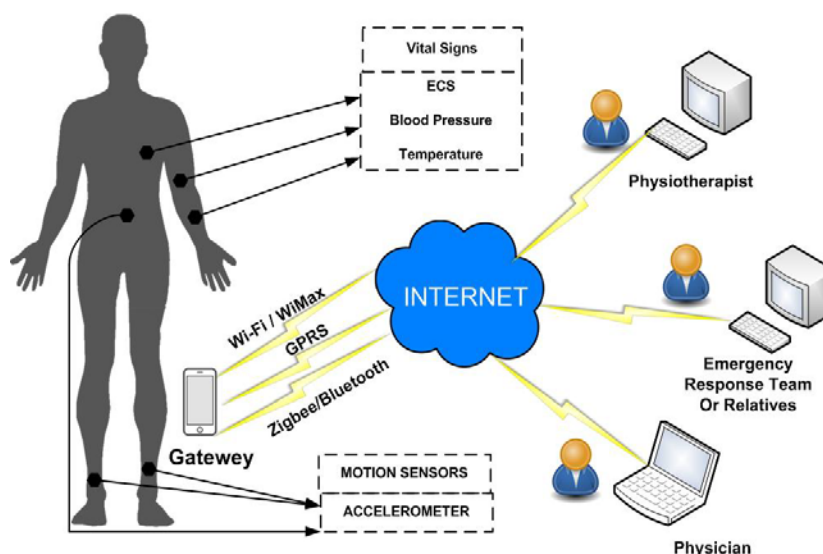
Според HealthIT Analytics анализирането на изображения е първата стъпка към персонализираната медицина.

Медицинската образна диагностика генерира и акумулира огромни количества рентгенови, кардиографски, ултразвукови и други изображения, помагачи в диагностиката на заболяването на отделния пациент. От решаващо значение е пълната автоматизация на получаването на медицинските образи и процеса на анализ, за да се поддържа висока прецизност на диагнозата.

Обработката на сигнали играе все по-важна роля в съвременността, най-вече поради все по-нарастващата популярност на IoT устройствата. Анализът на данните от IoT в медицината помага на лекарите да съставят надеждна и точна диагноза и прогноза на заболяването, за прогнозиране на прогресията на заболяването и ефекта от лечението, както и за идентифициране на целите на лекарствата.

Медицинските изображения се отнасят до техники и процеси, използвани за създаване на образи на различни части от човешкото тяло за диагностика и лечение в рамките на цифровото здраве. Медицинският образ в здравеопазването помага за ранна диагностика на заболявания и предписване на по-ефективна терапия. Интернет на Нещата на медицинските изображения са важна част от IoT технологиите и имат решаваща роля при образната диагностика.

Wireless Body Area Network



Фиг.4. Безжичната телесна мрежа (Body Area Network).

Европейската научно-изследователска инфраструктура за технологии за обработка на изображения в биологичните и биомедицинските науки (Euro-BioImaging European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences - Euro-BioImaging, EuBI) ще предостави на потребителите достъп до широк спектър от най-съвременни биологични и медицински технологии за биолози в Европа и в света. Тя ще предлага поддръжка и обучение на данни за изображения за потребители и доставчици на инфраструктура и ще оценява и ще включва нови технологии за изображения, за да осигури най-съвременни услуги по устойчив начин, като хранилища на изображения и инструменти за анализ. EuBI ще се състои от набор от допълнителни, силно взаимосвързани и географски разпределени възли - специални устройства за обработка на изображения - за достигане до европейски учени във всички държави-членки. Инфраструктурата ще се управлява от силна подкрепяща и координираща организация, която е в центъра на EuBI. Центърът ще предостави виртуалната точка за достъп, от която потребителите ще бъдат насочени към желаната технология за изображения, като обслужват съответните възли на EuBI. Центърът ще координира управлението на данните и дейностите за обучение, съобразени с нуждите на потребителите на инфраструктурата за изображения.

Концептуален модел на екосистемата на Интернет на медицинските Неща е представен на фиг.5.

