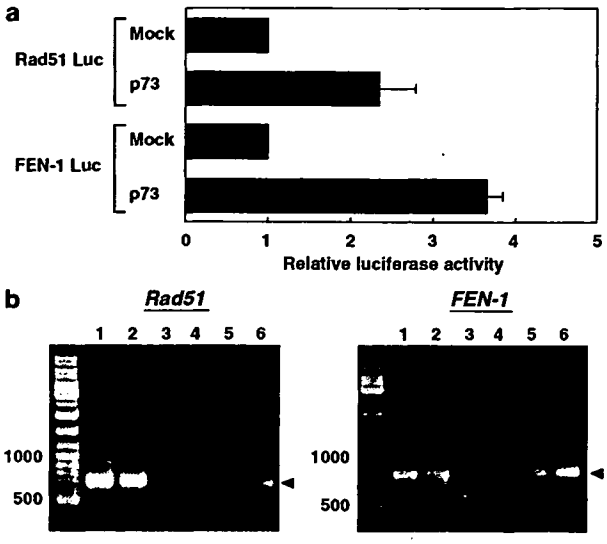


Figure 7 p73 and cisplatin treatment activate the expression of DNA repair genes. (a) Induction of the promoter activity by p73. Rad51 Luc or FEN-1 Luc was transiently co-transfected with p73 expression plasmid and CH110 plasmid (Amersham Biosciences, Piscataway, NJ, USA) expressing β-galactosidase as an internal control. The results are normalized to β-galactosidase activity and are representative of at least three independent experiments. Bars, ±s.d. (b) Enhancement of DNA binding activity of ZNF143 by cisplatin treatment. Soluble chromatin was prepared from Flag-ZNF143 stable transfectant untreated (lanes 1, 3 and 5) or treated (lanes 2, 4 and 6) with cisplatin, and immunoprecipitated with anti-mouse IgG (lanes 3 and 4) or anti-Flag (M2) antibodies (lanes 5 and 6). Extracted DNAs of immunoprecipitation (lanes 3–6) and soluble chromatin (lanes 1 and 2) were amplified using specific primer pairs for the *Rad51* and *FEN-1* promoter regions. Amplification products were subjected by electrophoresis as described in Figure 6d.

number of DNA repair genes contain potential binding site for ZNF143 in their regulatory regions (Supplementary Data). These results suggest that ZNF143 is a positive regulator like a master gene for the expression of DNA repair genes.

In conclusion, our data indicate that ZNF143 is a pivotal transcription factor that regulates the gene expression for DNA repair pathways together with p73 and is involved in cisplatin resistance. Our findings also raise the possibility that inhibition of ZNF143 function might be a target for therapeutic augmentation of cisplatin-based chemotherapy. Further investigation to define the molecular function of ZNF143 will greatly advance our understanding of cisplatin resistance.

## Materials and methods

### *Cell culture*

Human epidermoid cancer KB cells and human prostate cancer PC3 cells were cultured in Eagle's minimal essential medium (Nissui Seiyaku, Tokyo, Japan) containing 10% heat-inactivated fetal bovine serum. The cisplatin-resistant KB/CP4 and P/CDP6 cells were derived from KB and PC3 as described previously (Tanabe *et al.*, 2003). Seven lung cancer cell lines were obtained as described previously (Sugaya *et al.*, 2002). Cell lines were maintained in a 5% $CO_2$ atmosphere at 37°C.

### *Antibodies and drugs*

Anti-Flag (M2) monoclonal antibody and anti-Flag (M2) affinity gel were purchased from Sigma (St Louis, MO, USA). Anti-STAT3 (sc-482), anti-proliferating cell nuclear antigen (PCNA) (sc-56) and HA-probe (F-7) AC (agarose conjugate) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA). Anti-Rad51, anti-FEN-1 antibodies and anti-HA-peroxidase (3F10) were purchased from Calbiochem (Darmstadt, Germany), BD Biosciences (BD Biosciences Clontech, Palo Alto, CA, USA) and Roche molecular Biochemicals (Mannheim, Germany), respectively. The anti-ZNF143 antibody was kindly gifted by Dr GR Kunkel (Texas A&M University, TX, USA) (Ishiguchi *et al.*, 2004). Cisplatin, vincristine and etoposide were purchased from Sigma. Oxaliplatin was kindly provided from Yakult Honsha Co., Ltd., Tokyo, Japan.

### *Plasmid construction*

Plasmid construction of pGEX-p53, pGEX-p73 and pGEX-ZNF143 that express GST-p53, GST-p73 and GST-ZNF143 proteins in bacteria, respectively, and pcDNA3-HA-p53 that expresses HA-p53 protein in mammalian cells were described previously (Imamura *et al.*, 2001; Uramoto *et al.*, 2002, 2003). pcDNA3-HA-p73 expression plasmid in mammalian cells was kindly provided by Dr G Melino (University of Rome, Rome, Italy) (De Laurenzi *et al.*, 1998). For construction of pcDNA3–3 × Flag expression plasmid, the following double-stranded oligonucleotides were inserted to pcDNA3 expression plasmid (Invitrogen, San Diego, CA, USA) between *Bam*HI and *Eco*RI sites. Three times Flag oligonucleotides; 5′-ATGGACTACAAAGACCATGACGGTGATTATAAAGAT CATGACATCGATTACAAGGATGACGATGACAAGAAT TGG-3′. To obtain the pcDNA3–3 × Flag ZNF143, full length of ZNF143 cDNA was ligated at C-terminal of 3 × Flag in pcDNA3–3 × Flag mammalian expression plasmid. For construction of pIRES/hygro-3 × Flag ZNF143, *Bam*HI–*Not*I fragment containing 3 × Flag ZNF143 was obtained by digesting the pcDNA3–3 × Flag ZNF143 plasmid with *Bam*HI and *Not*I, and ligated in same sites of pIRES/hygro mammalian expression plasmid (BD Biosciences Clontech). To obtain Rad51 Luc and FEN-1 Luc, *Rad51* and *FEN-1* promoters were amplified by polymerase chain reaction (PCR) with the following primer pairs with restriction enzyme cleavage sites at the 5′ end; 5′-AGATCTGCGATGGTGAGAACTCGCGGA CC-3′ and 5′-AAGCTTCACCCCGCGGGCGTGGCACG-3′ for the *Rad51* promoter (−471 to +211); 5′-AGATCTGTA-CAGAGGCTGTGGGCGCTCC-3′ and 5′-AAGCTTGGTT CGGGGTTGCCCCGGGC-3′ for the *FEN-1* promoter (−525 to +319). These PCR products were cloned into the pGEM-T easy vector (Promega, Madison, WI, USA). The promoter fragments were gel-purified after *Bgl*II–*Hin*dIII digestion and ligated into the *Bgl*II–*Hin*dIII site of pGL3-basic vector (Promega).

### *Cloning of stable transfectants*

Two micrograms of pIRES/hygro vector or pIRES/hygro-3 × Flag ZNF143 was transfected into $1 \times 10^5$ cells of PC3 using 10 μl Superfect reagent (Qiagen, Hilden, Germany) according to the manufacturer's instructions. After 24 h of transfection, the cells were trypsinized and plated on 100-mm dishes with dilution to form colonies and cultured with maintenance medium containing 250 μg/ml hygromycin. The resulting colonies were isolated using cloning cylinders and transferred to 24-well plates. Cellular expression level of

3 × Flag ZNF143 in each clones was investigated by following Western blotting with anti-Flag antibody. Stable transfectant with expression of 3 × Flag ZNF143 was named PC3/3 × Flag ZNF143 (PC3/3FZ) and was used in this study. PC3/control vector (PC3/CV) was also selected by the transfection with pIRES/hygro vector alone.

*Northern blotting*

Northern blotting analysis was performed as described previously (Murakami *et al.*, 2001; Ishiguchi *et al.*, 2004). Briefly, total RNA was isolated using the Sepasol reagent (Nacalai Tesque, Kyoto, Japan). RNA samples (20 μg/lane) were separated on a 1% formaldehyde agarose gel and transferred to a hybond $N^+$ membrane (GE Healthcare Bio-Science, Piscataway, NJ, USA) with 10 × SSC. ZNF143 cDNA fragments were labeled with random primers using the Megaprime DNA labeling kit (GE Healthcare Bio-Science). After prehybridization and hybridization, signal intensities were quantified using a bio-imaging analyzer (FLA2000, Fuji Photo Film, Tokyo, Japan).

*Western blotting*

Whole-cell lysates and eluted nuclear extracts were prepared as described previously (Uramoto *et al.*, 2002). To prepare whole-nuclear protein, isolated nuclei was directly sonicated for 10 s and designated as nuclear fractions. The indicated amounts of whole-cell lysates, nuclear extracts or nuclear fractions were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). The proteins were transferred to a polyvinylidine difluoride microporous membranes (Millipore, Bedford, MA, USA) using a semi-dry blotter. The blotted membrane was treated with 5% (w/v) skimmed milk in 10 mM Tris, 150 mM NaCl, 0.2% (v/v) Tween 20 and incubated for 2 h at 4°C with a 1:5000 dilution of anti-ZNF143, a 1:10 000 dilution of anti-Flag (M2), a 1:2000 dilution of anti-PCNA, a 1:1000 dilution of anti-STAT3, a 1:500 dilution of anti-FEN-1 and a 1:1000 dilution of anti-Rad51 antibodies. The membrane was then incubated for 40 min at room temperature with a peroxidase-conjugated secondary antibody or a 1:5000 dilution of anti-HA-peroxidase. It was treated with an ECL kit (GE Healthcare Bio-Science) and exposed to Kodak X-OMAT film by autoradiography. The intensity in each signal was assessed numerically by NIH image program (NIH, Bethesda. MD, USA).

*Transient transfections and co-immunoprecipitation assay*

Transient transfection and immunoprecipitation assay were performed as described previously (Uramoto *et al.*, 2002; Izumi *et al.*, 2003). Briefly, 1 × $10^5$ PC3 cells were seeded into six-well tissue-culture plates. The following day, both 1 μg HA and 3 × Flag expression plasmids were transfected using Superfect reagent (Qiagen) according to the manufacturer's instructions. Three hours post-transfection, the cells were washed with phosphate-buffered saline, cultured at 37°C for 48 h in fresh medium and then lysed in buffer X containing 50 mM Tris-HCl (pH 8.0), 1 mM ethylenediaminetetraacetic acid (EDTA), 120 mM NaCl, 0.5% Nonidet P-40, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride and 1 μM $ZnCl_2$. After incubating for 30 min on ice, the lysates were centrifuged at 21 000 *g* for 10 min at 4°C. The supernatant (300 μg) was incubated for 2 h at 4°C with anti-Flag M2 affinity gel or HA-probe (F-7) AC, and the beads were washed three times with buffer X. Immunoprecipitated samples and pre-immunoprecipitated samples (50 μg) were separated by SDS–PAGE, and Western blotting analysis was performed with anti-Flag antibody and anti-HA-peroxidase as described above.

*Knockdown analysis using siRNAs*

Three kinds of double-stranded ZNF143 RNA 25 bp oligonucleotides were generated from Stealth Select RNAi (Invitrogen) No. 1. 5′-UAACCAUAGCAACAGAGUGCGUUCC-3′ and 5′-GGAACGCACUCUGUUGCUAUGGUUA-3′; No. 2. 5′-UAAUUUGUUGCACUGGCAAAUGCCC-3′ and 5′-GGGCAUUUGCCAGUGCAACAAAUUA-3′; and No. 3. 5′-AUAAGCUGUGGUACCAUCUUCCAGC-3′ and 5′-GC UGGAAGAUGGUACCACAGCUUAU-3′. siRNA transfections were performed according to the manufacturer's instructions (Invitrogen). Briefly, 1 μl lipofectamine transfection reagent (Invitrogen) was diluted in 250 μl Opti-minimum essential medium (MEM) I medium (Invitrogen) and incubated for 5 min at room temperature. Next, 50 or 100 pmol ZNF143 siRNA and Stealth RNAi negative control with medium GC (Invitrogen) diluted in 250 μl Opti-MEM I were added gently and incubated for 20 min at room temperature. Oligomer–Lipofectamine complexes and aliquots of 2 × $10^5$ PC3 or P/CDP6 cells in 2 ml culture medium were combined and incubated for 10 min at room temperature. Aliquots of 4 × $10^2$ PC3 cells or 6 × $10^2$ P/CDP6 cells were used for a colony-formation assay as described below. The remaining cells were seeded in 35 mm dishes with 2 ml culture medium and harvested after 72 h culture for Western blotting analysis as described above.

*Cytotoxicity assay*

For the colony-formation assay, 4 × $10^2$ PC3 or 6 × $10^2$ P/CDP6 cells transfected with siRNAs were seeded in 35 mm dishes with 2 ml culture medium. The following day, the cells were treated with indicated concentrations of cisplatin, oxaliplatin, etoposide and vincristine. After 7 days, the number of colonies was counted.

*Purification of GST fusion protein*

Induction and purification of GST fusion proteins were described previously (Ise *et al.*, 1999). Briefly, *Escherichia coli* cells transformed with GST expression plasmids were induced by isopropyl-1-thio-γ-D-galactopyranoside for 1 h and sonicated for 10 s in buffer X as described above. Soluble fractions were obtained by centrifugation at 21 000 *g* for 10 min at 4°C. GST fusion proteins were bound to 10 μl glutathione-sepharose 4B in a 50% slurry in buffer X for 4 h at 4°C, washed three times with buffer X and eluted with 50 mM Tris-HCl (pH 8.0) and 20 mM reduced glutathione according to the manufacturer's protocol (GE Healthcare Bio-Science).

