

Klinisch genetische databases en webtools voor de laboratoriumspecialist

A. GERRITS, M.J. VOS, A.P. ABBES, H. ENGEL en L.D. DIKKESCHEI

Moleculaire diagnostiek neemt een steeds grotere plaats in binnen de geneeskunde en dus ook binnen de klinische chemie. Dit vereist dat de laboratoriumspecialist zich ook op dit vlak moet (blijven) ontwikkelen. Of het nu gaat om het opzetten van een nieuwe bepaling, het interpreteren van resultaten, het op de juiste manier rapporteren van resultaten of om het plegen van intercollegiaal overleg: in alle gevallen zijn kennis en kunde op klinisch genetisch vlak een vereiste. Online bestaat er een groot scala aan waardevolle vrij toegankelijke databases die hierbij behulpzaam kunnen zijn. Door de groeiende hoeveelheid informatie en te analyseren data lijkt het er inmiddels op dat de term 'bioinformaticus' toegevoegd moet worden aan de lijst van kerncompetenties van de hedendaagse laboratoriumspecialist. Aan de hand van een overzicht van klinisch genetische databases en webtools bieden we een handvat voor gebruik in de klinisch-chemische praktijk, maar ook voor gebruik in een onderzoekssetting.

Moleculaire diagnostiek omvat onderzoek naar afwijkingen in DNA of RNA. Dit onderzoek kan plaatsvinden in het kader van erfelijke, maar ook van verworven aandoeningen. Onderzoek naar en advisering omtrent erfelijke aandoeningen is voorbehouden aan vergunninghouders van klinisch genetische diagnostiek. Op dit moment zijn alle academische centra vergunninghouder. Een deel van de moleculaire diagnostiek mag plaatsvinden in klinisch-chemische laboratoria van perifere ziekenhuizen, maar alleen in samenwerkingsverband met een academisch centrum. Hierbij fungeert het academisch centrum als regionaal kenniscentrum en spil in het netwerk (1).

De in 2013 vastgestelde professionele standaard 'Klinische Chemie en Laboratoriumgeneeskunde' hanteert twee streefnormen met betrekking tot moleculaire diagnostiek: de laboratoriumspecialist klinische chemie 1) is verantwoordelijk voor een goede uitvoering, interpretatie en eenduidige rapportage van moleculaire diagnostiek, en 2) pleegt overleg met een klinisch genetisch centrum (2). De laboratoriumspecialist treedt hierbij op als intermediair tussen aanvrager en klinisch genetisch centrum. Dit maakt dat de laboratoriumspecialist een gedegen achtergrond moet hebben in de moleculaire diagnostiek. Het hoge tempo van technologische vooruitgang op dit vlak,

gepaard gaande met geavanceerde technologie, maken dit tot een uitdaging.

Klinisch-chemische praktijk

Voor de klinisch-chemische praktijk is het belangrijk snel te kunnen achterhalen waar welke diagnostiek aangeboden wordt en waar achtergronden van bepaalde ziekten of genafwijkingen te raadplegen zijn. Hiervoor zijn verschillende databases beschikbaar.

Wie doet wat?

Zoals de Klinische Chemie de 'Wie doet wat' database heeft, heeft de Klinische Genetica de 'Landelijk Overleg DNA-diagnostiek' of LOD database. Elk klinisch genetisch centrum heeft zijn eigen specialisaties met betrekking tot het diagnostiekpakket, waardoor de expertise geconcentreerd is en testen zo efficiënt mogelijk kunnen worden uitgevoerd. Op dit moment kan in Nederland moleculaire diagnostiek ingezet worden voor >300 erfelijke aandoeningen. Met behulp van de LOD database kan nagegaan worden waar in Nederland diagnostiek wordt aangeboden voor een specifieke erfelijke aandoening. Er kan gezocht worden op ziekte of op gen (of OMIM#, Online Mendelian Inheritance in Man#). Daarnaast zijn de aanvraagformulieren voor de verschillende centra via deze website te downloaden.

