

Fig. 21 PacBioRS シークエンサーによる解析パフォーマンス

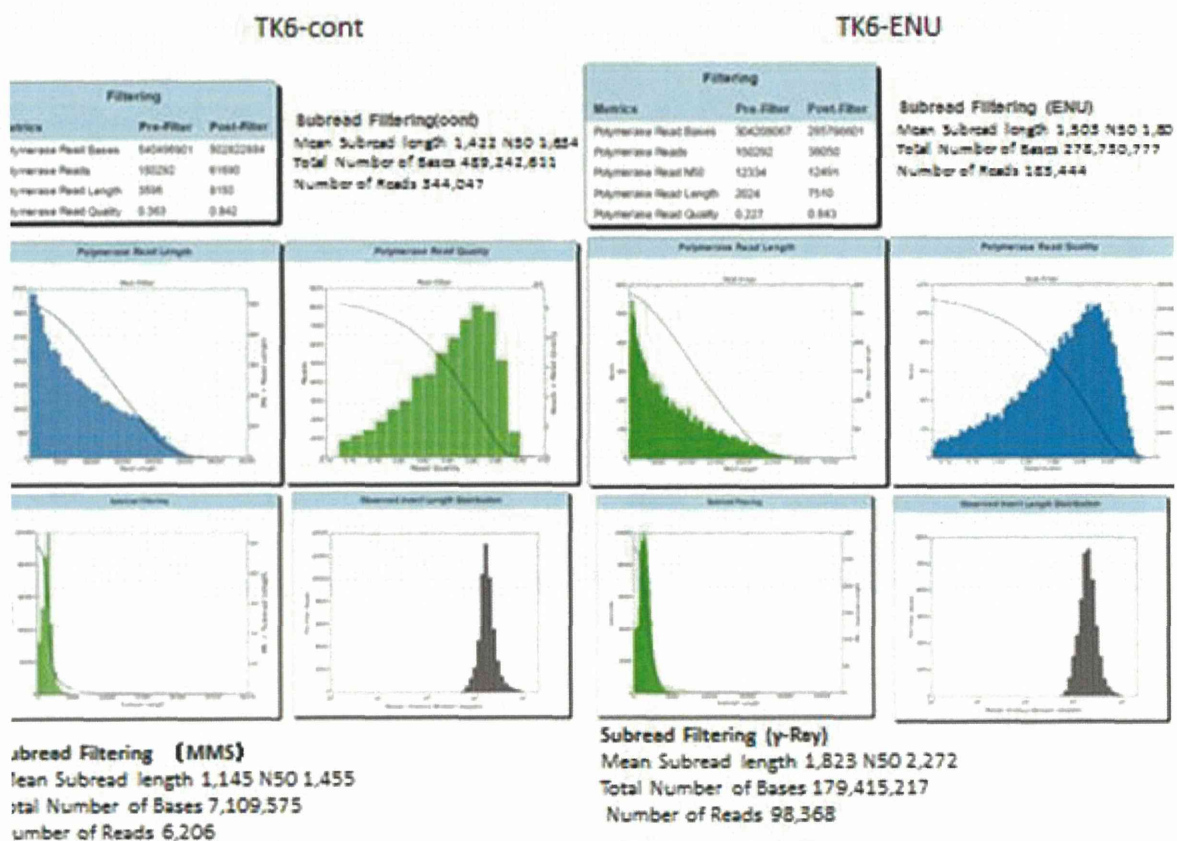
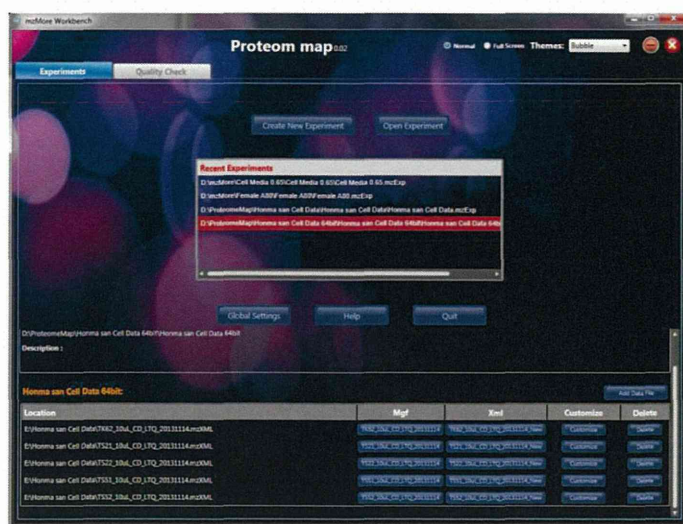


Fig. 22 ProteomeMapソフトウェアの概要

1) 生データの取り込み画面

ProteomeMap

- put your proteins on the map...

