Table 1: Continued

protein	peptides	glycopeptides					*N*-glycan				
							deduced monosaccharide composition				
	sequence[a,b]	elution position	Figure	peak no.[c]	scan in Figure 4A[d]	observed peptide-related ion[e]	observed *m/z* in SIM mode[b]	theoretical *m/z*[b]	dHex	Hex	HexNAc	NA	deduced structure[f] (diagnostic ion)
				r-4	3208 (C)	1585.7	1099.450 (3)	1099.448	2	5	4	0	H, CoreF(1731.7), $L^{a/x}$ (350.1, 512.1)
				r-5	3189	1584.7	1104.784 (3)	1104.780	1	6	4	0	H, CoreF(1730.6) or $L^{a/x}$ (350.4, 512.0)
				r-6	3144 (A)	1585.6	1113.127 (3)	1113.123	2	4	5	0	C, CoreF(1730.8), $L^{a/x}$ (350.1, 512.2)
					– (A)	1585.8	1134.118 (3)	1134.118	1	6	3	1	H, CoreF(1730.6) or 512(512.3)
					– (A)	1584.5	1153.466 (3)	1153.466	2	6	4	0	H, CoreF(1731.6), $L^{a/x}$ (350.1, 512.2)

[a] Theoretical peptide mass indicated in parentheses. [b] Monoisotopic values. [c] Peaks are numbered in decreasing order of their calculated mass. All glycopeptides are triply charged except for doubly charged ions indicated by (2) after the peak number. [d] Glycopeptides were characterized on the basis of alternative LC–MS^n runs with conditions indicated in parentheses (A, a C30 column, scan range of *m/z* 1000–2000; B, a C30 column, scan range of *m/z* 700–2000; C, a C18 column, scan range of *m/z* 1000–2000). [e] Y_1^{n+} or $Y_{1\alpha/\beta}^{n+}$, [(peptide + HexNAc + *n*H)/*n*]$^{n+}$; or $Y_{1\alpha}^{n+}$, [(peptide + HexNAc + dHex + *n*H)/*n*]$^{n+}$. All peptide-related ions are singly charged except for doubly or triply charged ions indicated by (2) or (3). [f] Structures are deduced by MS^n: C, complex-type oligosaccharide; H, hybrid-type oligosaccharide; Man-5-9, high mannose-type oligosaccharide containing 5–9 mannose residues; CoreF, trimannosylcore fucose; bisectGN, bisecting GlcNAc; diSia, disialic acid; $L^{a/x}$, Lewis a/x structure; $sL^{a/x}$, sialylated Lewis a/x structure; $L^{b/y}$, Lewis b/y structure; 512, glycan motif consisting of $dHex_1Hex_1HexNAc_1$. The structure in parentheses indicates the possible structures to be contained in the glycopeptide. [g] Glycosylation was confirmed by Asn–Asp conversion upon PNGase F digestion.

integrated mass spectrum (peaks f-1–9 and g-1–3 in panel F2 of Figure 5) and their MS/MS spectra suggested that complex-type oligosaccharides including $Le^{a/x}$ or $Le^{b/y}$ -modified and/or bisected oligosaccharides and BA-2 are attached to Asn272 (Table 1F).

(vii) Asn287. The MS/MS spectra of GPI-linked peptides were selected from all MS data on the basis of the GPI-characteristic oxonium ions, such as GlcN-Ino-PO_4^+ (*m/z* 422). The structures of the GPI moieties were characterized from their product ions appearing in the MS/MS spectra, and their peptide portions were identified by comparing their observed masses with the theoretical masses of predicted peptides. Figure 4B shows the TIC obtained by GCC-LC–MS^n for the hydrophilic glycopeptides. On the basis of the presence of GPI-characteristic oxonium ions, the MS data of GPI-linked peptides were located at position 26. The 9.5% of spectra generated at elution position 26 were assigned to those of GPI-linked peptides of LAMP, OBCAM, and neurotrimin.

Figure 5G shows one of the MS/MS spectra acquired at position 26 (precursor ion, $[M + 2H]^{2+}$ at *m/z* 902.5; peak L2 in Figure 4C). On the basis of the GPI-characteristic oxonium ions, such as NH_2Et-PO_4-Man-$GlcN^+$ (*m/z* 447.2), NH_2Et-PO_4-(HexNAc-)Man-$GlcN^+$ (*m/z* 650.3), NH_2Et-PO_4-(HexNAc-)Man-GlcN-Ino-PO_4^+ (*m/z* 910.2), NH_2Et-PO_4-(HexNAc-)(Hex-)Man-GlcN-Ino-PO_4^+ (*m/z* 1072.2), and GlcN-Ino-PO_4^+ (*m/z* 422.2), this peptide was identified as the GPI-linked peptide. The product ion at *m/z* 328.3 was assigned to GIN^{287}-NH-Et^+ on the basis of the fragments that arose by successive cleavages of HexNAc (*m/z* 1600.4), Ino-PO_4 (*m/z* 1340.5), GlcN (*m/z* 1178.3), Man-PO_4-$EtNH_2$ and Hex (*m/z* 732.2), Hex (*m/z* 570.2), and PO_4-Hex (*m/z* 328.3). In addition, the product ions at *m/z* 732.3 and 1072.2 suggested the existence of HexNAc-(NH_2Et-PO_4-)(Hex)-Man3 in the core structure of GPI (inset of Figure 5G). The presence of a positional isomer was inferred from the acquisition of two different MS/MS spectra of GPI-linked peptides (precursor ion $[M + 2H]^{2+}$, *m/z* 903) at different elution times (Table 2). The alternative runs also suggested the presence of a Hex-Man1 and HexNAc-(Hex-)(NH_2Et-PO_4-)Man3 (peak L1, data not shown, Table 2), and a nonsubstituted Man1 and HexNAc-(NH_2Et-PO_4-)Man3 (data not shown, Table 2) in the GPI core structure.

Glycosylation Analysis of OBCAM. OBCAM has six potential N-glycosylation sites at Asn17, -43, -113, -258, -266, and -279, and the predicted linkage site of GPI is Asn295. From the peptide-related ions, peptides eluted at positions 2, 25, and 7 were estimated to be glycopeptides containing Asn17, -258, and -266, respectively (panels A1–C1 of Figure 6). Panels A2–C2 of Figure 6 show the integrated mass spectrum of glycopeptides obtained from positions 2, 25, and 7, respectively. The glycopeptide containing Asn43 is identical to VAWLN^{38}R in LAMP. From the glycosylation at Asn38 in LAMP, Man-5-9 were inferred to be attached to Asn43 (panel A2 of Figure 5 and Table 1A). Although the MS/MS spectrum of the glycopeptide containing Asn113 (VHLIVQVPPQIMN^{113}ISSD) was not acquired, glycosylation at Asn113 was corroborated by detection of VHLIVQVPPQIMD^{113}ISSD after PNGase F treatment (data not shown). The feature of glycosylation at Asn279 was elucidated on the basis of the MS/MS spectra of glycosylated LGNTN^{279}ASITLYGPGAVID which was

Table 2: Summary of GPI Structure in LAMP, OBCAM, and Neurotrimin

protein	peptide (theoretical MW[b])	peak no. in Figure 4C	scan in Figure 4B	GPI-linked peptide: observed peptide-related ion[b] (charge state)	GPI-linked peptide: observed *m/z*[b] (charge state)	GPI-linked peptide: calculated mass	GPI moiety: deduced glycan composition: calculated mass	core	Man1	Man3			theoretical MW[b]
									Hex	Hex	HexNAc	P-EtNH$_2$	
LAMP	GIN[287] (302.3)	L1	3863	328.3 (1)	983.6 (2)	1965.1	1680.9	1	1	1	1	1	1681.3
		L2	**3828[c] (Figure5G)**	328.3 (1)	**902.5 (2)**	**1803.0**	**1518.8**	**1**	**0**	**1**	**1**	**1**	**1519.2**
			4040[c]	328.3 (1)	903.1 (2)	1804.2	1520.0	1	0	1	1	1	1519.2
			a	328.2 (1)	821.6 (2)	1641.1	1356.9	1	0	0	1	1	1357.0
OBCAM	GVN[295] (288.3)	**O1**	**3701 (Figure6D)**	314.3 (1)	**976.5 (2)**	**1951.0**	**1680.7**	**1**	**1**	**1**	**1**	**1**	**1681.3**
		O2	3633[d]	314.3 (1)	895.4 (2)	1788.7	1518.4	1	0	1	1	1	1519.2
			3853[d]	314.3 (1)	895.5 (2)	1788.9	1518.6	1	0	1	1	1	1519.2
		O3	3805	314.3 (1)	814.6 (2)	1627.1	1356.8	1	0	0	1	1	1357.0
neurotrimin	VNN[289] (345.4)	N1	3750	371.2 (1)	1004.8 (2)	2007.7	1680.3	1	1	1	1	1	1681.3
		N2	3741[e]	371.4 (1)	924.0 (2)	1846.1	1518.7	1	0	1	1	1	1519.2
			3896[e]	371.2 (1)	924.1 (2)	1846.1	1518.8	1	0	1	1	1	1519.2
		N3	**3873 (Figure7D)**	371.3 (1)	**842.8 (2)**	**1683.5**	**1356.1**	**1**	**0**	**0**	**1**	**1**	**1357.0**

[a] The structure of GPI was deduced by another LC−MS[n] run. [b] Average value. [c] Isomers. [d] Isomers. [e] Isomers.

acquired in an alternative run with the C30 column (scan range of *m/z* 1000−2000) (Table 1J).

(i) Asn 17. As shown in panel A1 of Figure 6, the glycopeptide that eluted at position 2 was assigned to AMDN[17]VTVR (and/or AMDN[12]VTVR in neurotrimin) glycosylated with dHex$_1$Hex$_5$HexNAc$_4$NeuAc$_4$ based on the $Y_{1\alpha}$ ion and the monoisotopic mass of the molecular ion. The attachment of three NeuAc residues in one branch of a biantennary complex type was suggested by the existence of characteristic B ions (*m/z* 495.2, 744.9, and 1239.2) (panel A1 of Figure 6). The molecular ions appearing in the integrated mass spectrum and their MS/MS spectra suggested that most of the glycans at Asn17 were disialic acid-conjugated oligosaccharides (peaks h-1−3 in panel A2 of Figure 6 and Table 1G).

(ii) Asn258. Panel B1 of Figure 6 shows the representative MS/MS spectrum of glycosylated ISTLTFFN[258]VSE that eluted at position 25. The monosaccharide composition (dHex$_2$Hex$_5$HexNAc$_6$NeuAc$_1$) implied two possible structures: a sLe$^{a/x}$ -modified core-fucosylated complex type and a Le$^{a/x}$ or antigen H-modified core-fucosylated and sialylated complex type (inset of panel B1 of Figure 6). The molecular ions (peaks i-1−2) in the integrated mass spectrum (panel B2 of Figure 6) and the detection of nonglycosylated ISTLTFFN[258]VSE revealed that Asn258 is partly glycosylated with the sLe$^{a/x}$ or Le$^{b/y}$-modified core-fucosylated complex type, and BA-2 (Table 1H).

(iii) Asn266. Panel C1 of Figure 6 shows the product ion spectra of the glycopeptide at position 7, the peptide portion of which was assigned to YGN[266]YTCVATNK on the basis of the $Y_{1\alpha/1\beta}$ ion in the MS/MS/MS spectrum. The glycan was characterized as the bisected and core-fucosylated complex-type oligosaccharide containing Le$^{a/x}$ structure from the monosaccharide composition (dHex$_2$Hex$_4$HexNAc$_5$), and the Le$^{a/x}$-, bisecting-, and core-fucose-related ions. The MS/MS spectra acquired with other glycoforms (peaks j-1−4 in panel C2 of Figure 6) together with the MS/MS spectra of the glycopeptides DYGN[266]YTCVATNK (position 13) and KDYGN[266]YTCVATNK (position 6) suggested that the Le$^{a/x}$-modified and/or bisected complex type and Man-5 were predominantly attached to Asn266 (Table 1I).

(iv) Asn295. On the basis of the GPI-characteristic oxonium ions and the peptide-related ion (*m/z* 314.3), the MS/MS spectrum of GPI-linked GVN[295] was picked out from position 26 (Figure 6D; precursor ion, *m/z* 976.5; peak O1 in Figure 4C). The fragments arising from the GPI moiety suggested the linkage of Hex to Man1, and HexNAc, Hex, and NH_2Et-PO_4 to Man3 in the core structure (Figure 6D, inset). Furthermore, the MS/MS spectrum of other GPI-linked GVN[295] (precursor ion, *m/z* 895; peak O2), which was picked out from position 26 based on the peptide-related ion, suggested that this GPI moiety contained HexNAc-(Hex)-(NH_2Et-PO_4-)Man3. Another MS/MS spectrum (precursor ion, *m/z* 814; peak O3) suggested the linkage of GPI moieties containing HexNAc-(NH_2Et-PO_4-)Man3 (Table 2). The existence of two isomers was suggested in peak O2 by the acquisition of two MS/MS spectra of GPI-GVN[295] (*m/z* 895) at different elution times.