Wanneer diagnostiek niet in Nederland wordt verricht, bestaat de mogelijkheid diagnostieklaboratoria binnen Europa of wereldwijd te zoeken. EDDNAL, de European Directory of DNA Diagnostic Laboratories, is een database doorzoekbaar op ziekte of OMIM# met als output alle laboratoria in 17 Europese landen die de betreffende DNA diagnostiek aanbieden. GeneTests is een database doorzoekbaar op ziekte, gen of type test en geeft als output alle laboratoria wereldwijd die de gevraagde diagnostiek aanbieden. Naast een overzicht van genetisch onderzoek biedt deze database tevens een overzicht van wereldwijd beschikbare biochemische onderzoeken met betrekking tot erfelijke ziekten.

Ziekten en geassocieerde genafwijkingen

Voor het zoeken naar achtergronden over specifieke erfelijke danwel verworven ziekten en geassocieerde genafwijkingen zijn verschillende databases beschikbaar. GeneReviews is een database met daarin een overzicht van door experts opgestelde peer-reviewed beschrijvingen ('hoofdstukken') van erfelijke ziekten. Deze worden regelmatig geupdate en gereviseerd wanneer er nieuwe inzichten zijn die daar aanleiding

Klinisch Chemisch Laboratorium, Isala, Zwolle

E-mail: a.gerrits@isala.nl

toe geven. GeneReviews bevat 622 verschillende hoofdstukken (nov 2014). Deze worden in gestandaardiseerd format gepresenteerd, met aandacht voor diagnose, klinische beschrijving, differentiële diagnose, behandeling, genetische counseling en moleculaire genetica. Het is mogelijk te zoeken op aandoening, maar ook op aangedaan gen (3). OMIM is een database met daarin een overzicht van genen en genetische aandoeningen en eigenschappen, elk met een uniek OMIM#. OMIM heeft in het bijzonder aandacht voor de moleculaire relatie tussen genetische variatie en fenotypische variatie. Waar GeneReviews zich beperkt tot informatie over erfelijke aandoeningen, gaat OMIM verder met ook informatie over verworven aandoeningen. Daarnaast bevat OMIM informatie over erfelijke eigenschappen zoals kleurenblindheid, haar- en oogkleur. OMIM is onder andere doorzoekbaar op aandoening en op gen (4). Orphanet is een portaalsite waarmee informatie over zeldzame (erfelijke, danwel niet erfelijke) ziekten kan worden gevonden. Orphanet bevat informatie over ~2.000 zogenaamde weesziekten (nov 2014). Er kan gezocht worden op aandoening of op gen. Daarnaast bevat de site een inventaris van weesgeneesmiddelen in alle ontwikkelingsstadia en een diagnose tool met behulp waarvan door het invullen van specifieke symptomen gezocht kan worden naar ziekten (5).

Voor het zoeken naar DNA-mutaties geassocieerd met specifieke ziekten zijn verschillende databases beschikbaar. HGMD, oftewel Human Gene Mutation Database, is een database waarin bekende genafwijkingen verantwoordelijk voor of geassocieerd met erfelijke ziekten zijn opgeslagen. Waar de publieke (uitgeklede) versie vrij toegankelijk is voor geregistreerde gebruikers, is de versie voor de professional enkel toegankelijk met een licentie (6). COSMIC, oftewel Catalogue Of Somatic Mutations In Cancer, is 's werelds grootste en meest uitgebreide database met somatische mutaties in relatie tot verschillende typen maligniteiten. COSMIC bevat informatie over mutaties in ~30.000 genen in >1 miljoen tumormonsters. Specifiek bevat het informatie over ~3 miljoen coderende puntmutaties, >10.000 gen fusies, >60.000 genomische herschikkingen, ~700.000 copy nummer varianten, >15.000 volledige genomesequenties en ~120 miljoen gen expressie varianten. Deze informatie is afkomstig uit ~20.000 publicaties (nov 2014). Deze enorme hoeveelheid data maakt het mogelijk om inzicht te krijgen in moleculaire biomarkers in de oncologie. Er kan via de 'Search box' gezocht worden op gen, met als output overeenkomende gen namen (aliassen), ziekte beschrijvingen, samples, mutaties en relevante publicaties. Ook kan via de 'Search box' of de 'Cancer Browser' gezocht worden op 1 van de 2.500 aanwezige verschillende maligniteit-classificaties, met als output fraaie visuele representaties van mutatiepatronen achter de geselecteerde ziekte (7).