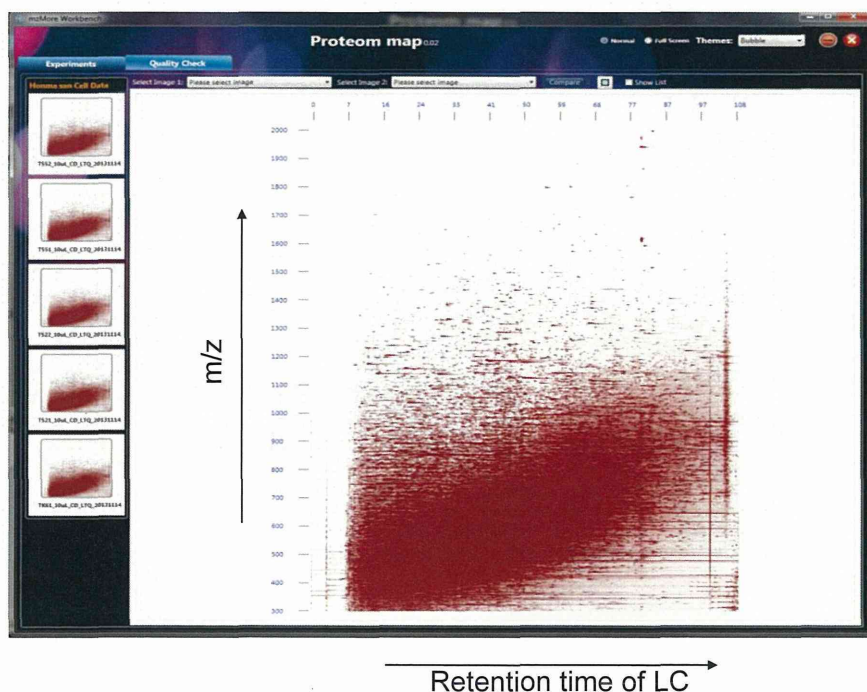


Compare

what proteins are identified,
where it is identified and
how its identified

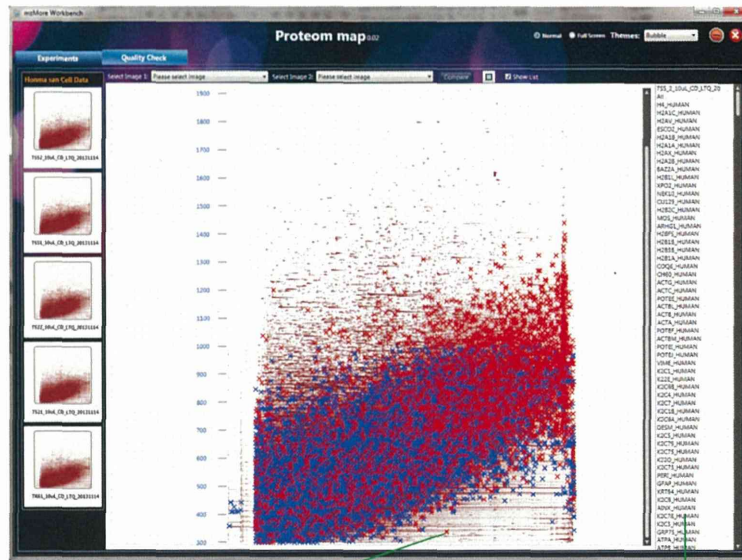
all in one screen

2) LC-MSデータの2次元マップ表示



3) MS/MSデータの取り込みと表示

Mapping of MS/MS Data on 3D Map & List of Identified Proteins

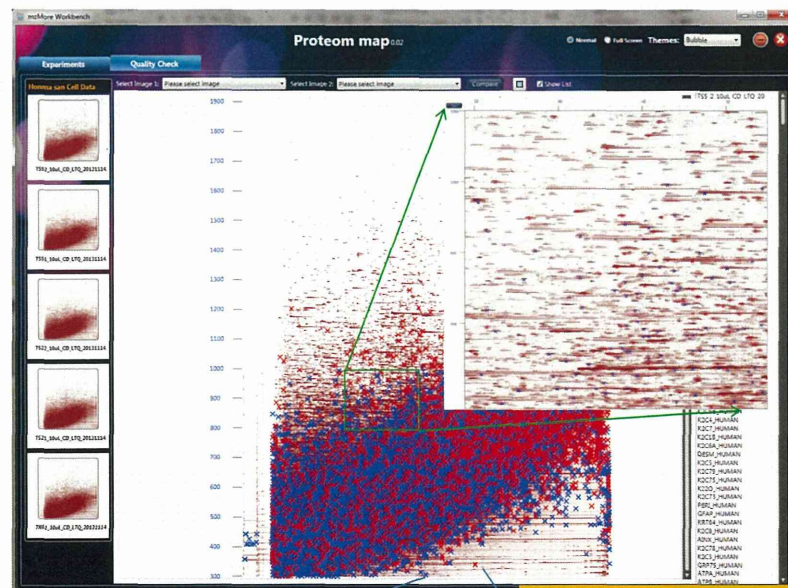


Location of MS/MS and identified proteins

List of identified proteins

4) ズームイン機能

Zoom IN Option of ProteomeMap

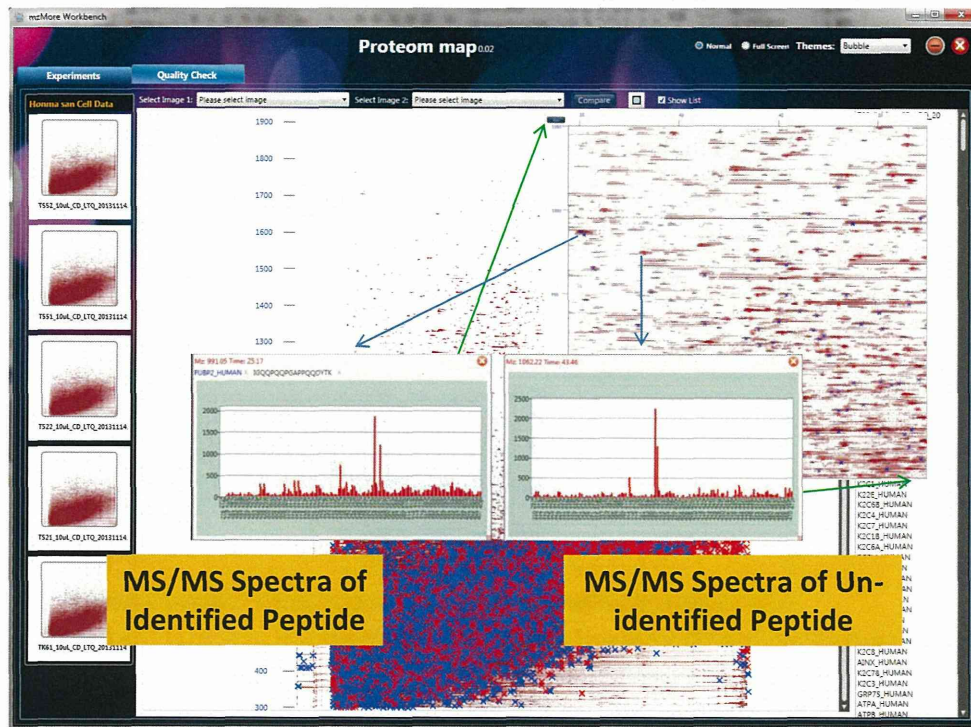


Blue X marks identified peptides

Red X marks un-identified MS/MS data

5) MS/MS スペクトルの表示

MS/MS View Option in ProteomeMap

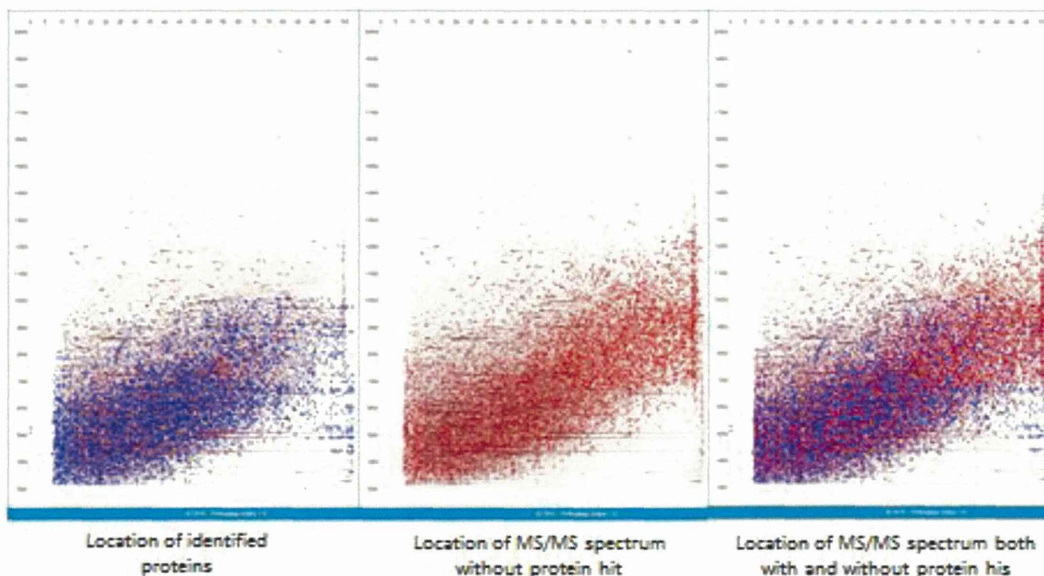


6) ProteoMap Online 操作画面

1. ProteoMapOnline 1.0 is a web based tool for sharing basic LC-MS/MS data such as 2D image of the LC-MS data, MS/MS spectrum and list of identified proteins and its peptides.



2. User can upload the 2D image of LC-MS, MS/MS spectrum as a .mgf file format and list of identified proteins as .XML file format (Desk top version of ProteoMap required).
3. ProteoMapOnline will display the MS/MS spectrum and identified proteins on the 2D image.
4. The location of MS/MS data with and without a peptide/protein hit are shown in two different colours, blue and red respectively.