Glycosylation Analysis of Neurotrimin. Neurotrimin contains seven potential N-glycosylation sites at Asn12, -38, -120, -184, -252, -260, and -273, and the predicted linkage site of GPI is Asn289. As the amino acid sequence in the

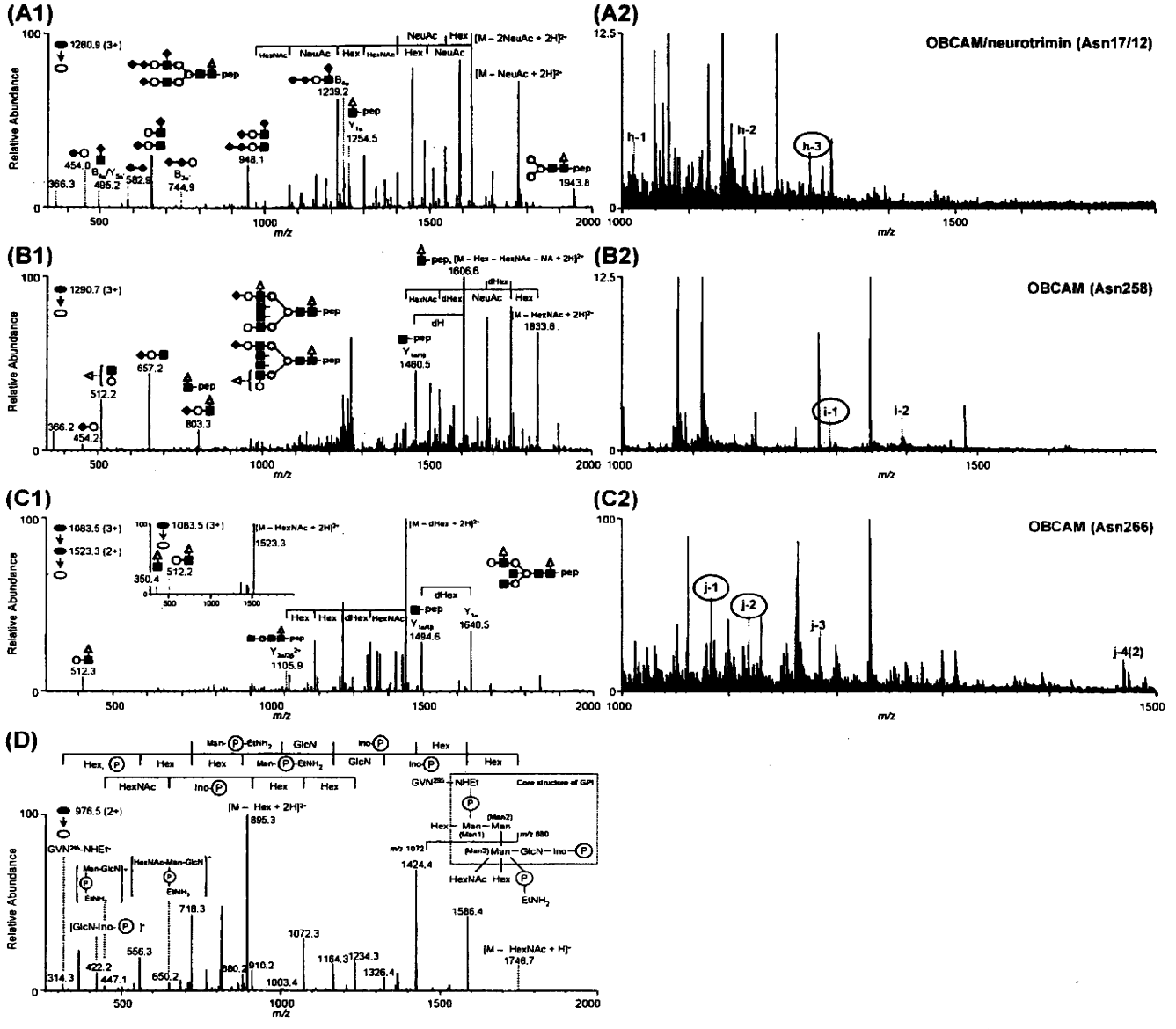


FIGURE 6: MS spectra of OBCAM glycopeptides. (A1) MS/MS spectra of glycopeptide AMDN[17]VTVR; elution position, 2; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1280.9). (A2) Integrated mass spectrum obtained from position 2. (B1) MS/MS spectrum of glycopeptide ISTLTFFN[258]VSE; elution position, 25; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1290.7). (B2) Integrated mass spectrum at position 25. (C1) MS/MS and MS/MS/MS spectra of glycopeptide YGN[266]YTCVATNK; elution position, 7; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1083.5). (C2) Integrated mass spectrum at position 7. (D) MS/MS spectrum of GPI-linked GVN[295]; elution position, 26; precursor ion, $[M + 2H]^{2+}$ (*m/z* 976.5). Symbols are as in Figure 9.

glycopeptide containing Asn12 (GTDN[12]ITVR) in neurotrimin is identical to GTDN[17]ITVR in OBCAM, the glycans at Asn12 are estimated to be hybrid and complex types containing disialic acid (panel A2 of Figure 6 and Table 1G). Likewise, the sequence of VAWLN[38]R in neurotrimin is identical to that of VAWLN[38]R in LAMP, and therefore, the linkage of Man-5-9 at Asn38 was inferred from the glycosylation at Asn38 in LAMP (panel A2 of Figure 5 and Table 1A). Although the MS/MS spectra of glycopeptides containing Asn120 were not acquired, glycosylation at Asn120 was confirmed by the identification of GND[120]ISLTCIATGR, GND[120]ISLTCIATGRPE, and GND[120]ISLTCIATGRPEPTVTWR after PNGase F digestion (data not shown). The substitution of Asn184 with a Lys or an Arg residue in neurotrimin was suggested as in case of SD rat by the identification of VTVNYPPYISE, which is a fragment of VN[184]VTVNYPPYISE (data not shown) (*33*).

The MS/MS spectra of glycopeptides containing Asn252, -260, -273, and -289 were located at positions 20, 5, 23, and 26 based on the peptide-related ions, respectively (panels A1−C1 and D of Figure 7). The integrated mass spectrum of the glycopeptides containing Asn252, -260, and -273 are shown in panels A2−C2 of Figure 7, respectively.

(i) Asn252. Panel A1 of Figure 7 shows the representative MS/MS spectra of glycopeptide LTFFN[252]VSE linked by $dHex_2Hex_6HexNAc_4$, acquired at position 20. A $Le^{a/x}$ -modified core-fucosylated and bisected hybrid-type oligosaccharide was deduced from the $Le^{a/x}$-related ions, and $Y_{1\beta/3\alpha/3\beta}^{2+}$ and $Y_{1\alpha}$. The majority of the glycans at Asn252 are estimated to be $Le^{a/x}$ or $Le^{b/y}$-modified complex- and hybrid-type oligosaccharides from the molecular ions (peaks k-1−9) in the integrated mass spectrum and their MS/MS spectra (panel A2 of Figure 7 and Table 1K).

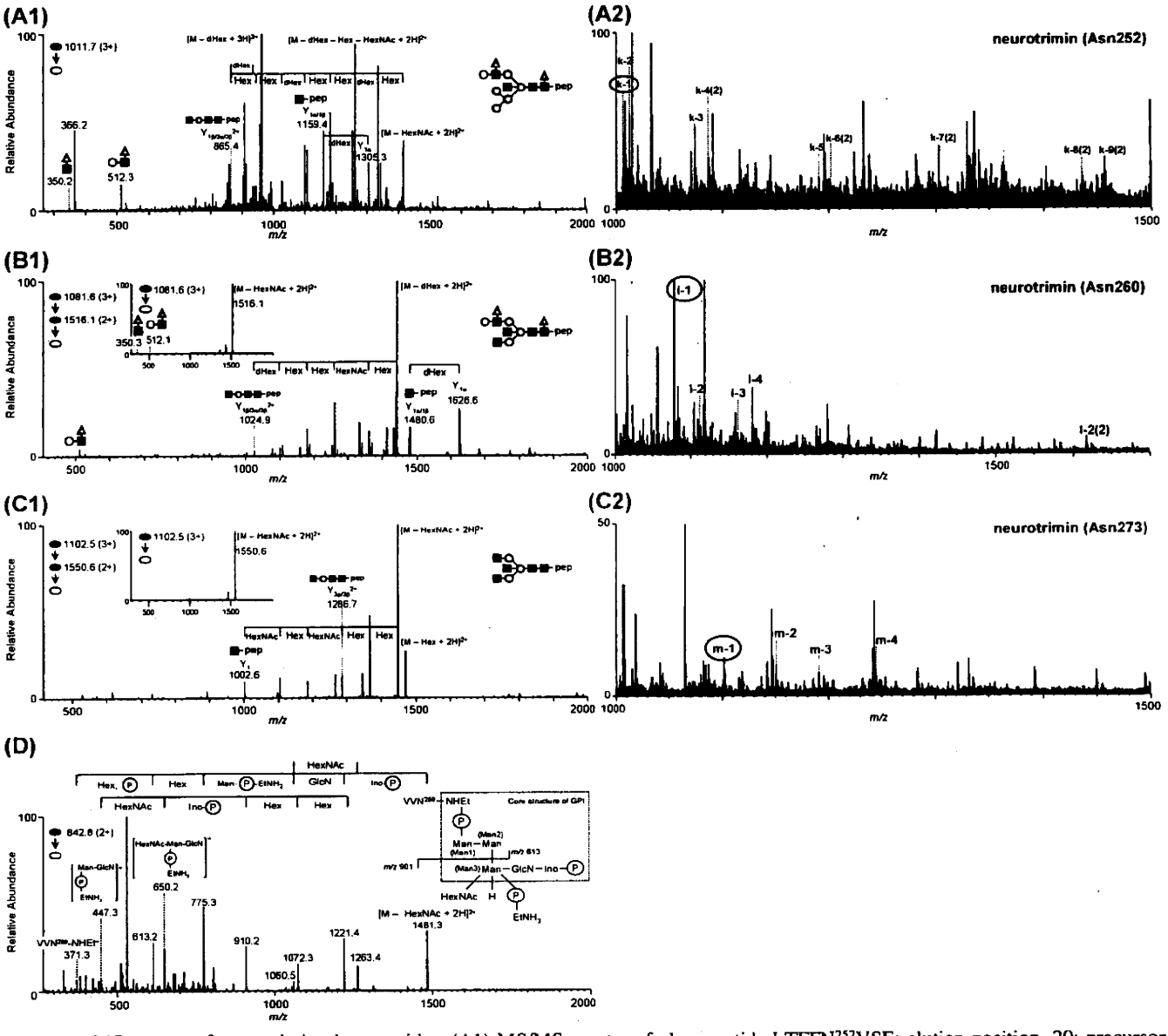


FIGURE 7: MS spectra of neurotrimin glycopeptides. (A1) MS/MS spectra of glycopeptide LTFFN^{252}VSE; elution position, 20; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1011.7). (A2) Integrated mass spectrum obtained from position 20. (B1) MS/MS and MS/MS/MS spectra of glycopeptide YGN^{260}YTCVASNK; elution position, 5; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1081.6). (B2) Integrated mass spectrum at position 5. (C1) MS/MS and MS/MS/MS spectra of glycopeptide LGHTN^{273}ASIMLFGPGAVSE; elution position, 23; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1102.5). (C2) Integrated mass spectrum at position 23. (D) MS/MS spectrum of GPI-linked VNN^{289}; elution position, 26; precursor ion, $[M + 2H]^{2+}$ (*m/z* 842.8). Symbols are as in Figure 9.

(ii) Asn260. Panel B1 of Figure 7 shows the representative product ion spectra of the glycopeptide eluted at position 5, the peptide portion of which was identified as YGN^{260}YTCVASNK on the basis of the $Y_{1\alpha/1\beta}$ ion in the MS/MS/MS spectrum. The monosaccharide composition ($dHex_2Hex_4HexNAc_5$), the Le$^{a/x}$-related ions in the MS/MS spectrum, and the presence of $Y_{1\beta/3\alpha/3\beta}{}^{2+}$ and $Y_{1\alpha}$ in the MS/MS/MS spectrum revealed the linkage of a Le$^{a/x}$-modified fucosylated and bisected complex-type oligosaccharide to this peptide (inset of panel B1 of Figure 7). The molecular ions in the integrated mass spectrum (peaks l-1−4 in panel B2 of Figure 7) together with the MS/MS spectra of glycosylated HDYGN^{260}YTCVASNK (position 8) suggested that Asn260 was predominantly glycosylated with the Le$^{a/x}$ or Le$^{b/y}$-modified bisected complex- and hybrid-type oligosaccharides and BA-2 (Table 1L).

(iii) Asn273. On the basis of the Y_1 ion and the monoisotopic mass, the glycopeptide eluted at position 23 was assigned to LGHTN^{273}ASIMLFGPGAVSE glycosylated with $Hex_3HexNAc_5$ (panel C1 of Figure 7). Its glycan moiety was characterized as a bisected agalacto-complex-type oligosaccharide based on $Y_{3\alpha/3\beta}{}^{2+}$. Other glycans at Asn273 were assigned to bisected complex- and hybrid-type oligosaccharides (peaks m-1−4 in panel C2 of Figure 7 and Table 1M).

(iv) Asn289. Figure 7D shows one of the MS/MS spectra of GPI-linked VNN^{289}, which was picked out from position 26 on the basis of the peptide-related ion (peptide-NH-Et$^+$, *m/z* 371.3). Three different MS/MS spectra of GPI-linked VNN^{289} were picked out from position 26 (Figure 4B). From the molecular ions [peaks N1 (*m/z* 1004), N2 (*m/z* 924), and N3 (*m/z* 842)] and their fragments, it was suggested that they contain Hex-Man1 and HexNAc-(Hex-)(NH_2Et-PO_4-)Man3,

