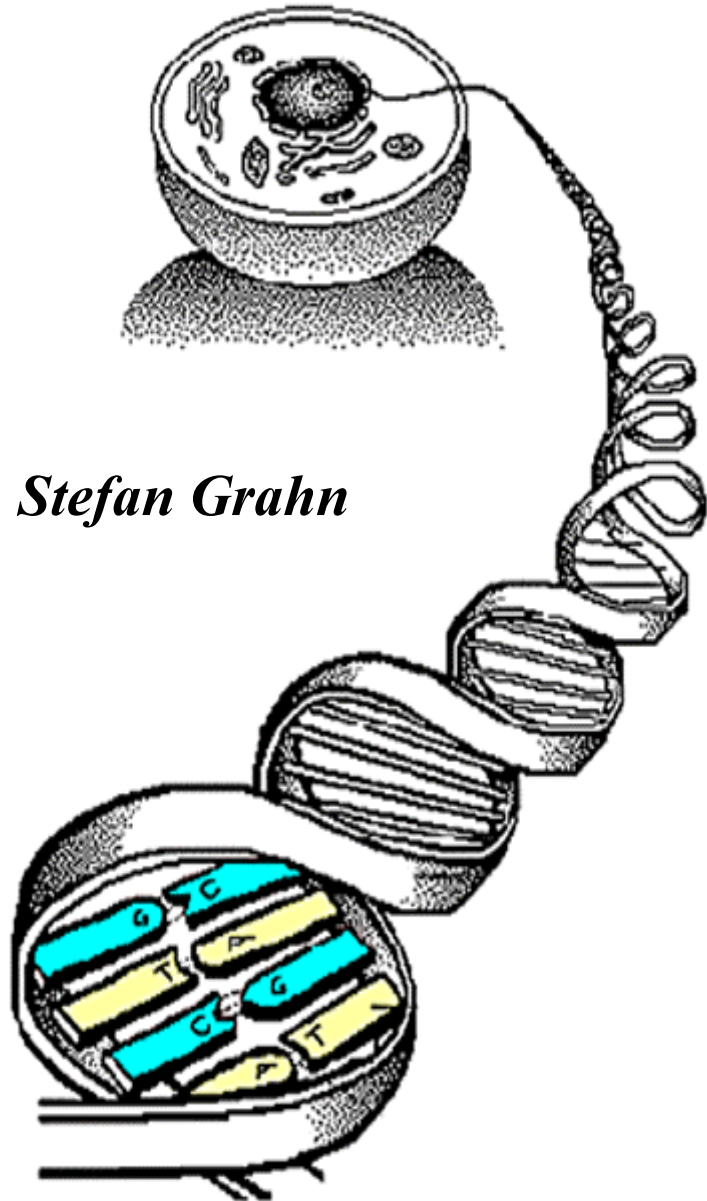


Analyser med Microarray

- ett förslag till databasmodell



Stefan Grahn



Institutionen för informatik, Handelshögskolan vid Göteborgs universitet
Examensarbete 10 p, ingående i ADB-Programmet, våren 2000
Handledare: Roy Corneliusson

Abstrakt

Den molekylärbiologiska forskningens stora problem idag är att kunna finna relevanta samband i all den information som produceras från laboratorier världen över. En av grundförutsättningarna för lösningen på detta problem ligger i lagringen av den biologiska informationen. Arbetet gav en introduktion till begreppet bioinformatik samt en populärvetenskaplig grund till molekylärbiologi. Som lösning på problematiken med lagring och inhämtande av information på ett forskningslaboratorium, redovisades en databasmodell att tillämpas vid laborationer för analys av genuttryck med microarray teknik. Modellen fungerar också som en mall för andra typer av molekylärbiologisk forskning. Arbetet rekommenderade också ett initiativ till ett gränsöverskridande samarbete mellan Institutionen för informatik och institutioner inom den medicinska forskningsvärlden.

Ett särskilt tack till

Thanos Magoulas som i ett tidigt skede ledde mig rätt i modelleringen med de bevingade orden: ”det är fel alltså”.

Tore Samuelson som gav mig möjligheten att studera en del av forskningen på institutionen för Medicinsk Biokemi.

Stefan Scheidl som genom hela mitt arbete, med ett gott humör och tålamod, svarat på alla frågor om molekylärbiologi och microarray.

INNEHÅLLSFÖRTECKNING

INLEDNING	6
BAKGRUND	7
VAD ÄR BIOINFORMATIK?	7
<i>Databaser</i>	7
<i>Informationsnätverk</i>	8
<i>Data mining</i>	8
VEM ÄR BIOINFORMATIKERN?	9
BIOLOGIN	10
<i>Livet</i>	10
<i>Cellen</i>	11
<i>Kromosomen</i>	12
<i>DNA molekylen</i>	12
<i>RNA</i>	13
<i>cDNA</i>	14
<i>Proteiner</i>	14
<i>PCR</i>	14
<i>Microarray</i>	15
SYFTET MED STUDIEN	17
METOD	17
RESULTAT	18
OBJEKTTABELLER	18
<i>SCIENTIST</i>	18
<i>LABORATION</i>	18
<i>MATERIA</i>	18
<i>SLIDE</i>	19
<i>PLATE</i>	19
<i>PLATEWELL</i>	19
<i>FILE</i>	19
PROCESSTABELLER.....	20
<i>TISSUE_PREPARATION</i>	20
<i>LASER_CAPTURING</i>	20
<i>RNA_PREPARATION</i>	21
<i>MRNA_PREPARATION</i>	21
<i>AMPLIFICATION</i>	21
<i>PROBE_PREPARATION</i>	21
<i>PCR</i>	22
<i>PRECIPITATION</i>	22
<i>COATING_SLIDES</i>	22
<i>PRINTING_SLIDES</i>	22
<i>HYBRIDIZATION</i>	22
<i>SCANNING</i>	23
ÖVRIGA TABELLER	23
<i>WELL_CLUSTERING</i>	23
<i>LABORATION_FILE</i>	23
DATABASMODELL	24

DATABASMODELL FÖR ANALYSER MED MICROARRAY

PROTOTYP	25
DISKUSSION	26
SLUTSATS.....	29
REFERENSER	30
TABELLBILAGOR.....	
BILAGA 1A.....	
BILAGA 1B.....	
BILAGA 1C.....	
BILAGA 1D.....	

Inledning

Informatikern kan idag använda sin kompetens inom de flesta verksamheter för att lösa olika problem. En starkt växande vetenskap, som kan sägas vara gränssnittet mellan biologi och informationsteknologi, är Bioinformatiken.

Molekylärbiologisk forskning ("genforskning"), har kommit så långt att det inte är något problem att samla information om livets byggstenar. Att lagra informationen är tack vare datorernas minneskapacitet heller inget problem. Det accelererande problemet är, att hantera all den biologiska information som produceras på laboratorier världen över, på ett sådant sätt att forskarna kan urskilja relevanta data. Höga krav ställs på informationens lagring, organisering, indexering och bra visualisering i användargränssnitten.

Det är här som Bioinformatikern med sin kunskap, å ena sidan om biologi, och å andra sidan om informatik, skall vara tongivande. Detta är en kunskapsprofil som är svår att matcha. Även om det idag ges tillfälle till dessa tvärvetenskapliga utbildningar, så kommer bristen på utbildade bioinformatiker bestå en lång tid framåt på grund av det ständigt ökande behovet av analys- och beslutsstödsystem inom området. Att arbeta tvärvetenskapligt betyder också att arbetet måste ske utanför tidigare givna ramar, det ställer därmed krav även på det personliga planet. Ett nytänkande och en ödmjukhet inför andra vetenskaper, istället för fyrkantighet och revirtänkande vid en arbetsansats och problemanalys är nödvändigt.

Detta examensarbete ger först en introduktion av begreppen bioinformatik och molekylärbiologi, med en populärvetenskaplig ansats på det senare. Därefter redovisas syftet, metoden och resultatet med en modell över föreslagen databas i nämnd ordning.

Examensarbetet är gjort vid Institutionen för Medicinsk Biokemi, Sahlgrenska Universitetssjukhuset, Göteborg.

Bakgrund

Vad är Bioinformatik?

Det finns flera olika definitioner : ”*utveckling och applicering av datasystem och datorstödd problemlösning, för behandling av resultat från experiment, modellering och databassökning som rör biologiska aspekter*” [1], ”*gränslandet mellan biologi, matematik och datalogi*”. [2] Andra definitioner är ”en teoretisk, datorbaserad forskning” och ”informationsteknologi tillämpat på biologi.”

Databaser

Grundstenen för Bioinformatik är databaserna. Både de publika, vilka började utvecklas i mitten på 1980-talet, och som är tillgängliga för alla som vill studera och jämföra resultat av tidigare forskning, samt de databaser som byggs upp av enskilda forskare i laboratorierna och innehåller information om laborationers resultat och parametrar.

De publika databaserna kan indelas i tre typer, beroende på vad de innehåller.

Primära databaser lagrar biomolekylära sekvenser, dvs delar av genomet (DNA) eller olika proteiner. De innehåller även associerad information som organism, art samt olika mutationer länkade till särskilda sjukdomar etc. De uppdateras av varandra dagligen.

Exempel på sådana databaser är för DNA sekvenser EMBL¹, GenBank² och DDBJ³. Exempel på databaser för proteinsekvenser är PIR⁴ och MIPS⁵.

Sekundära databaser, kallas även mönsterdatabaser, eftersom de innehåller resultaten av sekvenserna i de primära databaserna. På grund av de många olika primärdatabaserna med sina skilda sätt att analysera proteinsekvenser, så är både informationen och dess format olika. Detta kan ge forskaren en viss osäkerhet i de resultat som presenteras. En databas som har blivit populär är SWISS-PROT. Förmodligen beror det på att den har minimalt med redundant data och att den är korsrefererande med ca 30 andra databaser.

