Table 1: The glycosyltransferase reaction library and ID number of an oligonucleotide corresponding glycosyltransferase gene on the CFG microarray.

Reaction pattern	Entrez Gene ID (Common name)	EC number	KO number	CFG Array ID
Gal a1 4 Gal	53947 (A4GALT)	2.4.1.228	K01988	219488_at
Gal b1 3 Gal	126792 (B3GALT6)	2.4.1.134	K00734	NM_080605_s_at, H93450_s_at
Gal b1 3 GalNAc	8705 (B3GALT4)	2.4.1.62	K00715	210205_at
	29071 (C1GALT1C1)	-	-	-
	56913 (C1GALT1)	2.4.1.122	K00731	219439_at
Gal b1 3 GlcNAc	8706 (B3GALT3)	2.4.1.79	K00719	211379_x_at, 211812_s_at
	8707 (B3GALT2)	-	-	210121_at
	8708 (B3GALT1)	-	-	220829_s_at, 222969_at
	10317 (B3GALT5)	2.4.1.-	K03877	206947_at
Gal b1 4 Glc	9331 (B4GALT6)	2.4.1.-	-	206232_s_at, 206233_at
Gal b1 4 GlcNAc	2683 (B4GALT1)	2.4.1.-, 2.4.1.22, 2.4.1.38, 2.4.1.90	K00704, K00708	201882_x_at, 201883_s_at
	8702 (B4GALT4)	2.4.1.-, 2.4.1.90	-	210540_s_at
	8703 (B4GALT3)	2.4.1.-, 2.4.1.38, 2.4.1.90	k00708	210243_s_at
	8704 (B4GALT2)	2.4.1.-, 2.4.1.22, 2.4.1.38, 2.4.1.90	k00704, k00708	209413_at
	9331 (B4GALT6)	2.4.1.-	-	206232_s_at, 206233_at
	9334 (B4GALT5)	2.4.1.38	k00708	221484_at, 221485_at
Gal b1 4 Xyl	11285 (B4GALT7)	2.4.1.133	k00733	53076_at
Gal a1 3 Gal	28 (ABO)	2.4.1.37, 2.4.1.40	K00707, K00709	214504_at
GalNAc a1 3 Gal	28 (ABO)	2.4.1.37, 2.4.1.40	K00707, K00709	214504_at
GalNAc a1 3 GalNAc	26301 (GBGT1)	2.4.1.88	K00722	231780_at
GalNAc b1 4 Gal	2583 (GALGT)	2.4.1.92	K00725	206435_at
	124872 (GALGT2)	-	-	-
GalNAc b1 4 GlcA	22856 (CHSY1)	-	-	-
	55454 (GALNACT-2)	-	-	-
	55790 (ChGn)	-	-	-
	337876 (CSS3)	-	-	-
GlcA b1 3 GalNAc	22856 (CHSY1)	-	-	-
	337876 (CSS3)	-	-	-
GlcA b1 3 Gal	26229 (B3GAT3)	2.4.1.135	K00735	35179_at, 203452_at
	27087 (B3GAT1)	2.4.1.135	K00735	219521_at
	135152 (B3GAT2)	2.4.1.135	K00735	-
GlcNAc a1 4 GlcA	2131 (EXT1)	2.4.1.224, 2.4.1.225	K02366	201995_at
	2132 (EXT2)	2.4.1.224, 2.4.1.225	K02367	202012_s_at, 202013_s_at
	2134 (EXTL1)	2.4.1.224	K02368	206329_at
	2135 (EXTL2)	2.4.1.223	K02369	209537_at
	2137 (EXTL3)	2.4.1.223, 2.4.1.224	K02370	211051_s_at, AF083551_s_at
GlcA b1 4 GlcNAc	2131 (EXT1)	2.4.1.224, 2.4.1.225	K02366	201995_at
	2132 (EXT2)	2.4.1.224, 2.4.1.225	K02367	202012_s_at, 202013_s_at