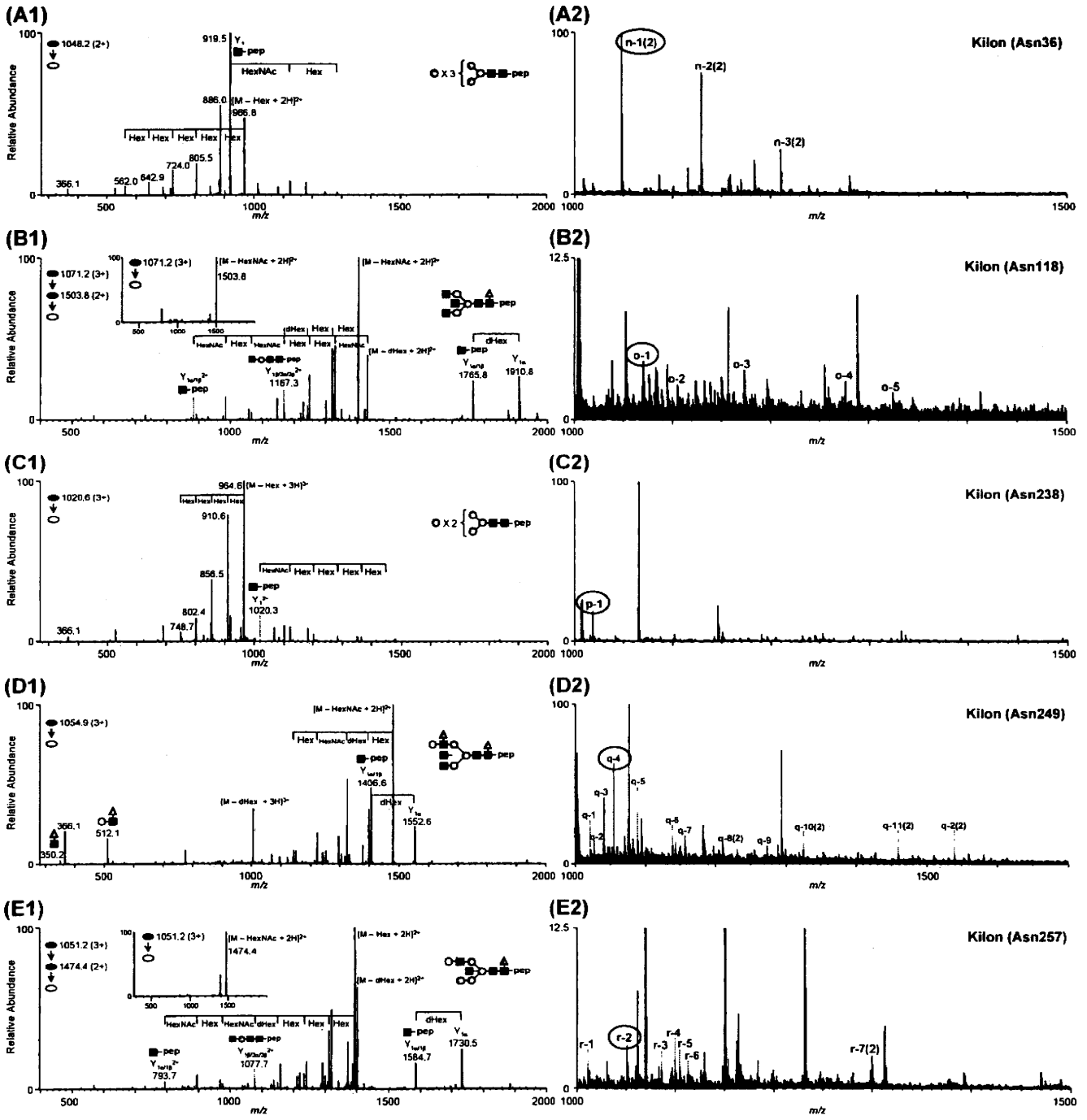


FIGURE 8: MS spectra of Kilon glycopeptides. (A1) MS/MS spectra of glycopeptide GAWLN^{36}R; elution position, 3; precursor ion, $[M + 2H]^{2+}$ (*m/z* 1048.2). (A2) Integrated mass spectrum obtained from position 3. (B1) MS/MS and MS/MS/MS spectra of glycopeptide GTN^{118}VTLTCLATGKPE; elution position, 16; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1071.2). (B2) Integrated mass spectrum at position 16. (C1) MS/MS spectrum of glycopeptide LFNGQQGIIIQN^{238}FSTR; elution position, 22; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1020.6). (C2) Integrated mass spectrum at position 22. (D1) MS/MS spectrum of glycopeptide SILTVTN^{249}VTQE; elution position, 17; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1054.9). (D2) Integrated mass spectrum at position 17. (E1) MS/MS and MS/MS/MS spectra of glycopeptide HFGN^{257}YTCVAANK; elution position, 10; precursor ion, $[M + 3H]^{3+}$ (*m/z* 1051.2). (E2) Integrated mass spectrum at position 10. Symbols are as in Figure 9.

HexNAc-(Hex-)(NH_2Et-PO_4-)Man3, and HexNAc-(NH_2Et-PO_4-)Man3, respectively. The existence of two isomers was suggested in peak N2 by the presence of two different MS/MS spectra at different elution times (Table 2).

Glycosylation Analysis of Kilon. Kilon has six potential N-glycosylation sites at Asn36, -118, -238, -249, -257, and -270. The predicted linkage site of GPI is Gly287. The typical MS/MS spectra and the integrated mass spectra of the glycopeptides containing Asn36, -118, -238, -249, and -257 are shown in panels A1−E1 and A2−E2 of Figure 8, respectively. The MS/MS spectra of the glycopeptide containing both Asn270 and Gly287 could not be picked out from the MS data.

(i) Asn36. Panel A1 of Figure 8 shows one of the MS/MS spectra acquired at position 3. This glycopeptide was identified as GAWLN^{36}R with Man-6 based on Y_1 ion and the monosaccharide composition. Other glycans at Asn36 were estimated as Man-5, -7, and -8 from the existence of

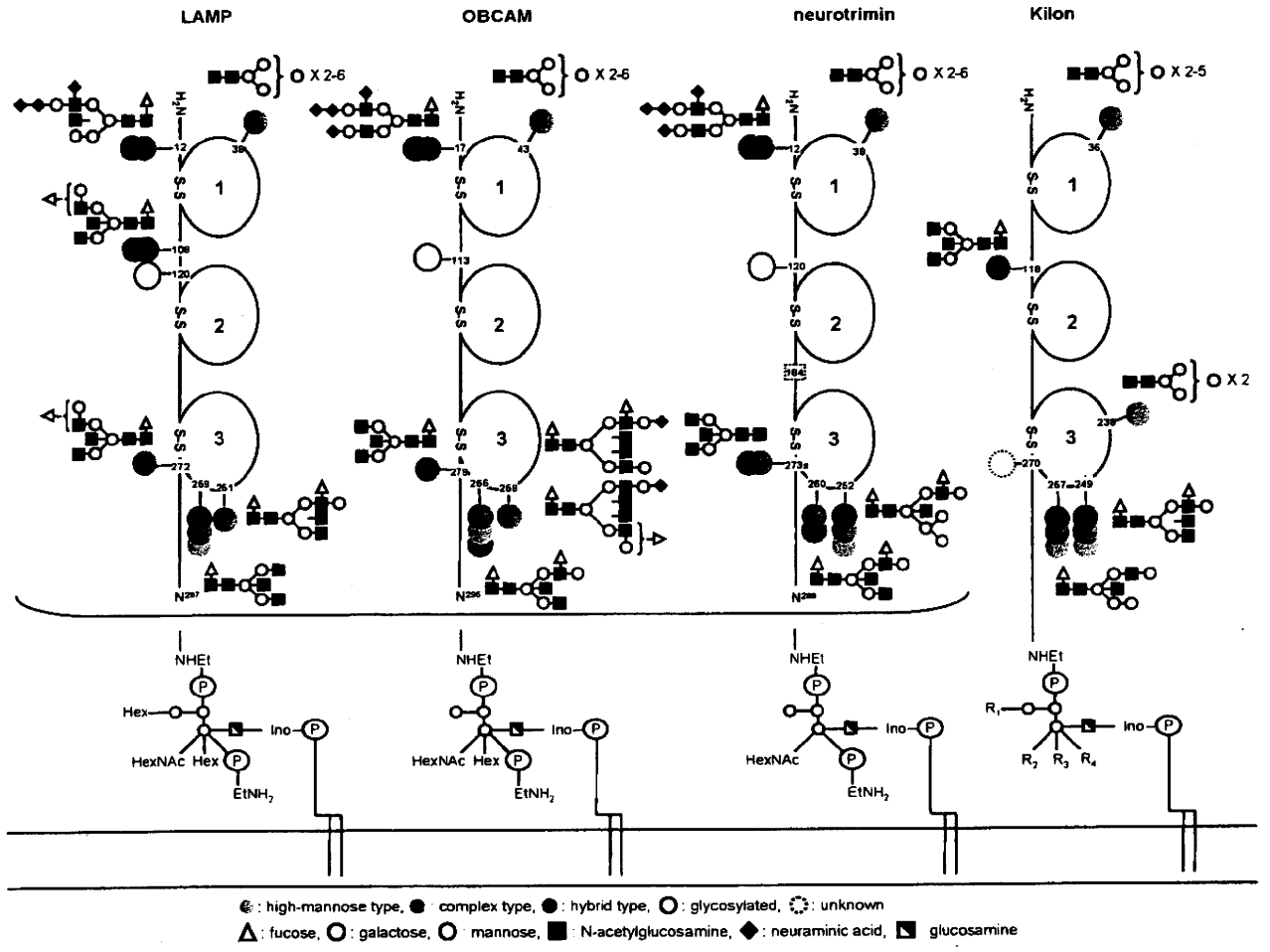


FIGURE 9: Summary of glycosylation of IgLON family proteins.

molecular ions with 81 *m/z* units intervals in the integrated mass spectrum (peaks n-1−3 in panel A2 of Figure 8) (Table 1N).

(ii) Asn118. As shown in panel B1 of Figure 8, the MS/MS/MS spectrum acquired at position 16 contained $Y_{1\alpha/1\beta}$, which suggested that the peptide portion is GTN[118]VTLTCLATGKPE. The linkage of BA-2 was deduced from the monosaccharide composition ($dHex_1Hex_3HexNAc_5$), and $Y_{1\beta/3\alpha/3\beta}^{2+}$ and $Y_{1\alpha}$ (inset of panel B1 of Figure 8). Additionally, the linkage of $Le^{a/x}$ or antigen H-modified and/or bisected complex type was suggested by the integrated mass spectrum (peaks o-1−5 in panel B2 of Figure 8 and Table 1O).

(iii) Asn238. The MS/MS spectra of glycopeptides that contain Asn238 were picked out from positions 22 [LFNGQQGIIIQN[238]FSTR (panel C1 of Figure 8)], 21 (RLFNGQQGIIIQN[238]FSTR), and 19 (KRLFNGQQGIII-QN[238]FSTR). These MS/MS spectra and molecular ions appearing in the integrated mass spectrum revealed that the only carbohydrate structure at Asn238 was Man-5 (peak p-1 in panel C2 of Figure 8 and Table 1P). Together with the results of the database search analysis, in which nonglycosylated peptide LFNGQQGIIIQN[238]FSTR was identified, it was suggested that Man-5 was partly attached to Asn238 (Table 1P).

(iv) Asn249. Panel D1 of Figure 8 shows the representative MS/MS spectrum of glycopeptide SILTVTN[249]VTQE at position 17. The carbohydrate structure was characterized as a $Le^{a/x}$-modified and core-fucosylated complex type by the existence of the $Le^{a/x}$-related ions and $Y_{1\alpha}$. The integrated mass spectrum and alternative LC−MS^n with the C30 column (scan ranges of *m/z* 700−2000 and 1000−2000) suggested that Asn249 is glycosylated with $Le^{a/x}$ or antigen H-modified core-fucosylated hybrid- and complex-type oligosaccharides, BA-2, and Man-5 (peaks q-1−11 in panel D2 of Figure 8 and Table 1Q).

(v) Asn257. As shown in panel E1 of Figure 8, one of the glycopeptides eluted at position 10 was identified as HFGN[257]YTCVAANK linked by $dHex_1Hex_5HexNAc_4$ based on $Y_{1\alpha/1\beta}$ ion in the MS/MS/MS spectra and monoisotopic mass. The carbohydrate structure was characterized as a bisected- and core-fucosylated hybrid-type oligosaccharide based on the presence of $Y_{1\beta/3\alpha/3\beta}^{2+}$ and $Y_{1\alpha}$ (inset of panel E2 of Figure 8). Other major glycans were estimated as Man-5, $Le^{a/x}$-modified complex- and hybrid-type oligosaccharides, and BA-2 (peaks r-1−7 in panel E2 of Figure 8 and Table 1R).

DISCUSSION

The cell adhesion molecules in the central nervous system play an essential role in the differentiation of neuronal cells and formation of neural circuits. Although glycosylation on the cell adhesion molecules is known to regulate cell−cell interactions (*2*−*4*), their carbohydrate structures remain unknown due to the difficulty with respect to their isolation and the limited sample amounts. The glycans in the IgLON family proteins are considered to be implicated in the

formation of neural circuits, including migration of neuronal cells, axonal guidance, and fasciculation. However, the high degree of homology of their amino acid sequences makes it difficult to isolate them from each other and to analyze their carbohydrate structures in detail.

In this study, we performed a site-specific glycosylation analysis of LAMP, OBCAM, neurotrimin, and Kilon simultaneously using SDS−PAGE and LC−MS^n. Enriched GPI-linked proteins were separated by SDS−PAGE, and four target proteins were extracted from a gel piece together with other contaminating proteins. The protein mixture was digested and analyzed by the C30 and C18-LC−MS^n runs via MS, data-dependent MS in SIM by the FT ICR-MS, and data-dependent MS/MS and MS/MS/MS. A set of MS data consisting of the mass spectrum, the mass spectrum acquired by the FT ICR-MS in SIM mode, the data-dependently acquired MS/MS, and the MS/MS/MS spectra of a glycopeptide was selected from all MS data on the basis of the existence of the oligosaccharide characteristic oxonium ions in the MS/MS spectrum. The carbohydrate structure and peptide sequence were deduced from the carbohydrate-related ions and peptide-related ions in the product ion spectra. The structural assignment of the glycopeptide was confirmed by the accurate mass acquired on the FT ICR-MS. The b- and y-ions arising from the peptide backbone in the MS/MS/MS spectra were also used for the peptide assignment. The carbohydrate heterogeneity at each glycosylation site was characterized by integrating the mass spectra of the glycopeptides which yielded identical peptide-related ions. We successfully determined the site-specific glycosylation in LAMP, OBCAM, neurotrimin, and Kilon with the exception of Asn120 in LAMP, Asn113 in OBCAM, Asn120 in neurotrimin, and Asn270 in Kilon. We also demonstrated the structure of the GPI moiety using LC−MS^n equipped with a GCC. A set of data was picked out from all MS data by using GPI-characteristic ions, and the structure of GPI and the linkage site were deduced from the product ions in the MS/MS spectra. Three different structures are commonly found in LAMP, OBCAM, and neurotrimin.

Figure 9 illustrates the site-specific glycosylation in the four proteins. N-Glycosylation sites near the N-terminus in LAMP, OBCAM, and neurotrimin were commonly occupied with biantennary complex-type and hybrid-type oligosaccharides containing disialic acids. Oligosialic acids and disialic acids, which are found in several glycoproteins, including NCAM, are considered to regulate the cell−cell interaction by changing their degree of polymerization (*6*). Disialic acids at the near N-terminus in LAMP, OBCAM, and neurotrimin might regulate the cell−cell interaction in a manner similar to that of other glycosylated adhesion molecules.

The first domains in IgLON family proteins are commonly glycosylated with Man-5, -6, -7, -8, and -9. The linkage of high-mannose-type oligosaccharides is found in several Ig superfamily proteins, including L1, MAG, and P0 (*3*). Since Horstkorte et al. have reported that L1 binds to NCAM through oligomannosidic carbohydrates in L1 (*34*), the high-mannose-type oligosaccharide in IgLON family proteins could interact with certain biological molecules.