*Electrophoretic mobility shift assay*

The sequences of the oligonucleotides used for the probes in EMSAs were as follows: human *U6 RNA* oligo, 5′-GCC TATTTCCCATGATTCCTTCATATTTGC-3′ and 5′-GGGC AAATATGAAGGAATCATGGGAAATAGG-3′; human *Rad51* oligo, 5′-GGTACATCTCCCGGCATGCATCGCCG GCG-3′ and 5′-GGCGCCGGCGATGCATGCCGGGAGAT GTA-3′; human *FEN-1* oligo, 5′-GGACCCGTGCCCATCC TACAATGCCCTGG-3′ and 5′-GGCCAGGGCATTGTAG GATGGGCACGGGT-3′; GC oligo for modification of cisplatin, 5′-GGCCGGGGCGGGGCGATCGGGGCGGGGGC-3′ and 5′-GGGCCCCCGCCCCGATCGCCCCGCCCCGG. The ZNF143 binding sites in these oligonucleotide probes were underlined (see Figure 6a). Preparation of the $^{32}P$-labeled oligonucleotide probes were described previously (Imamura *et al.*, 2001; Ishiguchi *et al.*, 2004). Briefly, the oligonucleotides were annealed with complementary strands. The double-stranded products were end-labeled with [α-$^{32}P$] deoxycytidimine triphosphate (GE Healthcare Bio-Science) using the

Klenow fragment (Fermentas, Vilnius, Lithuania) and purified from gel. For preparation of cispaltin-modified DNA, labeled GC oligonucleotide probe was treated with 0.3 mM cisplatin at 37°C for 6 h and purified by ethanol precipitation. EMSAs with purified GST fusion proteins were performed as described previously (Imamura *et al.*, 2001; Ishiguchi *et al.*, 2004). Briefly, GST fusion proteins were incubated for 5 min at room temperature in a final volume of 20 μl containing 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM EDTA, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, 10% glycerol, 0.05% Nonidet P-40 and 4 ng $^{32}$P-oligonucleotide probe. The reaction mixtures were resolved by electrophoresis on a 4% polyacrylamide gel (polyacrylamide/bisacrylamide, 80:1) by 10 V/cm for 90–120 min at room temperature with 0.5 × tris-borate-EDTA (TBE) buffer (45 mM Tris base, 45 mM boric acid and 1 mM EDTA). The gel was dried and analysed by a bio-imaging analyzer (FLA2000).

*ChIP assay*

The ChIP assay was performed as described previously (Uramoto *et al.*, 2002). Both PC3/CV and PC3/3FZ stable transfectants were treated with or without 20 μM cisplatin for 12 h. Briefly, protein–DNA crosslinking was performed by incubating PC3/control vector and PC3/3 × Flag-ZNF143 cells with formaldehyde. The cells were lysed in radioimmunoprecipitation assay buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Nonidet P-40 and 1 mM phenylmethylsulfonyl fluoride) and the lysates were sonicated. Soluble chromatin from $1 \times 10^6$ cells was incubated with 5 μg/ml anti-Flag (M2) affinity gel or anti-mouse immunoglobulin G (IgG) with protein A/G agarose (Santa Cruz) by rotation for 2 h at 4°C. Immune complexes were collected by centrifugation. They were then treated with 0.2 M NaCl to reverse protein–DNA crosslinking, and were digested with proteinase K and RNase A. The purified DNA was dissolved with 20 μl $dH_2O$. The DNA (1 μl) was then used for PCR analysis with the following primer pairs for the *Rad51* promoter region (−471 to +211): 5′-AG ATCTGCGATGGTGAGAACTCGCGGACC-3′ forward primer and 5′-AAGCTTCACCCCGCGGGCGTGGCACG-3′ reverse primer; for the *FEN-1* promoter region (−525 to +319): 5′-AGATCTGTACAGAGGCTGTGGGCGCTCC-3′ forward primer and 5′-AAGCTTGGTTCGGGGTTGCCCCGGGC-3′ reverse primer; the *PRDX4* promoter region (−121 to +36): 5′-AGATCTGCCCACGTGGCGGGGCGGGGAGC-3′ forward primer and 5′-CTCGAGCGCAGAAACACGTCCCTT GGCG-3′ reverse primer. The PCR products were separated by electrophoresis on a 2% agarose gel and stained with ethidium bromide.

*Transient transfection and luciferase assay*

Transient transfection and a luciferase assay were performed as described previously (Uramoto *et al.*, 2002). Briefly, $5 \times 10^4$ PC3 cells were seeded into 12-well tissue-culture plates. The following day, 0.2 μg of *Rad51* or *FEN-1* reporter plasmid was transfected with 1.2 μg of p73 expression plasmid using 3 μl/well Superfect reagent (Qiagen) according to the manufacturer's instructions. Three hours post-transfection, the cells were washed and cultured at 37°C for 48 h in fresh medium. Luciferase activity using cell lysated with lysis buffer and brief centrifugation was detected by a Picagene kit (Toyoinki, Tokyo, Japan), and the light intensity was measured with a luminometer (Luminescencer JNII RAB-2300; ATTO, Japan) according to the manusfacturer's instructions.

*ZNF143 binding site analysis*

The search for ZNF143 binding site was performed using the DataBase of Transcriptional Start Sites software version 5.2. 0 (http://dbtss.hgc.jp) developed by Dr Sumio Sugano and Dr Yutaka Suzuki and the Searching Transcription Factor Binding Sites software version 1.3 (http://mbs.cbrc.jp/research/db/TFSEARCH.html) developed by Dr Yutaka Akiyama. Briefly, each 1000 bp upstream region from transcriptional Start Site was obtained by the DataBase of Transcriptional start sites software with GeneID number of Entrez Gene database in National Center for Biotechnology Information. Then, staf binding site (19 bp consensus sequence), which is the same as ZNF143 binding site, was searched by the Searching Transcription Factor Binding Sites software with 70 threshold score. More than 150 genes associated with DNA repair were searched, and 83 ZNF143 binding sites in the 62 genes were found. These results were listed in a Supplementary Data.

**Acknowledgements**

This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (Mext), Kakenhi (13218132 and 18590307) and a Grant-in-Aid for Cancer Research from the Fukuoka Cancer Society, Japan.

**References**

Altaha R, Liang X, Yu JJ, Reed E. (2004). Excision repair cross complementing-group 1: gene expression and platinum resistance. *Int J Mol Med* **14**: 959–970.

Bhattacharyya A, Ear US, Koller BH, Weichselbaum RR, Bishop DK. (2000). The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. *J Biol Chem* **275**: 23899–23903.

Chaney SG, Sancar A. (1996). DNA repair: enzymatic mechanisms and relevance to drug response. *J Natl Cancer Inst* **88**: 1346–1360.

Cohen SM, Lippard SJ. (2001). Cisplatin: from DNA damage to cancer chemotherapy. *Prog Nucleic Acid Res Mol Biol* **67**: 93–130.

De Laurenzi V, Costanzo A, Barcaroli D, Terrinoni A, Falco M, Annicchiarico-Petruzzelli M *et al.* (1998). Two new p73 splice variants, γ and δ, with different transcriptional activity. *J Exp Med* **188**: 1763–1768.

Fujii R, Mutoh M, Niwa K, Yamada K, Aikou T, Nakagawa M *et al.* (1994). Active efflux system for cisplatin in cisplatin-resistant human KB cells. *Jpn J Cancer Res* **85**: 426–433.

Imamura T, Izumi H, Nagatani G, Ise T, Minoru N, Iwamoto Y *et al.* (2001). Interaction with p53 enhances binding of cisplatin-modified DNA by high mobility group 1 protein. *J Biol Chem* **276**: 7534–7540.

Ise T, Nagatani G, Imamura T, Kato K, Takano H, Nomoto M *et al.* (1999). Transcription factor Y-Box binding protein 1 binds preferentially to cisplatin-modified DNA and interacts with proliferating cell nuclear antigen. *Cancer Res* **59**: 342–346.

Ishiguchi H, Izumi H, Torigoe T, Yoshida Y, Kubota H, Tsuji S *et al.* (2004). ZNF143 activates gene expression in

response to DNA damage and binds to cisplatin-modified DNA. *Int J Cancer* **111**: 900–909.

Izumi H, Ohta R, Nagatani G, Ise T, Nakayama Y, Nomoto M *et al.* (2003). p300/CBP-associated factor (P/CAF) interacts with nuclear respiratory factor-1 to regulate the UDP-N-acetyl-alpha-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase-3 gene. *Biochem J* **373**: 713–722.

Keshelava N, Zuo JJ, Chen P, Waidyaratne SN, Luna MC, Gomer CJ *et al.* (2001). Loss of p53 function confers high-level multidrug resistance in neuroblastoma cell lines. *Cancer Res* **61**: 6185–6193.

Kohno K, Izumi H, Uchiumi T, Ashizuka T, Kuwano M. (2003). The pleiotropic functions of the Y-box-binding protein, YB-1. *Bioessays* **25**: 691–698.

Kohno K, Uchiumi T, Niina I, Wakasugi T, Igarashi T, Momii Y *et al.* (2005). Transcription factors and drug resistance. *Eur J Cancer* **41**: 2577–2586.

Kuwano M, Oda Y, Izumi H, Yang SJ, Uchiumi T, Iwamoto Y *et al.* (2004). The role of nuclear Y-box binding protein 1 as a global marker in drug resistance. *Mol Cancer Ther* **3**: 1485–1492.

Lieber MR. (1997). The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays* **19**: 233–240.

Murakami T, Shibuya I, Ise T, Chen AS, Akiyama S, Nakagawa M *et al.* (2001). Elevated expression of vacuolar proton pump genes and cellular pH in cisplatin resistance. *Int J Cancer* **93**: 869–874.

Myslinski E, Krol A, Carbon P. (1998). ZNF76 and ZNF143 are two human homologs of the transcriptional activator staf. *J Biol Chem* **34**: 21998–22006.

Ohga T, Koike K, Ono M, Makino Y, Itagaki Y, Tanimoto M *et al.* (1996). Role of the human Y box-binding protein YB-1 in cellular sensitivity to the DNA-damaging agents cisplatin, mitomycin C, and ultraviolet light. *Cancer Res* **56**: 4224–4228.

Raymond E, Faivre S, Chaney S, Woynarowski J, Cvitkovic E. (2002). Cellular and molecular pharmacology of oxaliplatin. *Mol Cancer Ther* **1**: 227–235.

Rincon JC, Engler SK, Hargrove BW, Kunkel GR. (1998). Molecular cloning of a cDNA encoding human SPH-binding factor, a conserved protein that binds to the enhancer-like region of the U6 small nuclear RNA gene promoter. *Nucleic Acids Res* **26**: 4846–4852.

Schaub M, Myslinski E, Schuster C, Krol A, Carbon P. (1997). Staf, a promiscuous activator for enhanced transcription by RNA polymerase II and III. *EMBO J* **16**: 173–181.

Spiro C, McMurray CT. (2003). Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. *Mol Cell Biol* **23**: 6063–6074.

Sugaya M, Takenoyama M, Osaki T, Yasuda M, Nagashima A, Sugio K *et al.* (2002). Establishment of 15 cancer cell lines from patients with lung cancer and the potential tools for immunotherapy. *Chest* **122**: 282–288.

Tanabe M, Izumi H, Ise T, Higuchi S, Yamori T, Yasumoto K *et al.* (2003). Activating transcription factor 4 increases the cisplatin resistance of human cancer cell lines. *Cancer Res* **63**: 8592–8595.

Tew KD. (1994). Glutathione-associated enzymes in anticancer drug resistance. *Cancer Res* **54**: 4313–4320.

Torigoe T, Izumi H, Ishiguchi H, Yoichiro Y, Mizuho T, Takeshi Y *et al.* (2005). Cisplatin resistance and transcription factors. *Curr Med Chem Anticancer Agents* **5**: 15–27.

Uramoto H, Izumi H, Ise T, Tada M, Uchiumi T, Kuwano M *et al.* (2002). p73 interacts with c-Myc to regulate Y-box-binding protein-1 expression. *J Biol Chem* **277**: 31694–31702.

Uramoto H, Izumi H, Nagatani G, Ohmori H, Nahasue N, Ise T *et al.* (2003). Physical interaction of tumour suppressor p53/p73 with CCAAT-binding transcription factor 2 (CTF2) and differential regulation of human high-mobility group 1 (HMG1) gene expression. *Biochem J* **371**: 301–310.

Vikhanskaya F, Marchini S, Marabese M, Galliera E, Broggini M. (2001). p73α overexpression is associated with resistance to treatment with DNA-damaging agents in a human ovarian cancer cell line. *Cancer Res* **61**: 935–938.

Wood RD, Mitchell M, Lindahl T. (2005). Human DNA repair genes, 2005. *Mut Res* **577**: 275–283.

Zamble DB, Lippard SJ. (1995). Cisplatin and DNA repair in cancer chemotherapy. *Trends Biochem Sci* **20**: 435–439.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc).

perspectives Journal of proteome research

# Personalized Medicine and Proteomics: Lessons from Non-Small Cell Lung Cancer

György Marko-Varga,[†,‡] Atsushi Ogiwara,[†,§,∥] Toshihide Nishimura,[§,∥] Takeshi Kawamura,[§,∥] Kiyonaga Fujii,[§,∥] Takao Kawakami,[§,∥] Yutaka Kyono,[∥] Hsiao-kun Tu,[∥] Hisae Anyoji,[∥] Mitsuhiro Kanazawa,[∥] Shingo Akimoto,[∥] Takashi Hirano,[⊥] Masahiro Tsuboi,[⊥] Kazuto Nishio,[§] Shuji Hada,[#] Haiyi Jiang,[×] Masahiro Fukuoka,[△] Kouichiro Nakata,[◆] Yutaka Nishiwaki,[+] Hideo Kunito,[§] Ian S. Peers,[◇] Chris G. Harbron,[◇] Marie C. South,[▽] Tim Higenbottam,[▽,£] Fredrik Nyberg,[*,∀,☆] Shoji Kudoh,[ƒ] and Harubumi Kato[⊥]

*Respiratory Biological Sciences, AstraZeneca R&D Lund, SE-221 87 Lund, Sweden, Clinical Proteome Center, Tokyo Medical University, Shinjuku Sumitomo Building 17F, 2-6-1 Nishishinjuku, Shinjuku, Tokyo 163-0217, Japan, Medical ProteoScope Company, Limited, Shinjuku Sumitomo Building 17F, 2-6-1 Nishishinjuku, Shinjuku, Tokyo 163-0217, Japan, Department of Surgery, Tokyo Medical University, 6-7-1 Nishishinjuku, Shinjuku, Tokyo 160-0023, Japan, Clinical Division, Research & Development, AstraZeneca K.K., Umeda Sky Building Tower East, 1-88, 1-chome, Ohyodo-naka, Kita-ku, Osaka 531-0076, Japan, Clinical Science Department, Research & Development, AstraZeneca K.K., Umeda Sky Building Tower East, 1-88, 1-chome, Ohyodo-naka, Kita-ku, Osaka 531-0076, Japan, Department of Medical Oncology, Kinki University School of Medicine, 377-2,Ohno-higashi, Osakasayama-city 589-8511, Osaka, Japan, Department of Respiratory Diseases, Toho University School of Medicine, 6-11-1, Omori-nishi, Ota-ku, Tokyo 143-8541, Japan, Dept. of Thoracic Oncology, National Cancer Centre Hospital East, 6-5-1, Kashiwanoha, Kashiwa-city, Chiba 277-8577, Japan, Statistical Sciences, AstraZeneca R&D Alderley Park, Cheshire, UK, Cancer & Infection Statistics, AstraZeneca R&D Alderley Park, Cheshire, UK, Medicine & Science, AstraZeneca R&D Charnwood, Loughborough LE11 5RH, Leicestershire, UK, Sheffield University, Sheffield, UK, Epidemiology, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden, Institute of Environmental Medicine, Karolinska Institute, Box 210, SE-171 77 Stockholm, Sweden, and 4th Department of Internal Medicine, Nippon Medical School, 1-1-5, Sendagi, Bunkyo-ku, Tokyo 113-8603, Japan*

**Received January 26, 2007**

Personalized medicine allows the selection of treatments best suited to an individual patient and disease phenotype. To implement personalized medicine, effective tests predictive of response to treatment or susceptibility to adverse events are needed, and to develop a personalized medicine test, both high quality samples and reliable data are required. We review key features of state-of-the-art proteomic profiling and introduce further analytic developments to build a proteomic toolkit for use in personalized medicine approaches. The combination of novel analytical approaches in proteomic data generation, alignment and comparison permit translation of identified biomarkers into practical assays. We further propose an expanded statistical analysis to understand the sources of variability between individuals in terms of both protein expression and clinical variables and utilize this understanding in a predictive test.

