Genetic analysis of the mouse brain proteome

Joachim Klose¹, Christina Nock^{1–3}, Marion Herrmann¹, Kai Stühler⁴, Katrin Marcus⁴, Martin Blüggel⁵, Eberhard Krause⁶, Leonard C. Schalkwyk⁷, Sohaila Rastan⁸, Steve D.M. Brown⁹, Konrad Büssow³, Heinz Himmelbauer³ & Hans Lehrach³

Published online: 25 March 2002, DOI: 10.1038/ng861

Proteome analysis is a fundamental step in systematic functional genomics. Here we have resolved 8,767 proteins from the mouse brain proteome by large-gel two-dimensional electrophoresis. We detected 1,324 polymorphic proteins from the European collaborative interspecific backcross. Of these, we mapped 665 proteins genetically and identified 466 proteins by mass spectrometry. Qualitatively polymorphic proteins, to 96%, reflect changes in conformation and/or mass. Quantitatively polymorphic proteins show a high frequency (73%) of allele-specific transmission in codominant heterozygotes. Variations in protein isoforms and protein quantity often mapped to chromosomal positions different from that of the structural gene, indicating that single proteins may act as polygenic traits. Genetic analysis of proteomes may detect the types of polymorphism that are most relevant in disease-association studies.

Introduction

In a strict sense, functional genomics means the determination of the specific function of each individual gene in a genome. Proteomics may offer an approach towards this goal. The experimental procedure would include separating and presenting all of the different proteins of an organism, detecting protein functions by characterizing each protein according to a broad spectrum of chemical and biological parameters, and matching each protein with its gene.

Two-dimensional electrophoresis (2-DE)^{1,2}, under conditions that allow maximum resolution^{3,4}, has the potential to reveal most, if not all, of the different proteins of a certain tissue or cell type of an organism. On the basis of 2-DE patterns, proteins can be characterized on a large scale⁵ with respect to many biological properties, including specificity for tissues, specificity for certain stages of development and aging, specificity for distinct cell organelles, sex specificity, quantitative characteristics (such as relative protein amount and turnover rate) and genetic features (such as polymorphisms and mode of transmission).

In combination with other techniques such as immunoblotting, specific staining and, in particular, techniques for determining peptide masses and peptide sequence tags, 2-DE also offers tools for characterizing chemical protein properties, such as amino-acid sequence and side-chain modifications. Matching proteins with their genes is most feasible when both the genome of an organism is sequenced completely and sequence data from 2-DE spots can be obtained by high-throughput mass spectrom-

etry. Proteins from 2-DE patterns can also be linked to their genes by a genetic approach that allows one to map genes encoding various characters of proteins.

The European collaborative interspecific backcross (EUCIB) has been established from two distantly related mouse strains, *Mus musculus* strain C57BL/6 (B6) and *Mus spretus* (SPR). These species diverged about three million years ago^6 . Compared with other species or subspecies from which hybrids can be produced under standard laboratory conditions, SPR shows the greatest degree of allelic variation⁷. The EUCIB experiment was undertaken to construct a high-resolution genetic map of the mouse genome^{8,9}. The backcross progeny (B₁) obtained comprised 982 animals and led to a genetic resolution of 0.3 cM.

Here we report the first results of a comprehensive genetic study of mouse proteins using the EUCIB resource. We prepared five organs (brain, liver, heart, kidney and muscle) from 200 B_1 animals and investigated the brain proteins as a first step. We separated the proteins by our large-gel 2-DE technique^{3,4} and used the patterns obtained to consider various protein characteristics: protein molecular mass and isoelectric point, determined by the electrophoretic position of protein spots; the relative abundance of proteins; the presence or absence of proteins and the occurrence of isoforms of proteins that might result from co- and post-translational modifications, alternative splicing and cleavage or degradation.

By comparing protein patterns from B6 and SPR and considering these various protein properties, we detected more than

 ¹Institut für Humangenetik, Humboldt-Universität zu Berlin, Charité, Campus Virchow-Klinikum, Augustenburger Platz 1, D-13353 Berlin, Germany.
 ²Fachbereich Biologie, Chemie, Pharmazie, Freie Universität Berlin, D-14195 Berlin, Germany. ³Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany. ⁴Institut für Physiologische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany. ⁵Protagen AG, D-44801 Bochum, Germany.
 ⁶Institut für Molekulare Pharmakologie, D-13125 Berlin, Germany. ⁷SGDP Research Centre, Institute of Psychiatry, London SE5 8AF, UK. ⁸MWB Business Exchange Suite 112/113 Wellington House, Cambridge CB1 1BH, UK. ⁹MRC Mammalian Genetics Unit and UK Mouse Genome Centre, Harwell, Oxford OX11 ORD, UK. Correspondence should be addressed to J.K. (e-mail: joachim.klose@charite.de) or H.H. (e-mail: himmelbauer@molgen.mpg.de).

	Number of variant protein spots/spot complexes (single				
	Categories of	of genetic sig	Total		
Types of genetic variation of protein spots in 2-DE patterns	Ă	В	C	no.	% ^c
Qualitative variation					
Variation in electrophoretic mobility (mV)	293 (574)	34 (46)	39 (48)	366 (668)	(7.62)
Quantitative variation					
Presence/absence (paV)	119 (137)	119 (132)	67 (72)	305 (341)	(3.89)
Variations in protein amount (aV)	29 (33)	132 (165)	104 (117)	265 (315)	(3.59)
Total	148 (170)	251 (297)	171 (189)	570 (656)	(7.49)
Total	441 (744)	285 (343)	210 (237)	936 (1324)	(15.10)

1,000 polymorphic proteins. We mapped several hundred of these onto the mouse genome and identified them by mass spectrometry. This genetic analysis of proteins resulted in mapping of chromosomal positions not only for structural genes, but also for regulatory loci and modifier genes. Our findings therefore show that different types of the same protein may map to different loci on the genome. Consequently, even a single protein can be considered as a polygenic trait.

Results

Large-scale detection of protein polymorphisms

We extracted proteins from mouse brain and resolved them by our large-gel technique, which detects more than 10,000 protein spots in a 2-DE pattern^{3,4}. According to our standard procedure, the total protein from a tissue was separated into three fractions (cytosol, membrane and chromatin fractions) in a way that avoids the loss of any proteins¹⁰. Here, we investigated the cytosol fraction of the brain (the brain supernatant fraction). A 2-DE standard pattern for this protein fraction has already been established¹¹. We documented 8,767 protein spots (8,595 B6 spots plus 172 spots absent in B6 but present in SPR) and assigned them each a 'spot number'.