The third domains of all IgLON proteins were highly heterogeneous due to a linkage of diverse oligosaccharides, including BA-2, the $Le^{a/x}$ or $Le^{b/y}$ motif, and Man-5. BA-2, a bisected agalacto-complex type, is known as a brain-specific glycan and is much more abundant in mammalian brains than in other tissues (*35, 36*). Recently, the Na^+/K^+-ATPase $\beta 1$ subunit was identified as a GlcNAc-binding protein in the mouse brain (*37*). The Na^+/K^+-ATPase $\beta 1$ subunit is a potassium-dependent lectin which binds to GlcNAc-terminating oligosaccharides and is involved in neural cell interactions in a trans-binding fashion. A 74 kDa protein was suggested to be the GlcNAc-terminating glycan carrier protein binding to the Na^+/K^+-ATPase $\beta 1$ subunit. The linkage of BA-2 to IgLON family proteins implies that these proteins might be the ligand proteins for the Na^+/K^+-ATPase $\beta 1$ subunit.

Glycosylation in a great number of membrane glycoproteins remains largely unknown. This is mainly because the limited amount of available sample and the low solubility of glycoproteins make their isolation quite difficult. Our strategy, which includes enrichment of the target glycoproteins, separation by SDS−PAGE, and LC−MS^n of digests of a protein mixture, can be applied to the site-specific glycosylation analysis of various membrane glycoproteins.

ACKNOWLEDGMENT

We thank Dr. Masayuki Kubota and Morihiko Yoshida (Thermo Fisher Scientific K.K.) for their technical support.

REFERENCES

1. Walsh, F. S., and Doherty, P. (1997) Neural cell adhesion molecules of the immunoglobulin superfamily: Role in axon growth and guidance. *Annu. Rev. Cell Dev. Biol. 13*, 425–456.
2. Kleene, R., and Schachner, M. (2004) Glycans and neural cell interactions. *Nat. Rev. Neurosci. 5*, 195–208.
3. Krog, L., and Bock, E. (1992) Glycosylation of neural cell adhesion molecules of the immunoglobulin superfamily. *APMIS, Suppl. 27*, 53–70.
4. Schachner, M., and Martini, R. (1995) Glycans and the modulation of neural-recognition molecule function. *Trends Neurosci. 18*, 183–191.
5. Liedtke, S., Geyer, H., Wuhrer, M., Geyer, R., Frank, G., Gerardy-Schahn, R., Zahringer, U., and Schachner, M. (2001) Characterization of N-glycans from mouse brain neural cell adhesion molecule. *Glycobiology 11*, 373–384.
6. Rutishauser, U. (1996) Polysialic acid and the regulation of cell interactions. *Curr. Opin. Cell Biol. 8*, 679–684.
7. Kunemund, V., Jungalwala, F. B., Fischer, G., Chou, D. K., Keilhauer, G., and Schachner, M. (1988) The L2/HNK-1 carbohydrate of neural cell adhesion molecules is involved in cell interactions. *J. Cell Biol. 106*, 213–223.
8. Cho, T. M., Hasegawa, J., Ge, B. L., and Loh, H. H. (1986) Purification to apparent homogeneity of a μ-type opioid receptor from rat brain. *Proc. Natl. Acad. Sci. U.S.A. 83*, 4138–4142.
9. Funatsu, N., Miyata, S., Kumanogoh, H., Shigeta, M., Hamada, K., Endo, Y., Sokawa, Y., and Maekawa, S. (1999) Characterization of a novel rat brain glycosylphosphatidylinositol-anchored protein (Kilon), a member of the IgLON cell adhesion molecule family. *J. Biol. Chem. 274*, 8224–8230.
10. Levitt, P. (1984) A monoclonal antibody to limbic system neurons. *Science 223*, 299–301.
11. Pimenta, A. F., Zhukareva, V., Barbe, M. F., Reinoso, B. S., Grimley, C., Henzel, W., Fischer, I., and Levitt, P. (1995) The limbic system-associated membrane protein is an Ig superfamily member that mediates selective neuronal growth and axon targeting. *Neuron 15*, 287–297.
12. Schofield, P. R., McFarland, K. C., Hayflick, J. S., Wilcox, J. N., Cho, T. M., Roy, S., Lee, N. M., Loh, H. H., and Seeburg, P. H. (1989) Molecular characterization of a new immunoglobulin superfamily protein with potential roles in opioid binding and cell contact. *EMBO J. 8*, 489–495.
13. Struyk, A. F., Canoll, P. D., Wolfgang, M. J., Rosen, C. L., D'Eustachio, P., and Salzer, J. L. (1995) Cloning of neurotrimin

defines a new subfamily of differentially expressed neural cell adhesion molecules. *J. Neurosci. 15*, 2141–2156.
14. Brauer, A. U., Savaskan, N. E., Plaschke, M., Prehn, S., Ninnemann, O., and Nitsch, R. (2000) IG-molecule Kilon shows differential expression pattern from LAMP in the developing and adult rat hippocampus. *Hippocampus 10*, 632–644.
15. Gil, O. D., Zhang, L., Chen, S., Ren, Y. Q., Pimenta, A., Zanazzi, G., Hillman, D., Levitt, P., and Salzer, J. L. (2002) Complementary expression and heterophilic interactions between IgLON family members neurotrimin and LAMP. *J. Neurobiol. 51*, 190–204.
16. Hachisuka, A., Nakajima, O., Yamazaki, T., and Sawada, J. (2000) Developmental expression of opioid-binding cell adhesion molecule (OBCAM) in rat brain. *Brain Res. Dev. Brain Res. 122*, 183–191.
17. Miyata, S., Matsumoto, N., Taguchi, K., Akagi, A., Iino, T., Funatsu, N., and Maekawa, S. (2003) Biochemical and ultrastructural analyses of IgLON cell adhesion molecules, Kilon and OBCAM in the rat brain. *Neuroscience 117*, 645–658.
18. Zacco, A., Cooper, V., Chantler, P. D., Fisher-Hyland, S., Horton, H. L., and Levitt, P. (1990) Isolation, biochemical characterization and ultrastructural analysis of the limbic system-associated membrane protein (LAMP), a protein expressed by neurons comprising functional neural circuits. *J. Neurosci. 10*, 73–90.
19. Hachisuka, A., Yamazaki, T., Sawada, J., and Terao, T. (1996) Characterization and tissue distribution of opioid-binding cell adhesion molecule (OBCAM) using monoclonal antibodies. *Neurochem. Int. 28*, 373–379.
20. Wada, Y., Tajiri, M., and Yoshida, S. (2004) Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal. Chem. 76*, 6560–6565.
21. Wuhrer, M., Hokke, C. H., and Deelder, A. M. (2004) Glycopeptide analysis by matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry reveals novel features of horseradish peroxidase glycosylation. *Rapid Commun. Mass Spectrom. 18*, 1741–1748.
22. Satomi, Y., Shimonishi, Y., and Takao, T. (2004) N-Glycosylation at Asn(491) in the Asn-Xaa-Cys motif of human transferrin. *FEBS Lett. 576*, 51–56.
23. Zaia, J. (2004) Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev. 23*, 161–227.
24. Wuhrer, M., Catalina, M. I., Deelder, A. M., and Hokke, C. H. (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci 849*, 115–128.
25. Wuhrer, M., Koeleman, C. A., Hokke, C. H., and Deelder, A. M. (2005) Protein glycosylation analyzed by normal-phase nano-liquid chromatography-mass spectrometry of glycopeptides. *Anal. Chem. 77*, 886–894.
26. Harazono, A., Kawasaki, N., Kawanishi, T., and Hayakawa, T. (2005) Site-specific glycosylation analysis of human apolipoprotein B100 using LC/ESI MS/MS. *Glycobiology 15*, 447–462.
27. Sandra, K., Devreese, B., Van Beeumen, J., Stals, I., and Claeyssens, M. (2004) The Q-Trap mass spectrometer, a novel tool in the study of protein glycosylation. *J. Am. Soc. Mass Spectrom. 15*, 413–423.
28. Itoh, S., Kawasaki, N., Harazono, A., Hashii, N., Matsuishi, Y., Kawanishi, T., and Hayakawa, T. (2005) Characterization of a gel-separated unknown glycoprotein by liquid chromatography/multistage tandem mass spectrometry: Analysis of rat brain Thy-1 separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *J. Chromatogr., A 1094*, 105–117.
29. Bordier, C. (1981) Phase separation of integral membrane proteins in Triton X-114 solution. *J. Biol. Chem. 256*, 1604–1607.
30. Lisanti, M. P., Sargiacomo, M., Graeve, L., Saltiel, A. R., and Rodriguez-Boulan, E. (1988) Polarized apical distribution of glycosyl-phosphatidylinositol-anchored proteins in a renal epithelial cell line. *Proc. Natl. Acad. Sci. U.S.A. 85*, 9557–9561.
31. Itoh, S., Kawasaki, N., Ohta, M., and Hayakawa, T. (2002) Structural analysis of a glycoprotein by liquid chromatography-mass spectrometry and liquid chromatography with tandem mass spectrometry. Application to recombinant human thrombomodulin. *J. Chromatogr., A 978*, 141–152.
32. Kikuchi, M., Hatano, N., Yokota, S., Shimozawa, N., Imanaka, T., and Taniguchi, H. (2004) Proteomic analysis of rat liver peroxisome: Presence of peroxisome-specific isozyme of Lon protease. *J. Biol. Chem. 279*, 421–428.
33. Nakajima, O., Hachisuka, A., Takagi, K., Yamazaki, T., Ikebuchi, H., and Sawada, J. (1997) Expression of opioid-binding cell adhesion molecule (OBCAM) and neurotrimin (NTM) in *E. coli* and their reactivity with monoclonal anti-OBCAM antibody. *NeuroReport 8*, 3005–3008.
34. Horstkorte, R., Schachner, M., Magyar, J. P., Vorherr, T., and Schmitz, B. (1993) The fourth immunoglobulin-like domain of NCAM contains a carbohydrate recognition domain for oligomannosidic glycans implicated in association with L1 and neurite outgrowth. *J. Cell Biol. 121*, 1409–1421.
35. Chen, Y. J., Wing, D. R., Guile, G. R., Dwek, R. A., Harvey, D. J., and Zamze, S. (1998) Neutral N-glycans in adult rat brain tissue: Complete characterisation reveals fucosylated hybrid and complex structures. *Eur. J. Biochem. 251*, 691–703.
36. Nakakita, S., Natsuka, S., Ikenaka, K., and Hase, S. (1998) Development-dependent expression of complex-type sugar chains specific to mouse brain. *J. Biochem. 123*, 1164–1168.
37. Kitamura, N., Ikekita, M., Sato, T., Akimoto, Y., Hatanaka, Y., Kawakami, H., Inomata, M., and Furukawa, K. (2005) Mouse Na^+/K^+-ATPase β1-subunit has a K^+-dependent cell adhesion activity for β-GlcNAc-terminating glycans. *Proc. Natl. Acad. Sci. U.S.A. 102*, 2796–2801.

BI8009778

Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 343 (2008) 2329–2335
Carbohydrate RESEARCH

Note

Glycosylation and ligand-binding activities of rat plasma fibronectin during liver regeneration after partial hepatectomy

Kotone Sano,[a] Miho Asahi,[a] Maiko Yanagibashi,[a] Noritaka Hashii,[c] Satsuki Itoh,[c] Nana Kawasaki[c] and Haruko Ogawa[a,b,*]

[a]*Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan*
[b]*Glycoscience Institute, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan*
[c]*Division of Biological Chemistry and Biologicals, National Institute of Health Sciences, Tokyo 158-8501, Japan*

Received 10 September 2007; received in revised form 18 February 2008; accepted 19 March 2008
Available online 27 March 2008

Abstract—Fibronectin (FN) is a multifunctional glycoprotein present in the extracellular matrix (ECM) and plasma. We previously reported that the glycosylation and ligand-binding of vitronectin (VN) change markedly after partial hepatectomy (PH). Here we show the changes of FN during liver regeneration. The yields of purified sham-operated (SH-) and PH-FN were higher than that of non-operated (NO)-FN, while binding activities of FNs to ECM ligands were changed only slightly by hepatectomy. The carbohydrate concentration of PH-FN decreased to 66% of that of NO- and SH-FN. By using LC/MS^n, eight kinds of complex-type *N*-glycan structures were found to be present in all FNs, and bi- and trisialobiantennary glycans were the major structures. Fucosylation was markedly increased, while O-acetylation of sialic acid was found to be decreased in PH-FN. The alterations in glycosylation and biological activities of FN after PH are different from those of VN, suggesting that these glycoproteins play different biological functions in tissue remodeling.

Keywords: Fibronectin; Glycosylation; Liver regeneration; Vitronectin; Partial hepatectomy

Numerous biological phenomena are mediated by the recognition of specific oligosaccharide signals. Among the functions of protein glycosylation are the stabilizing active conformations, protecting against proteolysis, and affording solubility to proteins.[1] Clarification of the molecular mechanisms by which glycosylation plays these roles would enable the use of glycosylation in molecular engineering for therapeutic purposes.