5. User can choose to display only the identified proteins or only the MS/MS spectrum of unidentified proteins or both.

Table 11 標準ゲノムDNAのアレル頻度

Gene	Chromosome Number	Reference Base	Observed Base	Amino Acid Change	Allelic Frequency	Codon Change	Exon ID	Validated by Digital PCR	Validated by NGS	Class of Mutation
BRAF	chr7	A	T	V600E	10.5	GTG/GAG	Exon 7 140453075,140453193	Yes	Yes	Missense
KIT	chr4	A	T	D816V	10.0	GAC/GTC	Exon 17 55599236,55599358	Yes	No	Missense
EGFR	chr7	AAGGAATTAAGAGAAGCA	AA	E746 - A750	2.0	N/A	Exon 19 55174772,55174870	Yes	No	Deletion
EGFR	chr7	T	G	L858R	3.0	CTG/CCG	Exon 21 55191719,55191874	Yes	No	Missense
EGFR	chr7	C	T	T790M	1.0	ACG/ATG	Exon 20 55181293,55181478	Yes	No	Missense
EGFR	chr7	G	A	G719S	24.5	GGC/AGC	Exon 18 55241614,55241736	Yes	Yes	Missense
KRAS	chr12	C	T	G13D	15.0	GGC/GAC	Exon 12 25398208,25398329	Yes	Yes	Missense
KRAS	chr12	G	A	G12D	6.0	GGT/GAT	Exon 2 2539869,2539748	Yes	No	Missense
NRAS	chr1	C	A	Q61K	12.5	CAA/AAA	Exon 3 115256599,115256777	Yes	No	Missense
PIK3CA	chr3	G	A	E545K	9.0	GAG/AAG	Exon 10 178935998-178936122	Yes	No	Missense
PIK3CA	chr3	A	G	H1047R	17.5	CAT/CGT	Exon 3 178951882,178957881	Yes	Yes	Missense
CDX2	chr13	AC	A	V306fs	41.5	N/A	N/A	Yes	Yes	Deletion
ARID1A	chr1	GC	G	P1562fs	33.5	N/A	N/A	Yes	Yes	Deletion
CCND2	chr12	AT	A	N/A	32.5	N/A	N/A	Yes	No	Deletion
