Microarrays and 2D gels: new tools for postgenomic muscle biology

Carlo Reggiani and Truus te Kronnie

Department of Anatomy and Physiology and Department of Pediatrics. University of Padova 35131 Padova, Italy

Abstract

The complete genome sequencing has opened the way to the investigation of the transcriptional activity of thousands of genes (i.e the transcriptome) through DNA microarrays. In parallel to these technological developments high-resolution two-dimensional gels combined with sensitive mass spectrometry and other new methods provide the possibility of studying all proteins present in cells (i.e. the proteome). Both transcriptome and proteome investigations are characterized by an unprecedented resolution power (high throughput) and by their totally descriptive character, so that a physiological condition or a specific tissue is explored without any preliminary hypothesis to be tested; on the other hand, the results of the high throughput studies are very often the starting point for the generation of new hypothesis to be validated in subsequent experiments. This paper aims to briefly review the most recent technological developments and their application to muscle physiology and pathology.
key words: microarray, proteome, transcriptome, two-dimensional electrophoresis

Basic Appl Myol 14(1): 7-12, 2004

The complete identification of coding sequences in the genome has opened the way to the utilization of new tools in biology. Advances in genome sequencing and bioinformatics have been followed by the development of methods, which allow the investigation of the transcriptional activity of thousands of genes (i.e the transcriptome) through DNA microarrays. However, a discrepancy may exist between the transcriptional profiles emerging from mRNA expression profiles and the actual protein levels in cells. Thus, it also becomes important to separate and identify the proteins expressed. High-resolution two-dimensional gels combined with sensitive mass spectrometry have been developed and new methods are on the way to study all proteins present in cells (i.e. the proteome). Both transcriptome and proteome investigations are characterized by an unprecedented resolution power (high throughput) and by their totally descriptive character, so that a physiological condition or a specific tissue is explored without any preliminary hypothesis to be tested; on the other hand, the results of the high throughput studies are very often the starting point for the generation of new hypothesis to be tested in subsequent experiments. This presentation aims to briefly review these recent technological developments and their application to muscle physiology and pathology.

Introduction to microarray technology

DNA microarrays, also referred to as DNA arrays, DNA chips, biochips, and Genechips, allow the determination of the genes expressed in a given cell type at a particular time and under particular conditions. They can also be used to compare the gene expression in two different cell types or tissue samples, for example healthy versus diseased tissue. Microarray technology represents a profiling strategy limited only by the probes imprinted onto the gene chips. However, this is at the same time a limitation of microarray studies in that they represent a closed approach to global screening; only probes that are spotted on the array are inquired for expression. Ideally, a chip contains probes that represent the complete transcriptome of the species studied, but in most studies published to date, microarray probes have been limited to a known segment of the transcriptome. In the case of muscle studies, the targets may be restricted to known "muscle genes". In such studies, the analysis is limited to testing alterations in gene expression of genes previously known to being expressed in muscle tissue. Genes not normally expressed in muscle are not considered, so the

hypothesis is that pathology and altered physiological conditions involve only quantitative gene expression alterations and no qualitative changes.

Microarray technology

DNA microarrays exploit the preferential binding of complementary single-stranded nucleic-acid sequences. The underlying principle is the same for all microarrays, no matter how they are made: the unknown sample is hybridized to an ordered array of immobilized DNA molecules whose sequence is known. A recent review [12] on the manufacturing and use of DNA microarrays discusses the most recent equipment, reagents and protocols available for biomedical research. The main types of microarrays are filter arrays, spotted glass slide arrays, and *in situ* synthesized oligonucleotide arrays. The first two are produced readily in academic facilities (so called 'do it yourself'), but can also be purchased from commercial manufacturers. For the moment, arrays of oligonucleotides that are synthesised *in situ* (e.g. Affymetrix GeneChip) require complex equipment and are only commercially produced.