Наред с геномиката, медицинските изображения са едни от най-бързо развиващите се големи клинични източници на данни в здравното пространство. Необходимият капацитет на дисковото пространство за съхранение на медицинските изображения се е увеличил повече от три пъти от 2005 г. насам и се очаква да се удвои като устойчива тенденция в близко бъдеще. Така здравните заведения често използват облачни услуги, за да осигурят висококачествено съхранение на медицински изображения.

Технологиите за облачни медицински изображения предлагат уеб-базирани платформи за анализ на изображения. Услугите за обработка на изображения в облака включват ултра бързи облачни изчислителни инфраструктури (суперкомпютри), визуализация, дълбоко обучение (deep learning) и съхранение и споделяне на медицински изображения в облака.

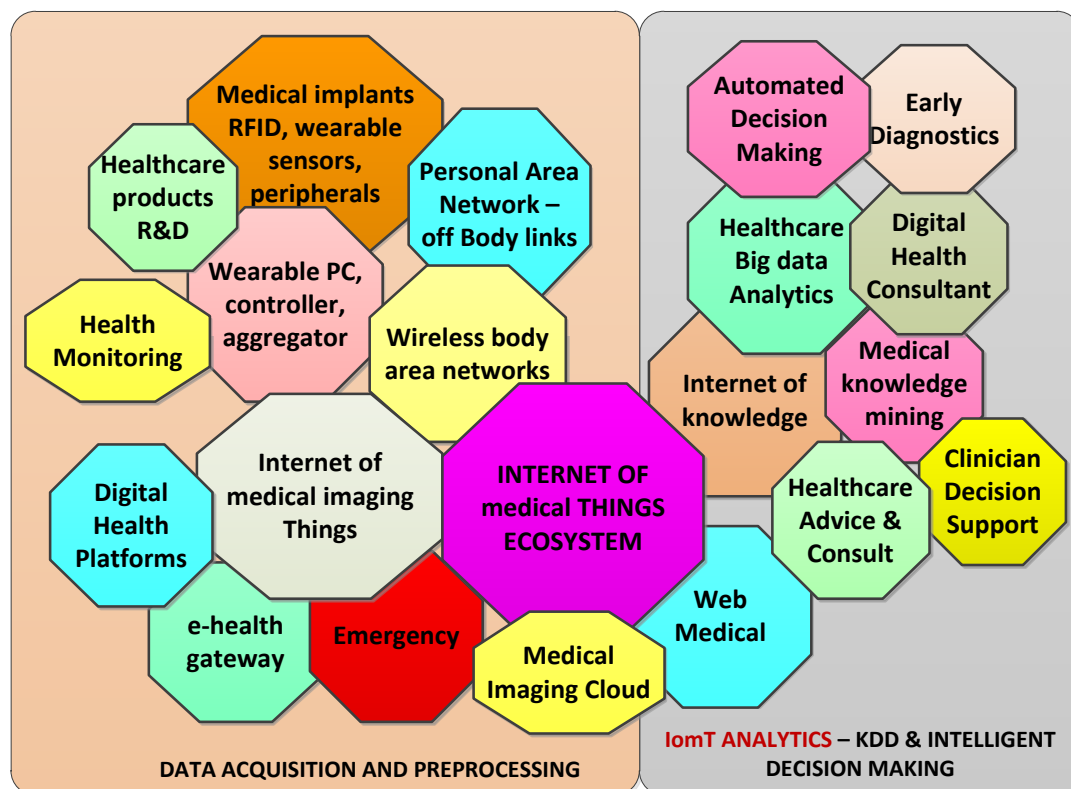
IBM Watson наскоро се увеличи с повече от 4 милиарда долара нови активи чрез придобиването на Phytel (здраве на населението), Explorys (облак), Merge (сливане) и Truven (анализи). Очевидно е, че събирането и натрупването на генерирани от пациента здравни данни от IomT устройствата е ключът към управлението на здравето на населението.

Релацията „Анализ на големи геномни данни – Прецизна медицина“

От 2000 г. медицината и биологията навлизат в постгеномната ера, която се свързва с появата и интензивното развитие на новата наука геномика. Предпоставките за появата на геномиката в световен мащаб обхващат новата технология на пълно геномно секвениране, т.е. определяне на нуклеотидната секвенция (първична структура) на всички геноми. Това доведе до бързо секвениране на десетки хиляди прокариотни и хиляди еукариотни геноми, включително човешкия геном. Базите данни ДНК и техният капацитет нарастват експоненциално през последното десетилетие. Геномните данни се удвоиха на всеки 5 месеца през последните 8 години.