Reaction pattern	Entrez Gene ID (Common name)	EC number	KO number	CFG Array ID
GlcNAc b1 4 GlcA	3036 (HAS1)	2.4.1.212	-	-
	3037 (HAS2)	2.4.1.212	-	-
	3038 (HAS3)	2.4.1.212	-	-
GlcA b1 3 GlcNAc	3036 (HAS1)	2.4.1.212	-	-
	3037 (HAS2)	2.4.1.212	-	-
	3038 (HAS3)	2.4.1.212	-	-
GlcNAc a1 4 Gal	51146 (A4GNT)	-	-	221131_at
GlcNAc b1 2 Man	4245 (MGAT1)	2.4.1.101	K00726	201126_s_at
	4247 (MGAT2)	2.4.1.143	K00736	203102_s_at
	55624 (FLJ20277)	-	-	217944_at
GlcNAc b1 3 Gal	10331 (B3GNT3)	2.4.1.-	-	204856_at, AB049584_x_at
	10678 (B3GNT1)	2.4.1.149	K00741	224154_at, AB049584_s_at
	11041 (B3GNT6)	2.4.1.149	K00741	203188_at
	79369 (B3GNT4)	-	-	221240_s_at
	84002 (B3GNT5)	2.4.1.206	K03766	225612_s_at
	93010 (B3GNT7)	-	-	XM_091928_at
GlcNAc b1 3 GalNAc	192134 (B3Gn-T6)	2.4.1.147	K00739	AB073740_at
GlcNAc b1 3 LFuc	3955 (LFNG)	-	K05948	215270_at
	4242 (MFNG)	2.4.1.222	K05948	204152_s_at, 204153_s_at
GlcNAc b1 4 Man	4248 (MGAT3)	2.4.1.144	K00737	209764_at
	11282 (MGAT4B)	2.4.1.145	K00738	220189_s_at
	11320 (MGAT4A)	2.4.1.145	K00738	219797_at
	25834 (HGNT-IV-H)	-	-	-
GlcNAc b1 6 Gal	2651 (GCNT2)	2.4.1.150	K00742	211020_at
GlcNAc b1 6 GalNAc	2650 (GCNT1)	2.4.1.102	K00727	205505_at
	9245 (GCNT3)	-	-	219508_at
	51301 (C2GNT3)	-	-	220831_at
GlcNAc b1 6 Man	4249 (MGAT5)	2.4.1.155	K00744	206720_at
	90693 (LOC90693)	-	-	-
Glc a1 3 Glc	79053 (ALG8)	2.4.1.-	K03849	-
Glc a1 3 Man	29929 (ALG6)	2.4.1.-	K03848	-
	55757 (UGCGL2)	2.4.1.-	-	218801_at
	56886 (UGCGL1)	-	-	-