So far the principal source of probe fragments used for spot arraying have been bacterial cDNA clone sets. For complex organisms, such as mice and humans, there are still some shortcomings in the libraries available; libraries contain a certain amount of redundancy, miss-annotation and contamination. In view of the logistical difficulties associated with handling large numbers of bacterial clones it has become very attractive to prepare large sets of oligonucleotide probes that take advantage of the rapid growth in genome sequence information. Because oligonucleotide sets, unlike cDNA sets, are not limited by the availability of physical clones, oligonucleotide targets could in principle be generated in-house from sequence information. But the design of oligonucleotides is a complicated procedure, and in practice most studies use commercially available oligonucleotide sets.

Expression analysis using glass slide microarrays is typically done by the competitive hybridisation of two targets, each labelled with a specific fluorescent dye such as Cy3 or Cy5. The comparative nature of gene expression measurements with spotted arrays creates problems in terms of archiving data and comparing data among different experiments and laboratories. In principle, some of these problems could be alleviated by adoption of universal references, but currently there is no consensus on reference RNA and targets. For a discussion of further items regarding informatics, microarray data storage and normalization of microarray data see [12].

Validation of microarray results

Methods are needed to sort out which changes in gene expression are really of importance. In most studies real-time quantitative PCR (RQ-PCR) assays are employed to provide an independent confirmation of changes in expression level seen in microarrays.

In general, microarray studies consider a twofold change in mRNA levels significant [2]. In some studies, however, gene expression changes greater than a factor of 1.7 are considered important [24]. The two-fold standard is arbitrary indeed. But even a two fold change may not be functionally significant: for example, mice heterozygous for a null allele are often without a phenotype [10].

Once gene expression changes that robustly correlate with tissue differences have been discovered, other tests will be needed to explore the functional significance of the expression level alterations. To this end large scale reverse genetics in mice is an option.

Baseline expression patterns

Establishment of baseline expression patterns in normal tissues is an essential element in the accurate interpretation of changes associated with pathological or other altered physiological states.

Muscle tissue, one third of body mass, is highly metabolically active: in the case of human muscle current information estimates that 30% of the genes (over a total of 40.000 genes) are expressed in muscle tissue [3] and more than 4000 EST's have been identified in human muscle samples [17] [20]. In an attempt to describe baseline expression profiles for a number of human tissues Hsiao et al [13] report the expression of 317 genes predominantly expressed in muscle among 7070 (HuGeneFL) genes tested. A smaller number of genes is muscle tissue specific, i.e. only expressed in muscle. Interestingly, striking quantitative differences in expression of so-called housekeeping genes are also noted when comparing tissues: both beta-actin and GAPDH commonly assumed to have constant expression levels, are among the most variably expressed genes. The muscle selective genes that have no overlap in expression level with any other tissue are genes associated with: cytoskeleton (alpha1 actin, alpha 2-3 actinin), contraction (tropomyosin, troponin, myosin) mitochondria (cytochrome C-1, creatine kinase), metabolism of glucose, glycogen and lipids (LDH, phosphoglucomutase 1) carbonic anhydrase III, creatine kinase), thermal regulation (neurotrophic tyrosine kinase receptor type 1).

Another interesting observation [13] is that several of these muscle specific genes show highly variable expression levels between different individuals. It is conceivable that such variations are due to differences in the precise cellular composition of the tissue sample, in addition to differences in age, gender, health and medication. Even so, the presence of genes variable among samples (of muscle tissues of different individuals) is consistent with the notion that the genotype and the inherent plasticity of muscle tissue of

an individual may contribute to gene-specific expression variation.

Experimental design

In human studies, one of the aims when designing a microarray experiment is to try to avoid inter-individual noise present between human subjects. Longitudinal experiments that, in the case of muscle, compare biopsies of the same individual under different conditions (e.g. training and untraining [11]) substantially eliminate inter-individual noise and thus allow the detection of more subtle changes in gene expression, namely those underlying the adaptation to the condition imposed on the muscle.

Applications of microarrays to muscle physiology and pathology

Microarray-based studies in muscle include analysis of gene expression in diseased versus normal tissue, gene expression regulation during development, atrophy, exercise and ageing, and gene expression in different muscles: white, red and extraocular.