**Keywords:** personalized medicine • gefitinib • therapy • interstitial lung disease • non-small cell lung cancer • biomarkers • predictive test • mass spectrometry • statistical analysis • proteomics

* To whom correspondence should be addressed. Epidemiology, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden; Tel, +46 31 706 5203; Fax, +46 31 776 3828; E-mail, Fredrik.Nyberg@astrazeneca.com.

† György Marko-Varga and Atsushi Ogiwara made equal contributions to this manuscript.

‡ Respiratory Biological Sciences, AstraZeneca R&D Lund.

§ Clinical Proteome Center, Tokyo Medical University.

∥ Medical ProteoScope Co., Ltd.

⊥ Department of Surgery, Tokyo Medical University.

# Clinical Division, Research & Development, AstraZeneca K.K.

× Clinical Science Department, Research & Development, AstraZeneca K.K.

△ Department of Medical Oncology, Kinki University School of Medicine.

◆ Department of Respiratory Diseases, Toho University School of Medicine.

+ Dept. of Thoracic Oncology, National Cancer Centre Hospital East.

◇ Statistical Sciences, AstraZeneca R&D Alderley Park.

▽ Cancer & Infection Statistics, AstraZeneca R&D Alderley Park.

∀ Medicine & Science, AstraZeneca R&D Charnwood.

£ Medical School, Sheffield University.

ν Epidemiology, AstraZeneca R&D Mölndal.

☆ Institute of Environmental Medicine, Karolinska Institute.

ƒ Fourth Department of Internal Medicine, Nippon Medical School.

10.1021/pr070046s CCC: $37.00 



Published on Web 07/17/2007

## Introduction

A personalized medicine approach uses appropriate biomarkers to select treatments best suited for an individual patient and disease phenotype. A multiple biomarker approach (e.g., proteomics) has the advantage over conventional single biomarkers of combining many different pieces of information. Here, we review the key features of state-of-the-art proteomic profiling and introduce recent analytic developments to build a proteomic toolkit for use in personalized medicine, and we describe how these may be applied in a viable method for exploiting predictive proteomic fingerprints in the clinic. The potential of our proteomics toolkit hopefully brings us one step closer to a practical personalized medicine.

Cancer therapy is moving toward individually selected treatments, chosen not only according to tumor cell type but also based on the patient's predicted responsiveness to different classes of therapy or susceptibility to therapeutic adverse events. This emerging personalized medicine approach draws on both genotype and phenotype information, including protein expression. To implement personalized medicine, we need to develop effective biomarker tests predictive of response to treatment or susceptibility to adverse events. The benefits of personalized medicine are exemplified by considering interstitial lung disease (ILD) among non-small cell lung cancer (NSCLC) patients, which is associated with various kinds of chemotherapy treatment. A personalized medicine approach, using a simple blood test to predict those NSCLC patients at risk of developing ILD, would clearly be of great value.

We review current thinking and present some novel developments in a number of areas that have to be integrated to develop and then practically apply such tests in a clinical setting:

- The large scale collection of reliable and high quality phenotypic and clinical data and blood samples.
- Protein analysis in blood.
- Data acquisition, handling, combining and analysis.
- Interpretation and utilization of results in a clinical setting.

## Clinical Background

**A Motivating Example: Gefitinib (IRESSA) Treatment of NSCLC.** The concepts of proteomics-based personalized medicine discussed in this article are very generally applicable. A motivating example that we will refer to in order to illustrate the potential benefits of personalized medicine is ongoing work in attempting to develop a simple blood test to address the potential occurrence of ILD in seriously ill NSCLC patients, the target group for the NSCLC treatment gefitinib.

Gefitinib is a "small molecule" inhibitor of the enzyme tyrosine kinase of the epidermal growth factor receptor (EGFR) family, such as erbB1. It is an approved therapy for advanced NSCLC in many countries and offers important clinical benefits (tumor shrinkage and improvement in disease-related symptoms) for "end-stage" patients. The large phase III ISEL (IRESSA Survival Evaluation in Lung Cancer) trial demonstrated some improvement in survival with gefitinib which failed to reach statistical significance compared with placebo in the overall population and in patients with adenocarcinoma.[1] However, in preplanned subgroup analyses, a significant increase in survival was shown with gefitinib in patients of Asian ethnicity and in patients who had never smoked.[1]

Analysis of the biomarker data from a subset of patients in the ISEL study suggested that patients with pretreated advanced NSCLC who have tumors with a high EGFR gene copy number (detected by fluorescent in situ hybridization [FISH]) have a higher likelihood of increased survival when treated with gefitinib compared with placebo.[2] Increased HER2 gene copy number has also been seen in tumors from patients who are responsive to gefitinib.[3] Somatic-activating mutations of EGFR in tumor tissue have also been associated with increased gefitinib responsiveness in patients with NSCLC.[4-7] Such mutations are more commonly found in tumor samples from patients of Asian origin and non-smokers.[8]

Following the ISEL subgroup analyses, and the biomarker evidence, it has become important to clarify which patients are more suitable for treatment with gefitinib. Analyses for both somatic-activating mutations and gene copy number require tumor tissue, which is not always available from the time of diagnosis; therefore, a blood test may represent a more versatile option and be of great value to clinicians.

With respect to tolerability, the search for a blood test that might include both genetic and proteomic biomarkers to define patients at risk of adverse effects from a drug, for example interstitial lung disease with gefitinib, is a focus of research.

**Interstitial Lung Disease as a Complication in NSCLC Patients.** ILD is a disease that afflicts the parenchyma or alveolar region of the lungs.[9] The alveolar septa (the walls of the alveoli) become thickened with fibrotic tissue. Associated with drug use, it can present precipitously with acute diffuse alveolar damage (DAD). The lungs show so-called "ground glass" shadowing on chest radiology, and patients complain of severe breathlessness. There are no effective treatments but patients can be supported by oxygen supplementation, corticosteroid therapy, or assisted ventilation. The process of alveolar damage is however fatal in some patients. ILD is a comorbidity in patients with NSCLC.[10-16] Both diseases are associated with cigarette smoking,[17-20] and ILD is also considered to be associated with various kinds of lung cancer chemotherapy.[21-26]

In the ISEL study of gefitinib in NSCLC mentioned above, ILD-type events occurred in 1% of both placebo and gefitinib-treated patients.[1] Most ILD-type events occurred in patients of Asian origin, where placebo and treated patients had similar prevalences of respectively 4% and 3%. The rate observed in the gefitinib-treated arm was in line with earlier safety data from Japan and a large gefitinib post-marketing surveillance study in Japan (3322 patients), where the reported rate of ILD-type events was 5.8%.[27]

A simple blood test to predict the potential occurrence of ILD in seriously ill NSCLC patients before initiating treatments would clearly be of great value. This article describes the personalized medicine approach, which could be used to provide such a test. Consequently, the proteomics objectives of the preliminary phase of the study we describe were to verify the protein expression alterations in blood plasma from case patients (who developed ILD) and control patients (without ILD) treated by gefitinib, using a liquid chromatography–mass spectrometry/mass spectrometry (LC–MS/MS) proteomics platform.

## Data and Sample Collection

To develop a personalized medicine test, it is essential to have access to an adequately sized collection of high quality tissue samples on which to perform proteomics analysis, with corresponding reliable diagnostic and clinical data.

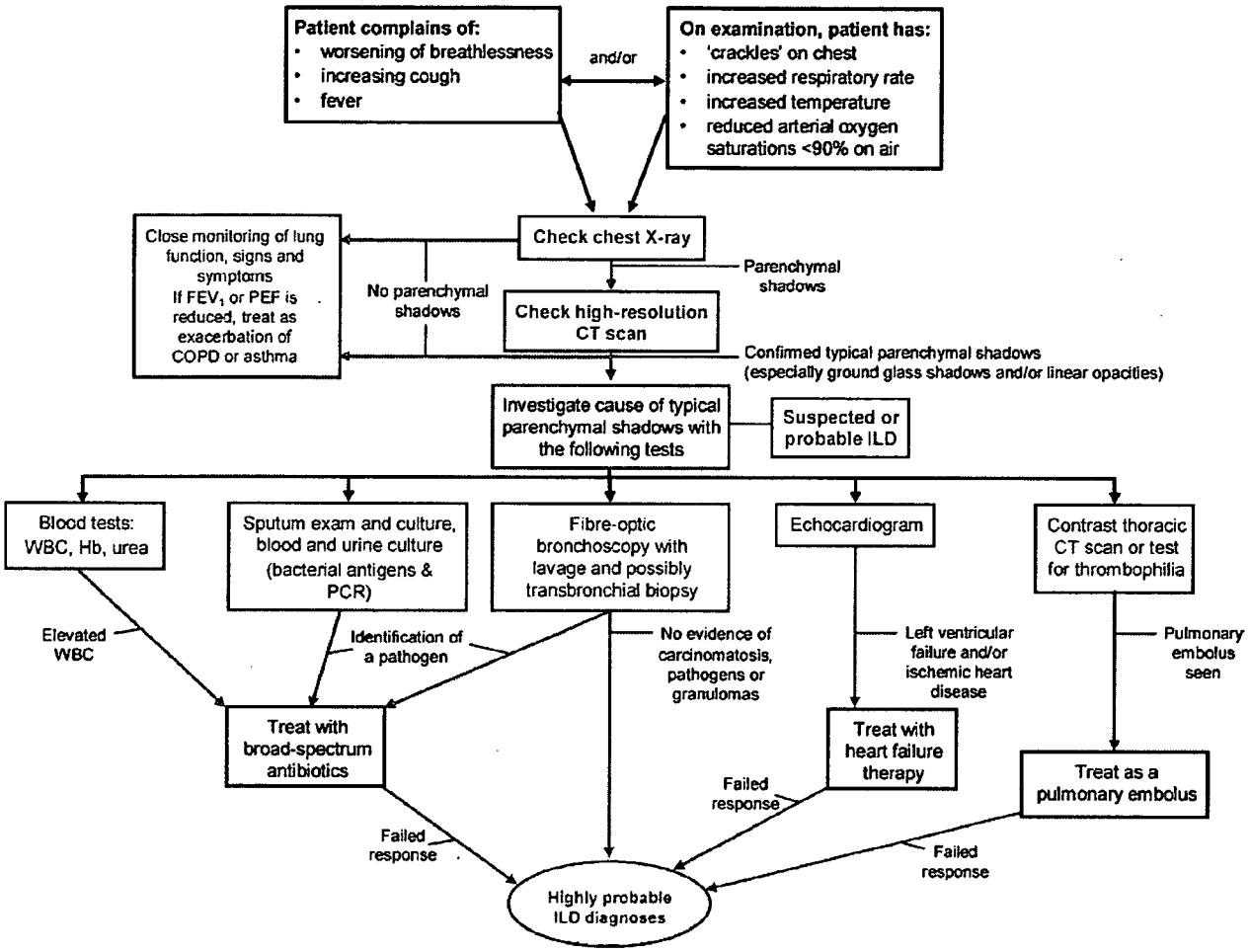


**Figure 1.** Algorithm for diagnosis of interstitial lung disease (ILD) in non-small cell lung cancer (NSCLC) patients.

As an example, in our work with gefitinib, samples were taken after obtaining informed consent from a nested case-control study, i.e., a case-control study performed within a prospective pharmacoepidemiological cohort of several thousand patients with advanced or recurring NSCLC who had received at least one prior chemotherapy regimen, and who were to be treated with gefitinib or chemotherapy. The main objective of this study was to measure the relative risk of ILD in Japanese patients with NSCLC using gefitinib compared with conventional therapy, with the associated aims of determining the incidence rate of ILD in late stage NSCLC patients and the principal risk factors for this complication.

Central to both the case-control study and the proteomics analysis was the use of internationally agreed criteria for the diagnosis of ILD and an algorithm of diagnostic tests to exclude alternative diseases.[28] Principal investigators in the study were asked to assess all patients for possible ILD using the diagnostic algorithm (Figure 1). Two case review boards of experts from oncology, radiology, and pulmonary medicine were set up to independently establish a consistent final diagnosis of ILD. In addition, extensive standard clinical and demographic risk factor data were collected on all registered cases and controls.

This degree of rigor in establishing accurate phenotypic diagnosis is critical to develop a robust and reliable personalized medicine test, as inaccuracies at this stage will affect all subsequent data analyses. The availability of clinical and risk factor data, and a rigorous epidemiological study design setting for the collection of proteomics samples is also of great value to fine-tune the statistical analysis.

## Is Proteomics Ready for Personalized Medicine Applications?

**The Human Proteome Map in Plasma.** The impetus to develop personalized medicine based on blood samples has encouraged proteomic profiling that identifies individual proteins and multiple "fingerprint" protein patterns. A remaining limitation has been the lack of integration of the technology of protein separation with bioinformatics and statistical methods. Extensive national and international[29,30] collaborations are being implemented to address some of these needs. An important component in this development is the Human Proteome Organization (HUPO; www.HUPO.org), a scientific consortium that supports various programmes to map the proteins expressed in various human tissues, disease states, etc.[31–33] One of these is the Plasma Proteome initiative started in 2002, aiming to annotate and catalog the many thousands of proteins and peptides[34–37] of the human plasma proteome. Recently results from the pilot phase with 35 collaborating laboratories from 13 countries[38–42] and multiple analytical

groups were made publicly available on the Internet (www.bioinformatics.med.umich.edu/hupo/ppp; www.ebi.ac.uk/pride). The combined efforts have generated 15 710 different MS/MS datasets that were linked to the International Protein Index (IPI) protein IDs, and an integration algorithm applied to multiple matches of peptide sequences yielded 9504 IPI proteins identified with one or more peptides[40] and characterized by Gene Ontology, InterPro, Novartis Atlas, and OMIM. Such advances provide an important platform for transforming proteomics from a technology to a useful biomarker tool applicable to personalized medicine.