Onderzoekssetting

De afgelopen 10 jaar heeft er een ware -omics revolutie plaatsgevonden. Door het beschikbaar komen van steeds geavanceerdere high-throughput technieken

kon onderzoek plaatsvinden op onder andere DNA-, DNA methylatie- en histonmodificatie-, RNA-, eiwit-, metaboliet- en zelfs interactieniveau, respectievelijk ook wel genomics, epigenomics, transcriptomics, proteomics, metabolomics en interactomics genoemd. Deze technieken brachten grote hoeveelheden data met zich mee, die een hoger niveau van datamanagement, -opslag en -visualisatie vereisten; een grote uitdaging voor bioinformatici.

Genomics

Het bestuderen van sequenties, functies en organisatie van het genoom wordt ook wel genomics genoemd. Er bestaan drie grote publieke databases waarin nucleotide sequenties worden opgeslagen: GenBank (National Center for Biotechnology Information, NCBI, VS), EMBL-EBI (European Molecular Biology Laboratory, European Bioinformatics Institute) en DDBJ (DNA Data Bank of Japan). Deze databases worden dagelijks gesynchroniseerd, zodat ze alle drie dezelfde ruwe data bevatten. De data worden echter in een ander format opgeslagen en annotaties kunnen verschillen. Sequenties worden direct door wetenschappers of genoom-sequencing groepen toegevoegd of indirect uit literatuur of patenten verkregen. Nucleotide sequentie databases kunnen gebruikt worden om sequenties in verschillende formats te downloaden of te analyseren; er kan bijvoorbeeld gezocht worden naar genen met als output de genomische sequentie met geannoteerde exonen en de eiwitsequentie. Daarnaast kan met behulp van de BLAST (Basic Local Alignment Search Tool) functie een ingevoerde DNA (of eiwit) sequentie vergeleken worden met alle sequenties in de database om homologen te identificeren (8-10).

Het in hoog tempo beschikbaar komen van DNA sequenties maakt dat er veel DNA varianten gevonden worden. De meest voorkomende DNA varianten betreffen de substitutie van één nucleotide door een ander nucleotide en worden ook wel 'Single Nucleotide Polymorphisms' of SNPs genoemd. Andere DNA varianten zijn deleties, inserties, inversies, duplicaties, translocaties, copy nummer varianten en varianten in het aantal herhalingen van kortere DNA sequenties (repeats). Het is relatief eenvoudig om te bepalen of een DNA sequentie afwijkt van een referentie sequentie. Het is echter veel moeilijker om vervolgens te bepalen of een eventuele DNA variant een effect heeft op de gezondheid. DNA varianten kunnen bepalen of mensen bepaalde ziekten ontwikkelen; een SNP in het APoE gen is bijvoorbeeld geassocieerd met een grotere kans op het ontwikkelen van Alzheimer. Daarnaast kunnen varianten bepalen hoe mensen reageren op bepaalde medicatie. Met behulp van dbSNP, database van SNPs en kleine varianten, en dbVAR, database van genomische structurele varianten, kan uitgebreide informatie ingewonnen worden over verschillende typen DNA varianten (11). ClinVar kan vervolgens gebruikt worden om relaties tussen klinisch belangrijke varianten en fenotypes te bestuderen (12). SNPedia betreft een wiki database met daarin SNPs en hun eventuele medische, fenotypische, forensische of genealogische associaties (13).