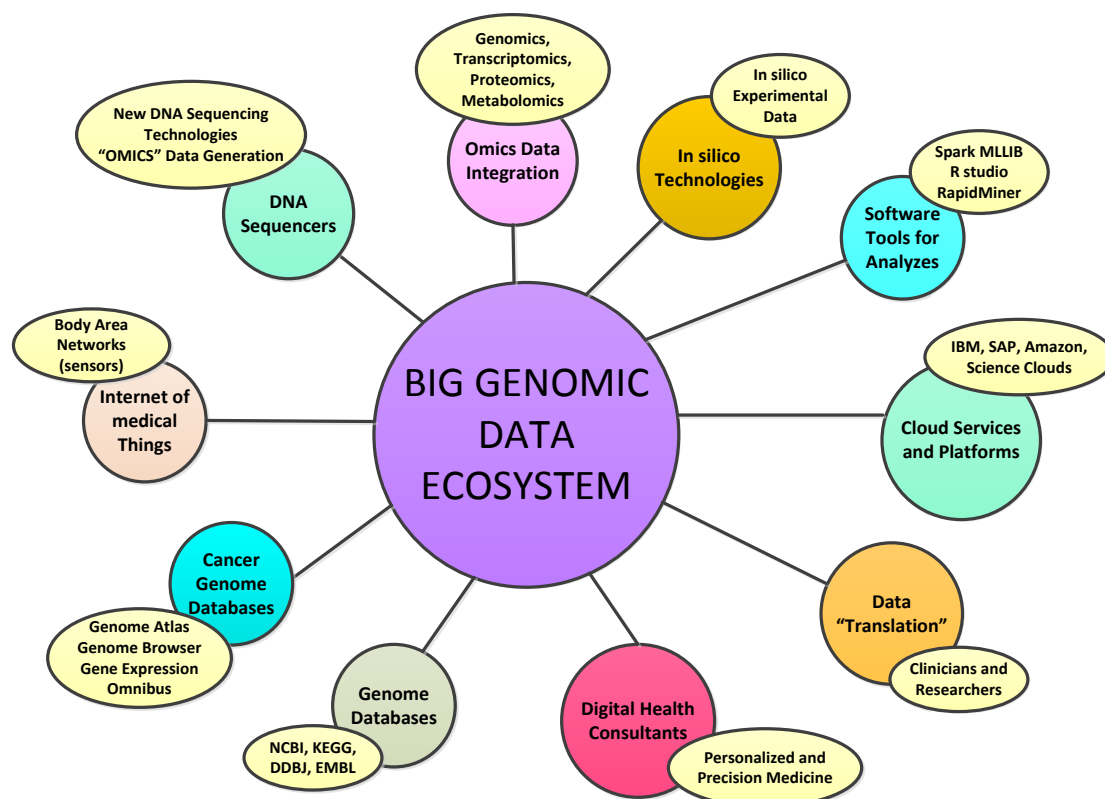
Концептуалният модел на екосистемата на големите геномни данни е представен на фиг.6. Основните източници на събиране на геномни данни са новите технологии за секвениране на ДНК, генериране на „omics“ данни (фиг.7), генериране на огромни количества експериментални данни *in silico*, геномни бази данни, бази данни за ракови

геноми, както и свързаните технологии Интернет на медицинските Неща (IomT) и облачните технологии. Следователно, има ново предизвикателство - капацитетът на геномните бази данни нараства по-бързо от капацитета на софтуерните инструменти за анализ.



Фиг.5 Концептуален модел на екосистемата на Интернет на медицинските Неща.

Прецизната медицина започва с геномиката и разчита на платформите „omics“ за анализ и интерпретация на широко мащабни данни. Днес има богат спектър от платформи за големи геномни данни като Google Genomics, референтната архитектура на IBM за геномика (IBM Reference Architecture), SAP® Connected Health платформа и т.н. Тези платформи обединяват усилията на разработчиците, изследователите и здравните организации за иновации в решенията, насочени към пациентите с цел подобряване на здравеопазването, намаляване на разходите и осигуряване на свързани здравни услуги.