**Protein Analysis in Blood—The Methods.** With respect to automated studies, multidimensional chromatography is the main technology used for protein analysis in blood. It is coupled to mass spectrometry either by electrospray ionization (ESI) for analysis in solution or matrix assisted laser desorption/ionization (MALDI) in solid phase applications.[39,41,43-47] Alternatively, ion-trap mass spectrometers are gaining recognition for high-throughput sequencing.[46,48-53] Linking a Fourier transform ion cyclotrone resonance (FTICR) unit to the linear trap can increase the resolution profoundly,[36,54-56] one of several novel principles for strengthening the assignment of protein annotations with the most commonly used protein search engines.[36,47,54-61] For protein annotation, the recent development of a human protein reference database complements these technologies.[61] Studies of protein expression in a variety of biological compartments ranging from sub-cellular to whole organisms have been undertaken with these analytic approaches.[62-70] Some key findings from the HUPO initiatives that impact on methodology include:

- For studies using blood samples, plasma rather than serum is preferred, with ethylenediaminetetraacetic acid (EDTA) as an anticoagulant.[40]
- The abundant proteins in plasma should be depleted prior to analysis.[40]
- Acceptance of protein annotation, i.e., accepted protein identities[39,40] should use standard criteria. These include having two identified peptide sequences from each protein, both with a statistical significance score high enough to ensure a correct sequence confirmation when compared with the corresponding gene sequence entity.[39]

Despite the advances in methodology, important hurdles to using proteomics in a personalized medicine context remain.

**Protein Expression Analysis in Blood—Some Important Hurdles.** Although protein profiling technology is highly automated and interfaced with database search engines to relate peptide sequences to protein identities and function,[39,40] there are many practical reasons why determining the relative abundance of proteins relevant for prediction purposes is difficult:

- About 90% of proteins are believed to be present only in low copy numbers, i.e., at medium and low abundance levels.[49]
- There can be variation both in the quantity and form of protein expression within normal physiological function.
- Between 300 000 and 3 million human protein species exist as direct gene products or post-translational modifications.[44]
- The relative abundance of the post-translational modifications occurring within the cell is called a Cell-Protein-Index Number (CPIN).[29,30] As an example, if one considers that there are 30 types of phosphorylation variants of a single phosphoprotein, and a hundred possible fold forms of glycosylation of a single glycoprotein, the theoretical CPIN varies considerably depending on the sample complexity.
- The dynamic range of protein expression within cells, between levels of most and least abundant proteins, is in the order of $10^8-10^{10}$.[34-36]
- In a typical clinical proteomics study the total cellular protein material in a sample seldom exceeds 10–20 milligrams. Therefore, the least abundant proteins would be present at starting levels not exceeding picograms.
- Recent studies use technology that can identify several thousand proteins in plasma samples,[29] but this still probably only represents a small fraction of the intermediate and processed protein forms. This is due to the current limitation of mass spectrometry not being able to ionize all amino acid sequences and protein modifications with equal efficiency. In most situations, a limited region of the full length protein is sequence annotated.
- The detection of differences in protein expression between groups of interest (e.g., cases and controls) takes place against a background of high variation between individuals within a group, within individuals over time and possible analytic run-to-run variation. Any method used to address this hurdle (which will involve "alignment" for spectral methods) directly impacts the ability to find good protein biomarkers.

Beyond the hurdles above, the fundamental challenge of protein biomarkers is to link the relative abundance of single markers or a fingerprint to clinically important biological processes based on some direct or indirect cause-effect link[29] related to normal or aberrant biological pathways.[47,49] In the following sections, we present the approach used for the identification of protein biomarkers potentially associated with development of ILD in NSCLC patients within the case-control study used as our motivating example. We build on the foundations described above and introduce further analytic developments and ideas relating to proteomic data generation, assaying and alignment to build a proteomics toolkit that can be applied today for personalized medicine approaches.

## A State of the Art Clinical Biomarker Analysis System

In the previous section, we described several challenges in proteomic analysis. Here we describe a system and analysis approaches that we have successfully implemented to address some of these issues.

**The Components of the Analysis System.** The analysis system (Figure 2) uses liquid chromatography-based high-resolution separation of peptides with an interface to tandem MS/MS, a technology which has been attracting great attention as the "shotgun" method of proteome analysis.[44,68-70] With this technology, after depletion of albumin and immunoglobulin G (IgG), all extracted plasma proteins are digested into their specific peptide components by proteolytic enzyme treatment.

The generated peptides are subjected to capillary reverse-phase submicro- to micro-flow liquid chromatography (capillary RP $\mu$LC), separated by retention times due to their physicochemical properties, and then detected and sequenced by a linear ion-trap tandem mass spectrometer[71] (LTQ, Thermo Fisher Scientific, San Jose, CA) interfaced with a spray needle tip for ESI of peptides.[70] A two-dimensional quadrupole ion trap mass spectrometer[71] is used, operated in a data-dependent acquisition mode with operational $m/z$ range limits set at 450–2000 (Figure 3, graphs A and B). Automatic switching to MS/MS acquisition mode is made in 1-second scanning cycles, controlled by the XCalibur software. The actual differences between annotated peptide fragment peaks shown in Figure 3, graph C, correspond to the amino acid residue mass, i.e.,

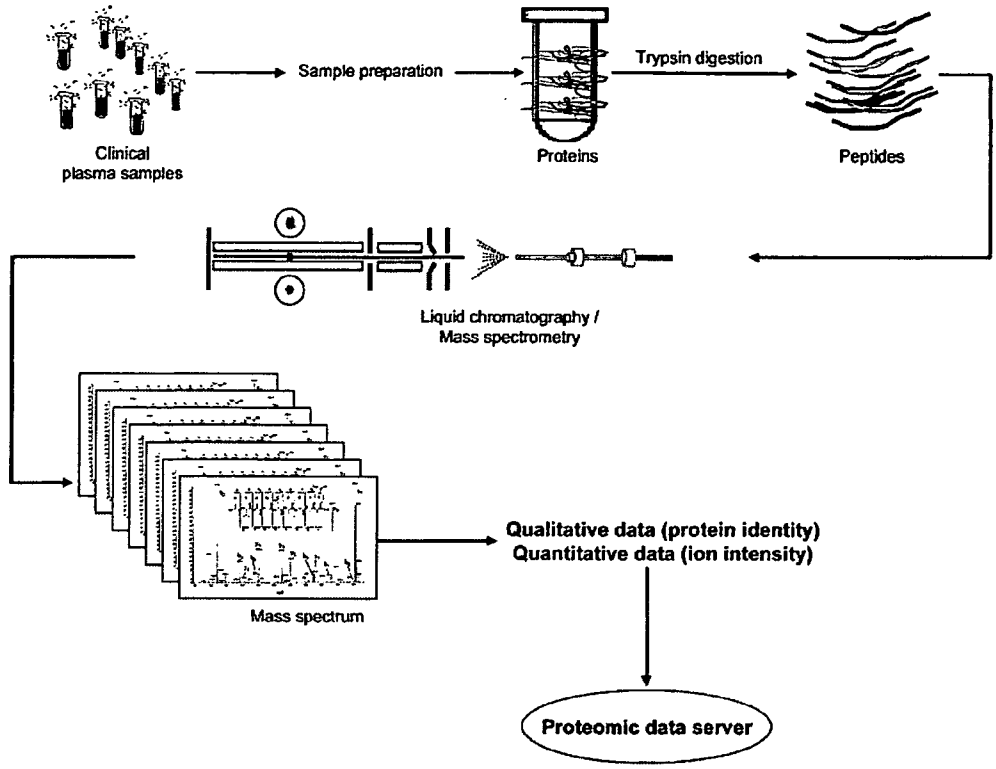


**Figure 2.** Schematic illustration of the clinical proteomics screening process.

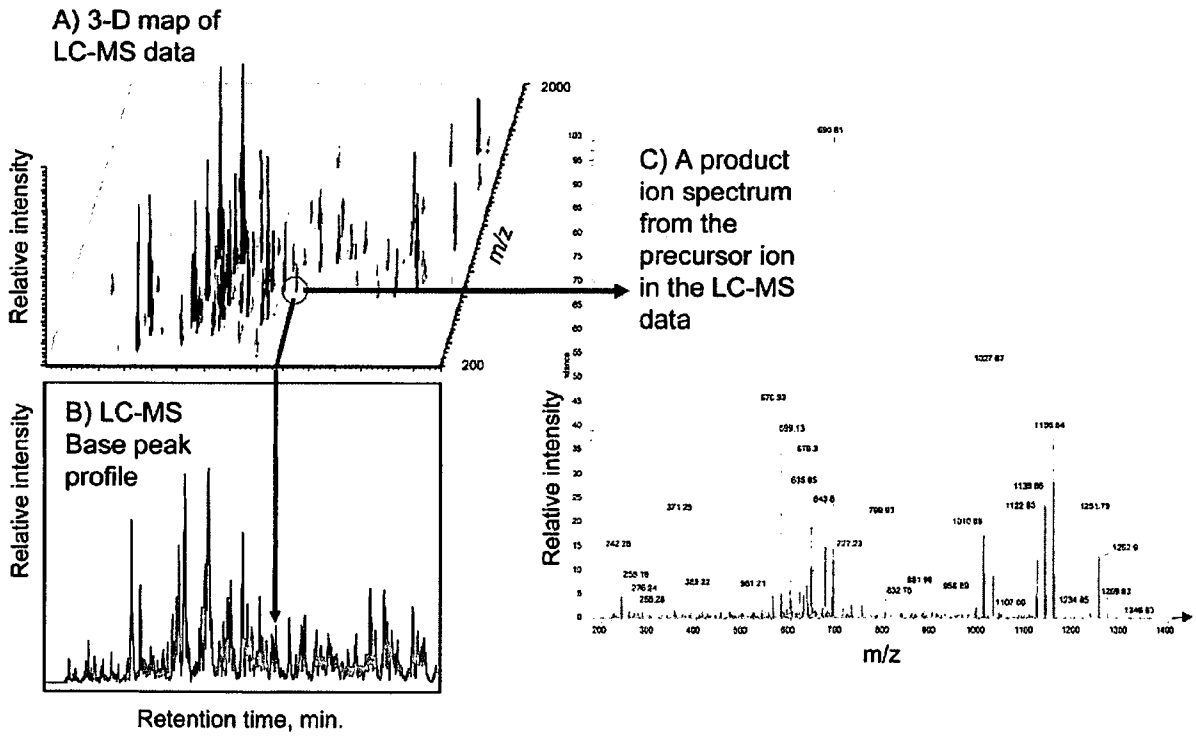


**Figure 3.** Profile of LC–MS data: (a) the three-dimensional view of LC-MS data, (b) the base-peak mass chromatogram, and (c) a product ion spectrum measured for a precursor ion in data-dependent acquisition mode (with MS acquisition operational *m/z* range set at 450–2000).

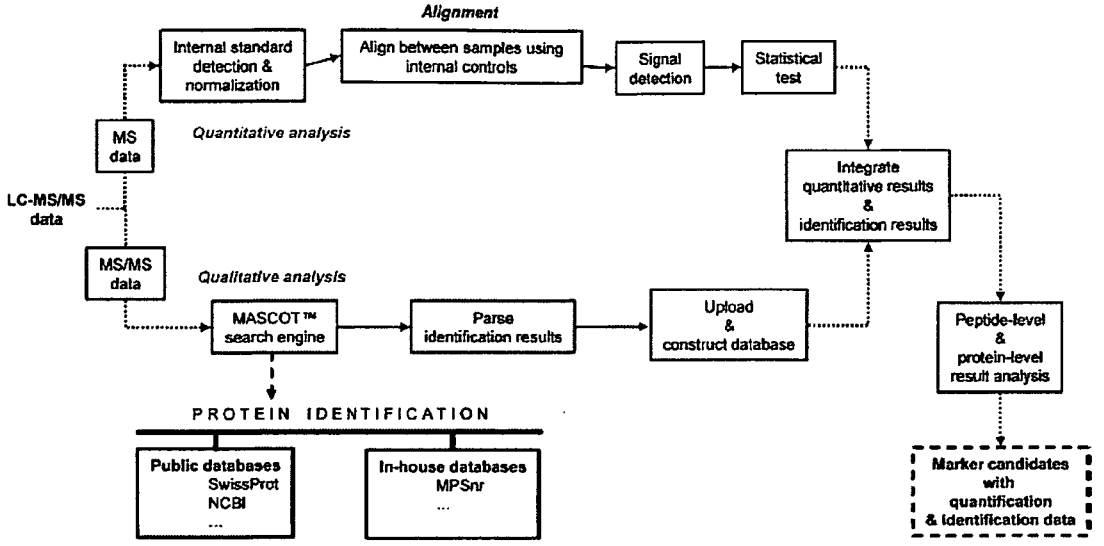


**Figure 4.** Overview of the data acquisition and database mining process developed within the gefitinib biomarker study.

identify the correct amino acid sequence. Internal standards are used for alignment of retention-times.

**How the Methodology Overcomes Some of the Hurdles.** The system described above addresses some of the hurdles noted previously. The digestion of all extracted plasma proteins into peptides will reduce the complexity by combining high-resolution nanoflow chromatographic fractionation with the separation power of modern mass spectrometry, performing automated and unattended shotgun sequencing in plasma.[35] Peptides are also more soluble and easier to handle than intact proteins. In addition, the two-dimensional quadrupole ion trap mass spectrometer[71] operates with a high-volume quadrupole electric field that makes it highly efficient to trap ions. The result is high sensitivity, high scanning speed, and better quantification over a wide dynamic range in comparison with the conventional three-dimensional ion-trap instruments.

Finding signals against a background of high variation is a further challenge, and the next section describes some approaches for addressing these.

## Initial Data Handling, Processing, and Analysis

Proteomic data analysis process can be considered as consisting of two components (Figure 4). *Quantitative analysis* is used to discover significant differences in peptide signal intensities by comparing two (or more) sample groups. This process uses data collected from an entire MS run to quantify the amount of peptide ions by their respective ion signal intensity. *Qualitative analysis* is used to identify the amino acid sequence of each peptide ion, from the respective product ion spectra. To maximize their value, the results from the two component analyses should be considered in combination.

A typical quantitative analysis may consist of several steps:

1. Normalization: To account for differences in the original sample concentrations. Typically, the total signal intensity is scaled to a constant value for each analyzed sample.

2. Alignment: Correcting for nonlinear fluctuation in retention time between different samples. A variety of methodologies are available for aligning LC–MS data sets. We have found the i-OPAL algorithm (Patent # WO 2004/090526 A1), which is based on the single linkage clustering algorithm[72] and which makes use of internal standard signals, to perform well. Other alignment algorithms include xcms.[73]

3. Peak picking or signal detection: Identifying individual peptide ions within the data.

4. Identify discriminating peptides: A number of methods can be used, often in combination. A common approach is to apply a Student's *t*-test and select peptides which are significant, i.e., with a *p*-value less than the chosen cutoff, and which also show a fold-change or intensity ratio greater than another criterion. Further developments of this aspect are discussed in the Principled Statistical Analysis section.