Genoombrowsers stellen ons in staat om over het genoom te 'surfen', net zoals we dat op internet kunnen doen met behulp van browsers als Internet Explorer, Chrome of Safari. Voor het doorzoeken van genomische data bestaan er op internet drie grote databases voor nucleair DNA: de UCSC Genome Browser, de Ensembl Genome Browser en de NCBI Mapviewer. Deze browsers bieden een grafische interface voor het zoeken naar onder andere sequenties, genen en moleculaire markers. Daarnaast bieden ze de mogelijkheid om sequenties van delen van het genoom te downloaden en PCR primers te ontwerpen. Ondanks dat deze databases veel vergelijkbare functionaliteiten bezitten, verschillen ze ook op een aantal punten, waaronder grafische presentatie en beschikbare tools voor verdere exploratie. Ook hanteren ze verschillende alignment-methoden, waardoor annotaties soms verschillen. Het is uiteindelijk de voorkeur voor grafische interface en aangeboden gelinkte informatie die de voorkeur voor één van deze drie genoombrowsers zal geven (11,14,15). Behalve genoombrowsers bieden UCSC, Ensembl en NCBI tools voor de analyse van SNPs, genen, DNA methylatie en histonmodificaties, genexpressie en bijvoorbeeld eiwitten. Hiermee zijn ze een handig startpunt voor velerlei analyses. Voor het doorzoeken van mitochondrieel DNA is er MITOMAP. Deze database rapporteert data met betrekking tot variaties in mitochondrieel DNA. Er kan gezocht worden naar mitochondriële genen, naar mitochondriële sequenties en naar nucleotide varianten in de algehele populatie of geassocieerd met ziekte. Met behulp van de MITOMASTER tool, geïncorporeerd in MITOMAP, kunnen zelfs gesubmitte mitochondriële DNA sequenties geanalyseerd worden (16). Genecards is een database doorzoekbaar op gen, met als output een overzichts'kaart' van dat gen. Deze kaart bevat onder andere informatie met betrekking tot aliases, genomische context, SNPs, genfunctie en -expressie, het gecodeerde eiwit, post-translationele modificaties, cellulaire localisatie, routes en interacties en geassocieerde ziekten (17).

Epigenomics

De studie van epigenetische modificaties wordt ook wel epigenomics genoemd. Deze modificaties betreffen reversibele modificaties die de expressie van genen beïnvloeden, zonder de daadwerkelijke DNA sequentie te veranderen. De twee meest gekarakteriseerde modificaties zijn DNA methylatie en histonmodificaties. Beide spelen een rol in velerlei processen, zoals differentiatie en ontwikkeling, maar ook tumorgenese. UCSC, Ensembl, NCBI, maar ook WashU, Washington University, beschikken over zogeheten 'epigenoom' browsers, waarmee epigenomische informatie gevisualiseerd en geanalyseerd kan worden (18). Daarnaast zijn er voor DNA methylatie specifieke databases, zoals PubMeth en MethyCancer, databases gericht op DNA-methylatiepatronen in maligniteiten (19,20). DiseaseMeth is gericht op DNA methylatie in maligniteiten, maar ook in genetische, metabole, neurologische en auto-immuunziekten (21). Alle drie databases zijn onder andere doorzoekbaar op gen en type maligniteit of ziekte. NGSmethDB bevat een

'methyloom'browser, waarmee genomwijde DNA-methylatiepatronen in verschillende weefsels en ziekten kunnen worden bestudeerd (22). Ook voor het bestuderen van histonmodificaties zijn er specifieke databases. HHMD, Human Histone Modification Database, is gericht op de opslag en integratie van experimentele data met betrekking tot histonmodificaties en is doorzoekbaar op onder andere type histonmodificatie, gen en type maligniteit. Daarnaast bevat het een 'histoom'browser, waarmee genomwijde histonmodificaties gevisualiseerd en geanalyseerd kunnen worden (23). Histome bevat een uitgebreider overzicht van histonen en histonmodificaties en ook informatie over histonmodificatoren: de enzymen die de histonen modificeren (24).

Transcriptomics

Het bestuderen van genexpressieprofielen met behulp van high-throughput technieken heeft de afgelopen jaren een grote vlucht genomen. GEO, oftewel Gene Expression Omnibus, bevat genexpressie data verkregen met behulp van array- of sequencing technologie. Wetenschappelijke bladen vragen bij publicatie van dergelijke data om deze te uploaden naar GEO. Gekoppeld aan deze data zijn de specificaties met betrekking tot de gebruikte technieken, gemeten monsters en de experimenten. Een experiment kan bijvoorbeeld het vergelijken van genexpressie profielen tussen twee ziektebeelden zijn. GEO is doorzoekbaar op verschillende manieren. Zo kan er via sleutelwoorden op experiment gezocht worden om genexpressie patronen van interesse te bestuderen. Deze worden weergegeven in de vorm van cluster heatmaps, waarbinnen gezocht kan worden op gen van interesse en naar significante verschillen. Daarnaast kan gezocht worden op gen om expressie van dat specifieke gen binnen verschillende experimenten te bestuderen (25). ArrayExpress is een database die deels dezelfde data en mogelijkheden bevat als GEO (26). BioGPS betreft een database die gebruikt kan worden om expressie van genen van interesse over verschillende weefsels te visualiseren (27).