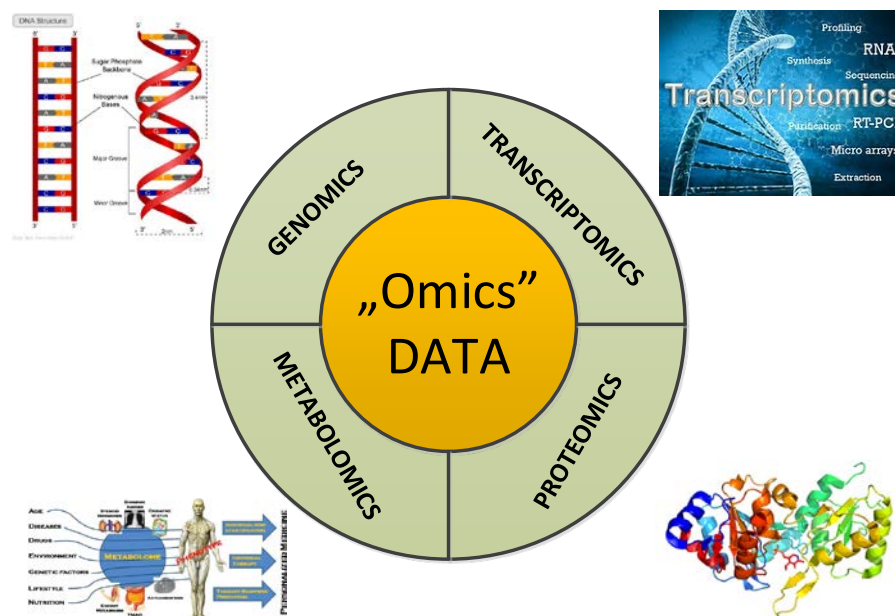
Фиг.6 Концептуален модел на екосистемата на големите геномни данни

Google Genomics предлага инфраструктура като услуга (IaaS), достъпна чрез тяхната облачна платформа, за изпълнение на масивни изчисления на виртуални машини на принципа на заплащане-при-ползване. Google предоставя уеб базиран API за геномика, който изпълнява функциите на портал за достъп и използване на софтуерни инструменти от портфолиото от решения на Google. Софтуерният инструмент BigQuery дава възможност за много бързи SQL-подобни заявки за масивни биологични и медицински набори данни. Базирайки се на модела за програмиране MapReduce, платформата предоставя възможности за откриване на взаимовръзки в геномните данни чрез методите на машинно обучение и други методи за откриване на знания (KDD).

Amazon's Genomics in the Cloud предлага инфраструктура, софтуерни инструменти и набори от данни за геномни анализи с цел улесняване на персонализираната медицина.

IBM® референтната архитектура от-край-до-край определя критичните средства за геномична обработка: Управление на данни (Datahub), оркестриране на натоварването (Orchestrator) и достъп до приложенията (AppCenter). Тя може да бъде разгърната с различни инфраструктурни и информационни технологии, спазвайки 3 основни