A popular choice for qualitative analysis is the MASCOT MS/MS ion search program.[74] This may be run against a number of different peptide sequence databases, for example the NCBI Nr, Refseq, Gene Ontology, HUGO, and Swiss-Prot sequence databases. The results of the quantitative analysis can then be combined with the qualitative analysis so that, for example, a peptide must be both discriminating and have annotation—i.e., have achieved a high MASCOT score showing confidence in identification—to be considered a candidate biomarker.

The approaches we have discussed above are focused on finding potentially discriminating proteins of clinical utility. In the following section, we describe the next stage in our thinking, namely how we could rapidly deploy in the clinic a viable method for exploiting a predictive proteomic fingerprint.

## A Proposal for Proteomics in the Clinical Setting: Mass Spectrometric Biomarker Assays - MSBA

Although today's technology allows for high-throughput analyses of many proteins rather than a single protein,[30] the details of how such multiplexing assays will be adapted for clinical use have not been well clarified. The Mass Spectrometric Biomarker Assay (MSBA) platform described here was conceived as one example of a rapid and seamless method to progress from identification of a diagnostic more directly to a clinically useful test. MSBA requires only a minute sample amount (5–20 μL) to obtain a read-out from a handful of quantified protein biomarkers (typically 3–35) and automatically analyzes proteins using liquid-phase separation and tandem mass spectrometry with simultaneous quantitation and identification.

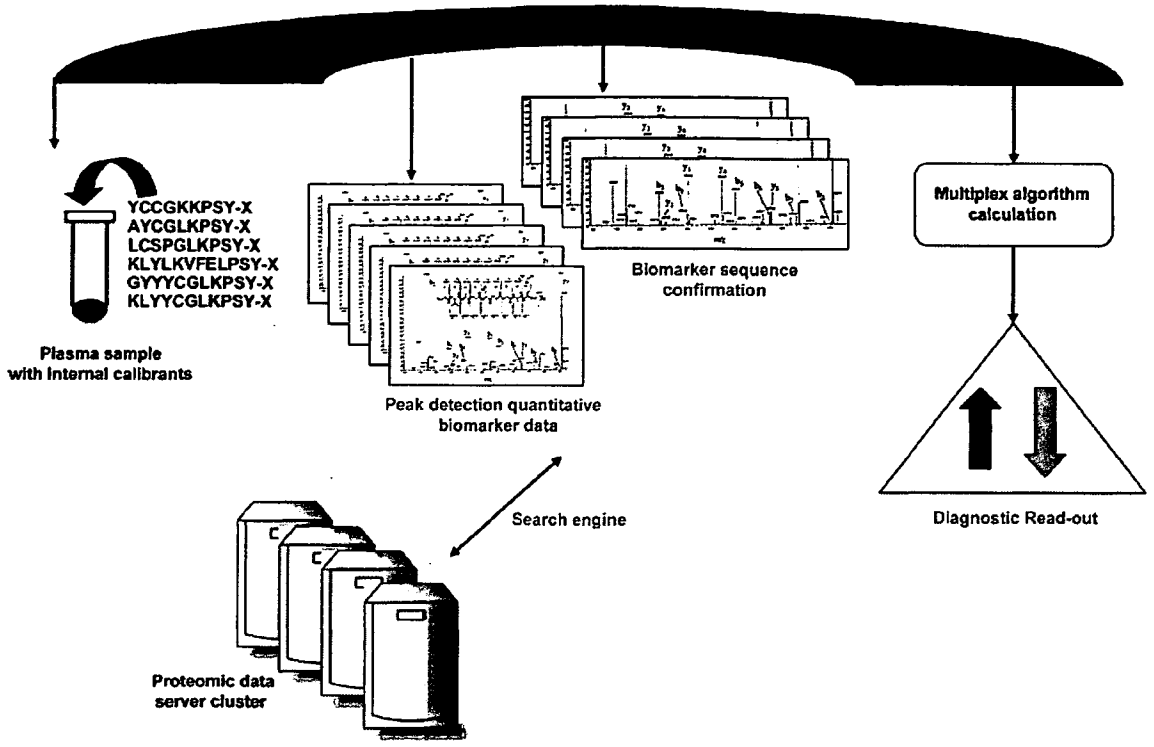


**Figure 5.** Entire flow of the operational components of Mass Spectrometric Biomarker Assays (MSBA).

The MSBA builds on a pre-defined Multiplex Biomarker list, which is stored within the MSBA database. Each marker entity has the values of masses and the relative retention time index with tolerance parameters. In running a patient sample, the predefined biomarker list is scanned to pick up patient sample signals that match with one of the predefined biomarker signals by satisfying the tolerance criteria (in general ±1 for $m/z$ value and ±2% for relative retention time index). The selected candidate signals are further confirmed using the product ion spectrum. That is, the product ion spectrum is represented as a vector by binning (grouping) the $m/z$ ratio values. Using the cosine correlation between the sample vectors and the reference vectors, we can confirm whether the selected candidate signals are truly assigned as target biomarkers. (A standard threshold value of the cosine correlation is 0.8.)

The process steps within the MSBA cycle are outlined in Figure 5. The calculation of the final multiplex biomarker assay read-out from all of the individual markers can be performed by a variety of applications, as discussed in more detail in the Principled Statistical Modeling Approach section. Figures 6A and B illustrate one approach, calculating a distance score which indicates to what extent a measured sample is distant from the case or control template in terms of predefined multiplex biomarkers.

$$S_{\substack{\text{case or}\\ \text{control}}} = \sqrt{\left[\frac{1}{n(n-2)}\right]\left[n\sum_i y_i^2 - \left(\sum_i y_i\right)^2 - \frac{\left[n\sum_i x_i y_i - \left(\sum_i x_i\right)\left(\sum_i y_i\right)\right]^2}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\right]}$$

If the ratio of $S_{case}$ and $S_{control}$ exceeds an MSBA threshold parameter, then the test sample is predicted to be a patient susceptible to develop ILD (ILD case); if not, the test sample is predicted to be a non-susceptible patient (control). We are currently evaluating the MSBA approach in practice.

## A Principled Statistical Modeling Approach

We have described an analytical approach based on proteomic data, with various novel developments. However, additional insight is needed to further improve model discrimination and to broaden the focus from the proteomic data to the ultimate goal of prediction using combinations of data. Statistical analysis can be used to provide further refinement by combining information from the full clinical and laboratory datasets.

An advantage of a multiple biomarker approach (e.g., proteomics) compared with standard single biomarker development is the capability to combine information from many different entities. An example is illustrated in Figure 7A. Considering each biomarker alone fails to separate the two groups of subjects, as there is considerable overlap for both biomarkers. Use of two biomarkers in combination completely separates the two groups.

We can also use clinical variables to advantage in the analysis of the peptide patterns. For example, the efficacy of gefitinib appears to be greater in non-smokers, women, patients of Asian origin, and patients with adenocarcinomas.[8] Figure 7B illustrates how, instead of two protein biomarkers, the combination of clinical data (e.g., age) and a proteomic biomarker is able to separate two groups.

On this basis, we propose using a principled statistical analysis approach to first explore and understand the data and

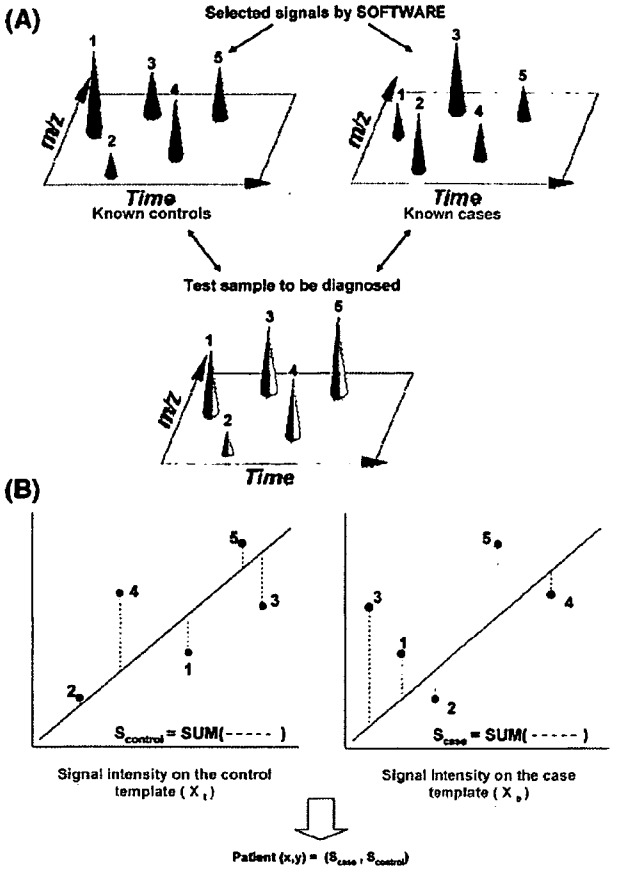


**Figure 6.** (A) Peptide signal comparison that MSBA (Mass Spectrometric Biomarker Assays) performs of the generated ions from the sample. The comparison is made both with the pattern of the controls and with the pattern of the case group for the corresponding signals. (B) Illustration of the regression model application of the MSBA where control templates and case templates are compared to that of the sample template generated in the analysis process.

then to model it and understand the quality of any models produced. A first step is to perform exploratory data analysis (EDA), for example using principal components analysis (PCA), to understand the major sources of data variation and the covariation between clinical parameters and protein intensity measures. The next step is univariate modeling for each protein marker individually, for example using analysis of covariance (ANCOVA), and an assessment of the effect of clinical parameters across the whole set of protein biomarkers using, for example, the False Discovery Rate as a tool.[75] This provides an understanding of key clinical variables and sources of variation within the data.

The next step is to perform multivariate predictive modeling using the proteins and clinical variables identified as being potentially important. There are a number of mathematical methods described in the literature for performing supervised classification, for example Support Vector Machines,[76] Random Forests,[77] PAM,[78] all of which have been successfully applied to high dimensional genomics data.[79] It remains an important unanswered question which modeling approach, or combination of modeling approaches, will generate the most predictive and robust models for data generated using this technology within a prospective study of this design.

Finally, to confirm that we have a practical prediction, the predictive power of a model must be assessed on a different set of patients from that used to generate the model. There are a number of approaches for external validation given a limited size dataset, for example the sequential approach of building a model based upon currently available data and testing on data from new patients when they become available, or withholding an arbitrary selection of subjects from the modeling as a test set and testing the model on these subjects. Internal validation approaches such as cross-validation or related bootstrapping methods may also be useful to assess the model selection *procedure*, but tend to overestimate the performance of a specific predictive model in subsequent external validation.[80,81] The key properties to consider when selecting an assessment method are to ensure that it will provide both precise and unbiased information regarding the prediction error rate of the potential model to be tested for clinical use. As well as assessing an overall predictive rate, it is also useful to separately assess the predictive rate for both the cases and controls and to consider the relative costs of making these false predictions within a clinical setting. Finally, the prevalence of the condition in question (here ILD) is also a critical factor in estimating what proportion of people predicted to be at risk are truly at risk, and this should also be borne in mind when evaluating a model for potential clinical use. The recently published FDA concept paper on drug-diagnostic co-

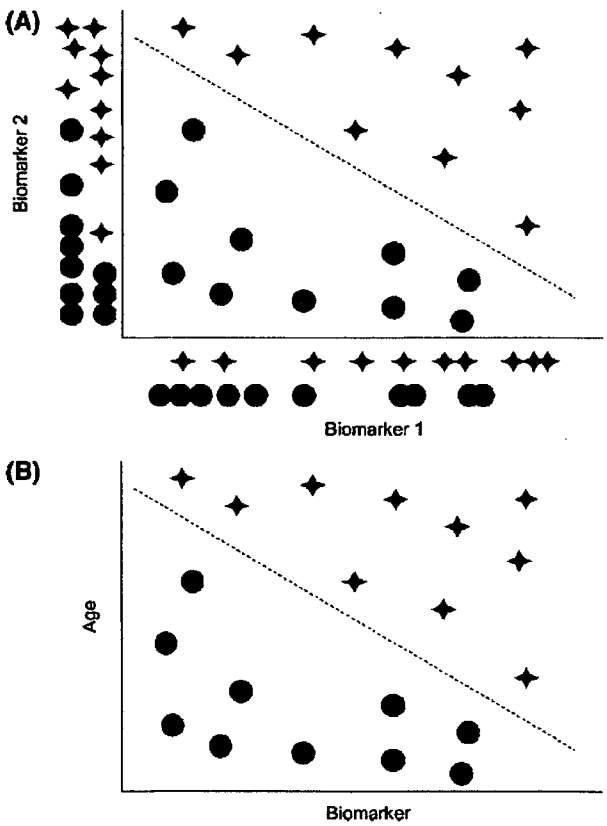


**Figure 7.** (A) Hypothetical example of the combined disease-linkage effect of two protein biomarkers. (Stars signify affected case individuals, circles non-affected control individuals). (B) Hypothetical example of the combined disease-linkage effect of a biomarker and a clinical variable. (Stars signify affected case individuals, circles non-affected control individuals).

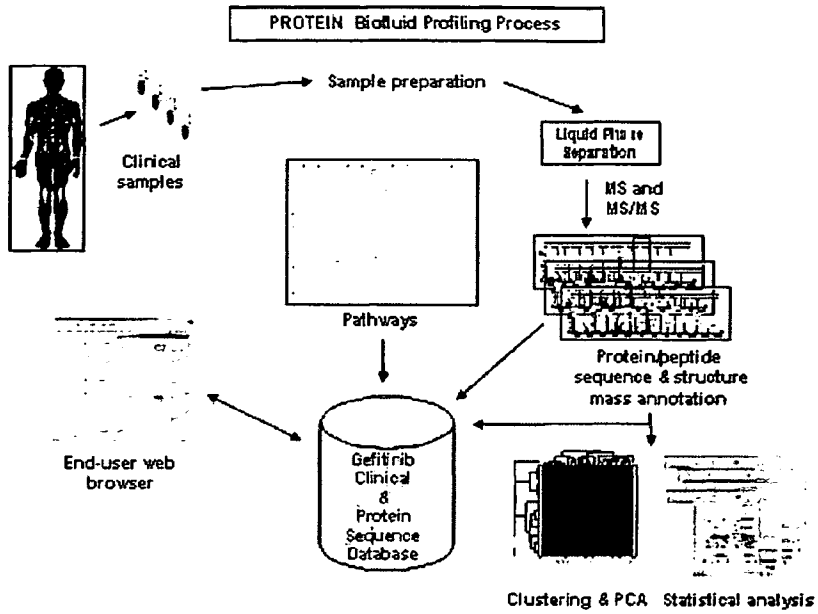


**Figure 8.** Illustration of the bioinformatics and data processing structure within which MSBA (Mass Spectrometric Biomarker Assays) data are captured, modified and analyzed.

development discusses many of the issues around validating predictive biomarkers.[82]

Finally, it is preferable to be able to assign a biological rationale to the biomarkers. Confidence in the reliability of a biomarker is greatly enhanced if we can correctly understand how it relates to the mechanism and progression of the disease of interest. Figure 8 illustrates a bioinformatics and data processing structure that we have developed to allow us to both conduct interactive exploratory and statistical analyses, and also investigate the disease and pathway linkage of discovered biomarker proteins through direct access to reference databases.