Tertiära databaser hämtar rådata från de sekundära databaserna. Syftet med de tertiära databaserna är att de automatiskt skall producera resultat som ger forskaren ett annat perspektiv på sambanden mellan den biologiska informationen, än vad de sekundära databaserna gör. På detta sätt kan samband hittas som inte kommer fram vid en normal analys av filtrerade data från sekundärdatabasen.

Exempel på tertiära databaser är OMIM⁶ och Entrez⁷ [8]

¹ European Molecular Biology Laboratory,

² USA, lyder under National Institute of Health

³ DNA Data Bank of Japan, samarbetar med EMBL och GenBank

⁴ Protein Information Resource. Ett samarbete mellan National Biomedical Research Foundation, International Protein Information Database of Japan och Munich Information Centre for Protein Sequences (MIPS)

⁵ Munich Information Centre for Protein Sequences, Tyskland. Bioinformatikgrupp för GSF (Forschungszentrum für Umwelt und Gesundheit)

⁶ Online Mendelian Inheritance in Man.

⁷ Arbetar på information som finns på databaser hos National Center for Biotechnology Information (NCBI)

En sammanställning av publika databaser är gjord av Andreas D. Baxevanis . [3]

Laboratoriernas experimentdatabaser innehåller informationen om alla de parametrar som hör till de olika experimenten som utförs lokalt.

Det är viktigt för forskningsgruppen att kunna härleda ett resultat för att försöket skall kunna upprepas med eller utan förändringar. Anledningar till förändringar i resultatet skall också kunna spåras bakåt. Resultatet av analysen från informationen i dessa databaser kommer i sin tur att leda till nya experiment, vilka genererar ytterligare information

Informationsnätverk

Informationsnätverk syftar till att underhålla databaser och tillhandahålla olika verktyg för biologiska beräkningar (biocomputing).

I Europa bildades 1988 EMBnet⁸. Det länkar samman de laboratorier i Europa som använder sig av bioinformatik. 1998 bestod det av 34 noder, 20 av dessa är nationella och sanktionerade av respektive land. Åtta av noderna är specialistnoder, dessa är akademiska-, industriella- eller rena forskningscentra som anses ha särskilt god kunskap inom området. Därutöver finns det sex stycken associerade icke Europeiska databaser som är av samma struktur som de nationella noderna. [8]

En motsvarighet i USA är NCBI⁹, det etablerades 1988 som en del i National Library of Medicine (NLM). NCBI:s roll är främst att skapa automatiserade system för lagring och analys av biologisk information. Sedan 1992 ansvarar man även för GenBank, en sekvensdatabas som hör till National Institute of Health (NIH).

I Asien är det Japan som är framträdande inom området och deras nätverk/informationsleverantör heter JIPID¹⁰.

Data mining

Bioinformatik handlar mycket om att utveckla programvaror för att kunna uttolka de stora datamängderna. Som jämförelse kan man säga att informationen i en cells DNA motsvarar en enda lång teckensträng som skulle behöva ett utrymme på 500 000 A4 sidor. I denna teckensträng skall forskarna extrahera relevanta data samt identifiera dess innebörd. Informationsnätverken har naturligtvis långtgående verktyg, men det finns också mjukvaruleverantörer som kan förse forskare med programvaror vilka kan hjälpa till med att bl.a. filtrera och ta fram samband mellan olika resultat som lagrats i laboratoriets databas, men många problem kvarstår.

Laboratorieutrustning med dess mjukvara är inte kompatibla med andra leverantörers utrustning och programvara. Programmen använder också olika algoritmer för beräkning, vilket gör att forskarvärlden inte fullt ut kan jämföra sina resultat på ett tillförlitligt sätt.

⁸ The European Molecular Biology network.

⁹ National Center for Biotechnology Information.

¹⁰ International Protein Information Database of Japan.

Vem är Bioinformatikern?

Bioinformatiken kan sägas ha sin födelse i mitten på 1970-talet. I den lilla skala som det då handlade om var det främst molekylärbiologer och biokemister som av intresse lärde sig lite om programmering och statistik. Med alltmer avancerade instrument samlar forskarna idag mer information om arvsmassan än tidigare. Detta ställer i sin tur allt högre krav på bioinformatikern.

Materialet och frågorna kommer från molekylärbiologin medan metoderna för att få fram ett beslutsunderlag kommer från datalogin och matematiken. Det är viktigt att påpeka att svaren alltid måste analyseras av forskaren eftersom datorn inte säger sanningen utan bara visar på ett statistiskt resultat.

Forskaren drar sålunda slutsatser med hjälp av sina kunskaper inom biologin och har till sin hjälp verktyg från informatiken och matematiken i samspel.
Docent Siv Andersson [2] illustrerar detta samspel enligt följande:

”En modern molekylärbiolog utan kontakter med matematiken/datalogin kan liknas vid ett barn utan pengar i en godisaffär, medan en matematiker/datalog utan kunskaper i biologi kan liknas vid en isolerad person i öknen – även med all världens pengar kan han inte få tag i något godis! Det är uppenbart att barnet i godisaffären (biologen/medicinaren) måste kombineras med plånboken i öknen (matematikern/datalogen). Redskapet som står till förfogande heter utbildning!”

Utbildning inom bioinformatik finns idag på universitet och högskolor. Men kommer det att finnas intresserade studenter? Problemet har varit och är att hitta personer med intresse och kompetens för detta gränsöverskridande område.

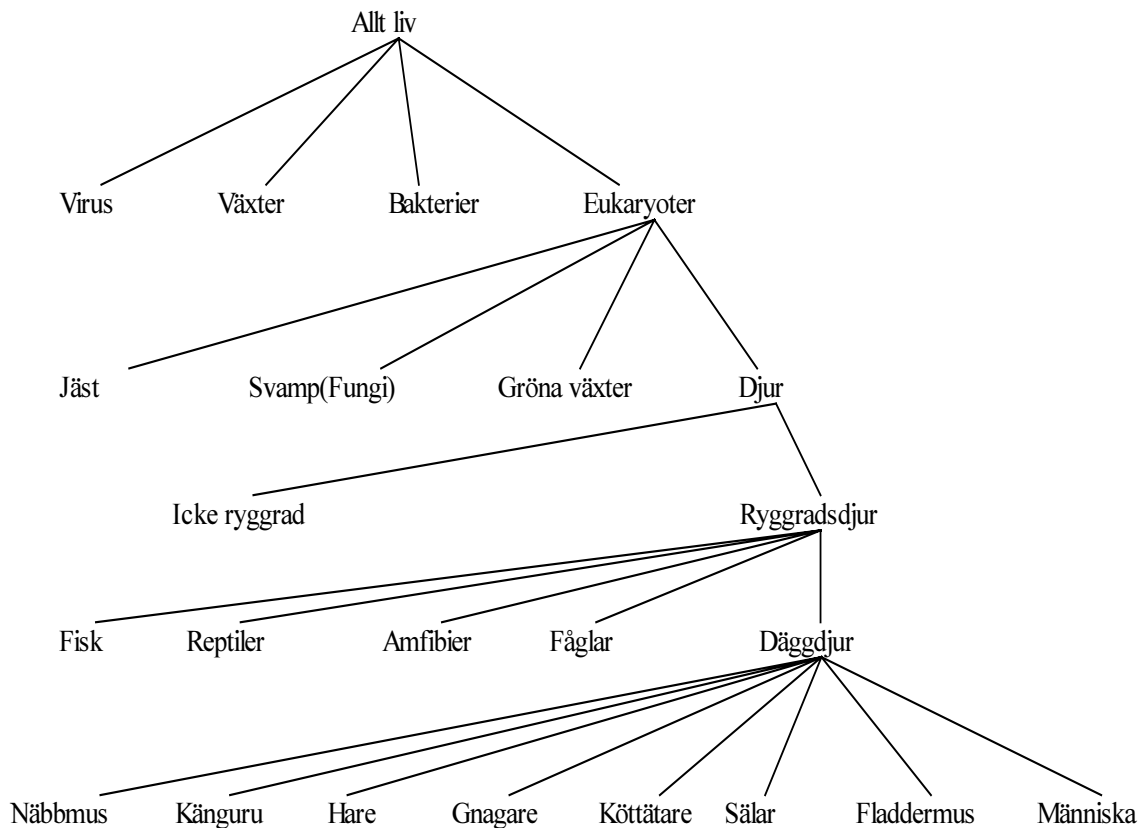
Biologin

Denna del är tänkt att ge en grundläggande kunskap för den som inte är så bevandrad inom biologin. Precis som andra vetenskaper använder sig biologin av fackuttryck som uteslutande är på engelska. Dessa uttryck är ibland svåra att översätta på ett korrekt sätt varför facktermer i min rapport kommer att skrivas på engelska, men förklaras.

Livet

Att definiera vad liv är inte enkelt, det centrala är att det måste kunna reproducera sig självt. En annan aspekt är att levande organismer har möjligheten att ta till sig av den energi som materian i organismens närmiljö innehåller, samt omvandla den till nya delar av sig själv eller sin avkomma.

Många menar att människan blev till vad den är idag genom evolutionen; Variation och selektionstryck.(se figur 1) Vissa teorier gör också gällande att vi långt tillbaka i tiden har samma anfader, vilket till vissa delar bevisas av genforskningen.