Comparing the protein patterns from B6 and SPR, we detected 1,324 consistently variant protein spots (Table 1 and see Web Table A on the *Nature Genetics* website). These variations were considered to be species-specific protein polymorphisms. We divided the polymorphic proteins into different categories according to the type of variation (Fig. 1): (i) variation in electrophoretic mobility (mV), (ii) variation in protein quantity or 'spot volume', which may be an increase/decrease in the protein amount (aV), or a presence/absence (paV) of protein and (iii) variation in the number and position of isoforms ('isospots') into which a protein may split (sV). Combinations of different types of variation were also

observed and are described here as, for example, mV + aV. The spot image that a polymorphic protein may create in a 2-DE pattern is considered to be a phenotype of that protein¹².

The variant spots detected were assigned a 'variation number', which was listed along with the spot number (see Web Table A) and identified in the standard 2-DE pattern. Frequently, spots occurred in dense arrays in 2-DE patterns, for example, in a horizontal spot series. We call such spot formations a 'spot complex' (Fig. 1). These complexes usually showed the same type of genetic variation, if polymorphic, and clearly represented groups of isospots of the same protein. Isospots of a variant spot complex were therefore given the same variation number. As a consequence, polymorphic spots were evaluated either as single spots (1 complex=*n* spots), or alternatively as spot complexes (1 complex=1 spot). Spots were



Fig. 1 Protein phenotypes determined by 2-DE. Brain proteins of the two mouse lines B6 and SPR and their F1 and B1 progeny are shown. a, Protein spot mV367BS (D10Jhu81e, chromosome 10): mobility polymorphisms of a single spot show that even changes in spot position as small as 1 mm (distance between spot centers) could be recognized unequivocally in large gels and used to distinguish heterozygotes (asterisk) from homozygotes throughout the backcross progeny (B1 patterns). This also shows the reproducibility of our 2-DE technique. b, Protein spot paV385BS (Cbr1m2, chromosome 16) shows that quantitative polymorphisms in proteins can be mapped genetically if the protein levels of homozygotes and heterozygotes are clearly distinguishable. c, Protein spot mV219BS+aV (Dia4, chromosome 8): allelic variation in protein amount in combination with a positional shift shows the allele-specific transmission of protein levels from the parental strains to the B₁ generation. d, Protein complex mV200BS (Suclg1, chromosome 6): mobility polymorphisms of a spot complex consisting of a series of four isospots. Variation in the number of isospots (spot splitting, sV) has been shown elsewhere¹¹. Usually, however, we scored variation in spot splitting as paV spots (see Web Table B). e, Pattern of GAPDH determined by B6-SPR hybrids. An electrophoretic mobility polymorphism of the parents produced a double spot in the hybrid characterized by a horizontal 0.04-pH shift and a vertical 600-dalton shift.

etermined in 2-DE Is by recombinant protein markers		Di		Number of var blecular mass (d		ic protein spots		Total
Markers	0	>0–100	>100–150	>150–250	>250-500	>500–1000	>1000–2500	
10–20 kD	1	6	6	7	13	7	2	42
20–30 kD	4	19	13	11	15	2	1	65
30–40 kD	2	14	2	17	22	10	1	68
Total	7	39	21	35	50	19	4	175
%	4	22	12	20	29	11	2	100

also considered in terms of 'spot families', which were defined as all the isospots that a protein may create in a 2-DE pattern. A spot family comprises single spots and possibly spot complexes and can be distributed arbitrarily over the whole 2-DE pattern. Spot families can be detected only by analyzing a 2-DE pattern by mass spectrometry but, in the case of polymorphisms, the very characteristic mV variations also allow the recognition of spot families (see below).

Dividing the polymorphic protein spots into qualitatively (mV) and quantitatively variant (aV and paV) spots showed that these two types of protein polymorphism occurred with similar frequencies (Table 1). However, quantitative variation was also frequently involved in qualitative polymorphisms. Quantitative variations were taken into account only if the difference between the two parental spots was quite obvious (Fig. 1). Densitometric evaluation of our spot patterns, which is not feasible at present, would certainly detect many more, less obvious quantitative variations. Thus, quantitative differences between allelic spots may be much more common than we have found.

To analyze the genetic determination of the protein polymorphisms detected between B6 and SPR (P generation), we investigated these polymorphisms in the F_1 (B6 × SPR) and the B_1 (F_1 × SPR) generations. Following the 1,324 variant protein spots throughout the three generations identified three categories of genetically significant variation (Table 1). Category A spots segregated into heterozygotes and SPR homozygotes among the 64 B_1 animals investigated, as expected from a backcross. Category B spots showed genetically interpretable patterns in F_1 but did not segregate in B_1 , usually because of dominance of the SPR phenotype. In some cases, these spots segregated irregularly, probably owing to quantitative trait loci¹³. Category C spots were difficult to evaluate, even in F_1 patterns, mainly because their intensities were not distinct enough for genetic interpretation.

We grouped the polymorphic spots according to the mode of inheritance of their phenotypes. The 327 qualitative variants (category A + B, Table 1) were transmitted codominantly without any exceptions and clearly showed double spots in heterozygotes. We evaluated the 570 quantitative variants (category A + B + C, Table 1) by estimating visually whether the volume of the hybrid spots was similar to that of B6, of SPR, or intermediate. The hybrid patterns showed 40% intermediary volumes, 19% dominance of the parental spot carrying the higher volume, and 11% dominance of the lower volume spot, including 5% dominance of spot absence. Of the aV and paV variant spots, 171 (30%) were category C spots.

Particular features of protein phenotypes

In our large 2D gels, the horizontal distance between allelic mV spots varied from 0.5 to 32 mm (0.3pH units). Electrophoretic

mobility changes of proteins in the horizontal direction is commonly explained by missense mutations that led to aminoacid substitutions involving changes in charge. Unexpectedly, however, most of the positional variants varied also, or exclusively, in the vertical direction, which suggested that there were changes in protein molecular mass. The difference in molecular mass that an amino-acid substitution can effect is, of the 380 possible amino-acid substitutions, only greater than 100 daltons in six cases (1.6%) and does not exceed 129 daltons (ref. 14).