## Future Perspectives

Within this paper we have discussed many of the issues that need to considered in developing a personalized medicine approach. A key starting point is that rigorous steps are taken to ensure accurate diagnosis and the careful gathering of both clinical and proteomic data to facilitate the search for peptide patterns.

There are many challenges in performing protein analysis in blood, but mass spectrometry equipment and methods can now be used to generate peptide data with high sensitivity, high scanning speed, and improved quantification. Data handling and processing techniques for steps such as peak alignment and the subsequent methodologies for statistical modeling and analysis are now far enough developed to generate high quality data and robustly analyze these data with confidence.

We have provided details of the MSBA method that can be used to easily translate protein intensities into a practical multiplex assay which can be exploited in the clinic without the need to develop anti-bodies for ELISA. We have also described how an expanded statistical analysis can be used to allow for the individual variance of protein expression to enable us to focus on the proteomic patterns that are actually related to ILD. Finally, we have emphasized the importance of validating the predictive power of a biomarker tool in a way that reflects the real-life setting of intended clinical use.

Hopefully, this combination of developments over a range of different areas brings us one step closer to a practical personalized medicine.

IRESSA is a trademark of the AstraZeneca group of companies.

**Acknowledgments.** We thank all involved in the Iressa study which provided the inspiration for this overview of personalized medicine approaches, including: the external Epidemiology Advisory Board (Kenneth J. Rothman, Jonathan M. Samet, Toshiro Takezaki, Kotaro Ozasa, Masahiko Ando) for their advice and scientific review of study design, conduct, and analysis; Professor Nestor Müller for his expert input into radiological aspects of ILD diagnosis; all Case Review Board members individually (M. Suga, T. Johkoh, M. Takahashi, Y. Ohno, S. Nagai, Y. Taguchi, Y. Inoue, T. Yana, M. Kusumoto, H. Arakawa, A. Yoshimura, M. Nishio, Y. Ohe, K. Yoshimura, H. Takahashi, Y. Sugiyama, M. Ebina) for their valuable work in blindly reviewing ILD diagnoses, as well as pre-study CT scans for pre-existing comorbidities, the Japan Thoracic Radiology Group, Shiga, Japan for their support of CRB work; and all Hospitals, Clinical Investigators, study monitors, nurses, data managers, other support staff, and the participating patients for providing and collecting the data in the study.

## References

(1) Thatcher, N.; Chang, A.; Parikh, P.; Pereira, J. R.; Ciuleanu, T.; von Pawel, J.; et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 2005, *366*, 1527–1537.

(2) Hirsch, F. R.; Varella-Garcia, M.; McCoy, J.; West, H.; Xavier, A. C.; Gumerlock, P.; et al. Increased epidermal growth factor receptor gene copy number detected by fluorescence in situ hybridization associates with increased sensitivity to gefitinib in patients with bronchioloalveolar carcinoma subtypes: a Southwest Oncology Group study. *J. Clin. Oncol.* 2005, *23*, 6838–6845.

(3) Cappuzzo, F.; Varella-Garcia, M.; Shigematsu, H.; Domenichini, I.; Bartolini, S.; Ceresoli, G. L.; et al. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J. Clin. Oncol.* 2005, *23*, 5007−5018.
(4) Araki, J.; Okamoto, I.; Suto, R.; Ichikawa, Y.; Sasaki, J. Efficacy of the tyrosine kinase inhibitor gefitinib in a patient with metastatic small cell lung cancer. *Lung Cancer* 2005, *48*, 141−144.
(5) Kim, K. S.; Jeong, J. Y.; Kim, Y. C.; Na, K. J.; Kim, Y. H.; Ahn, S. J.; et al. Predictors of the response to gefitinib in refractory non-small cell lung cancer. *Clin. Cancer Res.* 2005, *11*, 2244−2251.
(6) Lynch, T. J.; Bell, D. W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R. A.; Brannigan, B. W.; et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 2004, *350*, 2129−2139.
(7) Paez, J. G.; Jänne, P. A.; Lee, J. C.; Tracy, S.; Greulich, H.; Gabriel, S.; et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004, *304*, 1497−1500.
(8) Shigematsu, H.; Lin, L.; Takahashi, T.; Nomura, M.; Suzuki, M.; Wistuba II; et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl. Cancer Inst.* 2005, *97*, 339−346.
(9) American Thoracic Society: American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS Board of Directors, June 2001 and by The ERS Executive Committee, June 2001. *Am. J. Respir. Crit. Care Med.* 2002, *165*, 277−304.
(10) Raghu, G.; Nyberg, F.; Morgan, G. The epidemiology of interstitial lung disease and its association with lung cancer. *Br. J. Cancer* 2004, *91* (Suppl. 2), S3−S10.
(11) Asada, K.; Mukai, J.; Ougushi, F. Characteristics and management of lung cancer in patients with idiopathic pneumonia. *Jap. J. Thor. Dis.* 1992, *51*, 214−219.
(12) Hubbard, R.; Venn, A.; Lewis, S.; Britton, J. Lung cancer and cryptogenic fibrosing alveolitis. A population-based cohort study. *Am. J. Respir. Crit. Care Med.* 2000, *161*, 5−8.
(13) Matsushita, H.; Tanaka, S.; Saiki, Y.; Hara, M.; Nakata, K.; Tanimura, S.; et al. Lung cancer associated with usual interstitial pneumonia. *Pathol. Int.* 1995, *45*, 925−932.
(14) Ogura, T.; Kondo, A.; Sato, A.; Ando, M.; Tamura, M. Incidence and clinical features of lung cancer in patients with idiopathic interstitial pneumonia. *Nihon Kyobu Shikkan Gakkai Zasshi* 1997, *35*, 294−299.
(15) Takeuchi, E.; Yamaguchi, T.; Mori, M.; Tanaka, S.; Nakagawa, M.; Yokota, S.; et al. Characteristics and management of patients with lung cancer and idiopathic interstitial pneumonia. *Nihon Kyobu Shikkan Gakkai Zasshi* 1996, *34*, 653−658.
(16) Turner-Warwick, M.; Lebowitz, M.; Burrows, B.; Johnson, A. Cryptogenic fibrosing alveolitis and lung cancer. *Thorax* 1980, *35*, 496−499.
(17) Baumgartner, K. B.; Samet, J. M.; Stidley, C. A.; Colby, T. V.; Waldron, J. A. Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 1997, *155*, 242−248.
(18) Britton, J.; Hubbard, R. Recent advances in the aetiology of cryptogenic fibrosing alveolitis. *Histopathology* 2000, *37*, 387−392.
(19) Iwai, K.; Mori, T.; Yamada, N.; Yamaguchi, M.; Hosoda, Y. Idiopathic pulmonary fibrosis. Epidemiologic approaches to occupational exposure. *Am. J. Respir. Crit. Care Med.* 1994, *150*, 670−675.
(20) Nagai, S.; Hoshino, Y.; Hayashi, M.; Ito, I. Smoking-related interstitial lung diseases. *Curr. Opin. Pulm. Med.* 2000, *6*, 415−419.
(21) Lilly. Gemcitabine prescribing information. http://pi.lilly.com/gemzar.pdf, 2003.
(22) Kunitoh, H.; Watanabe, K.; Onoshi, T.; Furuse, K.; Niitani, H.; Taguchi, T. Phase II trial of docetaxel in previously untreated advanced non-small-cell lung cancer: a Japanese cooperative study. *J. Clin. Oncol.* 1996, *14*, 1649−1655.
(23) Merad, M.; Le Cesne, A.; Baldeyrou, P.; Mesurolle, B.; Le Chevalier, T. Docetaxel and interstitial pulmonary injury. *Ann. Oncol.* 1997, *8*, 191−194.
(24) Wang, G.-S.; Yan, K.-Y.; Perng, R.-P. Life-threatening hypersensitivity pneumonitis induced by docetaxel (taxotere). *Br. J. Cancer* 2001, *85*, 1247−1250.
(25) Erasmus, J. J.; McAdams, H. P.; Rossi, S. E. Drug-induced lung injury. *Semin. Roentgenol.* 2002, *37*, 72−81.
(26) Aviram, G.; Yu, E.; Tai, P.; Lefcoe, M. S. Computed tomography to assess pulmonary injury associated with concurrent chemo-radiotherapy for inoperable non-small cell lung cancer. *Can. Assoc. Radiol. J.* 2001, *52*, 385−391.
(27) Yoshida, S. The results of gefitinib prospective investigation. *Med. Drug J.* 2005, *41*, 772−789.
(28) Mueller, N. L.; White, D. A.; Jiang, H.; Gemma, A. Diagnosis and management of drug-associated interstitial lung disease. *Br. J. Cancer* 2004, *91*, S24−S30.
(29) Marko-Varga, G.; Fehniger, T. E. Proteomics and disease—the challenges for technology and discovery. *J. Proteome Res.* 2004, *3*, 167−178.
(30) Marko-Varga, G.; Lindberg, H.; Lofdahl, C. G.; Jonsson, P. H. L.; Dahlback, M.; Lindquist, E.; et al. Discovery of biomarker candidates within disease by protein profiling: principles and concepts. *J. Proteome Res.* 2005, *4*, 1200−1212.
(31) Omenn, G. S. The Human Proteome Organization Plasma Proteome Project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses. *Proteomics* 2004, *4*, 1235−1240.
(32) Omenn, G. S. Advancement of biomarker discovery and validation through the HUPO plasma proteome project. *Dis. Markers* 2004, *20*, 131−134.
(33) Orchard, S.; Hermjakob, H.; Binz, P. A.; Hoogland, C.; Taylor, C. F.; Zhu, W.; et al. Further steps towards data standardisation: the Proteomic Standards Initiative HUPO 3(rd) annual congress, Beijing 25−27(th) October, 2004. *Proteomics* 2005, *5*, 337−339.
(34) Anderson, N. G.; Matheson, A.; Anderson, N. L. Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* 2001, *1*, 3−12.
(35) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 2002, *1*, 845−867.
(36) Jacobs, J. M.; Adkins, J. N.; Qian, W. J.; Liu, T.; Shen, Y.; Camp, D. G.; et al. Utilizing human blood plasma for proteomic biomarker discovery. *J. Proteome Res.* 2005, *4*, 1073−1085.
(37) Anderson, N. G.; Anderson, L. The Human Protein Index. *Clin. Chem.* 1982, *28*, 739−748.
(38) Haab, B. B.; Geierstanger, B. H.; Michailidis, G.; Vitzthum, F.; Forrester, S.; Okon, R.; et al. Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data. *Proteomics* 2005, *5*, 3278−3291.
(39) Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; et al. PRIDE: the proteomics identifications database. *Proteomics* 2005, *5*, 3537−3545.
(40) Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 2005, *5*, 3226−3245.
(41) Patterson, S. D. Data analysis−the Achilles heel of proteomics. *Nat. Biotechnol.* 2003, *21*, 221−222.
(42) Rahbar, A. M.; Fenselau, C. Integration of Jacobson's pellicle method into proteomic strategies for plasma membrane proteins. *J. Proteome Res.* 2004, *3*, 1267−1277.
(43) Ho, Y.; Gruhler, A., Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 2002, *415*, 180−183.
(44) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198−207.
(45) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* 2004, *3*, 311−326.
(46) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* 2004, *101*, 13417−13422.
(47) Sadygov, R. G.; Liu, H.; Yates, J. R. Statistical models for protein validation using tandem mass spectral data and protein amino acid sequence databases. *Anal. Chem.* 2004, *76*, 1664−1671.
(48) Fujii, K.; Nakano, T.; Kanazawa, M.; Akimoto, S.; Hirano, T.; Kato, H.; et al. Clinical-scale high-throughput human plasma proteome analysis: lung adenocarcinoma. *Proteomics* 2005, *5*, 1150−1159.