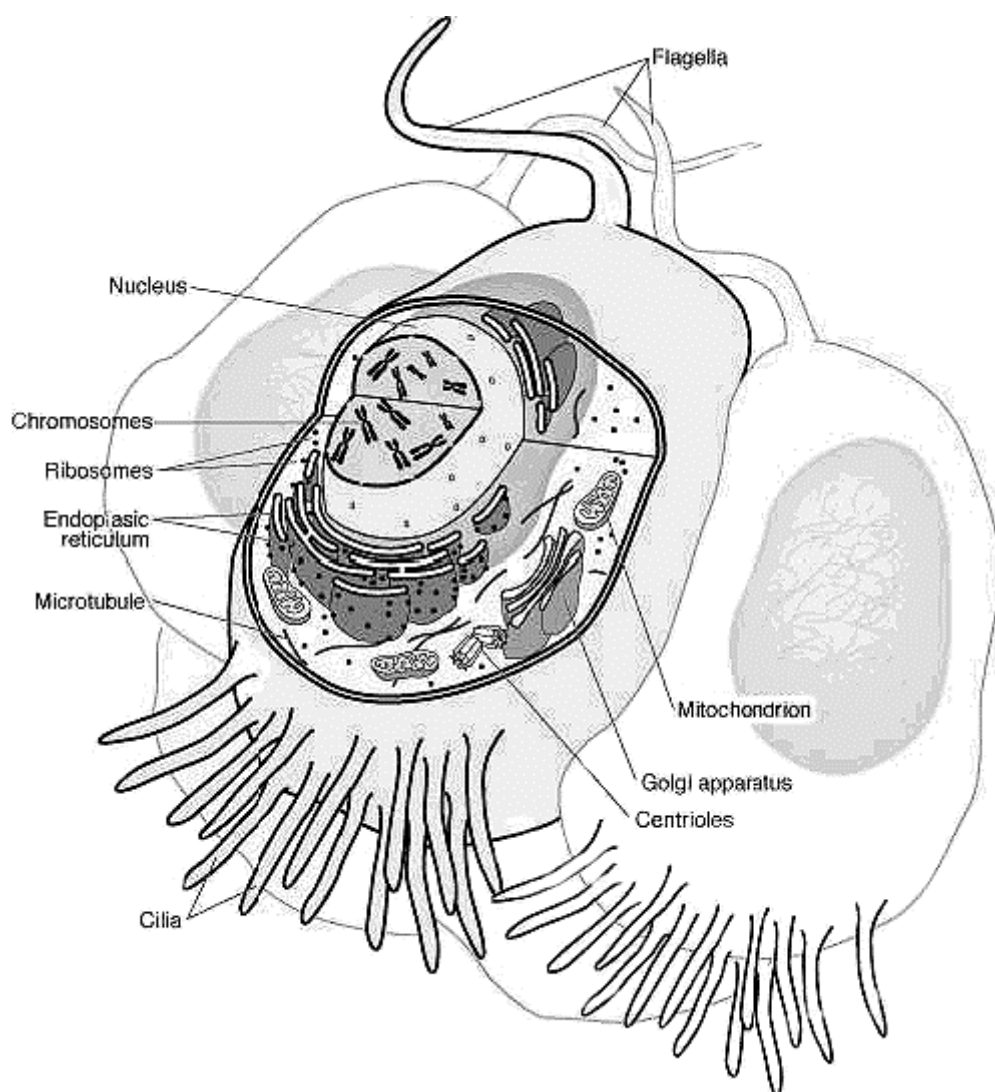


Figur 1: En av flera teorier om livets utveckling och samband.

Cellen

Människan har fjorton större vävnadstyper [5], det kan vara muskler, ben eller blod. Varje vävnad har sin typ av cell.

I cellen finns mängder av organeller¹¹, bl.a. mitokondrier¹² som har sitt eget DNA, men vad som skall fokuseras här är den del av cellen som kallas cellkärnan (nucleus). Kärnan har en diameter på ca 1 micrometer (0,000001 meter) och innehåller bl.a. 46 st kromosomer hos människan. Som en jämförelse kan sägas att hunden har 76 och tomaten 24.



Figur 2 En cell i genomskärning där cellkärnan med kromosomerna syns mitt i bilden

¹¹ Litet funktionellt system

¹² Energifabriken i cellen, har sitt eget DNA

Kromosomen

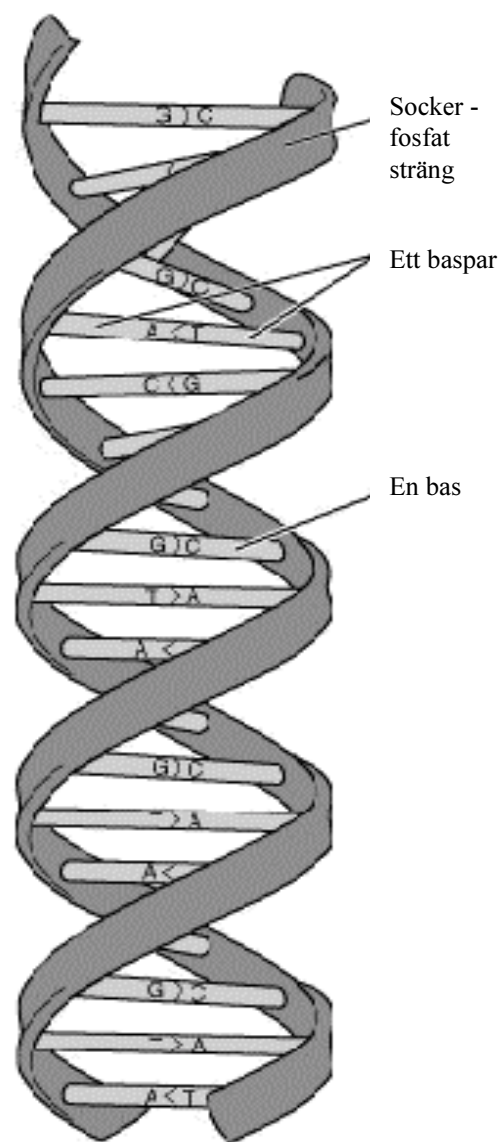
Kromosomen är uppbyggd av en deoxyribonukleinsyra (DNA) och är ca 1 meter lång. DNA har alltså en längd som är 1 miljon gånger större än cellkärnans diameter. Detta är möjligt tack vare att den är tvinnad (inför celledelning) och att den är upplöst. (Jämför med att lösa upp salt i vatten)

DNA molekylen

En hel DNA molekyl (se figur 3) består av delar som kallas för gener¹³. Varje cell i kroppen har en likadan DNA molekyl som består av ca 100 000 gener. Endast 20% av generna beräknas vara ”på” eller ge uttryck i cellen. Med andra ord är det ungefär 20 000 gener i varje cell som ser till att cellen överlever och avgör om cellen skall vara hår, öga eller hud. Allt på DNA-molekylen kodar inte för en gen. Det finns ”brussekvenser”, s.k. introner, mellan de kodande delarna som kallas exoner. En gen består vanligen av flera exoner som ligger utspridda på molekylen.

I maj månad i år publicerade M. Hattori et. al. [7], en artikel där de kartlagt generna för kromosom 21¹⁴. (tidigare kartlagd kromosom är nr 22) Det hävdas i artikeln att kromosomen består av 300 st urskiljningsbara gener och med bakgrund av att kromosomen står för 1-1,5 % av det mänskliga genomet så menar författarna att det totala antalet gener skulle uppskattas till ca 40 000, alltså färre än vad man tidigare trott.

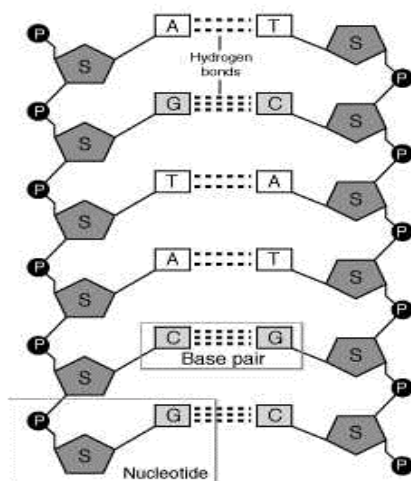
DNA molekylen är uppbyggd av ca 3 miljarder nukleotider. [4] De består främst av socker, fosfat och kväveinnehållande basiska föreningar s.k. baser. Dessa baser är fyra till antalet, Adenin (A), Tymin (T), Guanin (G) och Cytosin (C). Dubbelspiralen bildas genom att två strängar binds ihop där A endast binds ihop med T och där G endast binds ihop med C.(figur 4) Detta faktum utnyttjar naturen, och naturligtvis även forskarna.



Figur 3 Den dubbelsträngade DNA molekylen.

¹³ Genen kan ses som en instans av genomet (DNA)

¹⁴ En extra kopia av kromosom 21 orsakar Down's syndrom, den vanligaste genetiska orsaken till mental retardation



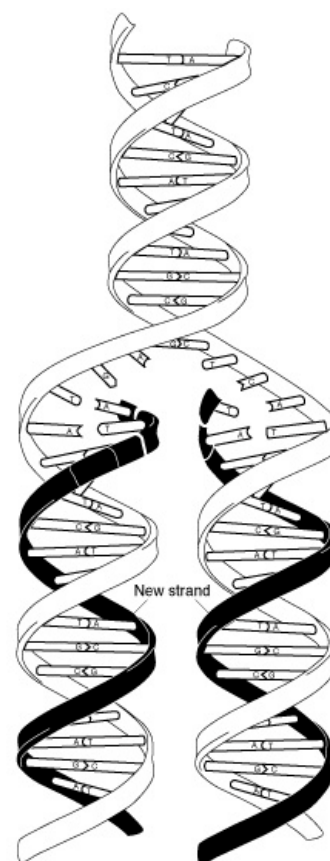
Figur 4. Baserna och deras bindningar

Varje gång en celledelning sker i kroppen delas DNA-strängen upp i två delar och fria baser kan binda in på den enkelsträngade DNA-molekylen. Tack vare bindningsreglerna skapas en kopia av den ursprungliga molekylen. (se figur 5) Forskare gör på liknande sätt med en metod som kallas PCR, vilken beskrivs längre fram. Genen i sig har ingen egen funktion utan är mallen för proteinsyntesen.