To estimate the range within which the molecular mass of polymorphic proteins varied, we investigated B6–SPR F_1 hybrid patterns with respect to the vertical distances between the two allelic spots of hybrid double spots (Fig. 1*e*). The distances varied between 1 and 10 mm. Of the 293 mobility changes detected in category A (Table 1), we evaluated 175 corresponding to those situated in the linear range of the molecular mass markers that were run on the 2-DE gels of hybrids. Spot complexes and families were taken as single spots.

About three-quarters of the proteins differed between B6 and SPR by more than 100 daltons, with a maximum difference of about 2,500 daltons (Table 2). Only 4% of the mobility variations investigated did not show a detectable difference in molecular mass between the two variant forms of allelic protein pairs. Although SDS gel electrophoresis is not precise enough for molecular mass determinations down to 100 daltons, we expected reliable results under our experimental conditions, because the proteins being compared were actually identical proteins (polymorphic forms of the same protein) that were run on the same gel and evaluated only with respect to differences in molecular mass and not absolute mass.

To explain why changes in protein molecular mass that were not reasonably due to amino-acid substitutions commonly accompanied polymorphisms, we examined glyceraldehyde 3phosphate dehydrogenase (GAPDH) in more detail as a representative case. The B6 GAPDH and SPR GAPDH spots (mV145BS) differed in position on the 2-DE gels by about 600 daltons (Fig. 1*e*); however, TOF (time-of-flight) mass spectrometry with on-blot desorption infrared matrix-assisted laser desoprtion-ionization (MALDI)¹⁵ revealed a nonsignificant difference of 98 daltons (J. Mattow and D. Schleuder, per. comm.).

Determinations of peptide mass for B6 GAPDH and SPR GAPDH with a sequence coverage of 80% identified a neutral amino acid substitution (Thr81Ala) with an alteration in mass of 30 daltons. Even if another substitution were to exist in the remaining 20% of sequences, this would not explain the 600-dalton shift between B6 and SPR GAPDH observed on the SDS gels. Thus, the vertical shift does not result from molecular mass alterations, but rather from changes in protein conformation that affect the constant ratio of 1.4 g of SDS per g of

protein or the shape of the SDS-protein complex. When protein samples were run on a 15% and a 10% acrylamide SDS gel for comparison, many protein spots situated on the 15% gel in the same molecular-weight range ran to different vertical positions on the 10% gel and vice versa (data not shown). By contrast, recombinant marker proteins migrated on 15% and 10% gels according to mass. It seems that, even under the conditions of an SDS gel, the migration of proteins may often depend on the individual shape of proteins, which in turn depends on protein conformation.

We found that proteins that were polymorphic in electrophoretic mobility were often also polymorphic in protein amount (Fig. 1). This type of protein phenotype could be used to study the genetic transmission of quantitative variation of single alleles. Of the 293 proteins variant in electrophoretic mobility (Table 1, category A), 221 showed clearly detectable differences in spot volume between B6 and SPR. Of these, 134 showed differences in quantity that were prominent enough to allow visual evaluation in F₁ and B₁ patterns (Fig. 1).

In analyzing these proteins genetically, we observed that the allelic differences in protein quantity between the homozygous parental strains B6 and SPR were mostly (73%, Table 3) retained in the heterozygotes of the F_1 and B_1 generations. Equalization of the allelic differences in protein levels, which is expected to occur in heterozygotes because of trans-acting regulation, only occurred in 6% of the proteins evaluated. The observations that quantitative variation of proteins often occurs in conjunction with electrophoretic mobility changes and that different protein expression levels are transmitted in an allele-specific manner to the next generation suggests that there is a strong link between the sequence of structural genes and the cellular concentration of proteins.

Genetic mapping of protein phenotypes

For each of the 441 polymorphic proteins of category A (Table 1), which comprised 744 single spots in the 2-DE standard pattern of the brain supernatant fraction, we scored the genetic segregation in the backcross progeny (Fig. 1) and used the data for genetic linkage analysis and gene mapping. We mapped 409 protein spots and/or complexes (665 single spots) chromosomally on the EUCIB cross, identifying 273 mV, 115 paV and 21 aV spots (Fig. 2 and see Web Fig A and Web Table A). Of these, we

analyzed 259 spots/complexes (466 single spots) by mass spectrometry (65 indirectly, see Methods). We also identified 39 nonvariant proteins. Allelic SPR spots usually correspond unambiguously with the allelic spots of B6 (ref. 16) and were identified only in a few cases.

In total, we identified 189 different proteins (see Web Table A): 67 were newly mapped on the mouse chromosomes (including 13 that matched only anonymous cDNA clones); 42 confirmed known chromosome positions; 39 were not variant in B6 or SPR but were mapped already (Fig. 2, reference chromosome) and 41 proteins were specific to spot families that were genetically heterogenous (see Web Table B on the Nature Genetics website). These families showed different mapping positions for isospots of the same protein, indicating that different genes act on a single protein. For example, the aldolase 3 C isoform mapped with spot mV159BS to chromosome 11, as known for this protein, but also mapped with spot paV419BS to chromosome 7. In other cases, 'one-spot families' were detected

Fig. 2 Genetic map of mouse chromosome 3 (MMU3). Polymorphisms occurring between the two mouse lines B6 and SPR in brain proteins were studied on 2-DE patterns from 64 mice of the EUCIB interspecific mouse backcross to establish a segregation pattern and the chromosomal position for each variant protein. . The positions of the mapped proteins are indicated by the variation number (composed of type of variation plus running number of variant protein spot plus brain supernatant fraction). If the protein was identified, then the gene symbol of the protein is given. Proteins identified by MALDI-TOF MS are boxed. For comparison, the reference map of MMU3 (Chromosome Committee map) is shown on the left side. Lines are drawn between genes on the reference map and mapped proteins that correspond to the product of these genes. Two nonpolymorphic protein spots (B3 279 and B2 200) were found to correspond to Atp5b. Chromosome panels for the completed set of 409 genetically mapped protein spots and protein complexes plus information on identified proteins are given in Web Fig. A. Markers on the EUCIB map are map anchors (Mit Markers, bold), secondary map anchors (Mit Markers, not bold) and IRS-PCR markers (green; see Methods). We assigned proteins to chromosomes if they linked to a marker on the map at a lod threshold of 4.0 in pairwise linkage calculations (see Web Table A). Proteins were then ordered with respect to mapped markers and with respect to each other. Proteins that could not be placed unambiguously on the map (map orders supported by log-likelihood ratios smaller than 2.5) were assigned to map intervals or put on the chromosome as 'linked'. Map dis-



© 2002 Nature Publishing Group http://genetics.nature.com

in which the isospot mapped to a position that was different from that known for the corresponding gene. At the maximum, we found 3 different positions for a single protein and, in total, 92 different positions for 41 proteins (see Web Table B).