BRCA2	chr13	CA	C	A1689fs	33.0	N/A	N/A	Yes	Yes	Deletion
ALK	chr2	G	A	P1543S	33.0	CCT/TCT	Exon 2 29415640,29416788	Yes	Yes	Missense
CTNNB1	chr3	C	A	S33Y	32.5	TCT/TAT	Exon 3 41266017,41266202	Yes	Yes	Missense
FBXW7	chr4	TC	T	G667fs	33.5	N/A	N/A	Yes	Yes	Deletion
PDGFRA	chr4	G	A	G426D	33.5	GGC/GAC	Exon 4 55138561,55138687	Yes	Yes	Missense
APC	chr5	C	T	R2714C	33.0	CGT/TGT	Exon 5 112173250,112181936	Yes	Yes	Missense
NOTCH1	chr9	G	A	P668S	31.5	CCG/TCG	Exon 9 139409742,139409852	Yes	Yes	Missense
FLT3	chr13	GGA	G	S985fs	10.5	N/A	N/A	Yes	Yes	Deletion
FLT3	chr13	A	G	V197A	11.5	GTG/GCG	Exon 13 28626882,28626811	Yes	Yes	Missense
IDH1	chr2	G	A	S261L	10.0	TCA/TTA	Exon 2 209106718,209106869	Yes	Yes	Missense
CTNNB1	chr3	CGCTT	C	S45del	10.0	N/A	N/A	Yes	Yes	Deletion
MET	chr7	GT	G	V237fs	6.5	N/A	N/A	Yes	Yes	Deletion
SH2D2A/NTRK1*	chr1	AC	A	N/A	8.5	N/A	N/A	Yes	Yes	Deletion
ABL2	chr1	TG	T	P986fs	8.0	N/A	N/A	Yes	Yes	Deletion
CDH1	chr16	A	G	N/A	7.5	N/A	N/A	Yes	No	None
FANCA	chr16	ACT	A	E345fs	7.5	N/A	N/A	Yes	Yes	Deletion
NF1	chr17	CT	C	L626fs	7.5	N/A	N/A	Yes	Yes	Deletion
NF2	chr22	AC	A	P275fs	8.0	N/A	N/A	Yes	Yes	Deletion
EP300	chr22	CA	C	K291fs	8.0	N/A	N/A	Yes	Yes	Deletion
MLH1	chr3	C	A	L187M**	8.5	CTG/ATG	Exon 3 37061801,37061954	Yes	Yes	Missense
FQFR1	chr8	G	A	P124L***	8.5	CCC/GTC	Exon 8 38285439,38285611	Yes	Yes	Missense

