<<基于Affymetrix芯片的基因 >

图书基本信息

书名：<<基于Affymetrix芯片的基因表达研究>>

13位ISBN编号：9787030329080

10位ISBN编号：7030329082

出版时间：2012-1

出版时间：科学出版社

作者：〔美〕Hinrich Göhlmann、Willem Talloen 著，张春秀 译

页数：327

内容概要

Affymetrix
GeneChip系统是目前应用最广泛的生物芯片平台。
但是由于Aflymetrix芯片含有超大量的信息，很多Affymetrix芯片用户趋向于使用默认的分析设置，得到的常常不是最优化的结论。
分子生物学家和生物统计学家根据十余年的基因表达谱实验研究和数据分析的实践经验编写了《基于Affymetrix芯片的基因表达研究》，从理论概念到实验结果，解释了使用Affymetrix芯片进行基因表达研究的全部过程，拆除了分子生物学、生物信息学和生物统计学之间无处不在的语言障碍。

本书权威实用，介绍了Affymetrix芯片的重要技术、统计学易犯的错误和问题，同时涉及其他芯片平台的一般规则和应用。
通过例证和全彩图例，描述了技术和统计方法的概念，为初学者提供详细指导。
本领域的专家则可以了解芯片所涉及的其他学科知识，拓展基因芯片表达谱研究的认识。

书籍目录

表格目录

章节摘录

Chapter 1Biological question All experimental work starts in principle with a question. This also applies to the field of molecular biology. A molecular scientist is using a certain technique to answer a specific question such as, “ Does the cell produce more of a given protein when treated in a certain way? ” Questions in molecular biology are indeed regularly focused on specific proteins or genes, often because the applied technique cannot measure more. Gene expression studies that make use of microarrays also start with a biological question. The largest difference to many other molecular biology approaches is, however, the type of question that is being asked. Scientists will typically not run arrays to find out whether the expression of a specific messenger RNA is altered in a certain condition. More often they will focus their question on the treatment or the condition of interest. Centering the question on a biological phenomenon or a treatment has the advantage of allowing the researcher to discover hitherto unknown alterations. On the other hand, it poses the problem that one needs to define when an “ interesting ” alteration occurs. 1.1 Why gene expression? 1.1.1 Biotechnological advancements Research evolves and advances not only through the compilation of knowledge but also through the development of new technologies. Traditionally, researchers were able to measure only a relatively small number of genes at a time. The emergence of microarrays (see BioBox 1.1) now allows scientists to analyze the expression of many genes in a single experiment quickly and efficiently. 1.1.2 Biological relevance Living organisms contain information on how to develop its form and structure and how to build the tools that are responsible for all biological processes that need to be carried out by the organism. This information ? the genetic .....Geneexpressionmicroarrays.Inmicroarrays,thousandstomillionsofprobesarefixedtoorsynthesizedonasolidsur-face,beingeitherglassorasiliconchip.Thelatterexplainswhymicroarraysarealsooftenreferredtoaschips.Thetar-getsoftheprobes,themRNAsamples,arelabelledwithfluo-rescentdyesandarehybridizedtotheirmatchingprobes.Thehybridizationintensity,whichestimatestherelativeamountsofthetargettranscripts,canafterwardsbemeasuredbytheamountoffluorescentemissionontheirrespectivespots.Therearevariousmicroarrayplatformsdifferinginarrayfabrication,thenatureandlengthoftheprobes,thenumberoffluorescentdyesthatarebeingused,etc.BioBox 1.1: Gene expression microarrays content ? is encoded in information units referred to as genes. The whole set of genes of an organism is referred to as its genome. The vast majority of genomes are encoded in the sequence of chemical building blocks made from deoxyribonucleic acid (DNA) and a smaller number of genomes are composed of ribonucleic acid (RNA), e.g., for certain types of viruses. The genetic information is encoded in a specific sequence made from four different nucleotide bases: adenine, cytosine, guanine and thymine. A slighlty different composition of building blocks is present in mRNA where the base thymine is replaced by uracil. Genetic information encoding the building plan for proteins is transferred from DNA to mRNA to proteins. The gene sequence can range in length typically between hundreds and thousands of nucleotides up to even millions of bases. The number of genes that contain protein-coding information is expected to be between 25,000 to 30,000 when looking at the human genome. A protein is made by constructing a string of protein building blocks (amino acids). The order of the amino acids in a protein matches the sequence of the nucleotides in the gene. In other words, messenger RNA interconnects DNA and protein, and also has some important practical advantages compared to both DNA and proteins (see BioBox 1.2). Increasing our knowlegde about the dynamics of the genome as manifested in the alterations in gene expression of a cell upon treatment, disease, development or other external stimuli, should enable us to transform this knowledge into better tools for the diagnosis and treatment of diseases. DNA is made of two strands forming together a chemical structure that is called “ double helix. ” The two strands are connected with one another via pairs of bases that form hydrogen bonds between both strands. Such pairing of so-called “ complementary ” bases occurs only between certain pairs. .....Centraldogmaofmolecularbiology.Thedogmaofmolec-ularbiologyexplainshowtheinformationtobuildproteinsistransferredinlivingorganisms.Thegeneralflowofbiologicalinformation(greenarrows)hasthreemajorcomponents:(1)DNAtoDNA(replication)occursinthecellnucleus(drawninyellow)priortocelldevision,(2)DNAtomRNA(transcrip-tion)takesplacewheneverthecell(drawninlightred)needstomakeaprotein(drawnaschainofreddots),and(3)mRNA