Slow/fast and extraocular muscles: The comparison between slow (soleus) and fast (white quadriceps) in the mouse has been carried out on 3000 genes present as probe sets on the Affymetrix Mu6500 Gene chip: about 20% of the genes were found actually expressed in the two muscles and 3% of these genes were differentially expressed between white and mixed red muscles [5]. Characterization of the genes increased the number of transcription factors known to being involved in red/white muscle regulation by 9 times and signalling proteins by 3 times.

Using the Mu74A Gene chip (Affymetrix) 400 genes differentially expressed were identified among 9.977 genes interrogated [21] [8]. These genes collectively provide a molecular signature for normal extraocular muscle, when compared to normal skeletal muscles and jaw muscles and reflect key aspects of muscle biology, including transcriptional regulation, sarcomeric organization, excitation contraction coupling and immune response.

To study the differential expression of different muscles (*psoas* versus *longissimus dorsi*) a cDNA microarray was constructed comprised of 5500 clones selected from two developmentally distinct cDNA libraries of *longissimus dorsi* of a 50-day porcine foetus and *gastrocnemius* of 3-day-old pig [2]. Due to presently insufficient sequence information on farm animals an oligonucleotide based microarray could not be designed for this study. Therefore an array of not yet identified cDNA clones was a good alternative: only clones of interest (differentially expressed) were eventually identified by DNA sequencing. Among clones examined 5% were differentially expressed between *psoas* and *longissimus dorsi* . Among these, 70 clones were sequenced and 47 clones turned out to being mitochondrial genes, while 21 did not show any homology with known mitochondrial or sarcomeric genes. A gene markedly over expressed in *psoas* encoded for fructose-1,6-biphosphatase, an enzyme involved in muscle glyconeogenesis.

Dystrophy: Following a previous study on expression in muscle dystrophy [7], Hoffman *et al* [3] used the most complete series of probes yet available to screen both normal and DMD human muscles (Affymetrix U95A,B,C,D and E Gene chips and a custom made probe set of 3600 muscle specific probes): together 65.000 probes which are expected to give nearly complete coverage of the human transcriptome. Overall evaluation of the data showed that 30% of all genes in the genome are transcribed in non-dystrophic human muscle and that Duchenne dystrophy muscles transcribe approximately 5-10% more genes than non-dystrophic muscle. Among the 30% of genes expressed in muscle, 3% are differentially expressed in dystrophin deficiency. Of the differentially expressed genes, 1324 probe sets correspond to characterized genes and 588 to expressed sequence tags or predicted genes. The paper elaborates on insulin-like growth factors and binding proteins which appear up-regulated in DMD muscles.

Recently, gene expression was studied in muscles of patients with well identified dysferlinopathy using a novel dedicated microarray platform made with 3'end skeletal muscle cDNAs [4]. The expression of genes codifying the sarcomeric proteins titin, nebulin, and telethonin was found down-regulated. Neither calpain-3 nor caveolin, a sarcolemmal protein interacting with dysferlin, were consistently reduced. Some genes of proteins interacting with calcium (S100 calcium-binding proteins and sarcolipin, a sarcoplasmic calcium regulator) appeared significantly up-regulated.

Insulin: A cDNA microarray consisting of 42.557 spots representing 29.308 genes from healthy volunteers was employed to study, the effect of hyperinsulinemic clamp in *vastus lateralis* by comparing expression before and after 3 hours of insulin infusion [23]. During insulin infusion, 762 genes among which 353 EST's were differentially expressed. Most of the genes with altered expression are novel targets of insulin.

Atrophy: Changes in gene expression related to muscle wasting due to inactivity were studied in rat *soleus* muscle using Affymetrix rat gene chip (8.799 probe sets). During the first days of inactivity, 330 genes showed altered expression (Stevenson and Kandarian, unpublished). Jagoe *et al.* [16] used a cDNA glass slide microarray to study gene expression in the *gastrocnemius* muscles of mice during muscle atrophy in response to food deprivation. Transcriptional changes indicate a complex adaptive program of protein degradation and suppression of glucose oxidation in muscle. Of 4000 genes on the array 5.9 % showed significantly altered expression after 48 hours of food deprivation.