Protein spots and number of proteins

The mV spots of category A (Table 1) could be grouped into spot families on the basis that spots of the same mV protein usually shifted precisely by the same distance and in the same direction and mapped to the same chromosome position. In addition, we confirmed the identity of spots in several cases by mass spectrometry. Figure 3 shows three of the largest spot families detected in the brain protein pattern. The total results (Table 4) show that about two-thirds of our standard pattern of brain proteins comprises isospots. This is certainly an underestimate, however. A completely identified 2-DE pattern would probably reveal additional variant spots belonging to a heterogeneous spot family that were not detectable by the mV characteristics of that family.

Table 4 shows that most of the proteins occurred in the form of single spots or small spot families, and that more than 50% of the protein spots resulted from a few (12) large spot families. As published elsewhere, we detected in 2-DE patterns from human brain about 100 isospots for the protein tau¹⁷ and in patterns from the heart more than 500 spots for the heavy chain of myosin¹¹. The relationship between number of spots and number of proteins in a 2-DE pattern depends largely on the properties of the proteins that compose the pattern.

Discussion

Separating mouse brain proteins by 2-DE allowed us to compare about 8,700 proteins between two mouse lines, B6 and SPR, that represent two different species. Among these proteins, 668 (7.6%) showed variation in electrophoretic mobility, which was caused most probably by single amino-acid polymorphisms (SAPs)^{16,18} resulting from single-nucleotide polymorphisms (SNPs)^{19–22}. From the small amount of public sequence data (GenBank) currently available for both B6 and SPR, we calculated a frequency of 0.6% SNPs in coding sequences (1 SNP every 170 bp, 10 genes, 11,790 coding bp),

The considerable discrepancy between the frequency of 7.6% qualitative protein polymorphisms actually detected by 2-DE and the frequency predicted on the basis of SNPs suggests that 2-DE is not sufficiently sensitive to reveal all of the changes that SAPs may induce in protein molecules. Numerous studies that tested the sensitivity of electrophoresis in detecting protein polymorphisms show, however, that isoelectric focusing detects even neutral amino-acid substitutions^{18,24} and that differences in a single amino acid can alter the electrophoretic mobility of proteins on SDS gels²⁵. These observations have been explained by amino-acid substitutions that induce changes in protein conformation or post-translational modifications²⁶⁻²⁸. Changes in protein structure have an effect on the SDS-binding capacity and the shape of the proteins, and therefore on the electrophoretic mobility of the SDS-protein complex. On the basis of these observations, we assume that most of the SAPs have no further consequences on the structural properties of proteins, and therefore result in electrophoretically silent polymorphisms. It is estimated that for interspecies variation, only 24% of polymorphic sites occur in structurally and functionally important regions of proteins²².

Using long-distance gels, we found that almost all (96%; Table 2: 100%–4% no difference in molecular mass) of the mobility polymorphisms included shifts in the vertical direction. True changes in the mass of proteins may have deleterious effects and, most likely, do not occur among established protein polymorphisms, which supports our assumption that the vertical shifts result from changes in protein conformation and shape. The deleterious effects of disease-causing missense mutations are frequently caused by alterations in protein conformation that result in misfolding, destabilization and degradation of the polypeptide^{19,29}. Thus, natural polymorphisms in protein conformation may be of the type that becomes deleterious under certain conditions and contributes to the genetic risk for common diseases. This suggests that 2-DE is a powerful approach for identifying highly predictive disease-related polymorphisms.

We found that at least 75% (221 of 293) of the mobility-variant proteins also showed quantitative variation, and that the parental differences in protein levels were mostly (73%) retained in the heterozygote protein patterns. It seems that the

and found that one SAP results from one SNP in a third of cases (that is, 1 SAP roughly every 500 bp). The mass of a protein encoded by 500 bp is about 19 kD.

In the standard pattern of brain proteins, 7,420 protein spots were in the 19-200 kD range, and thus should be polymorphic with 1-10 SAPs per spot. For the remaining 1,347 spots situated in the range of 9-19 kD, we calculated that 1,010 spots should contain one SAP and 337 spots should contain no SAPs. Ignoring the rather small portion of proteins that are completely conserved²¹ between mouse species²³, our rough estimation indicates that most (possibly up to 90%) proteins revealed in our B6-SPR 2-DE pattern should carry at least one SAP.

 Table 3 • Genetic determination of protein amount revealed by homologous B6–SPR

 proteins with variant electrophoretic mobility^a

Protein expression		Comparison of allelic expressi of proteins in B6 and SPR ^b			
	P generation	F ₁ generation	B ₁ generation (heterozygtes)		mber roteins
Determination in <i>cis</i> (allele-specific) ^c					73%
	B6 ^{↑d} , SPR↓ SPR↑, B6↓	B6↑, SPR↓ SPR↑, B6↓	B6↑, SPR↓ SPR↑, B6↓	33 65	, 3 / 6
Determination in <i>trans</i> (allele-unspecific)					6%
	B6↑, SPR↓ SPR↑, B6↓	B6=SPR SPR=B6	B6=SPR SPR=B6	7 1	
Irregular	B6↑, SPR↓	B6↓, SPR↑	B6↓, SPR↑	4	21%
	SPR↑, B6↓ B6=SPR	SPR↓, B6↑ B6↑, SPR↓	SPR↓, B6↑ B6↑, SPR↓	2 3	
	SPR=B6	SPR \uparrow , B6 \downarrow	SPR [↑] , B6↓	8	
	B6=SPR SPR=B6	B6=SPR SPR=B6	B6↑, SPR↓ SPR↑, B6↓	5 6	
Total				134	100%

^aFor examples see Fig. 1. ^bNumber of patterns investigated: 10 P, 10 F₁ and 64 B₁ patterns. ^cProtein expression was considered as allele-specific if the relationship in staining intensity of allelic protein spots, observed in the parental 2-DE patterns, was retained in the heterozygous F₁ and B₁ patterns. ^dArrows indicate that the spot volume (staining intensity) of a protein was higher (\uparrow) or lower (\downarrow) than in the other mouse strain.