Relatief nieuwe biomarkers in het veld van de klinische chemie zijn microRNAs (miRNAs). Langzaam maar zeker veroveren zij een plek binnen de diagnostiek, vooralsnog met name in de oncologie, maar ook al bij bijvoorbeeld hart- en vaatziekten. MiRNAs zijn evolutionair sterk geconserveerde kleine (~22-nucleotiden) niet-coderende RNAs. Ze zijn in staat genexpressie te fine-tunen door te binden aan target mRNAs, wat leidt tot mRNA destabilisatie of translationele onderdrukking. Elke miRNA kan tegelijkertijd honderden verschillende mRNAs targeten, en elke mRNA kan verschillende miRNA target sites bevatten. Dit resulteert in een complex interactienetwerk van miRNAs en mRNAs. Met behulp van miRBase, de microRNA database, kan gezocht worden op gepubliceerde miRNA sequenties. Als output worden zowel de stem-loop (hairpin) sequentie als ook de mature sequentie gegeven, de genomische context, eventuele geclusterde miRNAs, voorspelde en gevalideerde target mRNAs en literatuur en database links. miRBase bevat informatie over 2588 mature

Tabel 1. Overzicht databases en webtools.

Databases	Websites (hyperlinks in online-pdf versie)	Referentie
Wie doet wat		
LOD	http://www.dnadiagnostiek.nl/	
EDDNAL	http://www.eddnal.com/	
GeneTests	https://www.genetests.org/	
Ziekten & genafwijkingen		
GeneReviews	http://www.ncbi.nlm.nih.gov/books/NBK1116/	3
OMIM	http://www.ncbi.nlm.nih.gov/omim/	4
OrphaNet	http://www.orpha.net/	5
HGMD	http://www.hgmd.cf.ac.uk/ac/	6
COSMIC	http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/	7
Genomics		
GenBank	http://www.ncbi.nlm.nih.gov/genbank/	8
EMBL-EBI	http://www.ebi.ac.uk/embl/	9
DDBJ	http://www.ddbj.nig.ac.jp/	10
dbSNP	http://www.ncbi.nlm.nih.gov/snp/	11
dbVar	http://www.ncbi.nlm.nih.gov/dbvar/	11
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/	12
SNPedia	http://www.snpedia.com/index.php/SNPedia/	13
UCSC	http://genome.ucsc.edu/	14
Ensembl	http://www.ensembl.org/	15
NCBI	http://www.ncbi.nlm.nih.gov/	11
MITOMAP	http://www.mitomap.org/MITOMAP/	16
GeneCards	http://www.genecards.org/	17
Epigenomics		
Washu	http://epigenomegateway.wustl.edu/browser/	18
PubMeth	http://www.pubmeth.org/	19
MethyCancer	http://methycancer.psych.ac.cn/	20
DiseaseMeth	http://bioinfo.hrbmu.edu.cn/diseasemeth/	21
NGSmethDB	http://bioinfo2.ugr.es/NGSmethDB/	22
HHMD	http://bioinfo.hrbmu.edu.cn/hhmd/	23
HIstome	http://www.actrec.gov.in/histome/	24
Transcriptomics		
GEO	http://www.ncbi.nlm.nih.gov/geo/	25
ArrayExpress	https://www.ebi.ac.uk/arrayexpress/	26
BioGPS	http://biogps.org/	27
miRBase	http://www.mirbase.org/	28
TargetScan	http://www.targetscan.org/	29
miRTarBase	http://mirtarbase.mbc.nctu.edu.tw/	30
PolymiRTS	http://compbio.uthsc.edu/miRSNP/	31
Proteomics		
UniProt	http://www.uniprot.org/	32
PROSITE	http://prosite.expasy.org/	33
Pfam	http://pfam.xfam.org/	34
PDB	http://www.rcsb.org/pdb/	35
LOCATE	http://locate.imb.uq.edu.au/	36
GO	http://geneontology.org/	37
PRIDE	http://www.ebi.ac.uk/pride/	38
PPD	http://www.plasmaproteome.com/	39
Metabolomics		
HMDB	http://www.hmdb.ca/	40
Interactomics		
KEGG	http://www.genome.jp/kegg/	41
Reactome	http://www.reactome.org/	42
BioCyc	http://biocyc.org/	43
MetaCyc	http://www.metacyc.org/	43
STRING	http://string-db.org/	44
Overzicht websites		
ExPASy	http://www.expasy.org/	45