Reaction pattern	Entrez Gene ID (Common name)	EC number	KO number	CFG Array ID
LFuc a1 2 Gal	2523 (FUT1)	2.4.1.69	K00718	206109_at
	2524 (FUT2)	2.4.1.69	K00718	208505_s_at, D87942_at, D87942_s_at
LFuc a1 3 GlcNAc	2525 (FUT3)	2.4.1.65	K00716	214088_s_at
	2526 (FUT4)	2.4.1.-	-	209892_at, 209893_s_at
	2527 (FUT5)	2.4.1.65	K00716	211225_at, U27330_at
	2528 (FUT6)	2.4.1.65	K00716	210398_x_at, 210399_x_at, 211465_x_at, 211882_x_at, 211885_x_at
	2529 (FUT7)	2.4.1.-	-	210506_at
	10690 (FUT9)	-	-	207696_at
LFuc a1 4 GlcNAc	2525 (FUT3)	2.4.1.65	K00716	214088_s_at
LFuc a1 6 GlcNAc	2530 (FUT8)	2.4.1.68	K00717	203988_s_at
Man a1 2 Man	9488 (PIGB)	-	K05286	205452_at
	80235 (SMP3)	-	-	-
Man a1 3 Man	10195 (ALG3)	2.4.1.130	K03845	207396_s_at
	85365 (ALG2)	2.4.1.132	K03843	-
Man a1 4 GlcNAc	93183 (PIGM)	-	K05284	223470_at
Man a1 6 Man	79087 (ALG12)	2.4.1.130	K03847	218444_at
Man b1 4 GlcNAc	56052 (ALG1)	2.4.1.142	K03842	223355_at
Neu5Ac a2 3 Gal	6482 (ST3GAL1)	2.4.99.4	K00780	208322_s_at
	6483 (ST3GAL2)	2.4.99.4	K03368	205346_at
	6484 (ST3GAL4)	2.4.99.4	K03494	203759_at
	6487 (ST3GAL3)	2.4.99.6	K00781	225905_s_at
	8869 (ST3GAL5)	2.4.99.9	K03370	203217_s_at
	10402 (ST3GAL6)	2.4.99.10	K03792	210942_s_at
Neu5Ac a2 6 Gal	6480 (ST6GAL1)	2.4.99.1	K00778	201998_at, 214971_s_at
	84620 (ST6GAL2)	-	-	-
Neu5Ac a2 6 GalNAc	10610 (ST6GALNAC2)	2.4.99.-	-	204542_at
	27090 (ST6GALNAC4)	2.4.99.7	K03374	220937_s_at
	30815 (ST6GALNAC6)	2.4.99.-	K03376	222571_at
	55808 (ST6GALNAC1)	2.4.99.3	K03479	227725_at
	81849 (ST6GALNAC5)	-	-	BC001201_at
	256435 (ST6GALNAC3)	2.4.99.-	K03373	PROP_SIAT_at
Neu5Ac a2 8 Neu5Ac	6489 (ST8SIA1)	2.4.99.8	K03371	210073_at
	7903 (ST8SIA4)	2.4.99.-	-	206925_at
	8128 (ST8SIA2)	2.4.99.-	-	221285_at
	29906 (ST8SIA5)	2.4.99.-	K03369	206258_at
	51046 (ST8SIA3)	2.4.99.-	-	208064_s_at, 208065_at

I

Man a1 2 Man
Man a1 6 Man
Man a1 3 Man
LFuc a1 6 GlcNAc
GlcNAc b1 4 Man
GlcNAc b1 4 GlcNAc
GlcNAc b1 2 Man
Man b1 4 GlcNAc
Neu5Ac a2 3 Gal
Neu5Ac a2 6 Gal
Gal b1 4 GlcNAc
GlcNAc b1 6 Man

II

GlcNAc b1 4 GlcA
GlcA b1 3 GlcNAc

III

Gal b1 3 GalNAc
GlcNAc b1 6 GalNAc
LFuc a1 2 Gal
GalNAc a1 3 Gal
Neu5Ac a2 6 GlcNAc
Gal b1 3 GlcNAc
LFuc a1 4 GlcNAc
LFuc a1 3 GlcNAc
Gal b1 4 Glc
GlcNAc b1 3 Gal
GlcNAc b1 6 Gal

IV

Glc b1 4 Glc
Xyl a1 6 Glc

Man b1 4 Man
Gal a1 6 Man
Glc a1 4 Glc
Glc a1 6 Glc
Glc a1 2 Glc
Glc a1 3 Glc
Glc a1 3 Man
Glc b1 3 Glc
Glc b1 6 Glc
Gal a1 4 Gal
Gal a1 3 Gal
GalNAc b1 3 Gal
GlcNAc b1 3 GalNAc
Neu5Ac a2 6 GalNAc
GalNAc b1 4 Gal
Neu5Ac a2 8 Neu5Ac
Gal b1 4 Gal
Gal b1 3 Gal
Gal b1 6 Gal
GalNAc b1 4 GlcNAc
Man b1 4 Glc
Man b1 2 Man
Glc b1 2 Glc