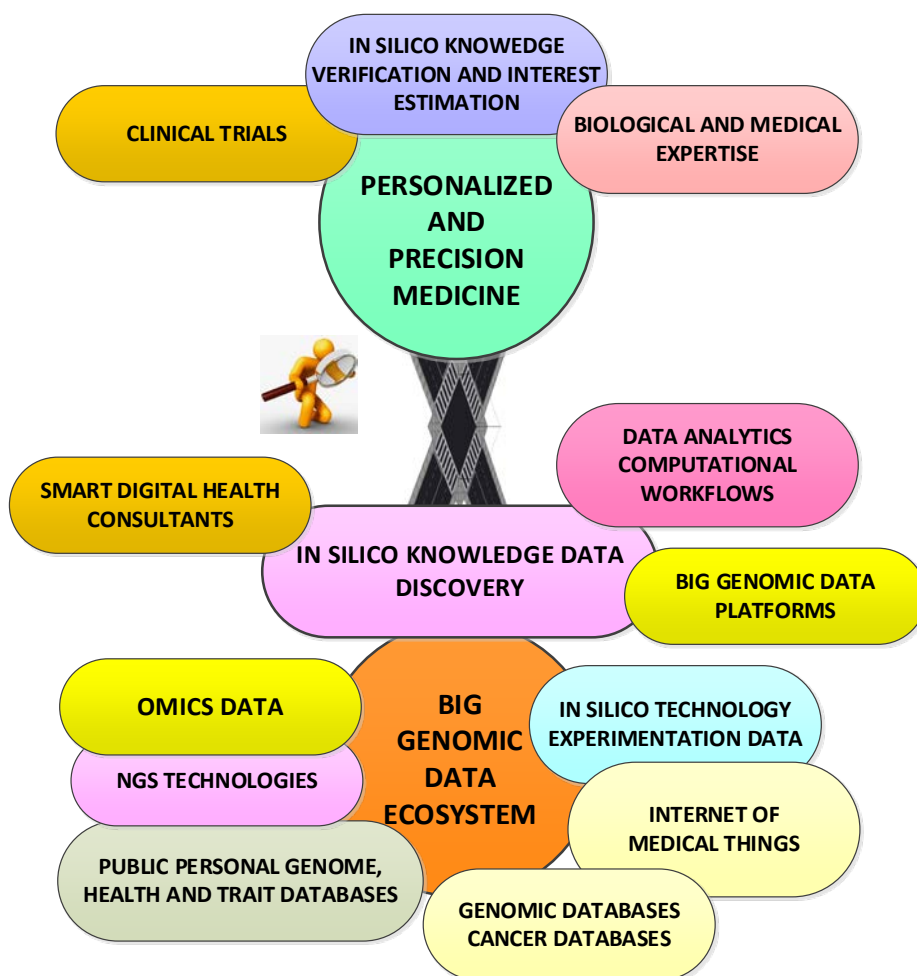
принципа: софтуерно дефинирани, ориентирани към данни и готови за приложение (software-defined, data-centric and application-ready).



Фиг.7 Спектър на “omics” данните

През 2015 г. Intel и Орегонският университет по здравеопазване и науката стартираха съвместен проект - Collaborative Cancer Cloud, който всъщност представлява високопроизводителна платформа за анализ, която натрупва и интегрира частни медицински данни, насочени към изследване на рака. Intel възнамерява да създаде мрежа от федерирани облаци към други институции, като разшири изследванията за болестта на Паркинсон.

Пресечната точка на Big data analytics и прецизната медицина се крие в откриването на *in silico* знания от данни (*in silico* KDD), което прави възможно проектирането и внедряването на интелигентни дигитални консултанти, оказващи помощ на медиците и изследователите да обработят огромното количество информация за пациентите с цел точна диагностика на заболяването и определяне на оптимална терапия за индивидуалния пациент. Пресечната точка на Big data analytics и прецизната медицина е показана на фиг.8. Тя започва с екосистемата на големите геномни данни, която придобива и акумулира от молекулярни omics данни, данни от IomT и данни от *in silico* експериментиране.