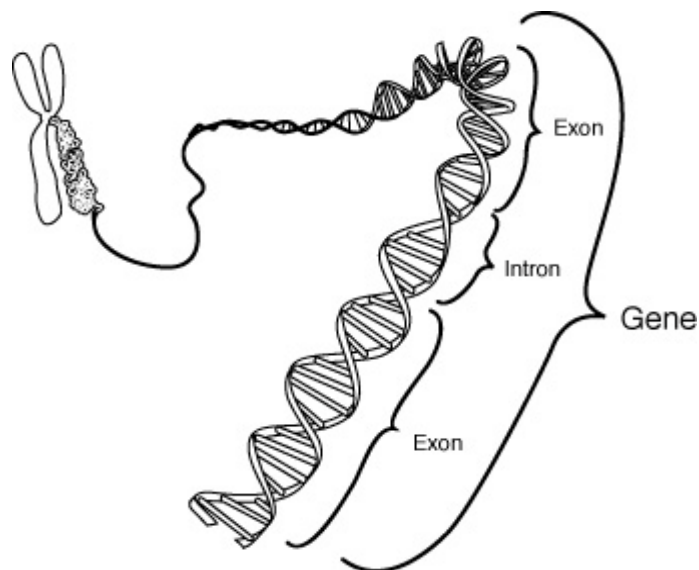
RNA

Ribonucleic Acid (RNA), som upptäcktes 1961 av fransmännen Francois Jacob och Jacques Monod [6], fungerar som en budbärare mellan cellkärnan och den proteinproducerande ribosomen. RNA är påfallande likt DNA med sina baser sammanbundna av socker och fosfatmolekyler, men RNA är enkelsträngat, inte dubbelsträngat som DNA. Dessutom är basen Tymin (T) ersatt av Uracil (U) och sockret i RNA består av ribos medan DNA-socker är deoxyribos.

Messenger RNA (mRNA) utgör en kopia av den genkodande DNA-strängen. Denna bildas av ett RNA-polymeras, vilket är ett enzym som gör en kopia av DNA-strängen på samma sätt som DNA-polymeras gör en DNA-kopia av DNA vid celledelningen. Innan mRNA kan fungera som en mall klipps intronsekvenserna bort och endast de kodande exonerna blir kvar. (jfr. figur 6)



Figur 5. En schematisk bild av DNA-replikering



Figur 6 En bild av kromosomen, DNA molekylen och genen

cDNA

Med hjälp av ett enzym från virus kan det göras en DNA-kopia av mRNA, denna kallas cDNA. Med hjälp av denna kopia kan DNA-tekniker användas på den enbart genkodande delen av DNA, dvs. utan introner.

Proteiner

Proteiner syntetiseras genom att mRNA ”läses av” i ribosomen,¹⁵ varvid en aminosyra tillförs för varje tripplett av nukleotider som utgör den genetiska koden. Det allra mesta i vår kropp består av proteiner och är livsnödvändiga. Exempel på proteiner är hormoner, signalsubstanser och enzymer.

PCR

Polymerase Chain Reaction (PCR) är den mest kända metoden för amplifiering *in vitro*¹⁶ av nukleinsyra. Det är alltså en metod för att kopiera upp stora mängder av en utvald sekvens på DNA-molekylen med hjälp av bl.a. syntetiska oligonukleotider (primers), vilka är nödvändiga som startsekvenser för DNA-syntesen.

Tekniken bygger på repetitiva temperaturskiftningar där steg ett innebär att vid hög temperatur separera det dubbelsträngade DNA som utgör mall (templat), till enkelsträngat DNA. Därefter, i steg två, vid en lägre temperatur få två primers att hybridisera (fästa) till respektive templatsträng. så att, i steg tre, ett särskilt enzym från bakterien *Thermus aquaticus*, som lever vid heta källor och klarar höga temperaturer, kan syntetisera nya komplementära strängar. Denna temperaturcykel i tre steg upprepas 20-40 gånger. Eftersom de nyligen

¹⁵ Cellens proteinfabrik

¹⁶ Betyder ”i glas”, alltså kontrollerade försök exempelvis i provrör. Ungefärlig motsats är ”in vivo”.

syntetiserade strängarna kommer att utgöra templat i nästföljande cykel, växer mängden av amplifierat DNA exponentiellt. [12]

Microarray

Tekniken med microarray är ett genombrott inom molekylärbiologin. Den ger möjlighet att studera tusentals gener vid ett analystillfälle.

Ett exempel på hur en analys med microarray tekniken kan gå till är följande. Plattor som består av ett antal brunnar (liten försänkning i plattan), var och en innehållande känt c-DNA¹⁷, placeras i en robot som skall fästa varje brunns innehåll på en glasskiva. Glasskivan har sedan tidigare behandlats för att det genetiska materialet skall fastna. Roboten har plats för tre plattor med cDNA och ca 50 glasskivor. Eftersom roboten hämtar cDNA från 4 st brunnar på en gång, bildas det ett mönster på glasskivan som kan indelas i metarader och metakolumner samt subrader och subkolumner. Detta beror på att avståndet mellan pinnarna på roboten är mycket större än avståndet mellan prickarna på glaset. Det innebär att koordinaterna för genen (brunnen) förändras när den appliceras på glaset och härledbarheten försvinner. Lyckligtvis finns det redan idag programvara som håller ordning på den saken eftersom det totalt i en laboration används 2000-5000 brunnar.

Från en annan process är det framtaget två stycken cDNA, från t.ex. djur eller cellkulturer. De är märkta med var sin fluorescerande färg och läggs på hela den sedan tidigare med cDNA behandlade glasskivan. Därefter läggs glasskivorna i värme över natt för att dessa två komplementära enkelsträngade cDNA skall fästa i varandra (hybridiseras). Inte alla färgade cDNA fastnar på den genpreparerade glasskivan. Endast de cDNA som uttrycks i provmaterialet binder ihop med det kända cDNA som finns på glasskivan. Nästa dag tvättas allt obundet cDNA bort.

Den tvättade glasskivan stoppas in i en scanner som aktiverar de fluorescerande färgerna med hjälp av en laser. En fil i .tif-format skapas där man kan se den ena färgen med olika intensitet för varje prick av cDNA som fästs på glasskivan tidigare. Intensiteten i färgen avgörs av hur mycket som bundit in. (Hur kraftigt uttrycket av den här genen är i cellen vid detta tillfälle.) Därefter scannas glasskivan en gång till för att läsa in den andra färgen.

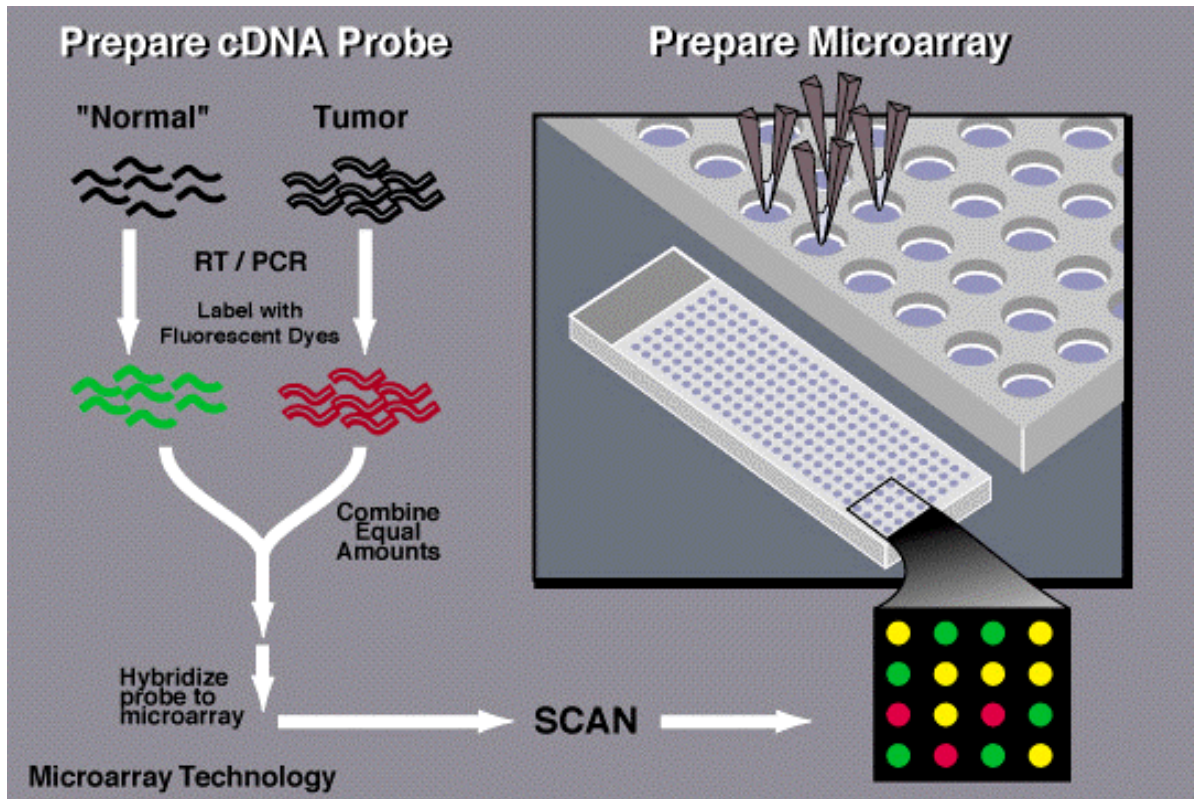
Bilden som forskaren nu ser, är ett antal olika prickar i olika färger som ligger i ett ruttmönster. För att mjukvaran skall kunna göra en beräkning av pixelintensiteten så behöver den veta vad som är en sann signal, bakgrundssignal eller enbart störning. Pixelintensiteten avgör hur kraftigt en hybridisering (genuttrycket) är.

Variationerna, prickarnas utseende och placering beror till stor del på mekaniska begränsningar i processen. Prickarnas storlek är 100–200 µm och har en feltolerans för placeringen på 10µm. För att hantera dessa variationer i lokaliseringen av prickarna krävs en avancerad algoritm som egentligen bara människan behärskar.

Först bestäms positionen på den egentliga pricken och därefter formen på den, dvs vilken del i denna kanske elliptiska form representerar den verkliga signalen. En metod är att forskaren anger de yttre gränserna för ruttmönstret samt gränserna för sub-ruttmönstret. Nästa steg är att bestämma vilka pixlar som är signal och vilka som är bakgrundssignal (image segmentation).