(49) Campbell, J. M.; Collings, B. A.; Douglas, D. J. A new linear ion trap time-of-flight system with tandem mass spectrometry capabilities. *Rapid Commun. Mass Spectrom.* 1998, *12*, 1463–1474.
(50) Cha, B. C.; Blades, M.; Douglas, D. J. An interface with a linear quadropole ion guide for an electrospray-ion trap mass spectrometer system. *Anal. Chem.* 2000, *72*, 5647–5654.
(51) Hager, J. W. Product ion spectral simplification using time-delayed fragment ion capture with tandem linear ion traps. *Rapid Commun. Mass Spectrom.* 2003, *17*, 1389–1398.
(52) Syka, J. E.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* 2004, *3*, 621–626.
(53) Shen, Y.; Zhao, R.; Belov, M. E.; Conrads, T. P.; Anderson, G. A.; Tang, K.; et al. Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.* 2001, *73*, 1766–1775.
(54) Wu, S. L.; Kim, J.; Hancock, W. S.; Karger, B. Extended Range Proteomic Analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J. Proteome Res.* 2005, *4*, 1155–1170.
(55) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* 2005, *4*, 2010–2021.
(56) Yates, J. R.; Cociorva, D.; Liao, L.; Zabrouskov, V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* 2006, *78*, 493–500.
(57) Anderson, D. C.; Li, W.; Payan, D. G.; Noble, W. S. A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J. Proteome Res.* 2003, *2*, 137–146.
(58) Carr, S.; Aebersold, R.; Baldwin, M.; Burlingame, A.; Clauser, K.; Nesvizhskil, A. The need for guidelines in publication of peptide and protein identification data: working group on publication guidelines for peptide and protein identification data. *Mol. Cell. Proteomics* 2004, *3*, 531–533.
(59) Fenyo, D.; Beavis, R. C. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* 2003, *75*, 768–774.
(60) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.
(61) Peri, S.; Navarro, J. D.; Kristiansen, T. Z.; Amanchy, R.; Surendranath, V.; Muthusamy, B.; et al. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.* 2004, *32*, D497–D501.
(62) Kratchmarova, I.; Blagoev, B.; Haack-Sorensen, M.; Kassem, M.; Mann, M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* 2005, *308*, 1472–1477.
(63) Dreger, M.; Bengtsson, L.; Schöneberg, T.; Otto, H.; Hucho, F. Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. *Proc. Natl. Acad. Sci. U.S.A.* 2001, *98*, 11943–11948.
(64) Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; et al. A protein interaction map of Drosophilia melanogaster. *Science* 2003, *302*, 1727–1736.
(65) Johnson, J. R.; Florens, L.; Carucci, D. J.; Yates, J. R., III. Proteomics in malaria. *J. Proteome Res.* 2004, *3*, 296–306.
(66) Hirsch, J.; Hansen, K. C.; Burlingame, A. L.; Matthay, M. A. Proteomics: current techniques and potential applications to lung disease. *Am. J. Physiol. Lung Cell Mol. Physiol.* 2004, *287*, L1–L23.
(67) Malmström, J.; Larsen, K.; Hansson, L.; Löfdahl, C.-G.; Norregard-Jensen, O.; Marko-Varga, G.; et al. Proteoglycan and proteome profiling of central human pulmonary fibrotic tissue utilizing minaturized sample preparation: A feasibility study. *Proteomics* 2002, *2*, 394–404.
(68) Malmström, J.; Larsen, K.; Malmström, L.; Tufvesson, E.; Parker, K.; Marchese, J.; et al. Proteome annotations and identifications of the human pulmonary fibroblast. *J. Proteome Res.* 2004, *3*, 525–537.
(69) Oh, P.; Li, Y.; Yu, J.; Durr, E.; Krasinska, K. M.; Carver, L. A.; et al. Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. *Nature* 2004, *429*, 629–635.
(70) Fujii, K.; Nakano, T.; Kawamura, T.; Usui, F.; Bando, Y.; Wang, R.; et al. Multidimensional protein profiling technology and its application to human plasma proteome. *J. Proteome Res.* 2004, *3*, 712–718.
(71) Schwartz, J. C.; Senko, M. W.; Syka, J. E. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 2002, *13*, 659–669.
(72) Sneath, P. H. A.; Sokai, R. R. *Numerical Taxonomy, The principles and practice of numerical classification*; W. H. Freeman and Co.: San Francisco, 1973.
(73) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak; G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 2006, *78*, 779–787.
(74) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence-databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551–3567.
(75) Storey, J. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B* 2002, *64*, 479.
(76) Vapnik, V. *Statistical Learning Theory*; Wiley: Chichester, UK, 1998.
(77) Breiman, L. Random forests. *Mach. Learn.* 2001, *45*, 5–32.
(78) Tibshirani, R.; Hastie, T.; Narasimhan, B.; Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 2002, *99*, 6567–6572.
(79) Lee, J. W.; Lee, J. B.; Park, M.; Song, S. H. An extensive comparison of recent classification tools applied to microarray data. *Comp. Stat. Data Anal.* 2005, *48*, 869–885.
(80) Steyerberg, E. W.; Harrell, F. E., Jr.; Borsboom, G. J.; Eijkemans, M. J.; Vergouwe, Y.; Habbema, J. D. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol.* 2001, *54*, 774–781.
(81) Bleeker, S. E.; Moll, H. A.; Steyerberg, E. W.; Donders, A. R.; Derksen-Lubsen, G.; Grobbee, D. E.; et al. External validation is necessary in prediction research: a clinical example. *J. Clin. Epidemiol.* 2003, *56*, 826–832.
(82) Food and Drug Administration (FDA): Drug-diagnostic co-development concept paper. Draft—not for implementation. http://www.fda.gov/cder/genomics/pharmacoconceptfn.pdf, 2005.

PR070046S

Lung Cancer (2007) 57, 302–310

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/lungcan

LUNG CANCER

# Abrogation of the interaction between osteopontin and αvβ3 integrin reduces tumor growth of human lung cancer cells in mice

**Ri Cui[a,b,*], Fumiyuki Takahashi[a,b], Rina Ohashi[a,b], Tao Gu[a,b], Masakata Yoshioka[a,b], Kazuto Nishio[c], Yuichiro Ohe[d], Shigeru Tominaga[a], Yumiko Takagi[a,b], Shinichi Sasaki[a], Yoshinosuke Fukuchi[a,b], Kazuhisa Takahashi[a,b]**

[a] *Department of Respiratory Medicine and Research Institute for Diseases of Old Ages, Juntendo University, School of Medicine, 2-1-1 Hongo, Bunkyo-Ku, Tokyo 113-8421, Japan*
[b] *Research Institute for Diseases of Old Ages, Juntendo University, School of Medicine, 2-1-1 Hongo, Bunkyo-Ku, Tokyo 113-8421, Japan*
[c] *Pharmacology Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-Ku, Tokyo 104-0045, Japan*
[d] *Department of Internal Medicine, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-Ku, Tokyo 104-0045, Japan*

Received 4 January 2007; received in revised form 15 March 2007; accepted 18 March 2007

**KEYWORDS**
Osteopontin;
Angiogenesis;
αvβ3 integrin;
Tumor growth;
Lung cancer

**Summary** Osteopontin (OPN) is a multifunctional cytokine involved in cell signaling by interacting with αvβ3 integrins. Recent clinical studies have indicated that OPN expression is associated with tumor progression and poor prognosis among patients with lung cancer. However, the biological role of OPN in human lung cancer has not yet been well-defined. The purpose of this study is to investigate and provide evidence for the causal role of OPN regarding tumor growth and angiogenesis in human lung cancer. In this study, we developed a stable OPN transfectant from human lung cancer cell line SBC-3 which does not express the intrinsic OPN mRNA. To reveal the in vivo effect of OPN on tumor growth of human lung cancer, we subcutaneously injected OPN-overexpressing SBC-3 cells (SBC-3/OPN) and control cells (SBC-3/NEO) into the nude mice. Transfection with the OPN gene significantly increased in vivo tumor growth and neovascularization of SBC-3 cells in mice. These in vivo effects of OPN were markedly suppressed with administration of anti-αvβ3 integrin monoclonal antibody or anti-angiogenic agent, TNP-470. Furthermore, recombinant OPN protein enhanced human umbilical vein endothelial cell (HUVEC) proliferation in vitro, and this enhancement was significantly inhibited with the

* Corresponding author at: Department of Respiratory Medicine, Juntendo University, School of Medicine, 2-1-1 Hongo, Bunkyo-Ku, Tokyo 113-8421, Japan. Tel.: +81 3 5802 1063; fax: +81 3 5802 1617.
*E-mail address:* cri@med.juntendo.ac.jp (R. Cui).


doi:10.1016/j.lungcan.2007.03.019

addition of anti-αvβ3 integrin antibody. Taken together, these results suggest that OPN plays a crucial role for tumor growth and angiogenesis of human lung cancer cells in vivo by interacting with αvβ3 integrin. Targeting the interaction between OPN and αvβ3 integrin could be effective for future development of anti-angiogenic therapeutic agents for patients with lung cancer.


## 1. Introduction

Lung cancer is one of the most frequently diagnosed solid tumors in the world, and is the leading cause of cancer-related deaths in Japan [1]. Despite advancement and improvements in surgical and medical treatments, the prognosis of lung cancer patients remains extremely poor [2]. These facts indicate how important it is to identify novel target molecules for the development of new anticancer therapies for human lung cancer.

Tumor growth and metastasis depend on blood supply and vessel formation. Therefore, anti-angiogenic therapy appears to be an attractive and rational approach for the treatment of solid tumors including lung cancer [3,4]. One approach to anti-angiogenic therapy is to inhibit the adhesive interactions required for tumor angiogenesis. The migration and proliferation of vascular endothelial cells is dependent on their adhesiveness to extracellular matrix (ECM) proteins through a variety of cell adhesion receptor including αvβ3 integrin [5,6]. Thus, the interaction between ECM and αvβ3 integrin may be a therapeutic target for lung cancer patients.

Osteopontin (OPN) is a multifunctional phosphoprotein that binds to αv integrin at the arginine-glycine-asparatic acid (RGD) motif of the central portion and exerts cell-adhesion and migration activity [7,8]. OPN is one of the ECM proteins produced by cancer cells, and is revealed to be overexpressed in various human tumors including the lung, breast, colon, ovary, and gastric cancers [9–14]. Previous studies suggested that OPN may be involved in the angiogenesis of cancer cells. For example, Senger et al. reported that OPN promotes vascular endothelial cell migration via αv integrin in cooperation with vascular endothelial growth factor (VEGF), suggesting that OPN may be involved in angiogenesis [15]. Shijubo et al. demonstrated that coexpression of OPN and VEGF is closely associated with angiogenesis and poor prognosis in stage I lung adenocarcinoma [16]. Thus, OPN is postulated to be related with tumor progression and angiogenesis in various cancers.

Recently, much interest has been focused on OPN expression in human lung cancer. Donati et al. investigated on the correlation between OPN expression in tumor tissues and survival of 136 patients with stage I non-small cell lung cancer (NSCLC), and indicated that OPN expression is a significant unfavorable prognostic factor for survival among patients with stage I NSCLC [17]. Hu et al. also reported that OPN expression was associated with tumor growth, tumor staging, and lymph node invasion of patients with NSCLC. They further analyzed OPN levels in plasma, and suggest that plasma OPN levels may serve as a biomarker for diagnosing or monitoring patients with NSCLC [18]. These findings from these clinical studies imply that OPN may be a therapeutic target and useful biomarker for human lung cancer. However, the biological and functional role of OPN in lung cancer animal model and therapeutic trials targeting OPN and its receptor, αvβ3 integrin, have not yet been reported.

In this study, we first developed stable transfectants from human small cell lung cancer (SCLC) cell line SBC-3 that constitutively secrete mouse OPN. We demonstrated that OPN transfected SBC-3 cells significantly increased in vivo tumorigenecity and neovascularization in comparison with the control cells in mice. In addition, we evaluated the therapeutic efficacy of anti-mouse αvβ3 integrin antibody (RMV-7) against OPN-overexpressing SBC-3 cells inoculated mice. The biological significance of OPN in tumor growth and angiogenesis of lung cancer and potential treatment using RMV-7 antibody are also discussed.

## 2. Materials and methods

### 2.1. Cell lines and reagents

Human small cell lung cancer cell line, SBC-3 cells, was kindly provided by Dr. I. Kimura (Okayama University, Okayama), and cultured in RPMI1640 (Koujin Bio, Saitama, Japan) medium containing 10% (v/v) fetal calf serum. HUVEC were purchased from Clonetics (San Diego, CA) and maintained with EGM-2 medium (Clonetics) on collagen-coated plastic flasks. The anti-mouse αvβ3 antibody (RMV-7) was kindly provided by Prof. Okumura (Department of Immunology, Juntendo University), and has been proven to interfere with OPN-mediated cell migration, adhesion, and proliferation [19,20]. Anti-human αvβ3 monoclonal antibody (LM609) was purchased from Chemicon International (Australia). The monoclonal antibody against murine CD31 was purchased from Pharmingen (San Diego, CA). The monoclonal antibody against murine OPN was purchased from Immuno-Biological Laboratories (Gunma, Japan). The polyclonal rabbit anti-single stranded DNA (ssDNA) was purchased from Dakocytomation (Tokyo, Japan). TNP-470 (6-*O*-(*N*-chloroacethyl-carbamoyl)-fumagillol), a semisynthetic analog of fumagillin derived from *Aspergillus fumigatus*, was kindly provided by Takeda Chemical Industries (Osaka, Japan).

### 2.2. Transfection

$5 \times 10^5$ SBC-3 cells were transfected with Lipofectamine Reagent (Invitrogen) using 8 μg of purified murine OPN cDNA cloned into the eukaryotic cDNA expression vector BMGneo as previously described [21]. This plasmid was designated as BMGneo-mOPN. Two days later, the cells were placed in G418 sulfate (Geneticin; Invitrogen) at 1 mg/ml for selection. Four weeks after transfection, G418-resistant colonies were expanded and isolated with limiting dilution. The

resulting selected and isolated SBC-3 cells transfected with BMGneo-mOPN and BMGneo were designated as SBC-3/OPN and SBC-3/NEO, respectively.

### 2.3. Detection of OPN and VEGF transcription by RT-PCR

Expression of OPN and VEGF mRNA were assessed by RT-PCR. Total RNAs were extracted from cultured cell lines with TRIzol reagent (Invitrogen). The primers for the RT-PCR were: OPN sense primer (5′-AGTCGACATGAGATTGGCAGTGATTTGC-3′), OPN anti-sense primer (5′-ACTCGAGGCCTCTTCTTTAGTTGACCTC-3′), VEGF sense primer (5′-TGCACCCATGGCAGAAGGAGG-3′), and VEGF anti-sense primer (5′-TCACCGCCTCGGCTTGTCACA-3′). RT-PCR was conducted using a Gene Amp RNA PCR kit (Applied Biosystems, Branchburg, NJ) according to the manufacture's instructions.

### 2.4. Determination of OPN protein secretion by ELISA

$5 \times 10^5$ SBC-3/OPN transfectants were cultured in 6-well plates with 2% FCS in RPMI 1640 medium overnight, followed by incubation in 3 ml serum free medium for an additional 24 h. Secreted murine OPN protein level in culture supernatant was measured with the commercial ELISA kit (Immuno-Biological Laboratories, Gunma, Japan) according to the manufacturer's instruction.

### 2.5. Western blot analysis

Conditioned medium from SBC-3/OPN and SBC-3/NEO cells were subjected to western blot analysis. Samples were separated on 10% acrylamide gels and transferred to a nitrocellulose filter with electroblotting at 4 °C. The filters were blocked in phosphate-buffered saline (PBS) containing 10% dry milk, washed in PBS containing 1% dry milk and 0.5% Tween-20, and then incubated with polyclonal rabbit anti-mouse OPN antibody (Immuno-Biological Laboratories, Gunma, Japan) at room temperature for 1 h. Filters were again washed and then incubated with horseradish-peroxidase-conjugated anti-rabbit antibody (Amersham Pharmacia Biotech) for 1 h. Filters were then washed with TBST (150 mM NaCl, 10 mM Tris, pH 8.0, 0.05% Tween-20), and specific proteins were detected using the enhanced chemiluminescence system (Amersham Pharmacia Biotech).

### 2.6. In vitro cell growth rates

SBC-3/NEO and SBC-3/OPN were placed onto 96-well plates at $2 \times 10^3$ cells/well in triplicate. At designated time points, the number of cells were quantified using a colormetric MTT assay as described previously [22].