miRNA sequenties (nov 2014). Bij het vinden van een miRNA die tot overexpressie komt bij een bepaald ziektebeeld, kan middels miRBase gezocht worden naar de mogelijke mRNA targets en daarmee de pathofysiologische verklaring. Andersom kan ook gezocht worden op sleutelwoord, bijvoorbeeld op een bepaald ziektebeeld, om te zien of er specifieke miRNAs mee geassocieerd zijn. Daarnaast kan gezocht worden op miRNAs die op een bepaalde plek in het genoom liggen, op miRNAs die in clusters liggen, op miRNAs die in bepaalde weefsels tot expressie komen en op sequentie (28). TargetScan kan gebruikt worden om mRNA targets van specifieke miRNAs te voorspellen of om miRNA target sites in het gen van interesse op te sporen (29). miRTarBase bevat experimenteel gevalideerde miRNA-target interacties (30). Met behulp van PolymiRTS kunnen polymorfismen in miRNAs of miRNA target sites gelinkt worden aan ziekten en biologische routes (31).

Proteomics

Uiteindelijk worden de biologische processen binnen cellen niet bepaald door mRNAs, maar door eiwitten. UniProt, de Universal Protein Resource, is een database die geraadpleegd kan worden om eiwitten te bestuderen. Het bevat eiwit sequenties en informatie over de biologische processen waarin het eiwit betrokken is, post-translationele modificaties, interacties met andere moleculen en de lokalisatie in cellen en organismen. UniProt betreft een samenvoeging van de Swiss-Prot, TrEMBL en PIR (Protein Information Resource) databases (32). De PROSITE en Pfam databases bevatten informatie over eiwitdomeinen, -families en functionele sites en kunnen gebruikt worden om de functie van eiwitten of om bijvoorbeeld de consequentie van een DNA mutatie op de functie van een eiwit te voorspellen (33,34). De functie van een eiwit is sterk afhankelijk van zijn 3D-structuur. De PDB, Protein Data Bank, bevat informatie over 3D configuraties van macromoleculen, waaronder nucleïnezuren, eiwitten en complexe structuren zoals ribosomen. Deze configuraties zijn vastgesteld met behulp van röntgenkristallografie en NMR. PDB kan gebruikt worden om 3D structuren te analyseren en visualiseren (35). LOCATE is een database waarmee de subcellulaire locatie van een eiwit achterhaald kan worden (36). In de GO, Gene Ontology, database zijn eiwitten ingedeeld in groepen, op basis van biologische processen, cellulaire componenten en moleculaire functies. Er kan gezocht worden op GO-groep om bijvoorbeeld alle eiwitten te vinden die een rol spelen in signaaltransductie, of op individueel eiwit om alle GO-groepen te vinden waar dit eiwit deel van uit maakt (37). Naast dat de sequenties en GO-groepen gebruikt kunnen worden om eiwitfunctie te voorspellen, kan tevens expressie van eiwitten in verschillende weefsels of ziekten iets zeggen over de functie. De PRIDE, PRoteomics IDentifications, database is een publieke opslag voor op massa spectrometrie (MS)-gebaseerde proteomics data. PRIDE is onderzoekbaar op onder andere type dataset (weefsel, ziekte of apparaat), eiwit ID en peptide sequentie. Het bevat daarnaast een tool om data verkregen met behulp van MS te visualiseren

en te analyseren (38). De PPD, Plasma Proteome Database, is een database waarin eiwitten gevonden in plasma (en serum) van gezonde individuen opgeslagen liggen. Inmiddels zijn er >9.000 eiwitten gevonden met behulp van op MS-gebaseerde methoden of immunoassays (nov 2014). PPD kan gebruikt worden als referentie database bij onderzoek naar nieuwe biomarkers (39).