¹⁷ Komplementärt DNA. Framtaget från mRNA med hjälp av ett enzym (reverse transcriptase)

En cirkel placeras över varje prick i bilden för att skilja signal från bakgrundssignal och ytterligare en cirkel för att särskilja vad som är kontamination (felaktiga pixlar).



Figur 7 *Principen för microarray tekniken*

Syftet med studien

Att modellera en databas som kan hantera experimentella parametrar och resultat från analyser med microarray i ett molekylärbiologiskt forskningslaboratorium och utvärdera modellen som en prototyp i MSAccess.

Metod

För insamling av det medicinska bakgrundsmaterialet användes intervjuer och sökningar på internet.

För databasmodelleringen användes lärare på institutionen samt kurslitteratur. Personliga kunskaper från informatikutbildning och ett intresse för molekylärbiologi har också legat till grund för arbetet.

Resultat

Ett antal tabeller redovisas nedan under tre rubriker. Tabellnamn skrivs med kursiva versaler och attributen med kursiva versaler och gemener. Alla attribut i tabellerna redovisas inte här, utan bara de som är intressanta för tabellernas uppkomst. För mer detaljer om tabellernas innehåll hänvisas till bilagorna 1A – 1D.

Under varje tabellnamn redovisas först den biologiska bakgrunden till tabellen (processen), därefter redovisas aspekter som varit vägledande vid modelleringen av databasen. Efter tabellerna redovisas databasmodellen.

Objekttabeller

Dessa tabeller innehåller information om de konkreta objekten i laboratoriets verksamhet. De är stommen i databasmodellen.

SCIENTIST

En enkel tabell som innehåller forskarens *name*, *scientistID*, *e-mail* och vilken *group*¹⁸ vederbörande tillhör. (Bilaga 1C)

LABORATION

Attributen här är också relativt få. *labName*, *labID* och en möjlighet till personliga kommentarer, *comment*, av den forskare som startar upp laborationen.

Kommentarerna är en viktig del som återkommer i nästan varje tabell. På grund av att många processer inte är helt utvecklade eller att forskaren vill uppmärksamma på vissa detaljer är kommentärmöjligheten särskilt viktig i processtabellerna. (Bilaga 1A)

MATERIA

Vid intervjuerna visade det sig att gruppen använde sig av olika typer av ursprung för framställandet av sitt c-DNA. Materia kan vara **odlade celler**, där det var viktigt att veta celltypen, hur många steg den odlats upp samt tillväxtkvalitén. Det kunde också vara **vävnad** från djur, främst laboriemus. Antingen är det ett organ eller ett helt embryo, i båda fallen är det intressant att veta bl. a. eventuella genetiska modifieringar (*genetic modification*)¹⁹, och genetisk bakgrund (*strain*)²⁰, vävnadstyp samt djurets ålder²¹. Materia kan sedan ta olika vägar (processer) in i laborationen.

Eftersom det från början var två olika typer av försöksmateria (cellinje och vävnad) skulle det resultera i två tabeller som hanterade materia. Det faktum att en laboration aldrig kommer att innehålla mer än en materia ledde tankarna till en generalisering varför jag införde attributet *category*, på så sätt kunde jag använda en tabell. (Bilaga 1A)

¹⁸ Namn på forskargruppen.

¹⁹ Exempelvis knockout som kan innebära att en gen undertrycks

²⁰ Exempelvis kan musen heta CE57 Black, man vet då vilken genetisk bakgrund musen har.

²¹ Två exempel på åldrar är 9,5E resp 9,5P. Vilket är ålder i dagar efter befruktning respektive födelse.

SLIDE

En ”slide” är en glasskiva ca 2 x 6 cm, som populärt också kallas för biochip.²² Det speciella med dessa glasskivor är att de måste vara belagda med ett tunt positivt laddat ytskikt för att kunna användas. Antingen köps dessa färdigbehandlade eller så behandlas de på laboratoriet. Det är på denna glasskiva som cirka 2000 gener kommer att fästas, och på dessa kommer experiment materia att appliceras och täckas med ytterligare en glasskiva (obehandlad)

I en laboration används ca 50 glasskivor, vilket gör det viktigt att kunna identifiera var och en. När en laboration förbereds, ”skapas”, numreras alla glasskivor vid ett tillfälle. Då är det viktigt att veta om glasskivan blivit ytbehandlad, och om så är fallet, vem som har gjort det och vad det var för sorts ytbehandling. Beroende på om attributet *coated* är sant eller falskt går glasskivan in i processen *PRINTING_SLIDES* eller *COATING_SLIDES*. (Bilaga 1C)

PLATE

En ”plate” är en mikrotiterplatta. Ett plasttråg med exempelvis 96 st brunnar (försänkningar) innehållande olika genetiskt material.

Viktigt här är att en platta existerar i två typer. Antingen är den en originalplatta med införskaffat cDNA. Eller så är den ett replikat²³. Anledningen till att man gör ett replikat är ekonomiskt. Man tar en mindre mängd och kopierar upp erforderlig mängd för försöket (*PCR*). Det är av mycket stor betydelse att plattan och dess innehåll kan härledas. Därför finns attributen *platename* och *plateID*. Tabellen är ett resultat av en generalisering, det finns två typer av plattor som har samma typ av egenskaper och därför har attributet *category* införts. Även här har därför två tabeller blivit en. (Bilaga 1B)

PLATEWELL

Detta är varje brunn i plattan som beskrivs i stycket ovanför. Varje brunn innehåller olika genetiskt material som senare skall fästas på glasskivan.

Brunnen identifieras med koordinater, ett system som laboratorieapparaternas mjukvaror kan hantera. Innehållet i varje brunn identifieras med *wellID*. Det finns också attribut som hanterar genens namn och genens nummer²⁴ på de publika databaserna. (Bilaga 1B)

FILE

Filer är ett komplement till attributet kommentar i vissa relationstabeller.

Anledningen är att vid vissa försök använder man sig av hela protokoll som har en storlek motsvarande en eller flera A4 sidor. Det är sällsynt att dessa blir aktuella vid en informationssökning.

En fil kan också vara innehållsdeklarationen till exempelvis en ytbehandling av glasskivor, tillhandahållen av en extern leverantör.

Det anses fullt tillfredsställande att protokollnamnet, *fileName*, finns sökbart i databasen. Ett protokoll används i flera laborationer och är mycket sällan ändrat. (Bilaga 1A)

²² Istället för transistorer på ett elektroniskt chip består det av mikroskopiska prickar av genetiskt material

²³ Platta med identiskt innehåll som originalplattan fast med mindre volym i varje brunn.

²⁴ *GeneAccessionNumber*, med detta nummer underlättas sökning efter genen på de publika databaserna där man exempelvis kan få reda på andra gener inom samma familj, ”grannar”.

Processtabeller

Jag väljer att redovisa dessa i grupp eftersom de beskriver processerna i varje laboration och har en konkret koppling till forskarens arbete. Alla attribut i tabellerna kommer inte att presenteras, men uppkomsten av de mest essentiella attributen kommer att förklaras.

Det faktum att dessa processer avbildas som kopplingstabeller kommer sig av att en forskare kan göra ett eller flera moment i en eller flera laborationer. De är alltså framtvungade av kardinaliteten²⁵ mellan tabellerna *SCIENTIST* och *LABORATION*.

TISSUE_PREPARATION

Som tabellnamnet antyder så innehåller den information om förberedelse av vävnad.

Intressant i denna tabell är *preparationType*, om man använt laser så måste det finnas en koppling till tabellen *LASERCAPTURING*. Den kopplingen finns via *LABORATION*. För säker identifiering av varje post finns här en tredelad nyckel; *laborationID*, *scientistID* och *materialID*. (Bilaga 1D)

LASER_CAPTURE

Den här metoden använder laser för att skära ut enskilda celler. Det kan göras på ett organ direkt eller på ett helt musembryo. För att förenkla denna process behandlas organet/embryot med ett ytskikt (*embedding*). Det hålls kontinuerligt fryst (*temperature*) och skivas sedan för att kunna delas upp i celler av lasern.

Aktuell *category* från *MATERIA* är enbart djur eller musembryo som kan ha gått igenom processen Tissue preparation, men inte nödvändigtvis eftersom det vid vissa tillfällen är lämpligt att behandla musembryot direkt i den här processen. Nycklar i denna tabell är *scientistID* och *labID* (Bilaga 1A)

²⁵ För att kunna navigera i databasen och omöjliggöra redundant data måste ”många till många förhållanden” elimineras. Det görs enklast med en eller flera kopplingstabeller.

RNA_PREPARATION

En isolering av det totala RNA²⁶ som finns i cellen. Cellen homogeniseras²⁷ och vätskan används sedan vanligtvis i ett kit²⁸. Efter olika delprocesser har man fått ut det totala RNA, dvs. alla typer av RNA, inte hela volymen som fanns i cellen.

Databasen måste kunna hantera det faktum att cellen inte alltid är hämtad från en cellkultur, utan lika gärna kan härstamma från vävnad direkt. Med andra ord måste tabellen kunna kopplas till antingen *MATERIA* eller till *LASERCAPTURING*. Samtidigt som den också likt andra tabeller måste kopplas till *LABORATION* och *SCIENTIST*. (Bil. 1C)

MRNA_PREPARATION

När det totala RNA är isolerat vill forskarna isolera det speciella RNA som kallas messenger-RNA (mRNA), eftersom det används för att skapa cDNA som är likt det vanliga DNA förutom att det är enkelsträngat och saknar introner (Jfr. informatikens defragmenterat)

Kopplingar måste säkerställas till både *RNA_PREPARATION* och *MATERIA* eftersom det är möjligt att använda sig av olika kit vilka antingen kan användas på celler direkt eller via det totala RNA. (Bilaga 1B)

AMPLIFICATION

När forskaren isolerat mRNA så måste volymen av det ökas, vilket genom olika metoder görs i denna process.