### 2.7. In vitro cell migration assay

SBC-3/OPN and SBC-3/NEO were transfered to 6-well culture plates at $5 \times 10^5$ cells/well and incubated with 2% FCS in RPMI 1640 medium overnight. The cells were washed in PBS, and 3 ml of serum free medium were added to each well. After 24 h, 3 ml of conditioned serum-free medium were collected and subjected to in vitro cell migration assay. In vitro cell migration was performed using a cell culture insert with 8 μm micropore membrane (Falcon; Becton Dickinson, Franklin Lake, NJ) as previously described [21]. Briefly, the suspension of HUVEC ($5 \times 10^4$ cells/200 μl in RPMI 1640 containing 0.1% BSA) was added to the upper chamber and the collected medium was added to the lower chamber. In order to confirm cell migration mediated by OPN, we conducted additional experiments by treating the cells with GRGDS peptide (Sigma) at the concentration of 100 μM or anti-human αvβ3 antibody at the concentration of 10 μg/ml. After incubation at 37 °C for 8 h, the filters were fixed with 10% formalin, and stained with crystal violet. The cells on the upper surface of the filters were removed by swabbing with a cotton swab, and the cells that had migrated to the lower surface were counted under a microscope at the magnification of 200×. All assays were performed in triplicate and at least three independent experiments were performed.

### 2.8. Soft agar colony formation assay

Six-well culture plates were covered with a layer of 0.5% agar in RPMI 1640 medium containing 20% (v/v) fetal calf serum to prevent the attachment of the cells to plastic substratum. Cell suspensions ($5 \times 10^3$ cells/well) of the SBC-3/OPN or SBC-3/NEO cells were prepared with 0.3% agar and poured into 6-well plates. After 2 weeks of incubation at 37 °C, the colonies containing at least 50 cells were counted. All assays were performed in triplicate.

### 2.9. Mice

Female athymic BALB/c nude mice, 6–7 weeks old, were purchased from Charlsriver Co., Ltd. (Tokyo, Japan) and maintained in our animal facilities under specific pathogen-free conditions. All animal experiments were performed according to the Guidelines on Animal Experimentation as established by Juntendo University, School of Medicine.

### 2.10. In vivo tumorigenicity

SBC-3/OPN and SBC-3/NEO cells were harvested from the culture flask with 0.05% Trypsin-EDTA (Invitrogen), washed three times, resuspended in PBS. Cell viability was determined by trypan blue dye exclusion test and cells were inoculated subcutaneously (s.c.) into the left flank of nude mice ($1 \times 10^7$ cells/mouse). To investigate whether tumor growth is mediated by the interaction between OPN and its receptor, the RMV-7 antibody was administered to SBC-3/OPN or SBC-3/NEO inoculated mice. Briefly, RMV-7 (200 μg/mouse) and control isotype-matched IgG (200 μg/mouse) were administered intraperitoneally from day 3 after inoculation three times a week for 3 weeks. TNP-470 (30 mg/kg) was also administered subcutaneously from day 7 twice a week for 3 weeks to reveal the involvement of angiogenesis in in vivo tumor growth. Tumor growth was measured with a digital caliper in two perpendicular diameters every week. Tumor volumes were calculated from the length ($a$) and width ($b$) by using the following formula:

volume ($mm^3$) = $ab^2/2$. Each group consisted of 10 mice. All experiments were performed twice.

### 2.11. Immunohistochemical staining

Histological sections were obtained from SBC-3/OPN and SBC-3/NEO tumor tissues resected from mice. After resection, tumor tissues were immediately embedded and frozen in Tissue-Tek OCD compound (Miles Laboratories, Elkhart, TN), and sections were cut at 4 μm thickness. Immunohistochemical staining for murine OPN and CD31 was performed as previously described [23]. To quantify apoptotic cell number in the tumor, we performed immunohistochemical staining for ssDNA. Briefly, the sections were fixed with 4% paraformaldehyde (PFA) and then incubated at 4°C overnight with rabbit anti-ssDNA antibody diluted to 1:400. Specific binding was detected through avidin-biotin peroxidase complex formation with biotin conjugated goat anti-rabbit IgG (Vectastain ABC kit, Vector, Burlingame, CA) and diaminobendizine (DAB) (Sigma, St. Louis, MI) as the substrate. Staining was absent when isotype-matched immunoglobulin was used as the control.

### 2.12. HUVEC proliferation assay

A 96-well flat bottom plastic assay plate (Corning, NY) was coated with recombinant mouse OPN (RD systems, Inc., CA; 10 μg/ml), polylysine (100 μg/ml) or BSA (10 mg/ml) in PBS and incubated overnight at 4°C. The plate was washed with PBS and non-specific adhesion sites were blocked with 1% BSA in PBS for 1 h at 37°C. After washing the wells with PBS, $5 \times 10^3$ cells in 100 μl of EGM-2 medium diluted with OPTI-MEM (Invitrogen) to 1/5 were seeded to each well. For some experiments, the HUVEC suspensions were pretreated with GRGDS peptide at the concentration of 100 μM or anti-human αvβ3 antibody at the concentration of 10 μg/ml for 1 h at 37°C. Then after 48 h incubation, 10 μl of 2-(2-methoxy-4-nitrophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium monosodium salt (WST-8, Dojindo, Kumamoto, Japan) was added to each well. The plate was further incubated at 37°C for 6 h for color development. Absorbance was measured at 450 nm on a microplate reader with microplate manager (Bio-Rad, Richmond, CA). All experiments were performed in triplicate.

### 2.13. Statistics

Statistical analysis was performed with analysis of variance (ANOVA). All data are presented as mean ± standard deviation. Differences between means were considered statistically significant at $P < 0.05$.

## 3. Results

### 3.1. Generation of stable transfectants that secretes murine OPN

BMGneo-mOPN or BMGneo were transfected into SBC-3 cells. Two OPN transfected SBC-3 clones (SBC-3/OPN#5 and SBC-3/OPN#6) and two control clones (SBC-3/NEO#1 and SBC-3/NEO#3) were constructed. To verify the expression of OPN and VEGF mRNA in transfectants, we conducted RT-PCR for OPN and VEGF, respectively. As shown in Fig. 1A, high levels of OPN mRNA expression were detected in the SBC-3/OPN cells, while there were no detectable expression levels observed in the SBC-3/NEO cells. For VEGF mRNA, there was no difference in the level of expression between SBC-3/OPN and SBC-3/NEO cells. Thus, transfection with OPN gene into SBC-3 cells does not affect the expression of other angiogenic inducers like VEGF mRNA. Secreted OPN protein from transfectants was confirmed with both western blot analysis and ELISA kit (Fig. 1B and C). OPN-transfected clones secreted significant amounts of OPN, while control clones did not. The clone SBC-3/OPN#5 secreted the high-

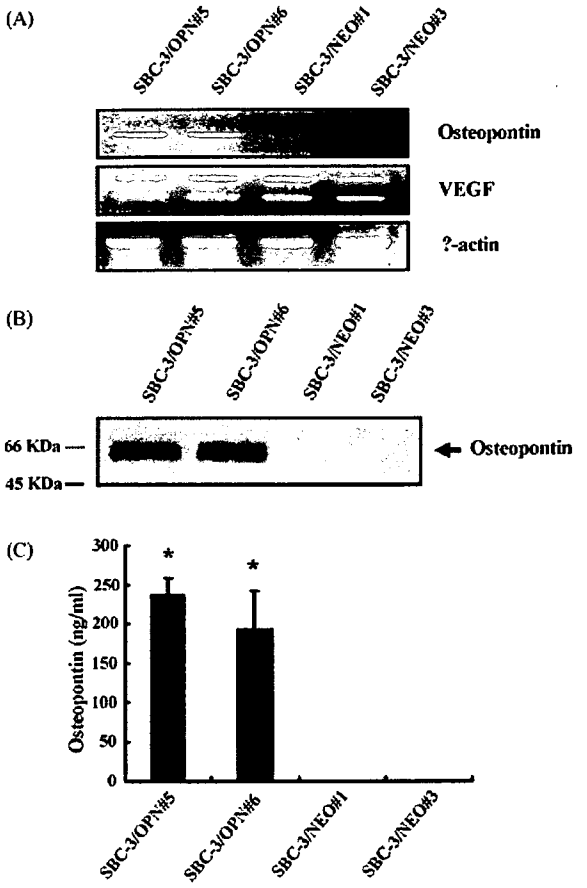


**Fig. 1** (A) Expression of OPN and VEGF mRNA determined with RT-PCR analysis. Total RNAs were extracted from each clone and 1 μg of RNAs were subjected to RT-PCR analysis for OPN (top panel), VEGF (middle panel) and β-actin (bottom panel) expression. (B) Western blot analysis of secreted mouse OPN protein. Conditioned mediums from SBC-3/OPN and SBC-3/NEO clones were subjected to western blot analysis using polyclonal antibody against OPN. The arrow indicates the expression of OPN, and molecular standards are shown on the left in KD. (C) Secretion of OPN protein from SBC-3/OPN and SBC-3/NEO cells. Conditioned medium from each clones were collected and subjected to ELISA analysis. Note that the clone SBC-3/OPN#5 secreted the highest level of OPN protein into the culture medium. *$P < 0.05$ vs. SBC-3/NEO#1 and SBC-3/NEO#3.

est level of OPN protein into the culture medium. Therefore, we utilized this clone in the subsequential experiments.

### 3.2. In vitro cell growth rate of stable OPN-transfectant

Cells were seeded onto 96-well plates and the number of cells was quantified at specific time intervals with MTT assay. Cultured SBC-3/OPN and SBC-3/NEO cells displayed similar in vitro growth rates (data not shown).

### 3.3. Biological activity of OPN protein secreted from the transfectant

Since endothelial cell migration is essential for tumor angiogenesis, we conducted migration assay using HUVEC. Conditioned medium from SBC-3/OPN cells significantly stimulated HUVEC migration as compared with conditioned medium from SBC-3/NEO cells. Moreover, HUVEC migration toward the culture medium of SBC-3/OPN cells was almost completely suppressed with the addition of GRGDS peptide and anti-human αvβ3 antibody (Fig. 2). These results suggest that OPN secreted from SBC-3/OPN is actually biological active and stimulates HUVEC migration by interacting with αvβ3 integrin.

### 3.4. Effect of OPN transfection on colony formation

We evaluated whether transfection with OPN gene affects colony formation of SBC-3 cells in vitro with soft agar colony formation assay. As shown in Fig. 2B, there was no significant difference in the number of colonies between SBC-3/OPN and SBC-3/NEO cells. Thus, colony formation of SBC-3 cells in vitro was not affected by transfection with the OPN gene.

### 3.5. In vivo tumorigenecity of OPN transfectant

To investigate whether OPN has any role in tumor growth in vivo, SBC-3/OPN#5 clone and SBC-3/NEO#1 clone were injected subcutaneously into the left flank of the nude mice. As shown in Fig. 3 A and B, in contrast to the absence of any significant changes in in vitro cell growth, the in vivo growth rate of SBC-3/OPN#5 was significantly faster than that of the SBC-3/NEO#1 cells. We also tested in vivo tumor growth of the other SBC-3/OPN clone, SBC-3/OPN#6, to confirm its enhanced in vivo tumorigenecity. As expected, SBC-3/OPN#6 demonstrated enhanced in vivo tumor growth compared to SBC-3/NEO#1 (data not shown).

### 3.6. Expression of OPN protein in SBC-3/OPN and SBC-3/NEO tumors

To investigate whether enhanced tumor growth of SBC-3/OPN clones in vivo was mediated by secreted OPN, immunohistochemical staining for OPN was conducted. The OPN-positive cell number was significantly greater in the SBC-3/OPN induced tumor in comparison with that of the SBC-3/NEO tumor (Fig. 3C). These results suggest that

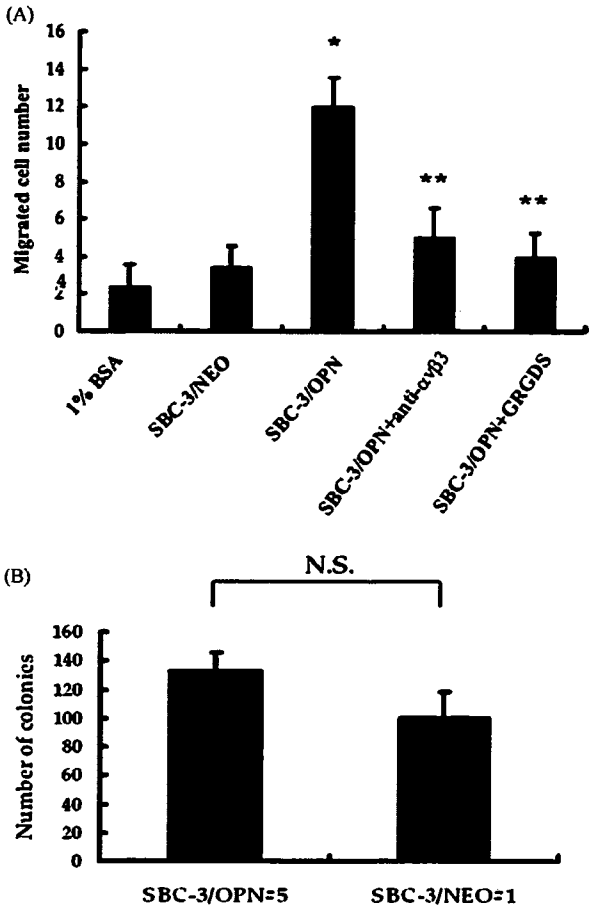


Fig. 2 (A) Migration of HUVEC toward conditioned medium from OPN-transfected cells. Cells were placed in the upper chamber and culture medium from SBC-3/NEO and SBC-3/OPN were added to the lower chamber. After 8 h incubation, cells that migrated through the porous filter were counted at ×200 magnification. Enhanced migration of HUVEC toward the culture medium from SBC-3/OPN was abrogated with the addition of either GRGDS peptide (100 μM) or anti-human αvβ3 antibody (10 μg/ml) to the upper chambers. Data are presented as mean ± S.D. $^{*}P < 0.0001$ vs. 1% BSA and SBC-3/NEO; $^{**}P < 0.001$ vs. SBC-3/OPN. (B) Soft-agar colony formation by SBC-3/OPN and SBC-3/NEO cells. Cells were seeded at an initial density of $5 \times 10^3$ cells into 6-well culture plates in triplicate in 0.3% agar. Colonies containing at least 50 cells were scored after 2 weeks of growth. Total colony per well were counted and presented as the mean ± S.D.

secreted OPN from SBC-3/OPN transfectants enhanced in vivo tumorigenesis.

### 3.7. Effect of OPN transfection on tumor angiogenesis

To investigate whether transfection with OPN gene results in increased tumor angiogenesis in vivo, we performed immunohistochemistry for CD31 and counted the microvessels in the SBC-3/OPN#5 and SBC-3/NEO#1 induced tumors of the nude mice. As shown in Fig. 4A, the number of CD31-