* This chromosome location was annotated to both SH2D2A and NTRK1 using hg19 and GRCH37.

** L187M is the correct annotation with reference to ENST00000383761. It may be observed as L323M in other transcripts.

*** P124L is the correct annotation with reference to ENST00000335922. It may be observed as P150L in other transcripts.

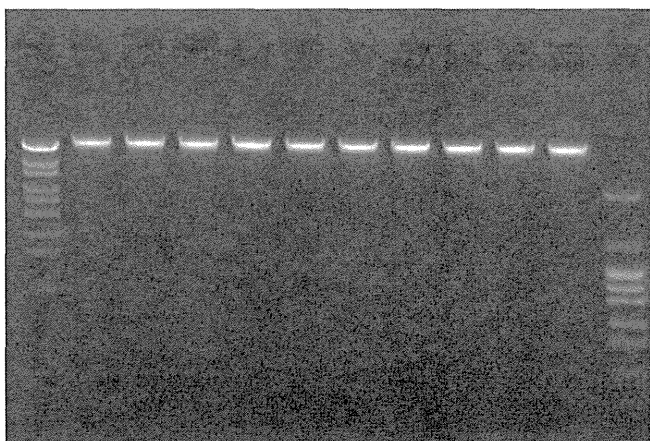
Table 12 標準ゲノムDNAにおける二本鎖DNA定量結果

サンプル	液量(ul)	濃度(ng/ul)	総量(ug)	吸光定量値/蛍光定量値
Standard DNA1	20	59.8	1.2	1
Standard DNA2	20	57.9	1.2	1
Standard DNA3	20	64.8	1.3	0.9
Standard DNA4	18	61.2	1.1	0.9
Standard DNA5	18	59.3	1.1	1
Standard DNA6	18	61.2	1.1	1
Standard DNA7	18	64.2	1.2	0.9
Standard DNA8	18	63.6	1.1	1
Standard DNA9	18	58.4	1.1	1
Standard DNA10	18	61.8	1.1	0.9
Standard DNA11	18	50.6	0.9	1.1
Standard DNA12	18	59	1.1	1
Standard DNA13	18	58.7	1.1	1
Standard DNA14	18	56.9	1	0.9
Standard DNA15	18	58	1	1
Standard DNA16	18	60.6	1.1	0.9
Standard DNA17	18	59.6	1.1	0.9
Standard DNA18	18	54.9	1	1
Standard DNA19	18	59.7	1.1	0.9
Standard DNA20	18	61.2	1.1	0.9
Total	510	40.8	20.8	-