Exercise/Aging: Roth et al. [24] queried a cDNA microarray filter representing 4000 human genes to study the role of age, sex and strength training (ST) on human muscle gene expression. Sex turned out to have the strongest effect on muscle gene expression: comparing men and women with a threshold for significant difference at 1.7 fold 200 genes were found to be differentially expressed. Only 50 genes were identified as differentially expressed in relation to age and 69 genes were identified as differentially regulated in all groups in response to ST.

Conclusive remarks on microarrays.

For the moment, microarray technology is still limited in that it is a closed approach to global screening; only probes that are spotted on the array are inquired for expression and, as long as complete genome information is not available, this remains a limitation. The analyses made so far will need to be repeated when complete genome arrays are available in order to arrive at "complete expression profiles". Furthermore, the development of methods that allow absolute, rather than relative measures of gene expression is a goal in view of a definitive description of gene expression patterns.

Proteome and Proteomics

The term proteomics used to denote the study of the "proteome", i.e. the entire protein complement of a genome, is rather recent as it was likely first used in the middle of the "90s [26]. The proteome is much more complex than the genome: contrary to the long believed correspondence "one-gene, one-protein" from each gene many proteins can be generated by alternative splicing of transcripts, by extensive post-translational processing, or by a combination of the two. How many proteins are present in a multicellular organism such as the human body is not known, but estimations of 10 times more than the gene number are considered likely. The ultimate goal of proteomic studies is the identification of all proteins and of their changes in relation with physiological conditions or diseases.

Protein separation is a key component of proteomic analysis. Until now the heart of protein separation has been two dimensional gel electrophoresis (2DE), which separates proteins based on their inherent isoelectric point (charge) in the first dimension (isoelectrophocusing or IEF) and mass (molecular weight) in the second dimension. 2DE is not a new technique but only recently the introduction of immobilized pH gradients (IPG) has allowed a high reproducibility of the results amongst different laboratories [22]. Commercially available IPG strips can cover large pH gradients (for example 3 to 10) or narrow gradients (for example 4 to 5 or 4 to 7).

The application of 2DE to muscle tissue encounters several limitations [19]. Although several hundred muscle proteins can be separated by 2DE, several key myofibrillar proteins are difficult to separate and/or quantify by 2DE due to their extreme acidic or basic isoelectric point (troponin C, pI 4.04 (mouse); troponin I, pI. 9.57 (rat)) or high molecular weight (titin and myosin heavy chain).

There are, moreover, several other limitations

--- high concentrations of denaturing agents or ionic detergents are required for complete disaggregation and solubilization of the myofilament proteins. Unfortunately, even low concentrations of charged detergents are intolerable to the IEF step of 2D electrophoresis.

--- not all proteins stain with silver and Coomassie Blue equivalently

--- the extent of phosphorylation of several proteins (for example, myosin light chains) may vary depending on whether the tissue is homogenized in the presence or absence of the kinase and phosphatase inhibitors (EDTA, sodium vanadate and sodium fluoride)

--- membrane associated proteins tend to aggregate during protein separation due to their high hydrophobicity.

Recently protein separation methodologies that exploit another intrinsic characteristic of proteins, hydrophobicity, have been proposed and may represent the future of proteomics. Reversed phase (RP) chromatography, especially high performance liquid chromatography (HPLC), offers a new way to achieve the required separation of cellular proteins[19].

Proteomics and skeletal muscle

Whereas many studies in the last years have applied proteomics methodologies to cardiac muscle [1] analysing large samples from explanted hearts, very small samples from cardiac biopsies [18] and myocardial samples from experimental animals, much less work has been done on skeletal muscle.

For skeletal muscle proteins, however, 2D maps are already available for mouse and human muscles. A murine map has been published [25] and is available on public data base (http://www.expasy.ch) which contains gel images, protein identities and accession numbers. Maps of 2D gels map of human vastus lateralis muscle with the detection of 500 spots and identification of 150 proteins (indicated with their accession number) have been recently published [9]. Maps have also been obtained for rat muscle proteins [27]. Denervation [14], immobilization [15] and mitochondrial aging [6] represent first examples of application of proteomics to skeletal muscle biology.