the correlation coefficient

Figure S3. The cluster dendrogram of the co-occurrence of the reaction patterns in the glycan database. The co-occurrence score of the reaction patterns appearing higher in the database (top 50 of 302 reaction patterns) was used. This dendrogram was drawn using the Ward method using negative value of the correlation coefficient as distance. The clusters were divided into five groups. Cluster I consists of *N*-glycan core reaction patterns and thier terminal/internal reaction patterns. Cluster III consists of *O*-glycan and glycolipid core reaction patterns. Clusters II and IV contain the components of proteoglycan and xyloglucan, respectively.
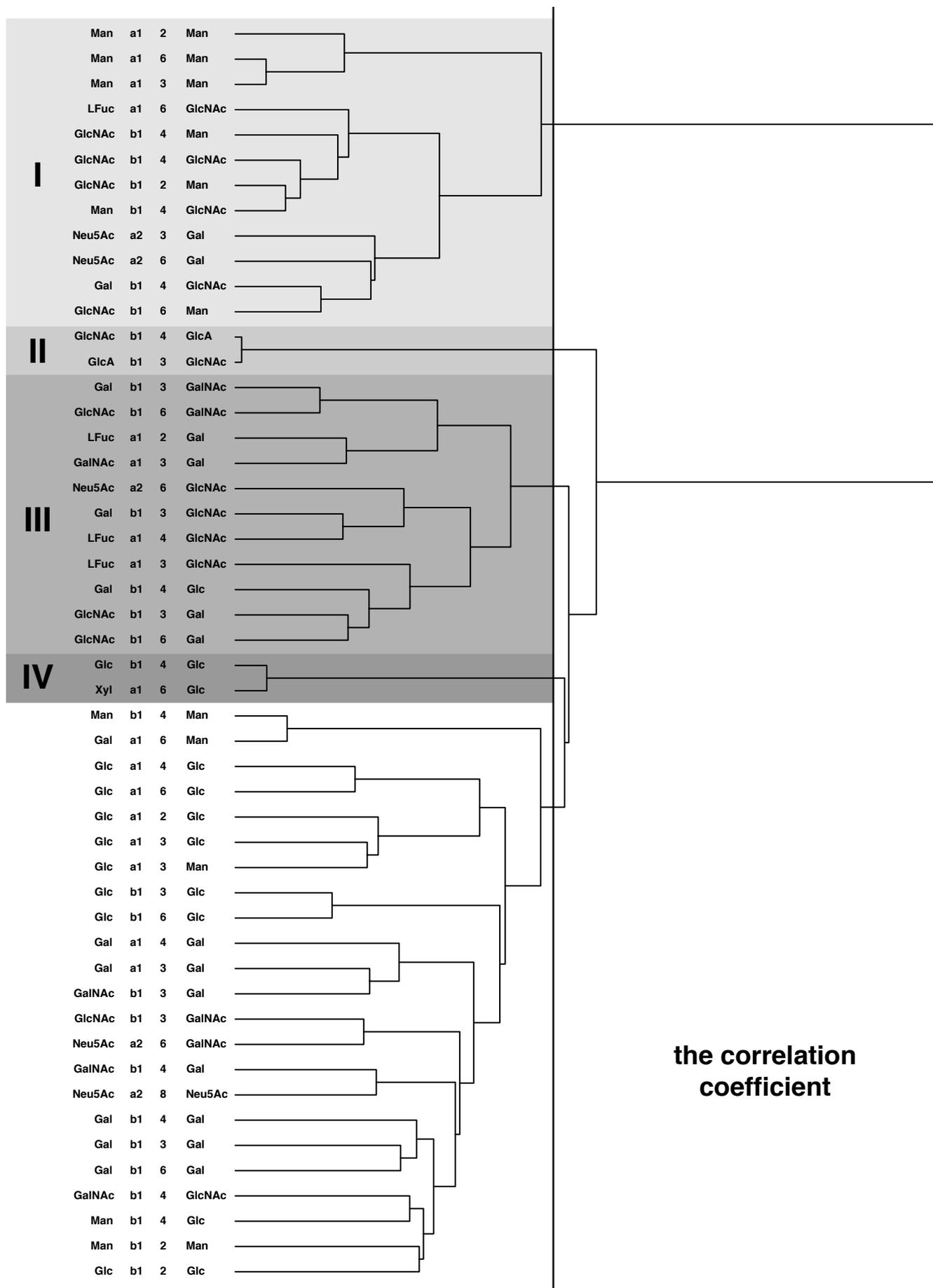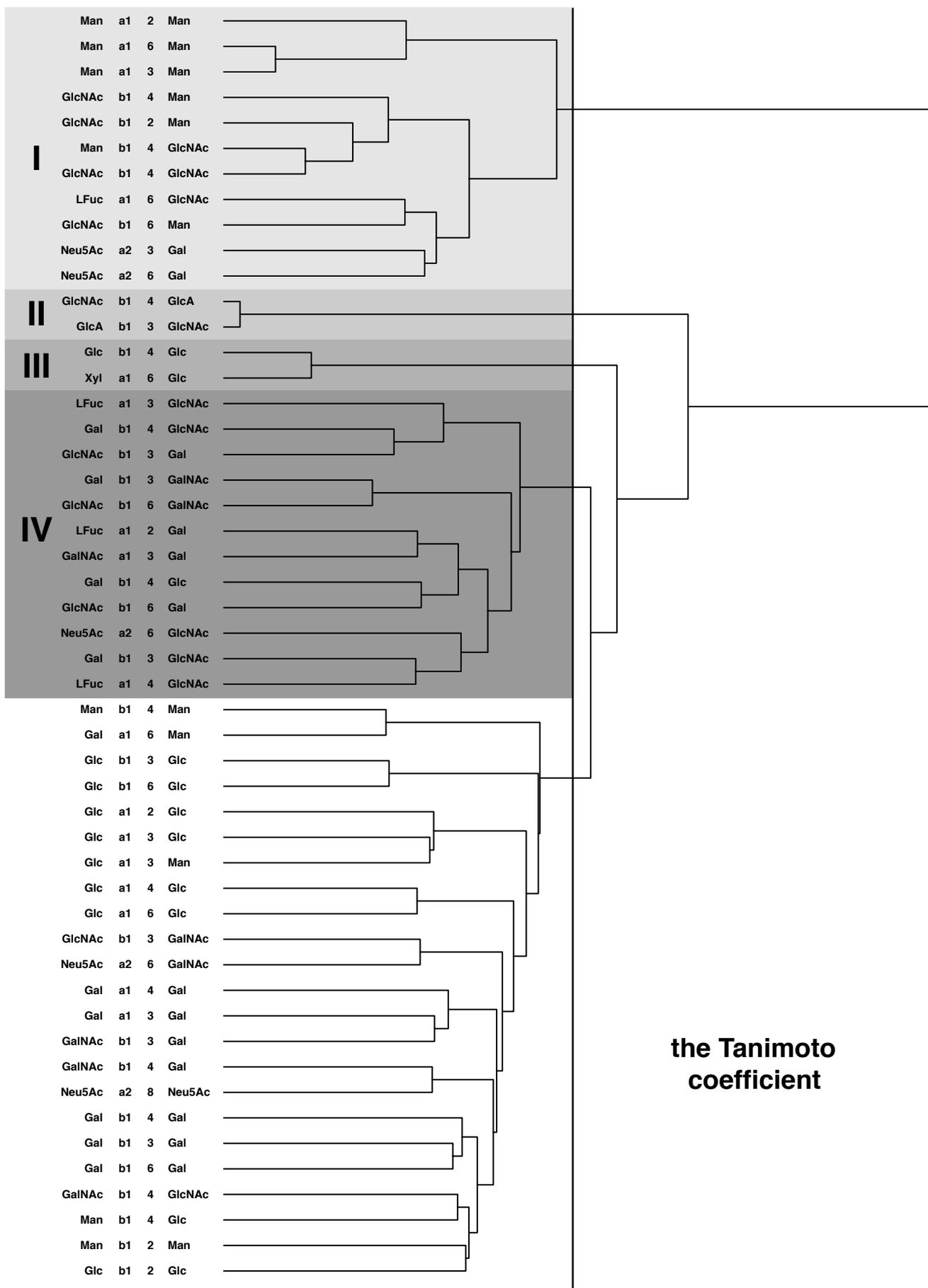
Figure S4. The cluster dendrogram of the co-occurrence of the reaction patterns in the glycan database. The co-occurrence score of the reaction patterns appearing higher in the database (top 50 of 302 reaction patterns) was used. This dendrogram was drawn using the Ward method using negative value of the Tanimoto coefficient as distance. The clusters were divided into five groups. Cluster I consists of *N*-glycan core reaction patterns and thier terminal/internal reaction patterns. Clusters II and III contain the components of proteoglycan and xyloglucan, respectively. Cluster IV consists of *O*-glycan and glycolipid core reaction patterns.