toproteins(translation)istheactualproteinsynthesisstepintheribosomes(drawningreen).Besidesthesegeneraltransferst hatoccurnormallyinmostcells,thereareals osomespecialinformationtransfersthatareknowntooccurinsomevirusesorin alaboratoryexperimentalsetting.BioBox 1.2: Central dogma of molecular biology

.....Hydrogen bonds can be formed between cytosine and guanine or between adenine and thymine. The pairing of the two strands occurs in a process called “ hybridization. ” Compared to DNA, mRNA is more dynamic and less redundant. The information that is encoded in the DNA is made available for processing in a step called “ gene expression ” or “ transcription. ” Gene expression is a highly complex and tightly regulated process by which a working copy of the original sequence information is made. This allows a cell to respond dynamically both to environmental stimuli and to its own changing needs, while DNA is relatively invariable. Furthermore, as mRNA constitutes only the expressed part of the DNA, it focuses more directly on processes underlying biological activity. This filtering is convenient as the functionality of most DNA sequences is irrelevant for the study at hand. Compared to proteins, mRNA is much more measurable. Proteins are 3D conglomerates of multiple molecules and cannot benefit from the hybridising nature of the base pairs in the 2D, single molecule, structure of mRNA and DNA. Furthermore, proteins are very unstable due to denaturation, and cannot be preserved even with very laborious methods for sample extraction and storage. When using microarrays to study alterations in gene expression, people normally will only want to study the types of RNA that code for proteins ? the messenger RNA (mRNA). It is however important to keep in mind that RNAnot only contains mRNA?acopyofa section of the genomic DNA carrying the information of how to build proteins. Besides the code for the synthesis of ribosomal RNA, there are other non-coding genes that, e.g., contain information for the synthesis of RNA molecules. These RNAs have different functions that range from enzymatic activities to regulating transcription of mRNAs and translation of mRNA sequences to proteins. The numbers of these functional RNAs that are encoded in the genome are not known. Initial studies looking at the overall transcriptional activity along the DNA are predicting that the number will most likely be larger than the number of protein-coding genes. People used to say that a large portion of the genomic information encoded in the DNA are useless (“ junk DNA ”). Over the last years scientific evidence has accumulated that a large proportion of the genome is being transcribed into RNAs of which a small portion constitutes messenger RNAs. All these other non-coding RNAs are divided into two main groups depending on their size. While short RNAs are defined to have sizes below 200 bases, the long RNAs are thought to be mere precursors for the generation of small RNAs, of which the function is currently still unknown ? in contrast to the known small RNAs such as microRNAs or siRNAs[6] (see BioBox 1.3 for an overview of different types of RNA). Microarrays are also being made to study differences in abundance of these kinds of RNA.