Partial (70%) hepatectomy is often used to study liver regeneration mechanisms. The remaining liver recovers its former weight within about two weeks in humans or 10 days in the rat.[2] Matrix degradation occurs in the early stage of this process, followed by biosynthesis of the matrix, cell proliferation, and cell differentiation. Previously, collagen-binding activities of VN were found to be significantly affected by alterations in N-glycosylation.[3] A marked increase in VN was found to be controlled by the increased multimerization induced by glycosylation changes during liver regeneration after PH of rats.[4]

FN is a typical ECM glycoprotein synthesized by various cells, including hepatocytes, that plays critical roles in many biological and pathological processes including embryogenesis, wound healing, metastasis, fibrosis, and thrombosis.[5] It exists primarily as a soluble dimer in body fluids such as plasma at high concentrations (0.3 mg/mL in humans)[6] that is assembled into insoluble fibrils with a fibrillar polymeric matrix during matrix assembly.[7] In addition to regulation of the biological activities of FN by alternative splicing of the III-type modules or V segment,[8,9] N- and O-glycosylation of human FN affects the interaction with biological ligands including cell receptors,[10] suggesting that glycosylation

* Corresponding author. Address: Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan. Tel./fax: +81 3 5978 5343; e-mail: ogawa.haruko@ocha.ac.jp

0008-6215/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2008.03.027

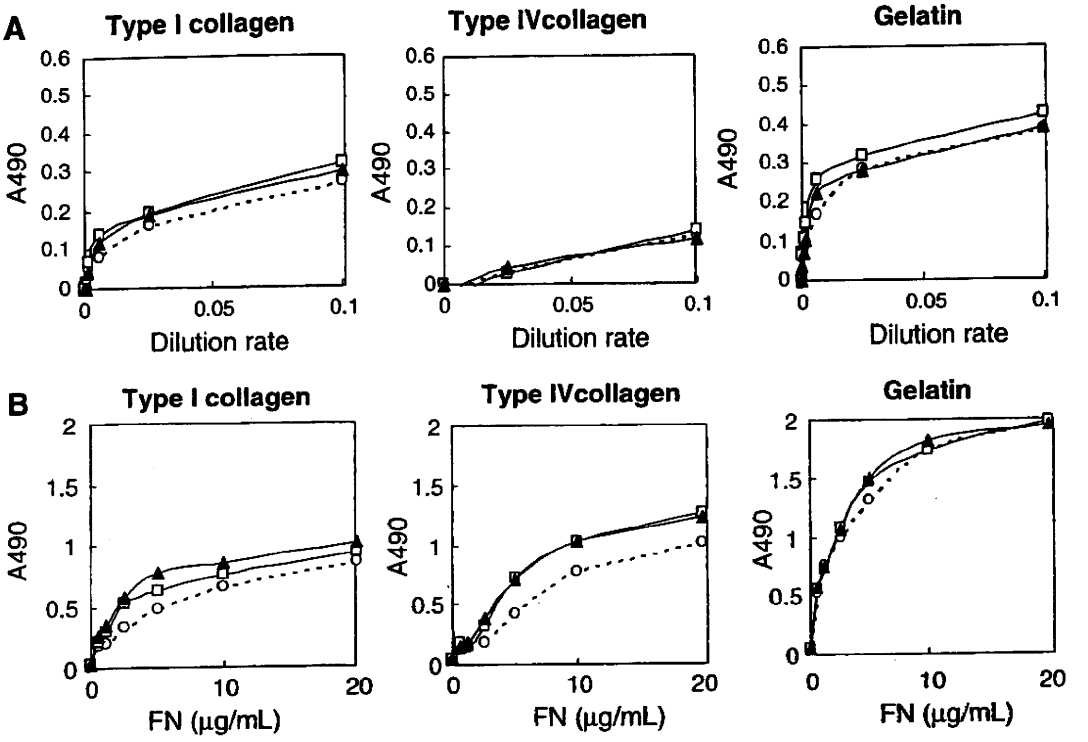


Figure 2. Binding activity of FNs in NO, SH, and PH plasma (A) and purified FNs (B) to collagen types I, and IV and gelatin by ELISA. (A) Binding of FNs in plasma. (B) Binding of purified FNs. Symbols used are circle, NO-FN; square, SH-FN; solid triangle, PH-FN.

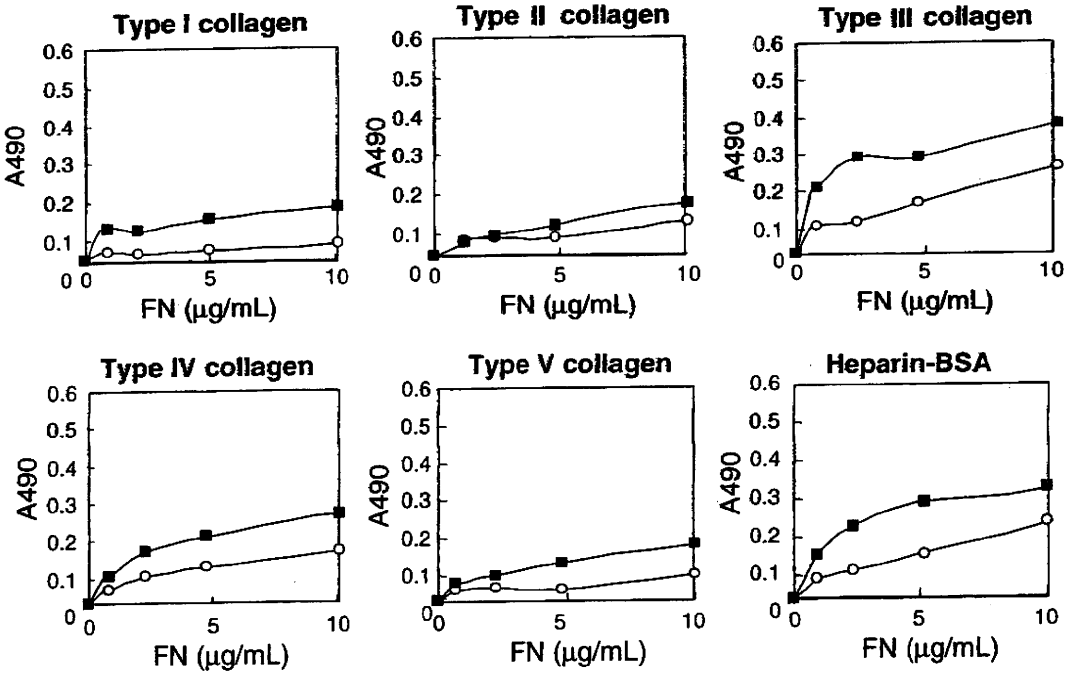


Figure 3. Effect of de-N-glycosylation on the binding activity of FN to collagen types I–V and heparin-BSA by ELISA. Symbols used are circle, control human FN treated without PNGase F; solid square, de-N-glycosylated human FN by PNGase F treatment.

lectin (PNA) after mild acid desialylation of FNs clearly indicated the presence of O-type glycans. The reactivity to *Psathyrella velutina* lectin (PVL) was increased in PH-FN but disappeared in all FNs after desialylation, indicating that sialylated glycans having more than trisialyl residues[15] increased after PH.

Table 1. Reactivity of various lectins with FNs

	NO-FN	SH-FN	PH-FN	Specificity
Con A	++	++	++	N-Linked biantennary or high Man-type
SNA	++	++	++	Sialyl α2-6 Gal
MAM	+	+	++	Sialyl α2-3 Gal
L-PHA	+++	+++	+++	N-Linked tri- or tetraantennary
E-PHA	++	++	+++	N-Linked bi- or triantennary
	(+++)	(++)	(+)	
AAL	++	+	+++	Core or outer fucosylated
LCA	++	++	++	N-Linked, core-fucosylated
PNA	−	−	−	O-Linked, unsialylated Galβ1-3GalNAc
	[++]	[+++]	[+++]	
PVL	++	++	+++	Non-reducing terminal GlcNAc or tri- or tetra sialyl
	[−]	[−]	[−]	
RCA	+++	+++	+++	Non-reducing terminal Galβ
WGA	++	++	++	GlcNAc

Reactivity of FNs with biotinyl lectins by dot blotting is expressed as staining intensity. +, indicates positive staining; −, negative staining; Reactivity after digestion with hexosaminidase is indicated in parentheses, and reactivity after mild acid desialylation is indicated in brackets. Abbreviations used are: Con A, Concanavalin A; SNA, *Sambucus nigra* agglutinin; MAM, *Maackia amurensis* mitogen; L-PHA, *Phaseolus vulgaris* leukoagglutinin; E-PHA, *Phaseolus vulgaris* erythroagglutinin; AAL, *Aleuria aurantia* lectin; LCA, *Lens culinaris* agglutinin; PNA, peanut agglutinin; PVL, *Psathyrella velutina* lectin; RCA, *Ricinus communis* agglutinin; WGA, wheat germ agglutinin.

As shown in Table 2, the total carbohydrate concentration of PH-FN was markedly decreased compared with those of NO- and SH-FN. The decrease in the ratio of the concentrations of GalNAc to total carbohydrates was notable in PH-FN, indicating that FN is highly O-glycosylated, but that the ratio is lowered in PH-FN. The Fuc concentration increased to about twice that in NO- and SH-FN, in contrast to plasma VN, in which the relative molar ratio of fucose was not changed by PH, although both of these glycoproteins are synthesized in the liver. The isoelectric points of NO-, SH-, and PH-FN, as determined by 2D-PAGE, showed no difference ($pI = 5.7$), suggesting that these FNs are sialylated to the same degree (data not shown).

The *N*-glycan structures of FN were determined by LC/MS to be the sugar alcohols. Rat plasma FN has seven potential N-glycosylation sites. As shown in Table 3 and Figure 4A, the most frequent *N*-glycan structures of rat plasma FN were biantennary complex type disialoglycans (BiNA(2)), which are very similar to human plasma FN,[16] and then biantennary trisialoglycans (BiNA(3)). As summarized in Table 3, eight kinds of *N*-glycan backbones were found. If variations of the number of O-acetylations of sialic acid residues were included, the varieties of structures would total more than 15 (Fig. 4). The increased ratios of fucosylation in PH-FN were in accordance with the carbohydrate composition results and reactivity with AAL. Because fucosylated glycans are known to participate in embryonic growth, differentiation, cell recognition, cancer formation, and inflammation,[17] the glycans of FN synthesized during liver regeneration may regulate tissue remodeling by increasing fucosylation through a change in binding to other ligands.

Table 2. Carbohydrate composition of rat FNs

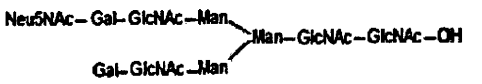
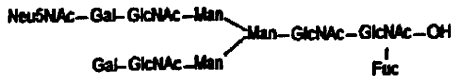
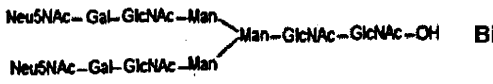
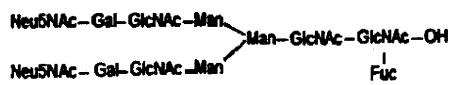
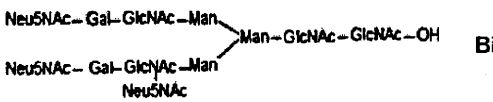
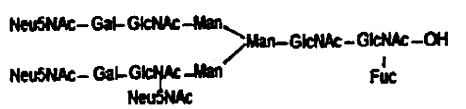
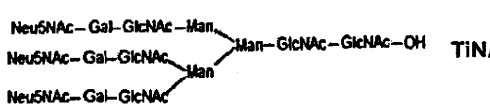
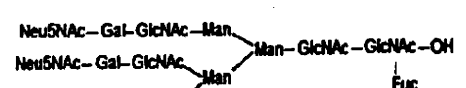
Carbohydrate	NO-FN	SH-FN	PH-FN
GalNAc	92 (9.3)	61 (5.3)	33 (4.1)
GlcNAc	33 (3.4)	37 (3.2)	21 (2.6)
Man	30 (3.0)	35 (3.0)	24 (3.0)
Fuc	1.9 (0.2)	2.5 (0.2)	4.5 (0.6)
Gal	37 (3.7)	57 (5.0)	44 (5.5)
Total	194	193	127

Values are expressed as mol/mol FN. Relative molar ratios of each carbohydrate are expressed in parentheses by taking the value of Man as 3.0.

The glycans contained various concentrations of O-acetylated Neu5NAc (Fig. 4B). The ratio of non-O-acetylated glycans in each backbone structure was higher in PH-FN than that in NO- and SH-FN. For example, Ac(0) in BiNA(2), a major glycan, was 70% in PH-FN, much higher than those of NO- (52%) and SH-FN (57%). This indicates that the O-acetylation of neuraminic acid is lowered after PH. Sialic acids from total membranes of rat liver have surprisingly high levels (approximately 20%) of O-acetylation at the 7- or 9-position.[18] These modifications modulate many biological interactions,[19–21] such as that the binding of Siglec-1 (sialoadhesin) and Siglec-2 (CD22) to glycoconjugates, which are hindered by 9-O-acetylation.[22,23] In the early stage of liver regeneration, the immune system may be regulated by the interaction of FN glycans with Siglecs via the change in O-acetylation.

These results indicate that the alterations in ligand-binding and glycosylation of FN and VN were significantly different in the early stage of liver regeneration and demonstrate that these glycoproteins play different biological roles in the promotion of tissue remodeling processes. This study provides insight into the regulation of individual ECM glycoproteins by glycosylation.

Table 3. Oligosaccharide structures of rat FN

Glycan structures	Abbreviation
	BiNA(1)
	BiNA(1)F
	BiNA(2)
	BiNA(2)F
	BiNA(3)
	BiNA(3)F
	TiNA(3)
	TriNA(3)F

Abbreviations used for glycan backbone structures are: BiNA(1), biantennary glycan with one *N*-acetylneuraminic acid; BiNA(1)F, biantennary glycan with one *N*-acetylneuraminic acid and a fucose linked to a penultimate GlcNAc; BiNA(2), biantennary glycan with two *N*-acetylneuraminic acids; BiNA(2)F, biantennary glycan with two *N*-acetylneuraminic acids and a fucose linked to a penultimate GlcNAc; BiNA(3), biantennary glycan with three *N*-acetylneuraminic acids; BiNA(3)F, biantennary glycan with three *N*-acetylneuraminic acids and a fucose linked to a penultimate GlcNAc; TiNA(3), triantennary glycan with three *N*-acetylneuraminic acids; and TiNA(3)F, triantennary glycan with three *N*-acetylneuraminic acids and a fucose linked to a penultimate GlcNAc.

1. Experimental

1.1. Animals and purification of rat FN from plasma

Male Wistar rats aged 5 weeks (weighing about 110 g; Nihon Clea, Tokyo, Japan) were subjected to two-thirds PH or SH as described previously.[3,24] Plasma was collected from NO rats, or 6, 24, 48, 120, and 168 h after SH and PH and stored at −80 °C until use. FN was purified from PH-, SH-, or NO-plasma using gelatin-Sepharose as described previously.[25] The purity of FNs was checked by SDS-PAGE.[10]

1.2. Determination of FN concentration in plasma (sandwich ELISA)

Wells of microtiter plates (Immulon 1, Dynatech Laboratories Inc., Chantilly, VA) were coated with rabbit anti-human FN IgGs (50 μL in 0.5% BSA/PBS) at 4 °C overnight. The wells were washed with TBS, and various concentrations of rat plasma or purified FN (50 μL) were added and incubated for 2 h at room temperature. After washing, the wells were blocked with 0.5% BSA/PBS for 1 h, then bound FN was detected with horseradish peroxidase (HRP)-anti-FN IgGs and visualized with *o*-phenylenediamine/0.007% H_2O_2. After stopping the reaction with 4 M H_2SO_4, absorbance was measured at 490 nm using a microplate reader.