Innehållet i denna tabell, likt alla andra, måste kunna härledas till den vävnad som använts och alla de processer som tidigare genomförts på det material som här amplificeras (förstärks) (Bilaga. 1A)

PROBE_PREPARATION

Enkelt beskrivet så är det i denna process som experimentmaterialet färgas för att senare kunna scannas in som en bild. Två prover märks upp med var sin fluorescerande färg, röd och grön (Cy3 och Cy5).

Två olika prober²⁹ används i samma laboration som identifieras med *probeID*. Övrigt att notera är mängden attribut (27 st.) som inte ställer till några problem i databasen men som bör beaktas vid utformande av användargränssnittet. (Bilaga 1C)

²⁶ Ribonucleic Acid. En molekyl som liknar DNA, fast mer instabil. Spelar en avgörande roll vid bildningen av protein. De olika typerna är messenger-RNA, transfer-RNA och ribosom-RNA.

²⁷ Cellen slås sönder för att frigöra alla ämnen.

²⁸ En köpt ”metod” som innehåller både instruktioner och förberedda provrör innehållande en reagens.

²⁹ En kort DNA sekvens som binder in till mRNA

PCR

Denna process har beskrivits tidigare varför jag inte kommer att gå in närmare på den här.

För databasens vidkommande är det nödvändigt att den ger möjlighet till härledning av materialets ursprung, vilken platta och brunn det är hämtat från och slutligen vilket nummer det har i de allmänna databaserna. (Bilaga 1B)

PRECIPITATION

Processen innebär att man förbereder provet för att fästas på glasskivorna. I korthet så används den pelletprodukt som producerats i PCR-processen. Pelleten löses upp och orenheter tvättas bort. Därefter torkas produkten och löses upp i en koncentration som lämpar sig för att användas i roboten som fäster cDNA:t på glasskivorna.

Vad det gäller härledbarhet och kopplingar till *SCIENTIST* och *LABORATION* skiljer sig tabellen inte från andra. (Bilaga 1C)

COATING_SLIDES

För att få cDNA att fästa och sitta kvar på glasskivan, måste den täckas med ett tunt, positivt laddat ytskikt.

När forskaren själv gör denna ytbehandling är det i denna tabell som parametrarna skrivs in. Särskilt viktiga är kopplingarna till *SLIDE* och *SCIENTIST*. Om forskaren vill göra förbättringar avseende prickarnas form på glasskivan så är ytskiktet en av felkällorna. Det är då viktigt att veta vilken typ av ytbehandling som använts. (Bilaga 1A)

PRINTING_SLIDES

Cirka 50 st. glasskivor läggs in i roboten samtidigt med ett mindre antal microtiterplattor (*PLATE*). Roboten fäster det material som finns i plattornas brunnar på glasskivorna. Ytterligare plattor läggs därefter in i roboten tills allt det genetiska material som är behövt, är fäst på glasskivorna. (20-52 st plattor motsvarande 2000-5000 brunnar.)

Vad som händer när plattornas innehåll fästs på glasskivan, är att koordinaterna som brunnarna hade i plattan inte blir desamma på glasskivan. Lyckligtvis finns det, som nämnts tidigare, en programvara tillhörande roboten som länkar samman koordinaterna i en fil. Vad som är viktigt i databasen är att det finns möjlighet att lagra sökvägen till denna fil och koppla ihop den med rätt laboration. (Bilaga 1B)

HYBRIDIZATION

På glasskivan med prickar av cDNA sprids det infärgade provmaterialet som tillåts reagera med kontrollmaterialet i en viss värme över natt. Dagen efter tvättas överskottet bort med, en som forskaren tycker, lämplig metod.

Förutom attributet *washing*, som måste tillåta personliga kommentarer då tvättning kan ske på flera sätt med små skillnader, har tabellen inga särskilt viktiga egenskaper jämför med övriga tabeller. (Bilaga 1A)

SCANNING

Varje glasskiva placeras i en scanner, som med hjälp av laser aktiverar en färg i taget av det provmaterial som tidigare lagts på glasskivan. En inläsning resulterar i en bild av glasskivan med dess färgade prickar. Proceduren upprepas för den andra färgen som också finns på glaset. Det halvt manuella arbetet med att lägga ett ruttmönster över prickarna vidtar. När detta är klart produceras en resultatfil med beräkningar av pixelintensiteten hos prickarna.

Databasen måste kunna lagra sökvägarna till de båda .tif filerna och apparatens fil för inställningar. Resultatfilen måste lagras i databasen uppdelad på sina attribut eftersom det är viktigt med sökbarheten på vissa data. Filen är en tabulerad textfil som kan läsas både av Word och Excel. Det speciella med filen är att den består av två huvuddelar som båda har samma attribut. I databasen löser jag detta med attributet *dataType*. (Bilaga 1D)

Övriga tabeller

Dessa tabeller är ingen representation av laboratoriets verksamhet, utan är med i modellen enbart av den anledning att de skall möjliggöra en identifiering av de kringliggande objekten.

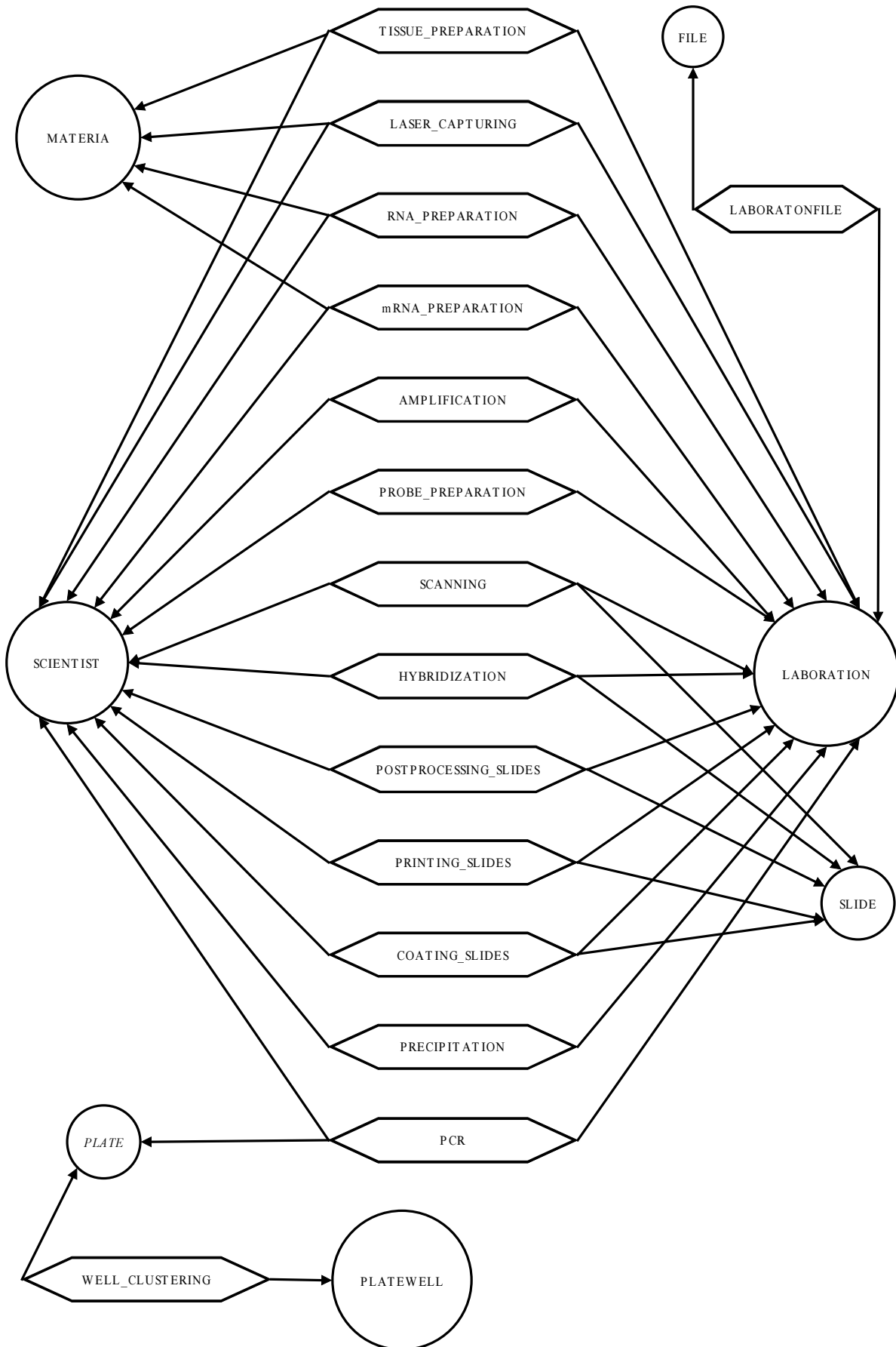
WELL_CLUSTERING

Denna kopplingstabell tillkom av det skälet att det fanns många brunnar (*PLATEWELL*) i en platta (*PLATE*), samt att en platta kunde ge upphov till många brunnar genom replikeringen av en originalplatta. Det är viktigt för forskaren att kunna härleda pricken i en .tif fil efter hybridisering till en verklig gen. (Bilaga 1D)

LABORATION_FILE

Liksom ovanstående tabell var även denna ett resultat av identifieringsbehovet. En fil kan ingå i många laborationer. T.ex. innehållsdeklarationen av en inköpt komponent. Motsvarande så innehåller en laboration minst 3 st. resultatfiler. (Bilaga 1A)

Databasmodell



Prototyp

För att göra en prototyp av databasen användes MsAccess. Det gjordes främst av den anledningen att verktyget snabbt ger en möjlighet att skapa både tabeller och gränssnitt, samt ger ett kvitto på att modellen fungerar.

Prototypen har också varit ett hjälpmedel i kommunikationen med de personer som skall använda sig av databasen. Jag har på ett enkelt sätt kunnat förvissa mig om att vi "talat samma språk" vad det gäller nödvändiga samband, och önskemål om möjlig funktionalitet. Det är endast den grundläggande funktionaliteten som byggts in i prototypen, men det var tillräckligt för att kunna bekräfta modellens funktionalitet.