Fig. 3 2-DE brain protein pattern from a B6-SPR hybrid. From the whole 2-DE pattern, consisting of an acid half and a basic half (see Methods), only the acid half is shown. The three protein spot families show that spot families can be recognized in 2-DE patterns on the basis of genetic variation. The spot family of heat-shock 70-kD protein 4 (HSP70, red) consists of 52 'isospots', which occur in 52 double spots (hybrid spots). The two allelic forms vary in the vertical direction in all 52 double spots, with a spacing of 0.5 mm and the B6 spot on top. The γ -enolase 2 family (blue) shows 24 double spots, which vary again in the vertical direction with the B6 spot on top, but with a spacing of 1 mm. The lactate dehydrogenase 2 B chain spot family (LDH2, yellow) includes 17 double spots that form in each case horizontal spot pairs spaced at 25 mm, with the B6 spots always on the left side. This family also includes 10 paV spots, which are interpreted as a difference in degradation rate between the two allelic forms¹¹. Four spots were present on the basic half of the 2-DE pattern (data not shown).

protein amount is determined by the allele-specific sequence of the coding DNA. Quantitative variation was not only observed in combination with electrophoretic mobility changes, but was a common feature of protein phenotypes. To explain this observation, we compared noncoding DNA sequences from *Mus musculus* and *Mus spretus* (GenBank) and found that SNPs occur at higher frequencies in non-coding (2.2% in introns (73 SNPs/3,264 bp), 1.4% in 5' untranslated regions (UTRs, 6/425) and 3' UTRs (44/3,231) than in coding (0.6%) sequences.

We constructed an average gene from available data^{30,31} with the following composition: 1,201 coding basepairs, 5,795 intron basepairs and 1,033 UTR basepairs (leaving out the adjacent promotor and enhancer sequences). According to these data, 4–5 SNPs would occur in the coding regions, with 135 SNPs in the non-coding regions of a gene. An important role in quantitative variation is attributed to polymorphisms in

noncoding sequences^{23,31–35}. This may explain a high genetic variability in protein quantity and support the conclusion that the cellular concentration of proteins is primarily allele-specific and is determined by the structural genes, the noncoding and the coding sequences. Given that the structural gene determines the basis concentration, however, several additional loci may contribute to the final cellular concentration of a protein¹³.

By genetically mapping various protein phenotypes as revealed by 2-DE, we observed in many cases that variation in a particular isospot of a protein did not map to the locus of the structural gene of this protein. This may be explained by polymorphisms in proteins that act on the variant protein observed and that are coded by 'protein modifier loci'. The presence/absence of isospots was the type of variation that frequently mapped to a 'protein modifier locus', whereas variation in spot series (which may reflect glycosylation, phosphorylation and other modifications)^{17,36} was observed rarely. Presence/absence of isospots may indicate variation in protein stability, limited proteolysis or protein degradation and would suggest that there are polymorphisms in proteins acting as chaperones, proteases or components of the proteasome complex³⁷. The dependence of the function of a protein on several genes may explain why even the phenotype of a monogenic disease can vary considerably in symptoms and severity among individuals¹⁹.

Methods

Interspecific mouse backcross. We used EUCIB mice^{8,9} produced at the Medical Research Council Clinical Research Centre, Harrow/London. We examined 10 females of the parental mouse species *Mus musculus*



domesticus, strain C57BL/6 (B6), 10 females of the parental species *Mus spretus* (SPR), 6 (B6 × SPR) F_1 females and 64 ($F_1 ×$ SPR) B_1 females. We excised the brain, liver, heart, kidney and muscle from these animals and investigated the brains in this study.

Protein extraction. We fractionated total mouse brains singly into a supernatant fraction containing the buffer-soluble proteins (cytosolic proteins), a pellet-extract fraction containing urea/CHAPS-soluble proteins (membrane proteins) and the final fraction (pellet suspension) obtained after extracting the first pellet and then releasing the chromatin-bound proteins from the final pellet by DNA digestion¹⁰. The brain supernatant fraction, containing most of the proteins, was used here.

2D electrophoresis. We separated the protein extracts by large-gel 2-DE^{3,4}. The gel format was 46.4 cm (isoelectric focusing direction) \times 30 cm (SDS–PAGE direction) \times 0.75 mm. We loaded 9 µl (0.1 mg of protein) of the protein sample prepared¹⁰ from the brain supernatant fraction on an isoelectric focusing tube gel (0.9 mm diameter). We prepared the gels with carrier ampholytes and used silver staining to visualize the proteins. The staining protocol has high sensitivity and reproducibility⁴ and requires the conductivity of the water to be kept below 0.1 µS cm⁻¹. We added thimerosal to the stop solution to get a nearly homogenous spot color. For sample comparisons, particularly the two samples of each B6–SPR pair, gels were run and stained in parallel. Proteins for analysis by mass spectrometry were run in 1.5-mm diameter gels loaded with 50 µl of sample and stained with colloidal Coomassie Brillant Blue G-250 (ref. 38). For recombinant marker proteins, a 10-kD protein ladder (GibcoBRL) was used.

Mass spectrometry. We excised protein spots from Coomassie blue–stained 2-DE gels and digested the proteins in-gel with trypsin³⁹. We analyzed tryptic peptides and identified proteins with MALDI-TOF MS using a REFLEX III spectrometer (Bruker-Daltonic) equipped with a