1.3. Assays of FN binding to immobilized ligands (ELISA)

Wells of microtiter plates were coated with gelatin, heparin-BSA, or collagen types I–V (10 μg/mL, 50 μL) in 0.1 M carbonate buffer (pH 9.0), and reacted with various concentrations of plasma or FN. The bound FN was detected by ELISA.[26,27]

1.4. Carbohydrate analyses of FNs

For lectin reactivity, FNs (0.5 μg) were dot-blotted onto a PVDF membrane and reacted with HRP-lectins as previously described.[28] The staining intensities were measured by the software program Scion Image. To determine carbohydrate composition, FNs (4.5 μg) were hydrolyzed, and the carbohydrates were analyzed according to the previously described method.[3]

1.5. Analysis of *N*-glycan structures by mass spectrometry (LC/MS^n)

FN (30 μg/about 200 μL) was reduced by boiling with 2% of 2-mercaptoethanol for 1 min. N-Linked oligosaccharides were released from the FN with *N*-glycosidase F (4.8 U) and reduced with 0.25 M $NaBH_4$. The oligosaccharides were desalted with ENVI-Carb C (Supelco Co., Ltd, Bellefonte, PA, USA) and subjected to LC/MS^n. The eluents were 2% acetonitrile/5 mM ammonium acetate (pH 9.6) (pump A), and 80% acetonitrile/5 mM ammonium acetate (pH 9.6) (pump B). The oligosaccharides were separated on a Hypercarb column (0.2 × 150 mm, Thermo Fisher Scientific, Waltham, MA, USA) with a 5–30% linear gradient of the pump B eluent over 60 min at a flow rate of 2 μL/ min. The MS spectra were acquired by using an LTQ-FT system (Thermo Fisher Scientific) with a single mass scan (*m/z* 7000–2000) on a Fourier transformation ion cyclotron resonance mass spectrometer, and data-dependent

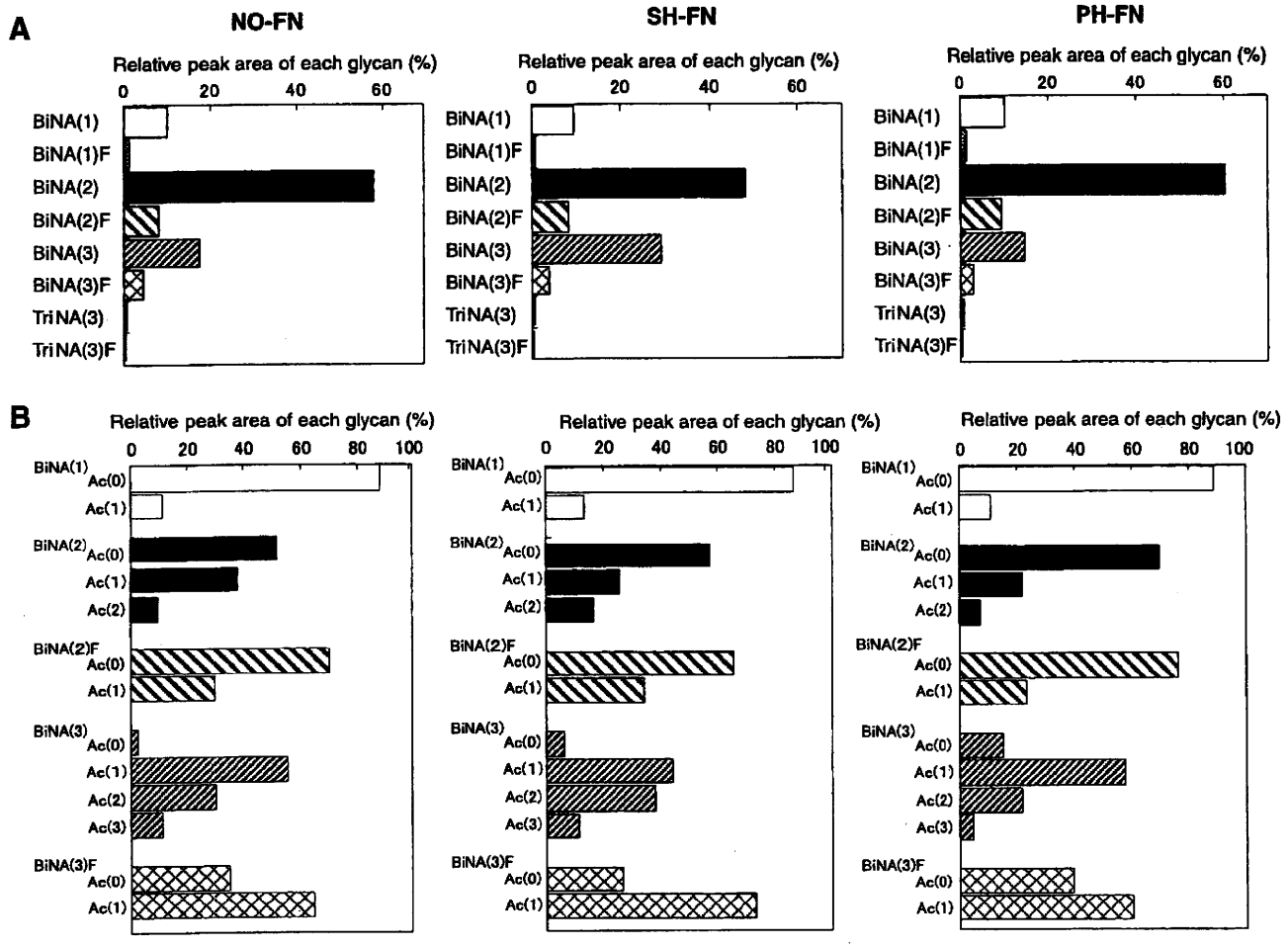


Figure 4. Ratios of each FN *N*-glycan. (A) The ratio of each glycan structure. (B) Ratios of glycans possessing various numbers of *O*-acetyl groups at the *N*-acetylneuraminic acid residues. The ratio of each glycan (the abbreviations are summarized in Table 3) was calculated from the relative peak area of mass chromatography signals in HPLC by taking the total peak area of each FN (A) or of each glycan (B) as 100%. Of the structures not expressed in (B), BiNA(1)F, TriNA(3), and TriNA(3)F, all without *O*-acetyl group (Ac(0)) were detected.

MS/MS and MS/MS/MS scans in both positive and negative ion modes.

Acknowledgments

This work was supported in part by a Grant from JSPS (12995 to KS) and a fellowship from Hayashi Memorial Foundation for Female Natural Scientists (KS) and Grants-in-aid for Scientific Research on Priority Areas 15040209 and 17046004 (HO) from the Ministry of Education, Culture, Sports, Science, and Technology. We thank our laboratory members for help in the operations and K. Ono for editing the English language of this article.

References

1. Varki, A. *Glycobiology* **1993**, *3*, 97–130.
2. Diehl, A. M.; Rai, R. M. *FASEB J.* **1996**, *10*, 215–227.
3. Uchibori-Iwaki, H.; Yoneda, A.; Oda-Tamai, S.; Kato, S.; Akamatsu, N.; Otsuka, M.; Murase, K.; Kojima, K.; Suzuki, R.; Maeya, Y.; Tanabe, M.; Ogawa, H. *Glycobiology* **2000**, *10*, 865–874.
4. Sano, K.; Asanuma-Date, K.; Arisaka, F.; Hattori, S.; Ogawa, H. *Glycobiology* **2007**, *17*, 784–794.
5. Potts, J. R.; Campbell, I. D. *Matrix Biol.* **1996**, *15*, 313–320, Discussion 321.
6. Dean, D. C.; Birkenmeier, T. M.; Rosen, G. D.; Weintraub, S. J. *Am. Rev. Respir. Dis.* **1991**, *144*, S25–S28.
7. McDonald, J. A. *Annu. Rev. Cell Biol.* **1988**, *4*, 183–207.
8. Chernousov, M. A.; Fogerty, F. J.; Koteliansky, V. E.; Mosher, D. F. *J. Biol. Chem.* **1991**, *266*, 10851–10858.
9. Brown, L. F.; Dubin, D.; Lavigne, L.; Logan, B.; Dvorak, H. F.; Van de Water, L. *Am. J. Pathol.* **1993**, *142*, 793–801.
10. Zheng, M.; Hakomori, S. *Arch. Biochem. Biophys.* **2000**, *374*, 93–99.
11. Singh, P.; Reimer, C. L.; Peters, J. H.; Stepp, M. A.; Hynes, R. O.; Van De Water, L. *J. Invest. Dermatol.* **2004**, *123*, 1176–1181.
12. Regnault, V.; Rivat, C.; Maugras, M.; Stoltz, J. F. *Rev. Fr. Transfus. Immunohematol.* **1988**, *31*, 19–34.
13. Chifflet, S.; Bolatto, C.; Tolosa, S. *J. Biochem. Biophys. Methods* **2004**, *59*, 139–143.

DNA Research **16**, 73–80, (2009) doi:10.1093/dnares/dsn035

Defining Developmental Potency and Cell Lineage Trajectories by Expression Profiling of Differentiating Mouse Embryonic Stem Cells

Kazuhiro Aiba[1,†,‡], Timur Nedorezov[1,†], Yulan Piao[1], Akira Nishiyama[1], Ryo Matoba[1,§], Lioudmila V. Sharova[1], Alexei A. Sharov[1], Shinya Yamanaka[2], Hitoshi Niwa[3], and Minoru S. H. Ko[1,*]

Developmental Genomics and Aging Section, Laboratory of Genetics, National Institute on Aging, NIH, Baltimore, MD 21224, USA[1]; Department of Stem Cell Biology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Kyoto 332-0012, Japan[2] and Laboratory for Pluripotent Cell Studies, RIKEN Center for Developmental Biology, Kobe, Hyogo 650-0047, Japan[3]

(Received 1 December 2008; accepted 11 December 2008; published online 26 December 2008)

Abstract

Biologists rely on morphology, function and specific markers to define the differentiation status of cells. Transcript profiling has expanded the repertoire of these markers by providing the snapshot of cellular status that reflects the activity of all genes. However, such data have been used only to assess relative similarities and differences of these cells. Here we show that principal component analysis of global gene expression profiles map cells in multidimensional transcript profile space and the positions of differentiating cells progress in a stepwise manner along trajectories starting from undifferentiated embryonic stem (ES) cells located in the apex. We present three 'cell lineage trajectories', which represent the differentiation of ES cells into the first three lineages in mammalian development: primitive endoderm, trophoblast and primitive ectoderm/neural ectoderm. The positions of the cells along these trajectories seem to reflect the developmental potency of cells and can be used as a scale for the potential of cells. Indeed, we show that embryonic germ cells and induced pluripotent cells are mapped near the origin of the trajectories, whereas mouse embryo fibroblast and fibroblast cell lines are mapped near the far end of the trajectories. We suggest that this method can be used as the non-operational semi-quantitative definition of cell differentiation status and developmental potency. Furthermore, the global expression profiles of cell lineages provide a framework for the future study of *in vitro* and *in vivo* cell differentiation.

Keywords: embryonic stem; embryonic germ; induced pluripotent stem; mouse embryo fibroblast; embryonal carcinoma; retinoic acids; neural stem/progenitor; trophoblast stem; principal component analysis; leukemia inhibitory factor; epigenetic landscape; Waddington; developmental potency; cell lineage trajectory; gene expression profiling; DNA microarray analysis

1. Introduction

Developmental biologists have long held a view that development naturally progresses from totipotent fertilized eggs with unlimited differentiation potential to terminally differentiated cells, like a ball rolling from high to low points on a slope as depicted in Waddington's epigenetic landscape.[1] The epigenetic landscape also points to another important aspect

Edited by Osamu Ohara

* To whom correspondence should be addressed. Tel. +1 410-558-8359. Fax. +1 410-558-8331. E-mail: kom@mail.nih.gov

† These authors contributed equally to this work.

‡ Present address: Stem Cell and Drug Discovery Institute, Kyoto Research Park, Kyoto, Kyoto 600-8813, Japan.

§ Present address: DNA Chip Research Inc., Yokohama, Kanagawa 230-0045, Japan.

of development, which is the emergence of different cell lineages during the development; cells in specific developmental lineages are thought to take discrete paths ('chreodes') on the imaginary slope. Analogy of cell's developmental potency to the potential energy is thus widely accepted[2,3] This analogy is also relevant to the fact that converting differentiated cells into pluripotent cells is difficult. In mammals, nuclear transplantation (cloning)[4,5] had been the only way to achieve such a 'up-hill battle' reprogramming, until the successful production of induced pluripotent stem (iPS) cells by infecting MEFs with retroviruses carrying expression cassettes for four genes (*Myc*, *Pou5f1*, *Sox2* and *Klf4*).[6] Notwithsanding its importance, the potency has only been defined operationally by *in vitro* and *in vivo* cell differentiation assays as 'the total of all fates of a cell or tissue region which can be achieved by any environmental manipulation'.[7] Nuclear transplantation experiments, where the success rate gradually decreases according to developmental stages of donor cells, provide yet another operational definition of developmental potential.[8–10] We previously showed a possibility to derive a scale of developmental potency from the global gene expression (transcript) profile data, but the data could not be that quantitative because of the use of a limited number of expressed sequence tags (ESTs) for the analysis.[11] The work also did not address the issue of cell linege separations.

Mouse embryonic stem (ES) cells[12,13] and embryonic germ (EG) cells[14,15] are prototypical stem cells. These cells can be maintained as undifferentiated state in culture (self-renewal) and have the capacity to differentiate into essentially all the cell types (pluripotency). Therefore, these pluripotent stem cells provide tractable systems to study the developmental potency and cell lineage separation. It has been shown that the manipulation of cell culture condition or a single-gene expression level can differentiate ES cells into relatively homogenous cell population that are similar to the first three lineages in mammalian development:[16] primitive ectoderm/neural ectoderm,[17,18] trophoblast[19,20] and primitive endoderm.[21] In the first system, ES cells are cultured in monolayer in N2B27 medium, which drives undifferentiated ES cells into neural lineages.[17] Previous DNA microarray analysis indicates that this *in vitro* ES cell differentiation process mimics *in vivo* cell differentiation to primitive ectoderm, neural ectoderm and subsequently neurons/glia cells.[18] In the second system, ES cells are engineered to downregulate *Pou5f1* (*Oct3/4*, *Oct4*) expression in a tetracycline-controllable manner (ZHBTc4 cell line[19]). It has been shown that repression of *Pou5f1* induces the differentiation of ES cells into trophoblast lineage.[19,20] In the third system, ES cells that are engineered to overexpress *Gata6* in a dexamethasone-inducible manner differentiate into primitive endoderm (extraembryonic endoderm).[21] Although the analyses of these ES cell differentiation systems have revealed the detailed changes of gene expression patterns, it remains to see whether the global comparison among these individual systems provide any further insights into developmental potency and cell lineage separation.