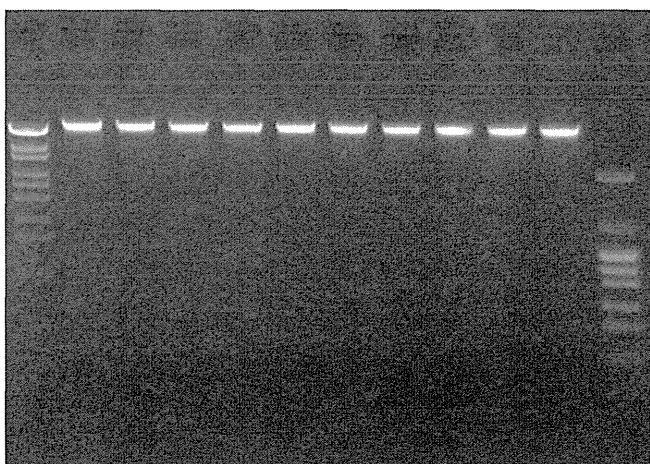
Quant-iT dsDNA BR Assay Kit (Thermo Fisher Scientific) を用いた蛍光定量が行われた。「吸光定量値/蛍光定量値」は核酸定量結果と二本鎖DNA 定量結果の乖離を示す(推奨値 ≤ 3)。評価項目として、解析に必要とされる核酸濃度、総量、及び核酸定量結果 (Table 11) と二本鎖DNA定量結果の乖離データを指標にして評価した。

Table 13 標準ゲノムDNAの定量結果 (Nanodrop を用いた吸光定量)

サンプル	液量(ul)	濃度(ng/ul)	総量(ug)	260/280	260/230
Standard DNA1	20	60.3	1.2	2.2	1.3
Standard DNA2	20	57.6	1.2	2.1	1.3
Standard DNA3	20	57.9	1.2	2.2	1.3
Standard DNA4	18	55.8	1	2.2	1.2
Standard DNA5	18	56.4	1	2.2	1.3
Standard DNA6	18	59.7	1.1	2.2	1.2
Standard DNA7	18	60.3	1.1	2	1.3
Standard DNA8	18	62.1	1.1	2.2	1.3
Standard DNA9	18	57	1	2.1	1.2
Standard DNA10	18	55.8	1	2.1	1.2
Standard DNA11	18	57.3	1	2.2	1.3
Standard DNA12	18	56.1	1	2.2	1.3
Standard DNA13	18	56.4	1	2.2	1.3
Standard DNA14	18	54	1	2.1	1.3
Standard DNA15	18	56.4	1	2.2	1.2
Standard DNA16	18	54.9	1	2.2	1.2
Standard DNA17	18	52.5	0.9	2.2	1.2
Standard DNA18	18	52.8	1	2.3	1.2
Standard DNA19	18	54.9	1	2.2	1.3
Standard DNA20	18	56.1	1	2.1	1.2

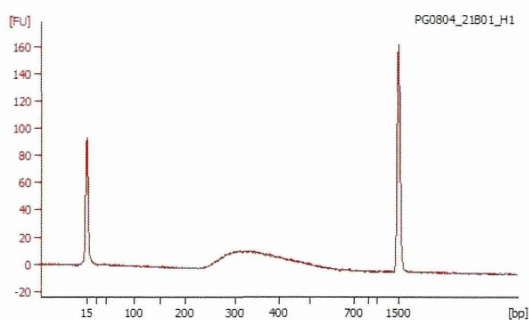


レーン (左から)	サンプル
1	λ -EcoT14 I digest
2	Standard DNA1
3	Standard DNA2
4	Standard DNA3
5	Standard DNA4
6	Standard DNA5
7	Standard DNA6
8	Standard DNA7
9	Standard DNA8
10	Standard DNA9
11	Standard DNA10
12	pHY Marker

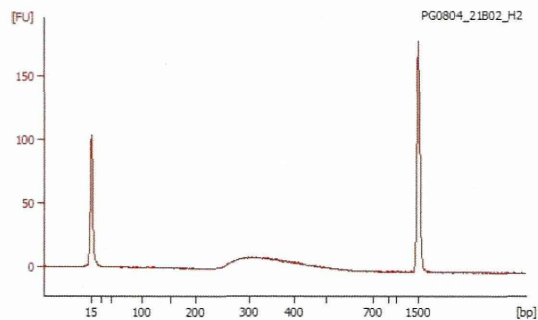


レーン (左から)	サンプル
1	λ -EcoT14 I digest
2	Standard DNA11
3	Standard DNA12
4	Standard DNA13
5	Standard DNA14
6	Standard DNA15
7	Standard DNA16
8	Standard DNA17
9	Standard DNA18
10	Standard DNA19
11	Standard DNA20
12	pHY Marker