Metabolomics

Het bestuderen van metabolieten en het identificeren van unieke chemische fingerprints wordt ook wel metabolomics genoemd. De HMDB, Human Metabolome Database, faciliteert de identificatie en karakterisatie van metabolieten middels NMR-, GC-MS-en LC/MS spectrometrie. HMDB bevat drie typen data: chemische, klinische en moleculair biologische/biochemische data. Ook bevat het informatie over medicijnen en gerelateerde metabolieten, toxinen en voedselcomponenten en -additieven. Er bestaat een mogelijkheid tot browsen door metabolieten, ziekten of metabole routes, maar ook tot specifiek zoeken op metaboliet, sequentie, chemische structuur, verwantschap, NMR- en MS(/MS)-spectra. Middels de SMPD, Small Molecule Pathway Database, (gelinkt via HMDB) kunnen metabole routes gevisualiseerd worden (40).

Interactomics

Om biologische of metabole processen of systemen te doorgronden, kunnen eiwitten bestudeerd worden in de context van een netwerk. KEGG, oftewel de Kyoto Encyclopedia of Genes and Genomes, Reactome, BioCyc en MetaCyc zijn databases voor de visualisatie, interpretatie en analyse van biologische en metabole routes ('pathways'). Deze worden weergegeven in de vorm van reactie- en interactienetwerken. In deze netwerken worden veranderingen in de status van moleculen weergegeven, zoals binding, activatie, translocatie en degradatie, naast klassieke biochemische reacties (41-43). STRING, Search Tool for the Retrieval of Interacting Genes/Proteins, is een database die bestaande en voorspelde eiwit interacties bevat. De interacties omvatten zowel directe (fysieke) als indirecte (functionele) associaties, afkomstig uit datamining van genomische context, high-throughput experimenten, (geconserveerde) coexpressie of uit textmining van databases en publicaties. STRING is doorzoekbaar op eiwitnaam of -sequentie en geeft als output een kleurrijk netwerk bestaande uit eiwitten die onderling verbonden zijn door lijnen. De verschillende kleuren lijnen geven verschillende soorten associaties aan. STRING importeert data uit andere databases als MINT, HPRD, BIND, DIP, BioGRID, KEGG, Reactome, IntAct, NCI en GO (44).

Systeembioologie

Het bestuderen van complexe systemen als geheel, onder meer door het integreren van alle hier genoemde data-levels, wordt systeembioologie genoemd. De systeembioologie kan behulpzaam zijn bij het ontrafelen van pathofysiologische processen, maar vereist wel dat medisch (laboratorium)specialisten, onderzoekers en bioinformatici dezelfde taal (leren) spreken.

Tot slot

Het hier beschreven overzicht (zie Tabel 1) is verre van compleet, maar bevat wel de meest praktische databases en webtools voor de hedendaagse klinisch-chemische praktijk of onderzoeksdoeleinden. Voor de echte klinisch bioinformatici onder ons zijn meer en specifiekere databases te vinden via websites als ExPASy, Bioinformatics Resource Portal of de speciaal aan databases gewijde edities van Nucleic Acids Research (45,46). Rest nog te vermelden dat in de huidige technologische en digitale revolutie de ontwikkelingen zich razendsnel opvolgen; wat vandaag nieuw is, is morgen alweer verouderd. Zo ook een overzicht van klinisch genetische databases.