It is likely that the coming years will see a large utilization of proteomic techniques to study muscle plasticity side-by-side with the traditional approaches based on immunohistochemistry and SDS-PAGE.

Address correspondence to:

Carlo Reggiani, Dept. of Anatomy and Physiology, Via Marzolo 3, I-35131 Padova, Italia;

phone +39 049 827 5313, fax +39 049 827 5301,
E-mail: carlo.reggiani@unipd.it

References

[1] Arrel Dk, Neverova I, And Van Eyk Je: Cardiovascular Proteomics. Evolution And Potential. *Circ Res* 2001; 88: 763-773.

[2] Bai Q, Mcgillivray C, Da Costa N, Dornan S, Evans G, Stear Mj, And Chang Kc: Development Of A Porcine Skeletal Muscle Cdna Microarray: Analysis Of Differential Transcript Expression In Phenotypically Distinct Muscles. *Bmc Genomics* 2003; 4: 8.

[3] Bakay M, Zhao P, Chen J, And Hoffman Ep**.** A Web-Accessible Complete Transcriptome Of Normal Human And Dmd Muscle: *Neuromuscul Disord* 2002; 12 Suppl 1: S125-141,

[4] Campanaro S, Romualdi C, Fanin M, Celegato B, Pacchioni B, Trevisan S, Laveder P, De Pitta C, Pegoraro E, Hayashi Yk, Valle G, Angelini C, And Lanfranchi G: Gene Expression Profiling In Dysferlinopathies Using A Dedicated Muscle Microarray. *Hum Mol Genet* 2002;11: 3283-3298.

[5] Campbell Wg, Gordon Se, Carlson Cj, Pattison Js, Hamilton Mt, And Booth Fw: Differential Global Gene Expression In Red And White Skeletal Muscle. *Am J Physiol Cell Physiol* 2001; 280: C763-768,.

[6] Chang J, Van Remmen H, Cornell J, Richardson A, And Ward Wf: Comparative Proteomics: Characterization Of A Two-Dimensional Gel Electrophoresis System To Study The Effect Of Aging On Mitochondrial Proteins. *Mechanisms Of Ageing And Development* 2003; 124: 33-41.

[7] Chen Jy-W, Zhao P, Borup R, And Eric P. Hoffman Ep: Expression Profiling In The Muscular Dystrophies: Identification Of Novel Aspects Of Molecular Pathophysiology. *J Cell Biol* 2000; 151: 1321-1336.

[8] Cheng G And Porter Jd: Transcriptional Profile Of Rat Extraocular Muscle By Serial Analysis Of Gene Expression. *Invest Ophthalmol Vis Sci* 2002; 43: 1048-1058.

[9] Gelfi C, De Palma S, Cerretelli P, Begum S, And Wait R: Two-Dimensional Protein Map Of Human Vastus Lateralis Muscle. *Electrophoresis* 2003; 24: 286-295,.

[10] Hampson R And Hughes Sm: Muscular Expression:Profiling Genes In Complex Tissue. *Genomebiology* 2001; 2: 1-3.

[11] Hittel Ds, Kraus We, And Hoffman Ep: Skeletal Muscle Dictates The Fibrinolytic State After Exercise Training In Overweight Men With Characteristics Of Metabolic Syndrome. *J Physiol* 2003; 548: 401-410.

[12] Holloway Aj, Van Laar Rk, Tothill Rw, And Bowtell Ddl: Options Available - From Start To Finish - For Obtaining Data From Dna Microarrays Ii. *Nature Genetics* 2002; 32 Supplement: 481-489.

[13] Hsiao Ll, Dangond F, Yoshida T, Hong R, Jensen Rv, Misra J, Dillon W, Lee Kf, Clark Ke, Haverty P, Weng Z, Mutter Gl, Frosch Mp, Macdonald Me, Milford El, Crum Cp, Bueno R, Pratt Re, Mahadevappa M, Warrington Ja, Stephanopoulos G, And Gullans Sr: A Compendium Of Gene Expression In Normal Human Tissues. *Physiol Genomics* 2001; 7: 97-104.