Number of spots per protein	Number of proteins (spot families)	Number of single spots
Unique spots 1 spot/protein	61% 111 proteins	111 spots
Spot families of low size 2 spots/protein 3 spots/protein	21% 32 proteins 7 proteins	64 spots 21 spots
Spot families of middle size 4 spots/protein 5 spots/protein 6 spots/protein 7 spots/protein 8 spots/protein 8 spots/protein 9 spots/protein	11% 6 proteins 4 proteins 4 proteins 1 protein 1 protein 1 protein	24 spots 21 spots 24 spots 28 spots 8 spots 8 spots 9 spots
Spot families of large size 12 spots/protein 13 spots/protein 14 spots/protein 17 spots/protein 18 spots/protein 19 spots/protein 22 spots/protein 27 spots/protein 23 spots/protein 38 spots/protein 52 spots/protein	7% 1 mV103BS protein 1 calcium-binding protein 1 tumor rejection antigen gp96 1 calpastatin 1 heat-shock 70-kD protein 8 1 aconitase 2 (modifier) 1 mV119BS protein 1 GAPDH 1 γ-enolase 2 1 LDH 2, B chain 1 synapsin I 1 heat-shock 70-kD protein 4	12 spots 13 spots 14 spots 14 spots 17 spots 18 spots 22 spots 24 spots 27 spots 38 spots 52 spots
Total	182 proteins	580 spots
Total 2-DE pattern (estimation):	2,770 proteins	8,767 spots
Mean value	1 protein	3 spots

Detection of protein polymorphisms. We compared visually the 2-DE patterns of brain proteins from B6 and SPR. Any changes observed between B6 and SPR patterns were categorized according to the three basic types of variation: variation in spot position (electrophoretic mobility), variation in presence/absence and variation in spot volume (the integral of spot area and staining intensity, reflecting the protein amount of the spot). Computer software for detecting genetic positional shifts in 2-DE images were not available, but these changes can be detected easily in the parental strains by overlaying the B6 and SPR gels and can be confirmed by the occurrence of double spots of half intensity in the hybrid patterns. Commercial programs available for quantitative evaluation of 2-DE patterns require manual correction of spot detection and spot matches to an extent that makes a comprehensive analysis of large-scale protein patterns not feasible.

We restricted the evaluation of spot volume to spots showing obvious differences between the two parental strains B6 and SPR—that is, changes distinct enough to enable us to recognize spot intensity levels ranging between those of the parental strains in the heterozygous $(F_1/B_1$

SCOUT 384 ion source. The acceleration voltage was 25 kV, the reflection voltage was 21.6 kV, the ion source acceleration voltage was 13.5 kV and the reflector-detector voltage was 1.6 kV. PSD (post source decay) spectra were obtained with a parent ion selection of around 30 daltons. We reduced the reflection voltage in 14 steps from 21.6 kV to 0.65 kV. Data were acquired on a SUN Ultra using XACQ software, version 4.0.2. Post-analysis processing was done using XMASS software, version 5.0.

We analyzed the obtained peptide mass fingerprint spectra by searching the nonredundant protein database of the National Center for Biotechnology Information (NCBI) with ProFound v. 4.8.2 (ref. 40). We restricted the parameters for the searches in the molecular weight (<250 kD) and taxonomy categories (animals only). The modification propionamide on cysteine residues was regarded as a complete modification, whereas the methionine oxidation was considered as a partial modification in the database search settings. One was used as the maximum missed tryptic cleavage sites in the peptide in the default search settings. We considered only the monoisotopic peptide masses in the searches with ProFound and set the mass tolerance for internal calibration to around 50 ppm.

For automatic interpretation of fragment ion spectra, we used the SEQUEST algorithm (v. 27) and screened the NCBI nonredundant protein database first and the expressed-sequence tags database³⁹ second. The chosen parameters for PFFSpectraSearch were monoisotopic masses, a mass tolerance of 0.5 daltons for the parent ions and 1.3 daltons for the fragment ions, tryptic digest and acrylamide modification of cysteine residues as a fixed modification. We set the maximal missed cleavages to 1 and the weighting factors for the B and Y ions to 1. The identification of proteins due to new entries in the sequence database was assured by repeating the search of uninterpretated mass spectrometry data automatically on a weekly basis (last search 8 October 2001). Proteins identified through our work are listed in ref. 11. A protein spot was considered to be indirectly through identified if identical in both phenotype (for example, the same distance and shift direction of the two allelic spots) and mapping position with a directly identified spot (for example, a single spot of a spot series).

generation) patterns. The genetic evaluation of spot volumes requires only the distinction of the two parental levels and the hybrid level, and not precise densitometrically measured spot volumes. To validate the quantitative polymorphisms detected between B6 and SPR, we screened six B6–SPR gel pairs independently for quantitative variations and then compared the six pairs side by side to recheck the variations, which were marked, in particular, if scoring results between spot pairs were inconsistent. The spot variants that were confirmed six times (one exception in six spot pairs was accepted if caused by a local technical problem, such as a gel-running artifact) were confirmed further by genetic analysis.

Genetic analysis of protein polymorphisms. We analyzed the protein polymorphisms detected genetically by investigating 4–6 2-DE patterns from F_1 animals and 64 patterns from B_1 animals. In case of monogenic codominant or intermediary (additive) transmission of spot phenotypes, the heterozygous F_1 pattern should show a double spot if variation concerns spot mobility, and an intermediate spot volume if variation in presence/absence or spot volume occurred in the parental spots. The phenotypes revealed by each variant spot in the F_1 pattern should segregate among the B_1 patterns with roughly 50% heterozygous and 50% SPR homozygous phenotypes. For each variant protein spot detected in the parental 2-DE patterns and found to follow this mode of inheritance, the corresponding genetic status was determined in each of the 64 B_1 patterns. We also evaluated B6-dominant spots. The segregation pattern obtained for each variant spot is of the gene underlying the polymorphism observed.

Mapping of DNA markers. The 982 B₁ animals of the EUCIB resource have been genotyped with a set of 93 anchor markers⁸ and also with 3,368 microsatellite markers⁹, resulting in a genetic resolution of 0.3 cM (95% c.i.). Among the 64 animals used in our investigation, 5 (LS570, LS580, LS581, LS689, LS724) had not been provided with anchor markers. We used interspersed repetitive sequence (IRS)–PCR technology both to construct a

framework of map anchors for these five animals and to increase the number of map anchors for a large section of the EUCIB B_1 progeny, including the other 59 animals used in our study.

IRS-PCR markers and access to clones. We selected 69 probes from a set of 230 IRS-PCR markers that detected a presence/absence polymorphism between B6 and SPR and that had been mapped on the EUCIB (B6 × SPR) × SPR backcross. Briefly, we derived markers from two IRS-PCR fragment plasmid libraries generated either from genomic B6 DNA or from the cell hybrid EJ167 that contains mouse chromosomes 17 and 3 on a human background⁴¹. We also generated markers by IRS-PCR fragment amplification from random BAC clones from the California Institute of Technology mouse 129/Sv BAC library obtained from Research Genetics⁴¹. Probe names consist of a prefix referring to the libraries (bir, B6; 173r, EJ167; mbacr, mouse BAC) followed by a plate address (plate, row, column). Bir and 173r probes can be obtained from the authors; mouse BAC clones are available through Research Genetics or the Resource Centre of the German Genome Project (see below). The amplification of IRS-PCR fragments from BAC clones has been described⁴¹. IRS markers have been designated D*Irs* (DNA segment, Chr*, interspersed repetitive sequence PCR marker*).