Фиг. 8 Пресечната точка на Big data analytics и прецизната медицина

Анализ на големи биомедицински данни и откриване на *in silico* знания

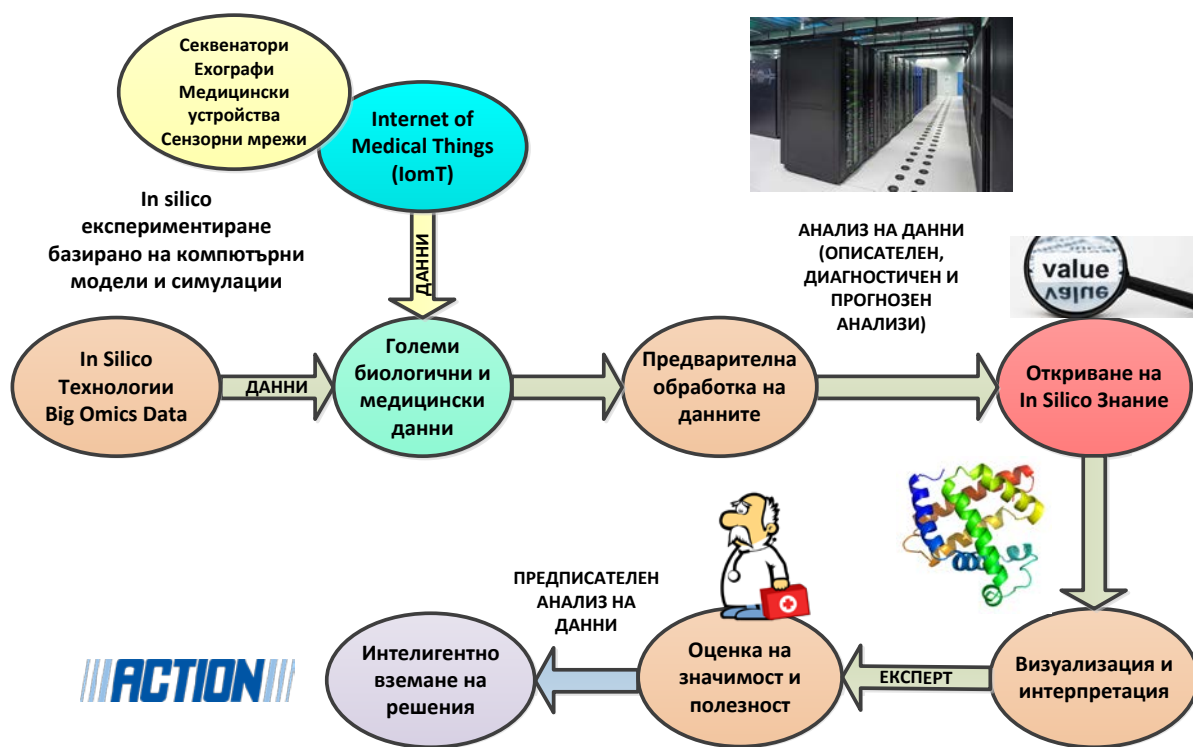
Изследователските методи и техники са заложи в технологичния процес за откриване на *in silico* знания и вземане на решения, и обхващат следните групи: (1) предварителна обработка на данни; (2) откриване на *in silico* знания и вземане на решения, и (3) последваща обработка, включваща представяне на резултати (знания, прогнози) по разбираем начин, подходящ за интерпретация и оценка на интереса към знанието (значимост).

Предварителната обработка на данни в случай на откриване на знания обхваща: интегриране на данни (от различни източници, например), филтриране на данни (по отношение на точност), дискретизация, избор на характеристики (по отношение на

значимостта). Селекцията на характеристиките се извършва чрез прилагане на метабвристични алгоритми за търсене и оптимизация, както и анализ с принципални компоненти - PCA, и множеството на характеристиките се оптимизира (редуцира) чрез итеративно изпълнение на алгоритъма за машинно обучение. Освен това, последващата обработка включва верификация, валидиране и оценка (полезност, стойност) на откритите знания, както и *in situ* визуализация.

Концептуалният модел на изчислителния конвейер за откриване на *in silico* знания и автоматизирано вземане на решения е представен на фиг.9.

Основните функционални единици за откриване на *in silico* знания посредством анализ на данни (**K**nowledge **D**ata **D**iscovery - KDD) всъщност са научни работни потоци (scientific workflows) за анализ, които са изключително полезни за стимулиране на е-науката. Всъщност, работният поток представлява изчислителния конвейер за *in silico* KDD (Фиг.9) на базата на специфичен метод на анализ (модел). Методът на работните потоци осигурява решение, подпомагащо проектирането и провеждането на експерименти, като използва наличните данни и софтуерни инструменти. Работният поток се дефинира като модел, определящ последователното изпълнение на процесите или потока от задачи, който се координира и планира въз основа на систематичен план. Научните работни процеси осигуряват метод за дефиниране на целите на експеримента на високо ниво, моделиран от работния поток на научните задачи. Различни видове задачи могат да се изпълняват в рамките на работния процес, особено когато изходите от една задача се използват като вход за следващата задача.



Фиг.9. Изчислителен конвейер за откриване на in silico знания и вземане на решения