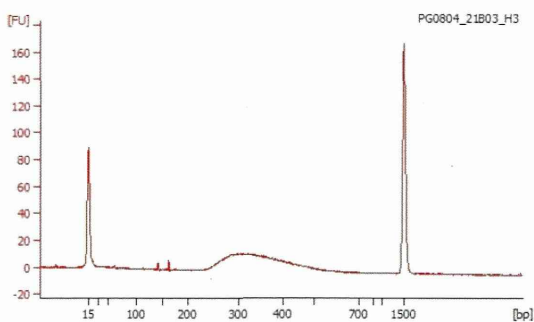
Fig. 23 アガロースゲル電気泳動によるゲノムDNAの品質確認 蛍光定量結果より50 ng 分の二本鎖DNA をアプライした。



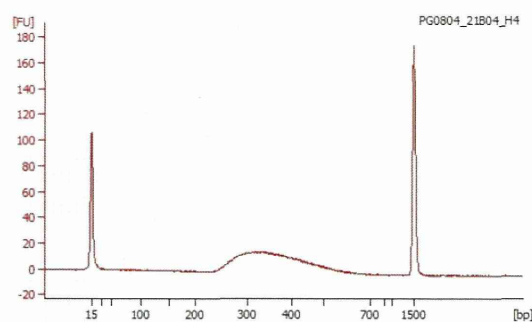
サンプル	Standard DNA Mix-1
ピークサイズ	310 bp
濃度	25.9 nmol/L



サンプル	Standard DNA Mix-2
ピークサイズ	305 bp
濃度	16.5 nmol/L



サンプル	Standard DNA Mix-3
ピークサイズ	304 bp
濃度	25.0 nmol/L



サンプル	Standard DNA Mix-4
ピークサイズ	312 bp
濃度	30.4 nmol/L

Fig. 24 Agilent 2100 Bioanalyzer を用いたリード長の検定結果 作製されたシーケンスライブラリーの品質をAgilent 2100 Bioanalyzer を用いて測定した。また、ライブラリーにはアダプター配列を付加されているため、ピークサイズからアダプターサイズ（約100base）を除いたサイズがクローニングサイズとなる。サンプルは、4種類のDepthについて解析を行うため、同サンプルで4解析分のライブラリーを独立して作製した。

Table 14 Phred クオリティスコアに基づいたシーケンス解析の精度評価

サンプル名	リード数	塩基数	Q30R1	Q30R2
Standard DNA#1	474,895,722	47,489,572,200	95.3	93.6
Standard DNA#2	529,964,716	52,996,471,600	94.8	92.0
Standard DNA#3	440,178,726	44,017,872,600	95.5	94.2
Standard DNA#4	469,051,582	46,905,158,200	95.0	93.5