Here we show that principal component analysis (PCA), which can reduce the dimensionality of the gene expression profiles,[22] maps cells in a multidimensional transcript profile space where the positions of differentiating cells progress in a stepwise manner along trajectories starting from undifferentiated ES cells located in the apex to the first three lineages in mammalian development: primitive endoderm, trophoblast and primitive ectoderm/neural ectoderm. Furthermore, EG cells and iPS cells are mapped near the origin of the trajectories, whereas mouse embryo fibroblast (MEF) and fibroblast cell lines are mapped near the far end of the trajectories.

2. Materials and methods

2.1 Cells and RNAs

For the majority of cells used in this study, we used a stock of Cy3-labeled cRNA samples that were used in our previous studies. The details of each cell types, their culture conditions, RNA extractions and Cy3-labeling can be found in the main text of this manuscript and in earlier publications.[18,20,23–25] Cells cultured for this study and the culture conditions are as follows. G0–G5 cells: Production and characterization of 5G6GR ES cell clones that are engineered to overexpress Gata6 in a dexamethasone-inducible manner are described previously.[21] ES cells were harvested every 24 h (Day 0–5) during differentiation in the presence of 100 mM dexamethasone (Sigma). F0–F5 cells: Undifferentiated F9 EC cells (ATCC number: CRL-1720) were treated with 100 nM all-*trans*-retinoic acid (RA, Sigma) and 1 mM dibutyryl cAMP (dbcAMP, Sigma) on adherent condition as reported.[26] Photos of F9 EC cells during differentiation are available as Supplementary Fig. S1 of the published paper.[27] F9 cells were harvested every 24 h (Day 0–5) during the differentiation. Both iPS-Fbxo15[6] and iPS-Nanog[28] were cultured on the STO feeder cells as described previously. To remove the feeder cells, the iPS cells were passaged twice on the gelatin-coated culture dish and harvested for RNA extraction. NIH3T3, STO, MEF_BL6 and MEF_DR4 were cultured under the standard condition and harvested for RNA extraction. Total RNAs isolated for this

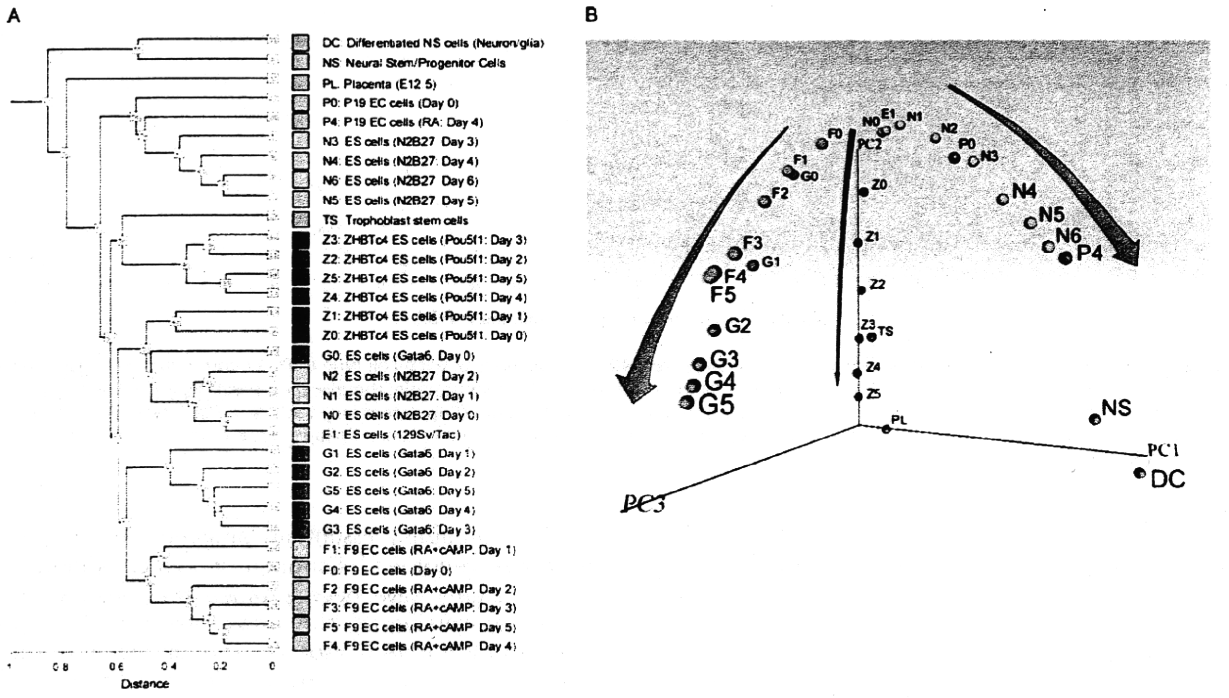


Figure 1. Global gene expression profiles of 32 different cell types. Names of the cells and their abbreviated forms are shown. Color coding corresponds with the colors in all other figures. DNA microarray analysis was carried out in duplicate: for each cell type, two independent cell cultures were prepared and used for separate DNA microarray analysis. Reproducibility between two microarray data for each cell type was very high as shown by the Pearson correlation coefficient (mean 0.9927; SD 0.0043; range 0.9788–0.9966). The average of two samples is shown here. (**A**) Hierarchical clustering analysis. (**B**) Principal component analysis (PCA). Individual cell types are mapped in the 3D space according to the first three principal components (PC1, PC2 and PC3). A movie file is available as Supplementary Video S1.

study were labeled with Cy3-dye and used for the DNA microarray hybridization.

2.2 *Microarray data analysis*

DNA microarray analysis was carried out as described previously,[18] except for the addition of ES cell total RNAs to the Universal Mouse Reference RNA (UMRR) and the use of a 4 × 44K microarray platform. In our previous DNA microarray studies,[23] we used the Universal Mouse Reference RNA (UMRR: Stratagene), which is a mixture of total RNAs from 11 different mouse cell lines. However, to increase the representation of genes expressed in ES cells for the current study, we mixed the UMRR with total RNAs from ES cells (MC1, derived from 129S6/SvEvTac strain) cultured in the undifferentiated condition with LIF at 2:1 ratio. These UMRR plus ES RNAs were labeled with Cy5-dye, mixed with Cy3-labeled samples and used for DNA microarray hybridization. To maximize the uniformity of the microarray data, all the samples, including the ones analyzed by DNA microarray previously, were hybridized to the same platform (the NIA Mouse 44K Microarray v3.0[23] manufactured by Agilent Technologies #015 087). The intensity of each gene feature per array was extracted from scanned microarray images using Feature Extraction *9.5.1.1* software (Agilent Technologies) as described previously.[23] Hierarchical clustering analysis and PCA (see Section 2.3) were carried out using an application developed in-house to perform ANOVA and other analyses (NIA Array Analysis software; http://lgsun.grc.nia.nih.gov/ANOVA/).[22] Three-dimensional PCA figures were generated using the virtual reality modeling language (VRML) function of the NIA Array Analysis software[22] and visualized by Cortona Vrml client (http://www.parallelgraphics.com/products/cortona/). To produce plots and lists of genes for Fig. 2, we used the R-statistical package[29] and only non-redundant set of oligonucleotide probes (a total of 25 164 probes/genes). Of these, 21 890 genes were significant in ANOVA with FDR < 0.05 and used for further analyses. Genes correlated to each cell lineage trajectory were identified based on the correlation of >0.90. All the DNA microarray data have been deposited in the NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible

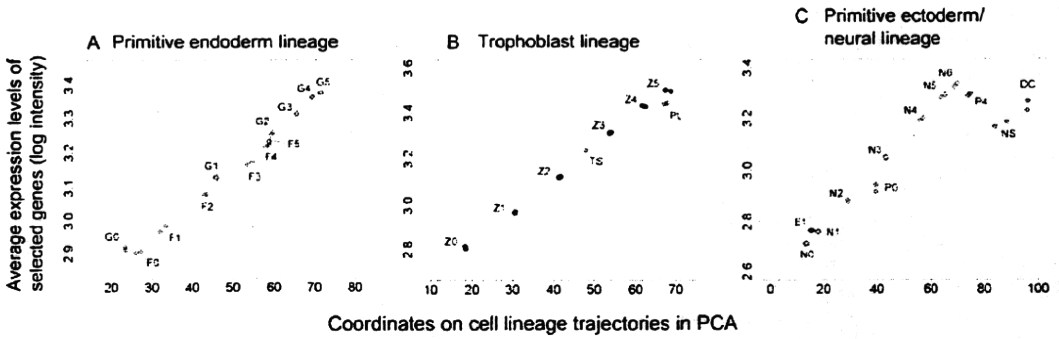


Figure 2. Average expression levels of selected genes correlate well with the locations of each cell type in PCA coordinates. See Fig. 1A for the abbreviated names of the cells. Each point shows DNA microarray data for each sample; therefore, each cell type is represented by two points. (**A**) Primitive endoderm lineage, represented by the list of 1738 genes (Supplementary Table S1). (**B**) Trophoblast lineage, represented by the list of 2311 genes (Supplementary Table S2). (**C**) Primitive ectoderm/neural lineage, represented by the list of 2463 genes (Supplementary Table S3).

through GEO Series accession number (GSE11 523) and the NIA Array Analysis software website (http://lgsun.grc.nia.nih.gov/ANOVA/).[22]

2.3 *Principal component analysis*

Principal component analysis is a statistical method to find major patterns in data variation, which is increasingly used for the analysis of gene expression microarray data.[22,30–32] Each principal component (PC) is a linear combination of log-transformed expression values of all genes, and all components are orthogonal, i.e. mutually independent. The first few PCs, which explain most of the observed variance, are the most important; the remaining PCs often represent random fluctuations. Therefore, by plotting data against the first two or three PCs, one can reduce the dimensionality of the data without losing much of information. Contribution of genes to each PC (which may be positive or negative) is determined by singular value decomposition algorithm applied to the covariance matrix. Knowing these contributions, the entire gene expression profile of a given cell type can be represented by a single point in a two- or three-dimensional space. If two cell types are represented by closely located points in the PCA plot, global gene expression profiles of these cells are very similar. In contrast, if corresponding points in the PCA plot are far apart, their global expression profiles are very different. The detailed discussion of PCA plots and interpretation of DNA microarray data can be found in the previous publication.[32]

3. Results and discussion

To obtain accurate, comprehensive and comparable expression profiles, we carried out all the DNA microarray analyses within a few weeks using the same platform containing essentially all genes encoded on the mouse genome.[23] Although two ES cell differentiation systems have been profiled previously using the earlier version of the array platform,[18,20] we carried out the DNA microarray analysis again for the data consistency. In the first system, ES cells are cultured in monolayer in N2B27 medium for 6 days, which drives undifferentiated ES cells into neural lineages.[17] RNAs were isolated and analyzed every day during the induction and RNAs (N0–N6: Fig. 1A). Previous DNA microarray analysis has been carried out using the array platform containing a limited number of genes, but indicates that this *in vitro* ES cell differentiation process mimics *in vivo* cell differentiation to primitive ectoderm, neural ectoderm and, subsequently, neurons/glia cells.[18] In the second system, ES cells that are engineered to downregulate *Pou5f1* (*Oct3/4*, *Oct4*) expression in a tetracycline-controllable manner (ZHBTc4 cell line differentiate into trophoblast lineage for 5 days).[19,20] RNAs were isolated and analyzed every day during the induction (Z0–Z5 in Fig. 1A). To include the first three lineages of mammalian development,[16] we added a third system: ES cells that are engineered to overexpress *Gata6* in a dexamethasone-inducible manner differentiate into primitive endoderm (extra-embryonic endoderm) for 5 days[21] (G0–G5 in Fig. 1A). We also carried out microarray analysis of a fourth system: F9 embryonal carcinoma (EC) cells differentiate into parietal endoderm—one of the extra-embryonic endoderm lineages for 5 days[26,27] (F0–F5 in Fig. 1A). For comparison, we also analyzed P1 9 EC cells before (P0) and after 4 days of RA induction (P4); neural stem/progenitor (NS) cells derived from adult mouse brain and their differentiated cells[21] (DC); and trophoblast stem (TS) cells[26] and E12.5 placenta (PL).[27] Altogether we obtained the DNA microarray data of 32 different cell types in duplicate.

We first carried out the hierarchical clustering analysis of these microarray data. Although the hierarchical clustering analysis grouped cells mostly into each differentiation system (Fig. 1A), it showed only the degree of similarity between different cell types and did not provide any insights into the relationship between the different cell lineages. To further analyze the data, we employed PCA, which can reduce the dimensionality of the gene expression profiles and map individual cells in a multidimensional space according to their global gene expression patterns (see Section 2.3 for the explanation of PCA). We first mapped the cells in the 3D transcript profile space (PC1, PC2 and PC3) (Fig. 1B). PCA of all these microarray data uncovered three major paths of differentiation from undifferentiated ES cells in 3D transcript profile space (Fig. 1B; Supplementary Video S1). Each path showed progressive changes in global expression patterns toward specific cell lineages: trophectoderm (Z0–Z5), extra-embryonic endoderm (G0–G5; F0–F5) and neural tissues (N0–N6). We call these paths 'cell lineage trajectories,[32]' because the cell differentiation process can be visualized as a trajectory in the multidimensional transcript profile space.