correlation
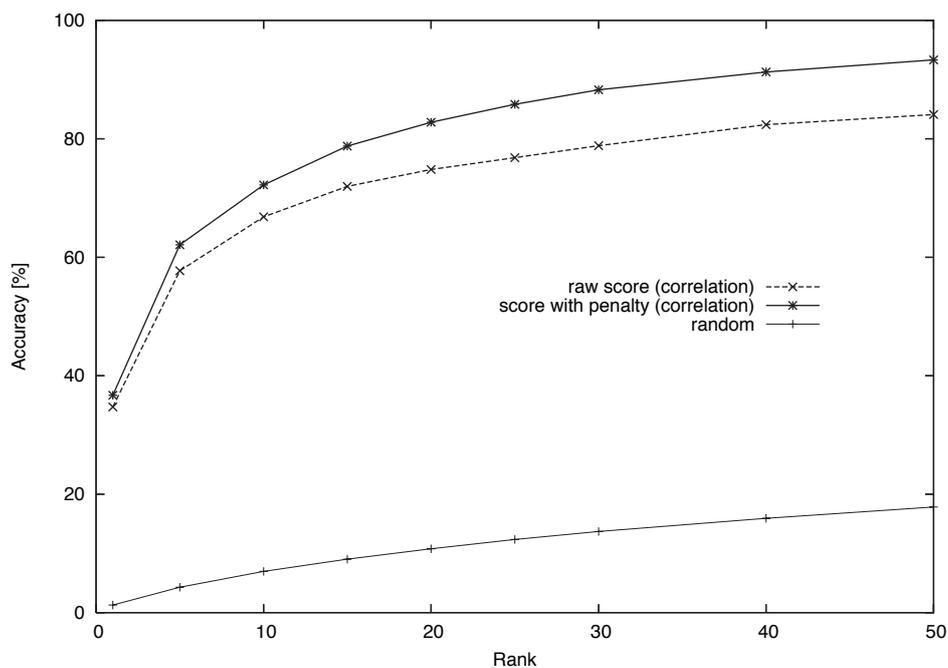
Tanimoto
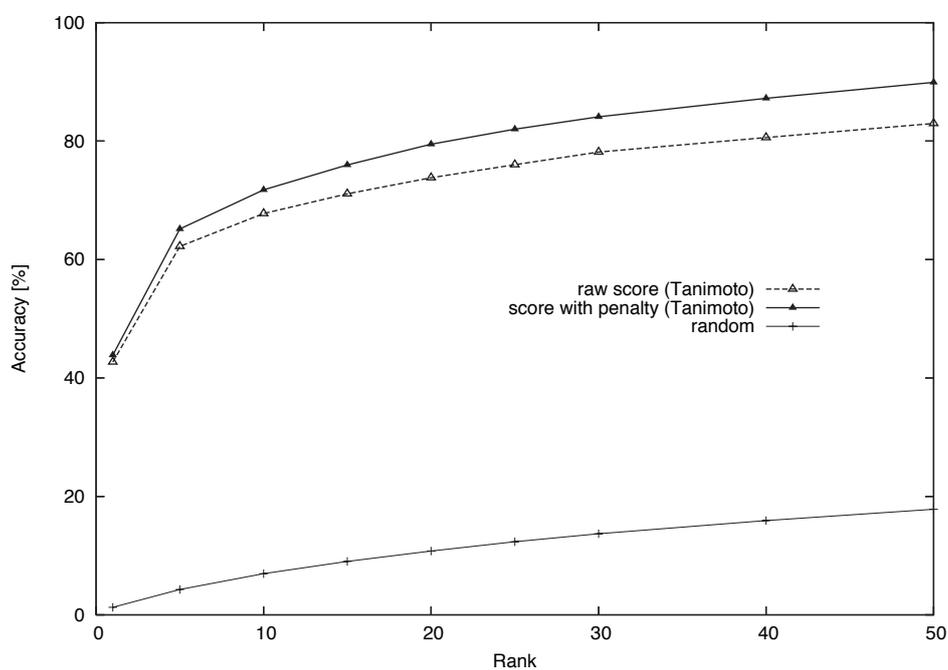
Figure S5. The accuracy of the prediction with the penalty score. The accuracy was calculated as the ratio of the query glycan predicted within the given ranking. The upper plot was obtaind using the correlation coefficient as scores, while bottom plot was obtained using the Tanimoto coefficient as scores. The solid line with ∗ and dashed line with × represents the accuracy using the correlation coefficient with and without the penalty score, respectively. The solid line with closed traiangle and dashed line with open triangle represents the accuracy using the Tanimoto coefficient with and without the penalty score, respevtively. The solid line with + is the accuracy from random data.