[14] Isfort Rj, Hinkle Rt, Jones Mb, Wang F, Greis Kd, Sun Y, Keough Tw, Anderson Nl, And Sheldon Rj: Proteomic Analysis Of The Atrophying Rat Soleus Muscle Following Denervation. *Electrophoresis* 2000; 21: 2228-2234.

[15] Isfort Rj, Wang F, Greis Kd, Sun Y, Keough Tw, Bodine Sc, And Anderson Nl: Proteomic Analysis Of Rat Soleus And Tibialis Anterior Muscle Following Immobilization. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002; 769: 323-332.

[16] Jagoe Rt, Stewart Hl, Lecker Mg, And Goldberg Al: Patterns Of Gene Expression In Atrophying Skeletal Muscles: Response To Food Deprivation. *Faseb J* 2002; 16: 1697–1712.

[17] Lanfranchi G, Muraro T, Caldara F, Pacchioni B, Pallavicini A, Pandolfo D, Toppo S, Trevisan S, Scarso S, And Valle G: Identification Of 4370 Expressed Sequence Tags From A 3'-End-Specific Cdna Library Of Human Skeletal Muscle By Dna Sequencing And Filter Hybridization. *Genome Res* 1996; 6: 35-42.

[18] Mcdonough Jl, Neverova I, And Van Eyk Je: Proteomic Analysis Of Human Biopsy Samples By Single Two-Dimensional Electrophoresis: Coomassie,Silver,Mass Spectrometry,And Western Blotting. *Proteomics* 2002; 2: 978 –987.

[19] Neverova I And Van Eyk Je: Application Of Reversed Phase High Performance Liquid Chromatography For Subproteomic Analysis Of Cardiac Muscle. *Proteomics* 2002; 2: 22-31.

[20] Noguchi S, Tsukahara T, Fujita M, Kurokawa R, Tachikawa M, Toda T, Tsujimoto A, Arahata K, And Nishino I: Cdna Microarray Analysis Of Individual Duchenne Muscular Dystrophy Patients. *Hum Mol Genet* 2003; 12: 595-600.

[21] Porter Jd, Khanna S, Kaminski Hj, Rao Js, Merriam Ap, Richmonds Cr, Leahy P, Li J, And Andrade Fh: Extraocular Muscle Is Defined By A Fundamentally Distinct Gene Expression

Profile. *Proc Natl Acad Sci U S A* 2001; 98: 12062-12067.

[22] Righetti Pg And Gianazza E: Immobilized Ph Gradients (Ipg). *Electrophoresis* 1992; 13: 185-186.

[23] Rome S, Clement K, Rabasa-Lhoret R, Loizon E, Poitou C, Barsh Gs, Riou Jp, Laville M, And Vidal H: Microarray Profiling Of Human Skeletal Muscle Reveals That Insulin Regulates ~ 800 Genes During An Hyperinsulinemic Clamp. *J Biol Chem*, 2003.

[24] Roth Sm, Ferrell Re, Peters Dg, Metter Ej, Hurley Bf, And Rogers Ma: Influence Of Age, Sex, And Strength Training On Human Muscle Gene Expression Determined By Microarray. *Physiol Genomics* 2002; 10: 181-190.

[25] Sanchez Jc, Chiappe D, Converset V, Hoogland C, Binz Pa, Paesano S, Appel R, Wang S, Sennitt M, Nolan A, Cawthorne Ma, And Hochstrasser Df: The Mouse Swiss-2d Page Database: A Tool For Proteomics Study Of Diabetes And Obesity. *Proteomics* 2001; 1: 136-163.

[26] Wasinger Vc, Cordwell Sj, Cerpa-Poljak A, Yan Jx, Gooley Aa, And Wilkins Mr: Progress With Gene-Product Mapping Of The Mollicutes: Mycoplasma Genitalium. *Electrophoresis* 1995; 16: 1090-1094.

[27] Yan Jx, Harry Ra, Wait R, Welson Sy, Emery Pw, Preedy Vr, And Dunn Mj: Separation And Identification Of Rat Skeletal Muscle Proteins Using Two-Dimensional Gel Electrophoresis And Mass Spectrometry. *Proteomics* 2001; 1: 424-434.