Diskussion

Vid försök där det används microarray teknik är det nödvändigt att ha en databas till sitt förfogande. Det finns inte tid och pengar till att söka igenom alla data manuellt, men om så vore, skulle det innebära en osäkerhet i forskningen som inte kan vara acceptabel.

Jag har gjort ett förslag till en fungerande databasmodell. I anknytning till det finns det frågor som ligger utanför detta arbetes gränser som bör utredas vidare. Vilken databasleverantör lämpar sig bäst med tanke på prestanda och pris? En annan fråga, som kanske är viktigast för forskaren: Hur skall ett användargränssnitt skall se ut.? Det är en viktig men ofta underskattad del som skulle kräva en djupare analys av arbetet vid laboratoriet än vad som varit möjligt för det här arbetet.

Modellen är gjord med tanke på att det kan tillkomma processer eller attribut. Den skall också kunna användas i andra typer av försök, naturligtvis med erforderliga modifieringar. Stommen i modellen är de tabeller som representerar forskare och laboration. Till detta läggs vad forskarna gör i sin laboration (de olika processerna). Slutligen fylls modellen på med det material som används i processerna och som har sina egna egenskaper, vilka inte påverkas inuti processen. Modellen gör det möjligt att addera processer så länge det skapas en koppling till tabellerna *SCIENTIST* och *LABORATION*. Att lägga till eller ta bort attribut i processerna är heller inget problem så länge nyckeln i tabellen finns kvar.

Databasmodellen i sin helhet är ganska enkel men ett par detaljer bör motiveras. I de allra flesta av tabellerna har jag tagit med attributet *comment*. Det är tänkt till att ge forskaren möjlighet att anteckna avvikelser från den normala proceduren, eller annat som bör noteras om just den aktuella processen. Forskarna ansåg detta som nödvändigt eftersom försöken till vissa delar ännu ligger på utvecklingsstadiet och varje liten förändring, som inte går att förutse i en modell, kan ge stora konsekvenser för ett experiment. Dessa kommentarer är främst tänkta för mindre och tillfälliga förändringar, som till exempel att en temperatur blivit för låg p.g.a. strömavbrott eller att tvättningen utförts på ett speciellt sätt.

För de större anteckningarna, exempelvis inställningar för roboten som fäster cDNA på glasskivorna, eller möjligheten att söka information om en leverantörs anvisningar om ett kit, var en annan lösning nödvändig. Efter diskussioner kom vi överens om att det enklaste sättet var att lagra en sökväg för varje fil i databasen. Dels var informationen så stor och sällan ändrad, dels ansågs det heller inte nödvändigt att kunna söka ut delinformation ur filerna. Möjligheten till kommentar tillgodoser behovet att ange förekommande avsteg från det ordinarie protokollet.

En laboration kommer som minst att behöva kopplas till tre stycken filer, men sannolikt är att även de flesta av processerna kommer behöva ha en protokollfil (anvisningar) kopplad till sig. Jag valde därför att inte göra ett attribut för filer i tabellerna, utan skapade en tabell enbart för filer (*FILE*). Fördelen med detta är att det nu är möjligt att koppla hur många filer som helst till en laboration eller flera filer till en process.

En annat område i modellen som kan vara ämne för diskussion, är tabellerna som rör plattor och brunnar med cDNA. Det är av ytterlig betydelse att det finns en härledbarhet från prickarna i de slutliga resultatfilerna till innehållet i den brunn som avses på resultatfilen. Det köps in flera olika plattor, alla med ett unikt innehåll i varje brunn. Från varje brunn i

plattorna används sedan en liten del som kopieras upp och placeras i nya plattor (replikat). Replikatets brunnar fördelas i sin tur på kanske 50 olika glasskivor. Varje ursprunglig platta ligger därmed till grund för flera replikat. För att undvika redundans valde jag att se brunnarna som objekt med sina egenskaper som rör dess innehåll. Med en tabell för plattor (*PLATE*) och en tabell för brunnar (*PLATEWELL*) kan varje brunn ingå i flera plattor, som sedan alla härstammar från en ursprunglig platta.

Medicinsk forskning är idag helt beroende av datorer. Den forskning som bedrivs inom det molekylärbiologiska området kräver dessutom väl fungerande databaser för beräkningar samt inhämtande och lagring av information. Tyvärr utnyttjas bara en liten del av datorernas kapacitet, naturligtvis beror detta på att i de flesta medicinska forskningslaboratorier finns det ingen som har någon kunskap om programmering. Det är inget konstigt i det, på laboratoriet utförs det medicinsk forskning och det är det som forskarna kan bäst, och det är det som de är där för att göra.

Forskarna har även ett annat behov av applikationer. Det behövs naturligtvis databaser av den typ som redovisas i det här arbetet, för att hantera information av stora kvantiteter, men också enklare applikationer som t.ex. kan kryptera personuppgifter om donatorer.

En integration av de olika vetenskaperna ser jag som helt nödvändig och kan med fördel bestå av flera delar. Den har, i en del, redan börjat med ett samarbete mellan Chalmers Tekniska Högskola och Sahlgrenska universitetssjukhuset. Det är i form av en utbildning inom bioinformatik och är viktigt steg i rätt riktning, men det behövs enligt mitt tycke också en annan del. En obligatorium inom doktorandutbildningen, som ger de medicinska doktoranderna grundläggande kunskaper hur de bäst kan utnyttja applikationer som exempelvis Word, Excel, PowerPoint och Access inom deras egna intresseområden.

Jag har, som tidigare nämnts, under examensarbetets gång märkt ett behov hos forskare att få hjälp med enklare applikationer. Det finns idag hos de flesta forskare ett fungerande system som är pappersbaserat. Nackdelen är att det ägnas mycket tid för att hämta och slå i pärmar när det behövs kompletterande information. Den tiden borde egentligen ägnas åt forskning. Det finns säkert många idéer som aldrig prövas därför att det tar för lång tid att hämta information för att se om idén är genomförbar eller ej. Att direkt kunna lägga in information på sin dator med hjälp av enkla applikationer skulle både underlätta för, och välkomnas av forskaren. Det blir då möjligt att snabbt kunna hämta informationen och på ett enkelt sätt dela den med olika samarbetsgrupper, både inom och utom landet.

För att ett gränsöverskridande arbete skall kunna bli fruktbart för en informatiker måste vederbörande vara uppmärksam på vissa problem som finns. I ett tidigt skede märkte jag att det uppstod kommunikationsproblem som hade sin grund i att jag ställde frågor som var direkt kopplade till databasmodelleringen. Detta innebar att forskaren försökte besvara frågorna på samma nivå, men på grund av begreppsförvirring försvårade det min uppfattning om hur hans verklighet såg ut. För att göra det hela enklare ställde jag istället frågor om vad de gjorde i sin verksamhet, hur de gjorde det, och varför. Sedan var det upp till mig att översätta vad jag fått veta till en databas.

Komplexiteten och flexibiliteten i forskarens laborationer som skulle avbildas i databasmodellen, innebar att jag i början av mitt arbete fick ägna mycket tid åt att lära mig deras verksamhet. Det var en något frustrerande upplevelse eftersom jag ville påbörja modelleringsarbetet snarast. Efterhand märkte jag att det var väldigt svårt att göra något i modellen utan att förstå deras verksamhet i sin helhet. Det innebar upprepade kontakter där jag fick inrikta mig på att förstå molekylärbiologi och forskningsmetodik. Det var ett otal frågor som jag ställde både till ”min” forskare, men också till andra med kunskap inom området.

Mina erfarenheter av att göra en databasmodell inom en annan vetenskap är dessa: Var **ödmjuk** inför andras kunskaper, men var också **bestämd** vad det gäller dina egna kunskaper. Arbetets första hälft består av att **fråga**. Den andra hälften består av att **kommunicera** med beställaren för att bekräfta svaren på frågorna som ställdes i första hälften.

Reflektioner som jag gjort vid min utbildning på institutionen för Informatik, var att de laborationer som hade en koppling till verkligheten alltid var mer utmanande och mer intressanta än de som var enbart fiktiva. Jag är därför övertygad om att ett samarbete mellan institutionen för Informatik och institutioner inom den medicinska världen, som faktiskt ligger under samma domän, kan vara fruktbart för båda parter. Informatik får verklighetsrelaterade laborationer och därmed också mer motiverade studenter. Medicinarna får applikationer som underlättar deras laborationer. På detta sätt kan även intresset för bioinformatik öka ytterligare, vilket i sin tur skulle ge en bredare kompetens i vår del av landet inom ett område som kommer att växa med en hastighet vilken idag bara kan anas.

Slutsats

Med denna databasmodell har forskaren en god möjlighet att på ett effektivt och säkert sätt hantera information genererad från försök med microarray

För att kunna göra en bra databasmodell inom en annan vetenskap är en väl utvecklad kommunikation mellan ”beställaren och utföraren” en nödvändighet

Det är fullt möjligt att jobba gränsöverskridande, och i framtiden kommer det att vara helt nödvändigt, eftersom forskningen blir allt mer specialiserad.