Table 6: A list of expressed glycosyltransferases and their reaction patterns using human carcinoma U937 cell line.

CFG Array ID	Entrez Gene ID (Common name)	Reaction pattern	Cluster number (Fig. 3)
219439_at	56913 (C1GALT1)	Gal b1 3 GalNAc	II
201883_s_at	2683 (B4GALT1)	Gal b1 4 GlcNAc	I
209413_at	8704 (B4GALT2)	Gal b1 4 GlcNAc	I
210243_s_at	8703 (B4GALT3)	Gal b1 4 GlcNAc	I
210540_s_at	8702 (B4GALT4)	Gal b1 4 GlcNAc	I
221484_at	9334 (B4GALT5)	Gal b1 4 GlcNAc	I
221485_at	9334 (B4GALT5)	Gal b1 4 GlcNAc	I
206233_at	9331 (B4GALT6)	Gal b1 4 GlcNAc, Gal b1 4 Glc	I, II
53076_at	11285 (B4GALT7)	Gal b1 4 Xyl	-
35179_at	26229 (B3GAT3)	GlcA b1 3 Gal	-
209537_at	2135 (EXTL2)	GlcNAc a1 4 GlcA	-
201995_at	2131 (EXT1)	GlcNAc a1 4 GlcA, GlcA b1 4 GlcNAc	-
202012_s_at	2132 (EXT2)	GlcNAc a1 4 GlcA, GlcA b1 4 GlcNAc	-
202013_s_at	2132 (EXT2)	GlcNAc a1 4 GlcA, GlcA b1 4 GlcNAc	-
201126_s_at	4245 (MGAT1)	GlcNAc b1 2 Man	I
203102_s_at	4247 (MGAT2)	GlcNAc b1 2 Man	I
204856_at	10331 (B3GNT3)	GlcNAc b1 3 Gal	II
AB049584_x_at	10331 (B3GNT3)	GlcNAc b1 3 Gal	II
225612_s_at	84002 (B3GNT5)	GlcNAc b1 3 Gal	II
203188_at	11041 (B3GNT6)	GlcNAc b1 3 Gal	II
220189_s_at	11282 (MGAT4B)	GlcNAc b1 4 Man	I
219508_at	9245 (GCNT3)	GlcNAc b1 6 GalNAc	II
218801_at	55757 (UGCGL2)	Glc a1 3 Man	-
206109_at	2523 (FUT1)	LFuc a1 2 Gal	II
D87942_at	2524 (FUT2)	LFuc a1 2 Gal	II
D87942_s_at	2524 (FUT2)	LFuc a1 2 Gal	II
214088_s_at	2525 (FUT3)	LFuc a1 3 GlcNAc, LFuc a1 4 GlcNAc	II
209892_at	2526 (FUT4)	LFuc a1 3 GlcNAc	II
210398_x_at	2528 (FUT6)	LFuc a1 3 GlcNAc	II
210399_x_at	2528 (FUT6)	LFuc a1 3 GlcNAc	II
211465_x_at	2528 (FUT6)	LFuc a1 3 GlcNAc	II
211882_x_at	2528 (FUT6)	LFuc a1 3 GlcNAc	II
211885_x_at	2528 (FUT6)	LFuc a1 3 GlcNAc	II
203988_s_at	2530 (FUT8)	LFuc a1 6 GlcNAc	I
205452_at	9488 (PIGB)	Man a1 2 Man	I
223470_at	93183 (PIGM)	Man a1 4 GlcNAc	-
203759_at	6484 (ST3GAL4)	Neu5Ac a2 3 Gal	I
201998_at	6480 (ST6GAL1)	Neu5Ac a2 6 Gal	I
220937_s_at	27090 (ST6GALNAC4)	Neu5Ac a2 6 GalNAc	-
227725_at	55808 (ST6GALNAC1)	Neu5Ac a2 6 GalNAc	-
40	32	21	