What do these trajectories represent? Because the transcript profile space is constructed from the expression patterns of genes, the changes in gene expression patterns are reflected in the changes in the position of a cell. To find gene expression changes that are correlated with these trajectories, we first drew a line for each cell lineage so that the squared distances from the cells of the lineage were minimized. The locations of other cell types, such as TS and NS, were projected onto the closest lineage trajectory (Fig. 1B). The *x*-axis of Fig. 2 shows the coordinates of these cell types on each trajectory (Fig. 2A–C). We then searched for sets of genes that were positively correlated with each trajectory line. We identified 1738 genes for the primitive endoderm lineage (Fig. 2A; Supplementary Table S1), 2311 genes for the trophoblast lineage (Fig. 2B; Supplementary Table S2) and 2463 genes for the primitive ectoderm/neural lineage (Fig. 2C; Supplementary Table S3) with a correlation of >0.90. The expression levels of these genes gradually increased as ES cells differentiated into each lineage (see Supplementary data for examples of gene expression changes). The average expression levels of these genes (*y*-axis in Fig. 2: cell lineage index[32]) can thus be used as surrogates for the coordinates of cell's positions on each trajectory (*x*-axis in Fig. 2). This lineage trajectory, therefore, not only defines a path to the fates of individual cell lineages, but also provides a scale for the extent of commitment/differentiation.

This notion was further supported by examining the locations of cells cultured and profiled independently in the transcript profile space: P19 EC cells before (P0) and after 4 days of RA induction (P4); NS cells derived from adult mouse brain and their differentiated cells[18] (DC); and TS cells[25,33] and E12.5 PL.[25] These cells were mapped onto the transcript profile space based on their global gene expression profiles. As we expected, we found that P0, P4, NS and DC mapped near the primitive ectoderm/neural lineage trajectory (Figs 1B and 2C). Interestingly, the position of P0 was consistent with the notion that in contrast to ES cells, undifferentiated P19 EC cells are more committed and equivalent to primitive ectoderm cells.[18] The position of P4 was also consistent with the notion that P19 EC cells differentiate into neuron-like cells after the induction with RA.[34] Consistent with their known differentiation status, TS and PL mapped near the trophoblast lineage trajectory (Figs 1B and 2B). Two extra-embryonic endoderm lineages, which were represented by F9 EC cell differentiation[26,35] (F0–F5) and ES cells driven by Gata6[21] (G0–G5), also showed lineage trajectories close to each other (Figs 1B and 2A). Taken together, these data support the notion that cell differentiation can be visualized as cell lineage trajectories in 3D transcript profile space.

We also noticed that these cell lineage trajectories represent the gradual loss of developmental potency or potential of cells (Fig. 1B). To examine this point further, we carried out microarray analysis of 12 additional cell types, including three undifferentiated ES cell lines (MC1, derived from 129S6/SvEvTac; MC2, derived from C57BL/6J; D3_GL; and EBRTcH3) and two undifferentiated EG cell lines [TGC8.5_5, derived from the primordial germ cell (PGC) of E8.5 C57BL/6J embryo; and EG1, derived from the PGC of E9 129 embryo].[24] We also analyzed the expression profiles of MEF cells, stromal cells (STO) and a fibroblast cell line (NIH3T3). We then analyzed iPS cells, which have been established by infecting MEFs with retroviruses carrying expression cassettes for four genes (*Myc*, *Pou5f1*, *Sox2* and *Klf4*).[6,28] We used two different iPS lines: one was established by using Fbxo15 promoter-driven GFP as an indicator of pluripotency[6] and the other was established by using Nanog promoter-GFP as an indicator of pluripotency.[28]

Hierarchical clustering analysis of the microarray data of all 44 cell types (original 32 cell types and 12 additional cell types; see Supplementary data for detail) showed that undifferentiated ES and EG cells were clustered together with iPS cells, whereas NIH3T3 fibroblast cells, STO stromal cells and MEF cells were clustered together (Fig. 3A). However, it was difficult to assess the overall relationship among all the cell types analyzed here. We, therefore, analyzed the microarray data of all 44 cell types by PCA

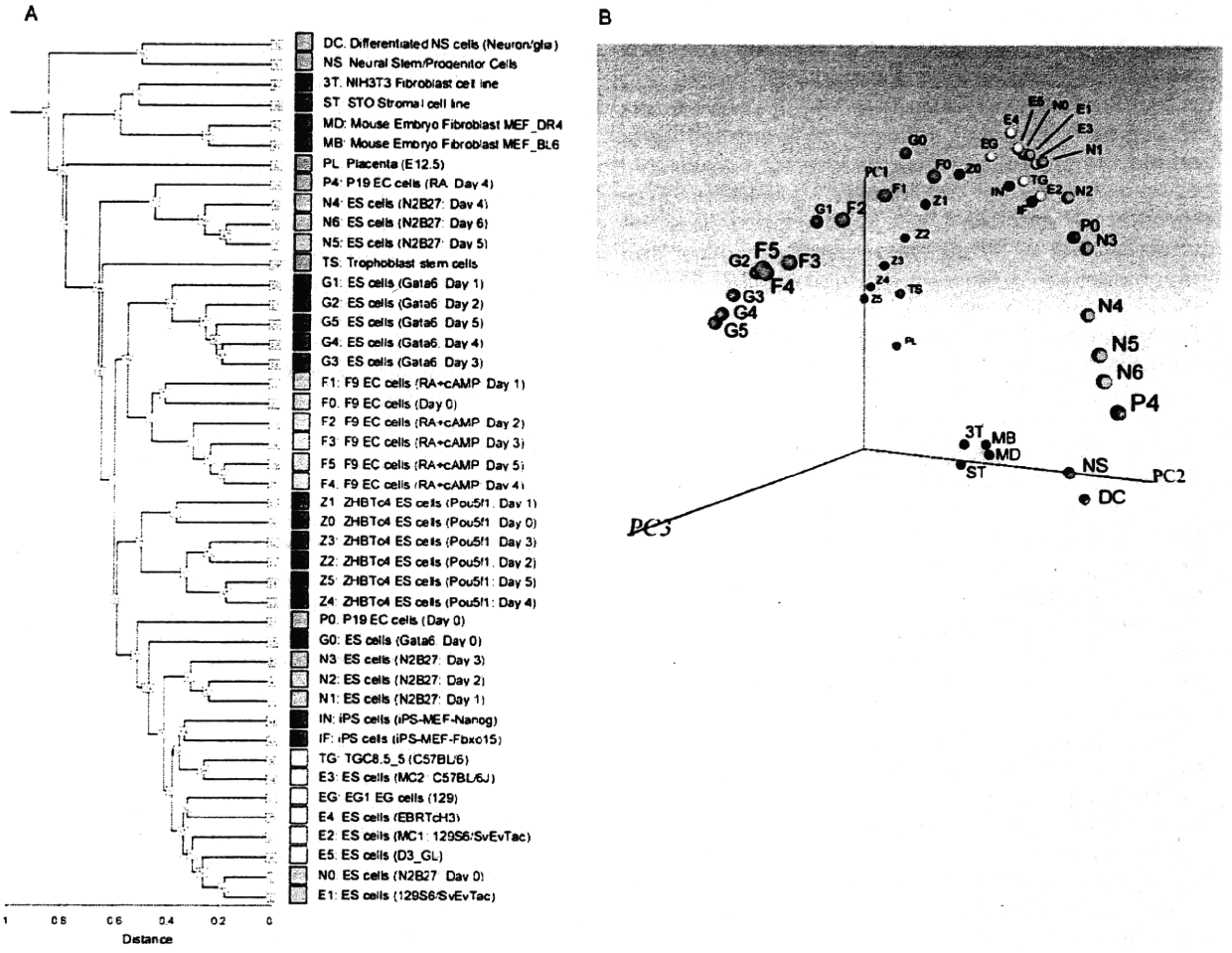


Figure 3. Global gene expression profiles of 44 different cell types. In addition to 32 cell types in the Fig. 1, 12 cell types are added for the analysis. Names of the cells and their abbreviated forms are shown. Color coding of cells are consistent throughout the paper. Average of two samples is shown. **(A)** Hierarchical clustering analysis. **(B)** Principal component analysis (PCA). Individual cell types are mapped in the 3D space according to the first three principal components (PC1, PC2 and PC3). A movie file is available Supplementary Video S2.

(Fig. 3B; Supplementary Video S2). Interestingly, an overall structure of the transcript profile space remained essentially the same as that obtained by 32 cell types (Fig. 1B). The majority of the cells were aligned along the three major cell lineage trajectories, except for fibroblast and stromal cells (NIH3T3, STO and MEFs). The fact that these cells did not fall into any of the three cell lineage trajectories is consistent with their mesodermal origin. Their locations near the bottom of all three trajectory lines also support the general trend that these cell lineage trajectories represent the gradual loss of developmental potency of cells (Fig. 1B). Indeed, undifferentiated ES and EG cells mapped near the apex of the trajectory figures, as expected (Fig. 3B). Similarly, the iPS cells mapped near the apex of the trajectories. It has been demonstrated that these iPS cells have the complete phenotype of pluripotent cells, including germline competency in the case of iPS-Nanog.[28] Considering that the iPS cells are derived from MEFs, changing the locations of cells from positions with low developmental potency (MEFs) to those with high developmental potency (iPSs) strongly supports our notion that PCA analysis can visualize and predict the developmental potency of cells.

The work presented here provides a concept and a method to visualize and quantitate the differentiation status and developmental potency of cells. The PCA figures (Figs 1B and 3B) show remarkable similarity to the conceptual picture of Waddington's epigenetic landscape, where cell lineage trajectories represent creodes.[1] Such a concept has been taken for granted by embryologists for over a century, but it now seems to be substantiated by the global gene expression patterns. The locations of cells in the multidimensional transcript profile space can be used to

identify cells of unknown origin, determine the differentiation status of cells and to assess the developmental potency of cells, possibly without going through experimental manipulation and testing of the cells. The present work also provides a framework for the future study of ES cell differentiation by visualizing the direction to which cells are going to differentiate.

Acknowledgements: We thank Drs. Tilo Kunath and Janet Rossant for providing RNAs of TS, and Dr. Angelo Vescovi for providing RNAs of NS and DC, Dr. Brigid Hogan for providing TG cells and Dr. Colin Stewart for providing EG1 cells. We also thank Dawood Dudekula and Yong Qian for assisting data analysis.

Supplementary Data: Supplementary data are available online at www.dnaresearch.oxfordjournals.org.

Funding

This work was entirely supported by the Intramural Research Program of National Institutes of Health, National Institute on Aging (Z01 AG000662).

References

1. Waddington, C. H.. 1957, *The Strategy of the Genes*, George Allen & Unwin Ltd: London.
2. Slack, J. M.. 2002, Conrad Hal Waddington: the last Renaissance biologist?, *Nat. Rev. Genet.*, **3**, 889–895.
3. Andrews, P. W.. 2002, From teratocarcinomas to embryonic stem cells, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **357**, 405–417.
4. Campbell, K. H., McWhir, J., Ritchie, W. A. and Wilmut, I.. 1996, Sheep cloned by nuclear transfer from a cultured cell line, *Nature*, **380**, 64–66.
5. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. and Yanagimachi, R.. 1998, Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei, *Nature*, **394**, 369–374.
6. Takahashi, K. and Yamanaka, S.. 2006, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, *Cell*, **126**, 663–676.
7. Slack, J. M. W.. 1991, *From Egg to Embryo: Determinative Events in Early Development*, Cambridge University Press: Cambridge, UK.
8. Gurdon, J. B. and Byrne, J. A.. 2004, The first half-century of nuclear transplantation, *Biosci. Rep.*, **24**, 545–557.
9. Hochedlinger, K. and Jaenisch, R.. 2002, Nuclear transplantation: lessons from frogs and mice, *Curr. Opin. Cell Biol.*, **14**, 741–748.
10. Yanagimachi, R.. 2002, Cloning: experience from the mouse and other animals, *Mol. Cell. Endocrinol.*, **187**, 241–248.
11. Sharov, A. A., Piao, Y., Matoba, R., Dudekula, D. B., Qian, Y., VanBuren, V., Falco, G., Martin, P. R., Stagg, C. A., Bassey, U. C., et al.. 2003, Transcriptome analysis of mouse stem cells and early embryos, *PLoS Biol.*, **1**, E74.
12. Evans, M. J. and Kaufman, M. H.. 1981, Establishment in culture of pluripotential cells from mouse embryos, *Nature*, **292**, 154–156.
13. Martin, G. R.. 1981, Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells, *Proc. Natl Acad. Sci. USA*, **78**, 7634–7638.
14. Matsui, Y., Zsebo, K. and Hogan, B. L.. 1992, Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture, *Cell*, **70**, 841–847.
15. Resnick, J. L., Bixler, L. S., Cheng, L. and Donovan, P. J.. 1992, Long-term proliferation of mouse primordial germ cells in culture, *Nature*, **359**, 550–551.
16. Pfister, S., Steiner, K. A. and Tam, P. P.. 2007, Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development, *Gene Expr. Patterns*, **7**, 558–573.
17. Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. and Smith, A.. 2003, Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture, *Nat. Biotechnol.*, **21**, 183–186.
18. Aiba, K., Sharov, A. A., Carter, M. G., Foroni, C., Vescovi, A. L. and Ko, M. S.. 2006, Defining a developmental path to neural fate by global expression profiling of mouse embryonic stem cells and adult neural stem/progenitor cells, *Stem Cells*, **24**, 889–895.
19. Niwa, H., Miyazaki, J. and Smith, A. G.. 2000, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells, *Nat. Genet.*, **24**, 372–376.
20. Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M. G., Sharov, A. A. and Ko, M. S.. 2006, Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling, *PLoS One*, **1**, e26.
21. Shimosato, D., Shiki, M. and Niwa, H.. 2007, Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells, *BMC Dev. Biol.*, **7**, 80.
22. Sharov, A. A., Dudekula, D. B. and Ko, M. S.. 2005, A web-based tool for principal component and significance analysis of microarray data, *Bioinformatics*, **21**, 2548–2549.
23. Carter, M. G., Sharov, A. A., VanBuren, V., Dudekula, D. B., Carmack, C. E., Nelson, C. and Ko, M. S.. 2005, Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray, *Genome Biol.*, **6**, R61.
24. Sharova, L. V., Sharov, A. A., Piao, Y., Shaik, N., Sullivan, T., Stewart, C. L., Hogan, B. L. and Ko, M. S.. 2007, Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains, *Dev. Biol.*, **307**, 446–459.
25. Tanaka, T. S., Kunath, T., Kimber, W. L., Jaradat, S. A., Stagg, C. A., Usuda, M., Yokota, T., Niwa, H., Rossant, J. and Ko, M. S.. 2002, Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity, *Genome Res.*, **12**, 1921–1928.
26. Strickland, S., Smith, K. K. and Marotti, K. R.. 1980, Hormonal induction of differentiation in