.....RNA.IncontrasttomRNA(messengerRNA)whichcontainstheinformationofhowtoassembleaprotein,thereare alsodifferenttypesofnon-codingRNA(sometimesabbreviatedasncRNA)a.Herearethetypesthataremostrelevantinthe contextofthisbook:miRNAinlength,whichregulategeneexpression.longncRNA(longnon-codingRNA)arelongRNA moleculesthatperformregulatoryroles.AnexampleisXIST,whichcanalsobeusedfordataqualitycontroltoidentifythege nderofasubject(seeBioBox3.5).rRNA(ribosomalRNA)arelongRNAmoleculesthatmakeupthecentralcomponentoft heribosomeb.Theyareresponsiblefordecodingm RNAintoaminoacidsandareusedforRNAqualitycontrolpurposes(se eSection3.1.2.8).siRNA(smallinterferingRNA)aresmalldouble-strandedRNAmoleculesofabout20-25nucleotidesinl engthandplayavarietyofrolesinbiology.Themostcommonlyknownfunctionisaprocesscalled RNAinterference(RNAi).InthisprocesssiRNAsinterferewiththeex-pressionofaspecificgene,leadingtoadownregulationofthesynthesisofnewpr oteinencodedbythatgenec.tRNA(transferRNA)aresmallsingle-strandedRNAmoleculesofabout74-95nucleotidesinl enghts,whichtransferasingleaminoacidtoagrowingpolypeptidechainattheribosomalsiteofproteinsynthesis.Eachtype oftRNAmoleculecanbeattachedtoonlyonetypeofaminoacid.aNon-codingRNAreferstoRNAmoleculesthataretransc ribedfromDNAbutnottranslatedintoprotein.bRibosomescanbeseenastheproteinmanufacturingmachineryofallliving cells.cThereare,however,alsoprocessesknownassmallRNA-inducedgeneactivationwherebydouble-strandedRNAsta rgetgenepromoterstoinducetranscriptionalactivationofassociatedgenes.BioBox 1.3: siRNA

.....In this book we will focus on studying mRNA. However, most likely many remarks given

on the experimental design and the data analysis will apply to the study of small RNA as well. 1.2 Research question The key to optimal data analysis lies in a clear formulation of the research question. Being aware of having to define what one considers to be a “ relevant ” finding in the data analysis step will help in asking the right question and in designing the experiment properly so that the question can really be answered. A well-thought-out and focused research question leads directly into hypotheses, which are both testable and measurable by proposed experiments. Furthermore, a well-formulated hypothesis helps to choose the most appropriate test statistic out of the plethora of available statistical procedures and helps to set up the design of the study in a carefully considered manner. To formulate the right question, one needs to disentangle the research topic into testable hypotheses and to put it in a wider framework to reflect on potentially confounding factors. Some of the most commonly used study designs in microarray research will be introduced here by means of real-life examples. For each type of study, research questions are formulated and example datasets described. These datasets will be used troughout the book to illustrate some technical and statistical issues. 1.2.1 Correlational vs. experimental research Microarray research can either be correlational or experimental. In correlational research, scientists generally do not apply a treatment or stimulus to provoke an effect on, e.g., gene expression (influence variables), but measure them and look for correlations with mRNA (see StatsBox 1.1). A typical example are cohort studies, where individuals of populations with specific characteristics (like diseased patients and healthy controls) are sampled and analysed. In experimental research, scientists manipulate certain variables (e.g., apply a compound to a cell line) and then measure the effects of this manipulation on mRNA. Experiments are designed studies where individuals are assigned to specifically chosen conditions, and mRNA is afterwards collected and compared. It is important to comprehend that only experimental data can conclusively demonstrate causal relations between variables. For example, if we found that a certain treatment A affects the expression levels of gene X, then we can conclude that treatment A influences the expression of gene X. Data from

版权说明