Gal b1 - 4 GlcNAc b1~4 Man
Neu5Ac a2 - 6 Gal b1 - 4 GlcNAc b1~2 Man

G05467
Score = 2.627 (1st)

Neu5Ac a2 - 6 Gal b1 - 4 GlcNAc b1~4 Man
Gal b1 - 4 GlcNAc b1~2 Man

G05814
Score = 2.627 (1st)

Neu5Ac a2 - 6 Gal b1 - 4 GlcNAc b1~4 Man
Neu5Ac a2 - 6 Gal b1 - 4 GlcNAc b1~2 Man

G04844
Score = 2.568 (3rd)

Neu5Ac a2 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1~6 Man a1
LFuc a1~3 ... LFuc a1~3 ...
Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1~2 Man a1
Gal b1~4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1~4 Man a1
LFuc a1~3 ... LFuc a1~3 ...
Neu5Ac a2 - 6 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1~2 Man a1
LFuc a1~6 GlcNAc
6,3 Man b1 - 4 GlcNAc b1~4 GlcNAc

G11846
Score = 2.549 (4th)

Neu5Ac a2 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1~6 Man a1
LFuc a1~3 ...
Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1~2 Man a1
Gal b1~4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1~4 Man a1
LFuc a1~3 ... LFuc a1~3 ...
Neu5Ac a2 - 6 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1~3 Man a1
LFuc a1~6 GlcNAc
6,3 Man b1 - 4 GlcNAc b1~4 GlcNAc

G04206
Score = 2.442 (5th)

Neu5Ac a2~6 Gal b1~4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 Glc
Neu5Ac a2~3 ...
LFuc a1~3 ... LFuc a1~3 ...

G04056
Score = 2.421 (6th)

Neu5Ac a2 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 Glc
LFuc a1~3

G00197
Score = 2.419 (7th)

Neu5Ac a2 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 Glc
LFuc a1~3

G10278
Score = 2.419 (7th)

Neu5Ac a2 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 Glc
LFuc a1~3

G10271
Score = 2.419 (7th)

Neu5Ac a2 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 Glc
LFuc a1~3 ... LFuc a1~3

G00193
Score = 2.413 (10th)

Neu5Ac a2 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1~4 GlcNAc b1 - 3 Gal b1 - 4 GlcNAc b1 - 3 Gal b1 - 4 Glc
LFuc a1~3 ... LFuc a1~3

G00194
Score = 2.413 (10th)

Figure S7. Predicted glycan structures from the gene expression profile of human carcinoma U937 cell line. Predicted structures with high score (top 10) are shown. While thy have different core structures, they share the terminal sialic acid (Neu5Ac) and fucose, and some glycans contain sialyl Lewis X structure (Neu5Ac a2-3 Gal b1-4 (LFuc a1-3) GlcNAc). Note that the structure of G04056 (6th), which is not in nature, is miss-predicted due to annotation error in the KEGG GLYCAN database.