IRS-PCR and genotyping. We amplified genomic mouse DNA (50 ng) in 60-µl reactions containing 1 × PCR buffer (35 mM Tris-Base, 15 mM Tris-HCl, 0.1% Tween 20, 50 mM KCl, 1.5 mM MgCl, 125 μM of each dNTP) and 1 µg of primer B1R. Cycling conditions were 94 °C (3 min), followed by 35 cycles of 94 °C (30 s), 65 °C (30 s) and 72 °C (180 s). IRS fragment amplification from bacterial suspensions and colony material was done as described⁴¹. For filter production, we added bromphenol blue to a concentration of 0.5% to complex IRS-PCR reaction products from mouse backcross progeny and from parental strains. We spotted 1 µl of reaction products onto Hybond N+ nylon membranes (Amersham) using the Hydra-96 microdispenser (Robbins Scientific) at a density of 384 spots on a surface of 7×11 cm. The membranes were incubated in 1.5 mM NaCl and 0.5 mM NaOH for complete denaturation for 1 min, neutralized in 0.5 M sodium phosphate pH 7.2 for 5 min, air-dried and then ultraviolet-crosslinked in a Stratalinker (Stratagene). Genotyping of simple sequence length polymorphism (SSLP) markers D1Mit211, D1Mit234, D8Mit4, D16Mit4, D19Mit12 and DXMit81 was done in standard conditions $^{42,43}\!\!.$ We ran PCR products on 4% gels prepared from 'small DNA' agarose (Biozym).

Probe labeling and filter hybridization. We labeled probes radioactively with 10 μ Ci of a [³²P]dCTP at room temperature for 16 h using random priming⁴⁴. Filters were hybridized in 15-ml Falcon tubes in Church buffer⁴⁵, to which probes were added directly after denaturation. After incubation in a water bath at 65 °C over night, we washed the filters at 65 °C for 30 min once with 2 × SSC/0.1% SDS and once with 0.1×SSC/0.1% SDS. We then exposed the filters to Kodak XAR film at -80 °C for 1–5 d.

Linkage map construction. For the linkage of genes underlying protein 2-DE polymorphisms, we used the program package MAPMAKER/EXP 3.0 (ref. 46). EUCIB backcross map anchor data^{8,9} were downloaded from the EUCIB project website (see below) and used in combination with our IRS–PCR marker and SSLP marker genotyping data for the construction of a robust 163-marker framework map. Pairwise lod score calculation with the Mapmaker 'group' command on the complete data set, comprising all proteins plus chromosome-specific framework markers (88 EUCIB anchors, 69 IRS PCR markers and 6 SSLPs), identified linkage groups at a threshold lod of 4.0.

We placed proteins relative to the map intervals defined by the framework markers in a multipoint linkage analysis. Localization of a protein within a map interval was assumed to be correct when a log-likelihood ratio exceeding 2.5 relative to the next best order was observed⁴⁷. Proteins that failed to integrate into the map unambiguously were either assigned to map intervals that describe the regions for which the threshold of 2.5 was not met or merely listed as 'linked to chromosome'. Once proteins had been placed on the map, an additional 53 EUCIB MSO markers (microsatellite ordered; secondary map anchors) were integrated into the map to provide additional landmark reference points that subdivided intervals specified by map anchors. In the course of the EUCIB project,

these MSO markers had been selected from the mouse genetic map⁴² and had been typed only on a subset of mice. The MSO markers used for our work represent those secondary anchors that had at least genotyping data for 10 mice of the 64 used for protein analysis.

URLs. ProFound version 4.8.2 http://prowl.rockefeller.edu/cgi-bin/Pro-Found; list of Bir and 173r probes, http://www.molgen.mpg.de/~rodent; Resource Centre of the German Genome Project, http://www.rzpd.de; EUCIB project website, http://www.hgmp.mrc.ac.uk/MBx.

Note Supplementary information is available on the Nature Genetics website.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 27 July 2001; accepted 22 February 2002.