Referenties

1. Planningsbesluit klinisch genetisch onderzoek en erfelijkheidsadviesing. VWS. Gewijzigd 16-02-2009; Wet bijzondere Medische Verrichtingen, artikelen 2 en 5.
2. Professionele standaard 'Klinische Chemie en Laboratoriumgeneeskunde'. 2013.
3. Pagon RA, et al., editors. GeneReviews® (Internet). Seattle (WA): University of Washington, Seattle; 1993-2014. 21 nov 2014.
4. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). 21 nov 2014.
5. Orphanet: an online database of rare diseases and orphan drugs. Copyright, INSERM 1997. 21 nov 2014.
6. Stenson PD, et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet. 2014; 133: 1-9.
7. Forbes SA, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015. (Epub ahead of print).
8. Benson DA, et al. GenBank. Nucleic Acids Res. 2015. (Epub ahead of print).
9. Brooksbank C, et al. The European Bioinformatics Institute's data resources 2014. Nucleic Acids Res. 2014; 42: D18-25.
10. Kosuge T, et al. DDBJ progress report: a new submission system for leading to a correct annotation. Nucleic Acids Res. 2014; 42: D44-49.
11. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2015. (Epub ahead of print)
12. Landrum MJ, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42: D980-985.
13. Cariaso M, et al. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. Nucleic Acids Res. 2012; 40: D1308-1312.
14. Rosenbloom KR, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 2015. (Epub ahead of print)
15. Cunningham F, et al. Ensembl 2015. Nucleic Acids Res. 2015. (Epub ahead of print)
16. Lott MT, et al. mtDNA variation and analysis using MITO-MAP en MITOMASTER. Current Protocols in Bioinformatics. 2013; 44: 1.23.1-26.
17. Stenzer G, et al. In-silico human genomics with GeneCards. Hum Genomics. 2011; 5: 709-717.
18. Zhou X, et al. The human epigenome browser at Washington University. Nature Methods. 2011; 8: 989-990.
19. Ongenaert M, et al. PubMeth: a cancer methylation database combining text-mining and expert annotation. Nucleic Acids Res. 2008; 36: D842-846.
20. He X, et al. MethyCancer: the database of human DNA methylation and cancer. Nucleic Acids Res. 2008; 36: D836-41.
21. Lv J, et al. DiseaseMeth: a human disease methylation database. Nucleic Acids Res. 2012; 40: D1030-1035.
22. Geisen S, et al. NGSmethDB: an updated genome resource for high quality, single-cytosine resolution methylomes. Nucleic Acids Res. 2014; 42: D53-59.
23. Zhang Y, et al. HHMD: the human histone modification database. Nucleic Acids Res. 2010; 38: D149-154.
24. Khare SP, et al. H1stome—a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res. 2012; 40: D337-242.
25. Barrett T, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013; 41: D991-995.
26. Rustici G, et al. ArrayExpress update—trends in database growth and links to data analysis tools. Nucleic Acids Res. 2013; 41: D987-990.
27. Wu C, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 2009; 10: R130.
28. Kozomara A, et al. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011; 39: D152-157.
29. Lewis BP, et al. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 120: 15-20.
30. Hsu SD, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2014; 42: D78-85.
31. Bhattacharya A, et al. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res. 2014 ;42: D86-91.
32. The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014; 42: D191-8.
33. Sigrist CJ, et al. New and continuing developments at PROSITE. Nucleic Acids Res. 2013; 41: D344-347.
34. Finn RD, et al. The Pfam protein families database. Nucleic Acids Res. 2014; 42: D222-230.
35. Berman HM, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28: 235-242.
36. Sprenger J, et al. LOCATE: a mammalian protein subcellular localization database. Nucleic Acids Res. 2008; 36: D230-233.
37. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. Nat Genet. 2000; 25: 25-29.
38. Vizcaino JA, et al. The Proteomics Identification (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2013; 41: D1063-1069.
39. Nanjappa V, et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic Acids Res. 2014; 42: D959-965.
40. Wishart DS, et al. HMDB 3.0—The Human Metabolome Database in 2013. Nucleic Acids Res. 2013; 41: D801-807.
41. Kanehisa M, et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014; 42: D199-205.
42. Croft D, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2014; 42: D472-477.
43. Caspi R, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2014; 42: D459-471.
44. Franceschini A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41: D808-815.
45. Artimo P, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012; 40: W597-603.
46. Nucleic Acids Res. Database issues. 2013/ 2014/ 2015