Referenser

- [1] *Bioinformatik, Genteknik* (2000-05-23)
Högskolan i Kalmar
<http://www.ng.hik.se/~nstma/bioovn.htm>
- [2] *Linnaeus Centre for Bioinformatics* (2000-05-23)
<http://www.linnaeus.bmc.uu.se/swellsum.html>
- [3] Baxevanis, Andreas D.
The Molecular Biology Database Collection: an online compilation of relevant database resources.
2000 Oxford University Press
Nucleic Acids Research, 2000, Vol. 28, No.1, 1-7
http://www.oup.co.uk/nar/Volume_28/Issue_01/html/gkd115_gml.html
- [4] Thinkquest (2000-05-23)
<http://library.thinkquest.org/17498/>
- [5] Hunter, Lawrence
Molecular Biology for Computer Scientists
- [6] Lloyd Bellenson, Joel
Expression data and the bioinformatics challenges
- [7] Hattori, M et al
The DNA sequence of human chromosome 21
Nature, vol 405, 18 maj 2000, 311
- [8] Attwood, Teresa K. and Parry Smith, David J.
Introduction to Bioinformatics
Addison Wesley Longman Limited, 1999
ISBN 0-582-32788-1
- [9] Stryer, Lubert
Biochemistry
W.H. Freeman & Co, New York, 1981
- [10] Schena, Mark
Microarray Biochip Technology
Eaton Publishing, Biotechniques Books; 2000
- [11] Connolly, Thomas, Begg, Carolyn och Strachan, Anne
Database Systems
Adison-Wesley, 1998, Andra upplagan
ISBN 0-201-34287-1
- [12] Laboratoriet nr 1; 2000
ISSN 0345-697x

Tabellbilagor

Bilaga 1A

AMPLIFICATION	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
date	<i>Date</i>
startmaterialName	<i>varchar(10)</i>
startmaterialAmount	<i>float</i>
method	<i>varchar(25)</i>
productAmount	<i>float</i>
productConc	<i>float</i>
comment	<i>varchar(100)</i>

COATING_SLIDES	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>slideID</u>	<i>varchar(10)</i>
date	<i>date</i>
coatingName	<i>varchar(50)</i>
chemical	<i>varchar(25)</i>
coatingTime	<i>varchar(10)</i>
coatingTemp	<i>varchar(10)</i>
comment	<i>varchar(100)</i>

FILE	
<u>fileNumber</u>	<i>varchar(10)</i>
fileName	<i>varchar(50)</i>

HYBRIDIZATION	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>slideID</u>	<i>varchar(10)</i>
date	<i>date</i>
hybridizationTime	<i>varchar(10)</i>
temperature	<i>varchar(10)</i>
probe1ID	<i>varchar(10)</i>
probe2ID	<i>varchar(10)</i>
probe1Amount	<i>float</i>
probe2Amount	<i>float</i>
washing	<i>varchar(50)</i>
comment	<i>varchar(100)</i>

LABORATION	
<u>labID</u>	<i>varchar(10)</i>
labName	<i>varchar(25)</i>
comment	<i>varchar(100)</i>

LABORATION_FILE	
<u>labID</u>	<i>varchar(10)</i>
<u>fileNumber</u>	<i>varchar(10)</i>
path	<i>varchar(50)</i>

LASER_CAPTUREING	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>materialID</u>	<i>varchar(10)</i>
date	<i>date</i>
fixation	<i>varchar(50)</i>
temperature	<i>varchar(10)</i>
time	<i>varchar(10)</i>
embedding	<i>varchar(50)</i>
sliceThickness	<i>varchar(10)</i>
nameGlassCoating	<i>varchar(25)</i>
staining	<i>varchar(25)</i>
cellName	<i>varchar(25)</i>
comment	<i>varchar(100)</i>

MATERIA	
<u>materialID</u>	<i>varchar(10)</i>
category	<i>varchar(50)</i>
tissueName	<i>varchar(50)</i>
species	<i>varchar(25)</i>
age	<i>varchar(10)</i>
geneticModification	<i>varchar(25)</i>
strain	<i>varchar(25)</i>
celtype	<i>varchar(25)</i>
passageNumber	<i>varchar(10)</i>
stageOfGrowth	<i>varchar(25)</i>
medium	<i>varchar(50)</i>
incubationTemperature	<i>varchar(10)</i>
oxygenConcentration	<i>float</i>
carbonDiOxidConcentration	<i>float</i>
comment	<i>varchar(100)</i>

Bilaga 1B

mRNA_PREPARATION	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
date	<i>date</i>
startmaterialType	<i>varchar(10)</i>
startmaterialAmount	<i>float</i>
kit	<i>varchar(25)</i>
rnaAmount	<i>float</i>
totalAmount	<i>float</i>
concentration	<i>float</i>
measuringMethod	<i>varchar(25)</i>
comment	<i>varchar(100)</i>

PLATE	
<u>plateID</u>	<i>varchar(10)</i>
category	<i>varchar(25)</i>
origin	<i>varchar(50)</i>
platename	<i>varchar(50)</i>
location	<i>varchar(50)</i>

PLATEWELL	
<u>wellID</u>	<i>varchar(50)</i>
xCoordinate	<i>varchar(10)</i>
yCoordinate	<i>varchar(10)</i>
geneName	<i>varchar(50)</i>
geneAccessionNumber	<i>varchar(50)</i>

POSTPROCESSING_SLIDES	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>slideID</u>	<i>varchar(10)</i>
comment	<i>varchar(100)</i>

PCR	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>plateID</u>	<i>varchar(10)</i>
pcrMachineName	<i>varchar(25)</i>
concMgCl	<i>float</i>
amountMgCl	<i>float</i>
concPCRBuffer	<i>float</i>
amountPCRBuffer	<i>float</i>
concdNTP	<i>float</i>
amountdNTP	<i>float</i>
concForwardPrimer	<i>float</i>
amountForwardPrimer	<i>float</i>
concBackwardPrimer	<i>float</i>
amountBackwardPrimer	<i>float</i>
unitsTaqPolymerase	<i>float</i>
nameTaqPolymerase	<i>varchar(25)</i>
ngDNA	<i>float</i>
water	<i>float</i>
finalConcentration	<i>float</i>
gel	<i>varchar(25)</i>
amountPCRProduct	<i>float</i>
Voltage	<i>float</i>
minutes	<i>float</i>
comment	<i>varchar(100)</i>

PRINTING_SLIDES	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>slideID</u>	<i>varchar(10)</i>
comment	<i>varchar(100)</i>

Bilaga 1C

PROBE_PREPARATION	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>probeID</u>	<i>varchar(10)</i>
date	<i>date</i>
amplified	<i>Boolean</i>
fluorescenceLabeling	<i>varchar(10)</i>
amountLabeling	<i>float</i>
startmaterial	<i>varchar(10)</i>
primer	<i>varchar(25)</i>
amountPrimer	<i>float</i>
unlabeledNucleotide	<i>varchar(10)</i>
concentrationA	<i>float</i>
concentrationT	<i>float</i>
concentrationC	<i>float</i>
concentrationG	<i>float</i>
nameRT	<i>varchar(25)</i>
amountRT	<i>float</i>
amountBuffer	<i>float</i>
totalVolume	<i>float</i>
denaturationTime	<i>varchar(10)</i>
denaturationTemp	<i>varchar(10)</i>
annealingTime	<i>varchar(10)</i>
annealingTemp	<i>varchar(10)</i>
reactionTime	<i>varchar(10)</i>
reactionTemp	<i>varchar(10)</i>
postPurification	<i>varchar(100)</i>
comment	<i>varchar(100)</i>

PRECIPITATION	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
saltName	<i>varchar(25)</i>
saltAmount	<i>float</i>
saltStockConc	<i>float</i>
alcoholName	<i>varchar(25)</i>
alcoholStockConc	<i>float</i>
alcoholAmount	<i>float</i>
time	<i>varchar(10)</i>
temperature	<i>varchar(10)</i>
centrifugeTime	<i>varchar(10)</i>
centrifugeTemp	<i>varchar(10)</i>
centrifugeRpm	<i>varchar(10)</i>
ulEtOHWash	<i>float</i>
concEtOHWash	<i>float</i>
timeEtOHWash	<i>varchar(10)</i>
centrifugeWashRpm	<i>varchar(10)</i>
centrifugeWashTime	<i>varchar(10)</i>
centrifugeWashTemp	<i>varchar(10)</i>
dryingTime	<i>varchar(10)</i>
resuspSolutionName	<i>varchar(25)</i>
resuspSolutionAmount	<i>float</i>
resuspTime	<i>varchar(10)</i>
resuspTemp	<i>varchar(10)</i>
comment	<i>varchar(100)</i>
date	<i>date</i>

RNA_PREPARATION	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>materialID</u>	<i>varchar(10)</i>
date	<i>date</i>
startAmount	<i>float</i>
homogenizer	<i>varchar(50)</i>
kit	<i>varchar(50)</i>
concentration	<i>float</i>
totalAmount	<i>float</i>
rnaAmount	<i>float</i>
measuringMethod	<i>varchar(50)</i>
comment	<i>varchar(100)</i>

SCIENTIST	
<u>scientistID</u>	<i>varchar(10)</i>
name	<i>varchar(50)</i>
e-mail	<i>varchar(50)</i>
group	<i>varchar(25)</i>

SLIDE	
<u>slideID</u>	<i>varchar(10)</i>
brand	<i>varchar(50)</i>
slideProductName	<i>varchar(50)</i>
coated	<i>boolean</i>

Bilaga 1D

SCANNING	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>slideID</u>	<i>varchar(10)</i>
resultType	<i>varchar(25)</i>
date	<i>date</i>
comment	<i>varchar(100)</i>
dataType	<i>varchar(25)</i>
Mode_sig	<i>float</i>
Mode_bkg	<i>float</i>
Total_sig	<i>float</i>
Total_bkg	<i>float</i>
Std_sig	<i>float</i>
Std_bkg	<i>float</i>
Xcoord	<i>number</i>
Ycoord	<i>number</i>
Mean_sig	<i>float</i>
Mean_bkg	<i>float</i>
Median_sig	<i>float</i>
Median_bkg	<i>float</i>
Meta_row	<i>number</i>
Meta_col	<i>number</i>
Sub_row	<i>number</i>
Sub_col	<i>number</i>
Name	<i>varchar(50)</i>
Selected	<i>number</i>

TISSUE_PREPARATION	
<u>scientistID</u>	<i>varchar(10)</i>
<u>labID</u>	<i>varchar(10)</i>
<u>materialID</u>	<i>varchar(10)</i>
date	<i>date</i>
tissueType	<i>varchar(25)</i>
preparationType	<i>varchar(25)</i>
comment	<i>varchar(100)</i>

WELL_CLUSTERING	
<u>wellID</u>	<i>varchar(10)</i>
<u>plateID</u>	<i>varchar(10)</i>