- Klose, J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* 26, 231–243 (1975).
- O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250, 4007–4021 (1975).
- Klose, J. & Kobalz, U. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* 16, 1034–1059 (1995).
- Klose, J. Large-gel 2-D electrophoresis. Methods Mol. Biol. 112, 147–172 (1999).
- Anderson, N.G., Metheson, A. & Anderson, N.L. Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* 1, 3–12 (2001).
- Ferris, S.D., Sage, R.D., Prager, E.M., Ritte, U. & Wilson, A.C. Mitochondrial DNA evolution in mice. *Genetics* 107, 681–721 (1983).
- Avner, P., Amar, L., Dandolo, L. & Guenet, J. L. Genetic analysis of the mouse using interspecific crosses. *Trends. Genet.* 4, 18–23 (1988).
- Breen, M. et al. Towards high resolution maps of the mouse and human genomes—a facility for ordering markers to 0.1 cM resolution. Hum. Mol. Genet. 3, 621–627 (1994).
- Rhodes, M. et al. A high-resolution microsatellite map of the mouse genome. Genome Res. 8, 531–542 (1998).
- Klose, J. Fractionated extraction of total tissue proteins from mouse and human for 2-D electrophoresis. *Methods Mol. Biol.* 112, 67–85 (1999).
- Gauss, C., Kalkum, M., Löwe, M., Lehrach, H. & Klose, J. Analysis of the mouse proteome. (I) Brain proteins: separation by two-dimensional electrophoresis and identification by mass spectrometry and genetic variation. *Electrophoresis* 20, 575–600 (1999).
- 12. Klose, J. Genotypes and phenotypes. Electrophoresis 20, 643–652 (1999).
- Damerval, C., Maurice, A., Josse, J. M. & de Vienne, D. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* 137, 289–301 (1994).
 Lehmann, W.D. in *Massenspektrometrie in der Biochemie* 410 (Spektrum,
- Lehmann, W.D. in Massenspektrometrie in der Biochemie 410 (Spektrum, Heidelberg, 1996).
- Strupat, K., Laras, M., Hillenkamp, F., Eckerskorn, C. & Lottspeich, F. Matrixassisted laser-desorption ionization mass spectrometry of proteins electroblotted after polyacrylamide gel electrophoresis. *Anal. Chem.* 66, 464–470 (1994).
- Thiellement, H. et al. Proteomics for genetic and physiological studies in plants. Electrophoresis 20, 2013–2026 (1999).
- Janke, C., Holzer, M., Klose, J. & Arendt, T. Distribution of isoforms of the microtubule-associated protein tau in grey and white matter areas of human brain: a two-dimensional gelelectrophoretic analysis. *FEBS Lett.* **379**, 222–226 (1996).
- Whitney, J.B., Cobb, R.R., Popp, R.A. & O'Rourke, T.W. Detection of neutral amino acid substitutions in proteins. *Proc. Natl Acad. Sci. USA* 82, 7646–7650 (1985).
- Scriver, C.R. & Waters, P.J. Monogenic traits are not simple: lessons from phenylketonuria. *Trends. Genet.* 15, 267–272 (1999).
 Cargill, M. et al. Characterization of single-nucleotide polymorphisms in coding
- Cargill, M. et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nature Genet. 22, 231–238 (1999).
- Halushka, M.K. et al. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nature Genet. 22, 239–247 (1999).
- Sunyaev, S., Ramensky, V. & Bork, P. Towards a structural basis of human nonsynonymous single nucleotide polymorphisms. *Trends Genet.* 16, 198–200 (2000).
 Lindblad-Toh, K. et al. Large-scale discovery and genotyping of single- divergence
- Lindblad-Toh, K. et al. Large-scale discovery and genotyping of single- divergence of B2m alleles of wild Mus musculus and Mus spretus implies positive selection. Immunogenetics 38, 106–116 (1993).
 Lacombe, C. et al. in Protides of the Biological Fluids; Proceedings of the
- Lacombe, C. et al. In Protides of the Biological Fillids, Proceedings of the Colloqium Vol. 32 (ed. Peeters, H.) 1001–1003 (Pergamon, Oxford, 1985).
 Noel, D., Nikaido, K. & Ames, G.F. A single amino acid substitution in a histidine-
- Noel, D., Nakado, K. & Anles, G.F. A single amino actu substitution in a installine transport protein drastically alters its mobility in sodium dodecyl sulfatepolyacrylamide gel electrophoresis. *Biochemistry* 18, 4159–4165 (1979).
- Literský, J.M., Scott, C.W. & Johnson, G.V. Phosphorylation, calpain proteolysis and tubulin binding of recombinant human tau isoforms. *Brain Res.* 604, 32–40 (1993).
- Gerstner, A., Csapo, Z., Sasvari-Szekely, M. & Guttman, A. Ultrathin-layer sodium dodecyl sulfate gel electrophoresis of proteins: effects of gel composition and temperature on the separation of sodium dodecyl sulfate-protein complexes. *Electrophoresis* 21, 834–840 (2000).
- Wiltfang, J. et al. Improved electrophoretic separation and immunoblotting of βamyloid (A-β) peptides 1-40, 1-42, and 1-43. Electrophoresis 18, 527-532 (1997).
- Lai, K., Langley, S.D., Dembure, P.P., Hjelm, L.N. & Elsas, L.J. Duarte allele impairs biostability of galactose-1-phosphate uridyltransferase in human lymphoblasts. *Hum. Mutat.* **11**, 28–38 (1998).

- Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94 (1997).
 Pesole, G. et al. UTRdb and UTRsite: specialized databases of sequences and
- Pesole, G. et al. UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. *Nucleic. Acids Res.* 28, 193–196 (2000).
 Phillips, P.C. From complex traits to complex alleles. *Trends. Genet.* 15, 6–8 (1999).
- Phillips, P.C. From complex traits to complex alleles. *Trends. Genet.* 15, 6–8 (1999).
 Le Stunff, C., Fallin, D., Schork, N.J. & Bougneres, P. The insulin gene VNTR is associated with fasting insulin levels and development of juvenile obesity. *Nature Genet.* 26, 444–446 (2000).
- Weiss, K.M. & Terwilliger, J.D. How many diseases does it take to map a gene with SNPs? *Nature Genet.* 26, 151–157 (2000).
 Silva, A.J. & White, R. Inheritance of allelic blueprints for methylation patterns.
- Silva, A.J. & White, R. Inheritance of allelic blueprints for methylation patterns. Cell 54, 145–152 (1988).
- Gooley, A.A. & Packer, N.H. in Proteome Research: New Frontiers in Functional Genomics (eds Wilkens, M.R., Williams, K.L., Appel, R.D. & Hochstrasser, D.F.) 65–91 (Springer, Berlin, 2001).
- Kato, G.J. Human genetic disease of proteolysis. *Hum. Mutat.* 13, 87–98 (1999).
 Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in
- Neuriorr, V., Aroia, N., Iaube, D. & Ehrnardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9, 255–262 (1988).
- 39. Marcus, K., Immler, D., Sternberger, J. & Meyer, H.E. Identification of platelet proteins separated by two-dimensional gel electrophoresis and analyzed by

matrix assisted laser desorption/ionization-time of flight-mass spectrometry and detection of tyrosine-phosphorylated proteins. *Electrophoresis* **21**, 2622–2636 (2000).

- Zhang, W. & Chait, B.T. ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem.* 72, 2482–2489 (2000).
- Schalkwyk, L.C. et al. Advanced integrated mouse YAC map including BAC framework. Genome Res. 11, 2142–2150 (2001).
- Dietrich, W.F. et al. A comprehensive genetic map of the mouse genome. Nature 380, 149–152 (1996).
- Schalkwyk, L.C. *et al.* Panel of microsattelite markers for whole-genome scans and radiation hybrid mapping and a mouse family tree. *Genome Res.* 9, 878–888 (1999).
- 44. Feinberg, A.P. & Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13 (1983).
- Church, G.M. & Gilbert, W. The genomic sequencing technique. Prog. Clin. Biol. Res. 177, 17–21 (1985).
- Lander, E.S. *et al.* MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1, 174–181 (1987).
- 47. Dietrich, W. et al. A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131, 423–447 (1992).