Q30R1: 片鎖・両鎖解析において 1 回目に読み取られるリードの品質

Q30R2: 両鎖解析において 2 回目に読み取られるリードの品質

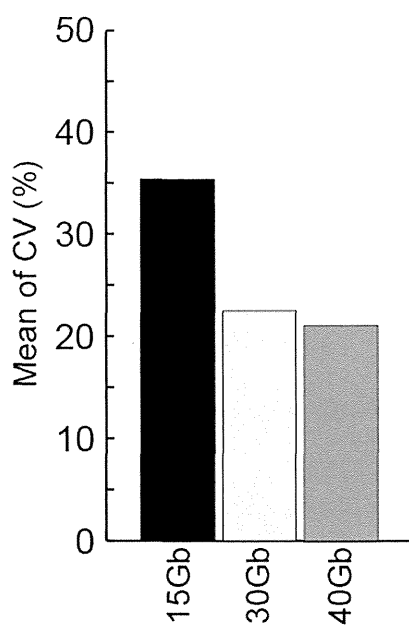


Fig. 25 シーケンス量が 15Gb, 30Gb, 40Gb の場合における変動係数 (CV) の平均

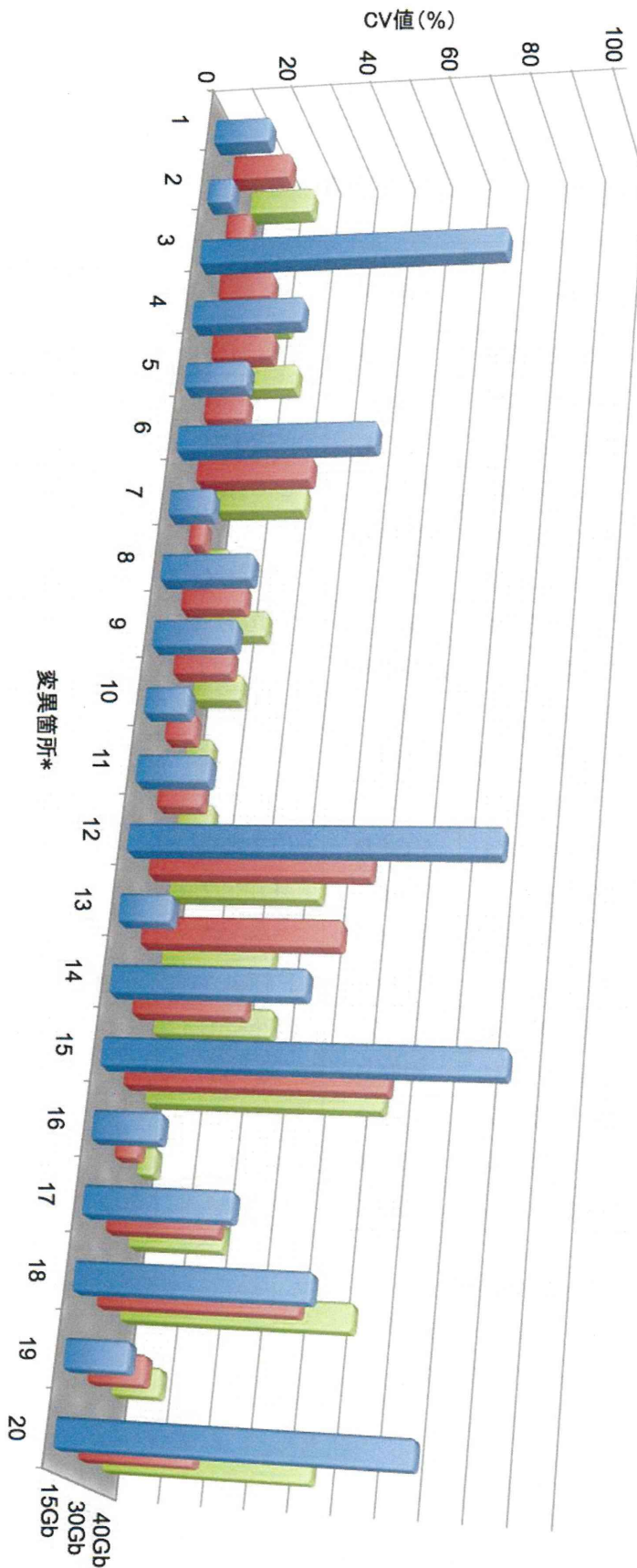


Fig. 26 15Gb、30Gb、40Gbの量でシーケンスされたときの各変異箇所における変異塩基の変動係数(CV) CV値は4回の独立したシーケンス解析から得られた変異頻度に基づいて算出された。*横軸の変異箇所については、Table 15に示した。

Table 15 標準ゲノムDNA上で確認された変異箇所

	リファレンス ID	位置	遺伝子	アミノ酸変化	リファレンス配列 (塩基)	変異塩基
1	chr1	115,256,530	NRAS	Q61K	G	T
2	chr2	29,416,326	ALK	P1543S	G	A
3	chr2	209,106,786	IDH1	S261L	G	A
4	chr3	37,061,883	MLH1	L187M**	C	A
5	chr3	41,266,101	CTNNB1	S33Y	C	A
6	chr3	178,936,091	PIK3CA	E545K	G	A
7	chr3	178,952,085	PIK3CA	H1047R	A	G
8	chr4	55,138,600	PDGFRA	G426D	G	A
9	chr4	55,599,321	KIT	D816V	A	T
10	chr5	112,179,431	APC	R2714C	C	T
11	chr7	55,241,707	EGFR	G719S	G	A
12	chr7	55,249,071	EGFR	T790M	C	T
13	chr7	55,259,515	EGFR	L858R	T	G
14	chr7	140,453,136	BRAF	V600E	A	T
15	chr8	38,285,611	FGFR1	P124L***	G	A
16	chr9	139,409,754	NOTCH1	P668S	G	A
17	chr12	25,398,281	KRAS	G13D	C	T
18	chr12	25,398,284	KRAS	G12D	C	T
19	chr13	28,626,706	FLT3	V197A	A	G
20	chr16	68,867,462	CDH1	N/A	T	C

変異頻度については、Table 11を参照。

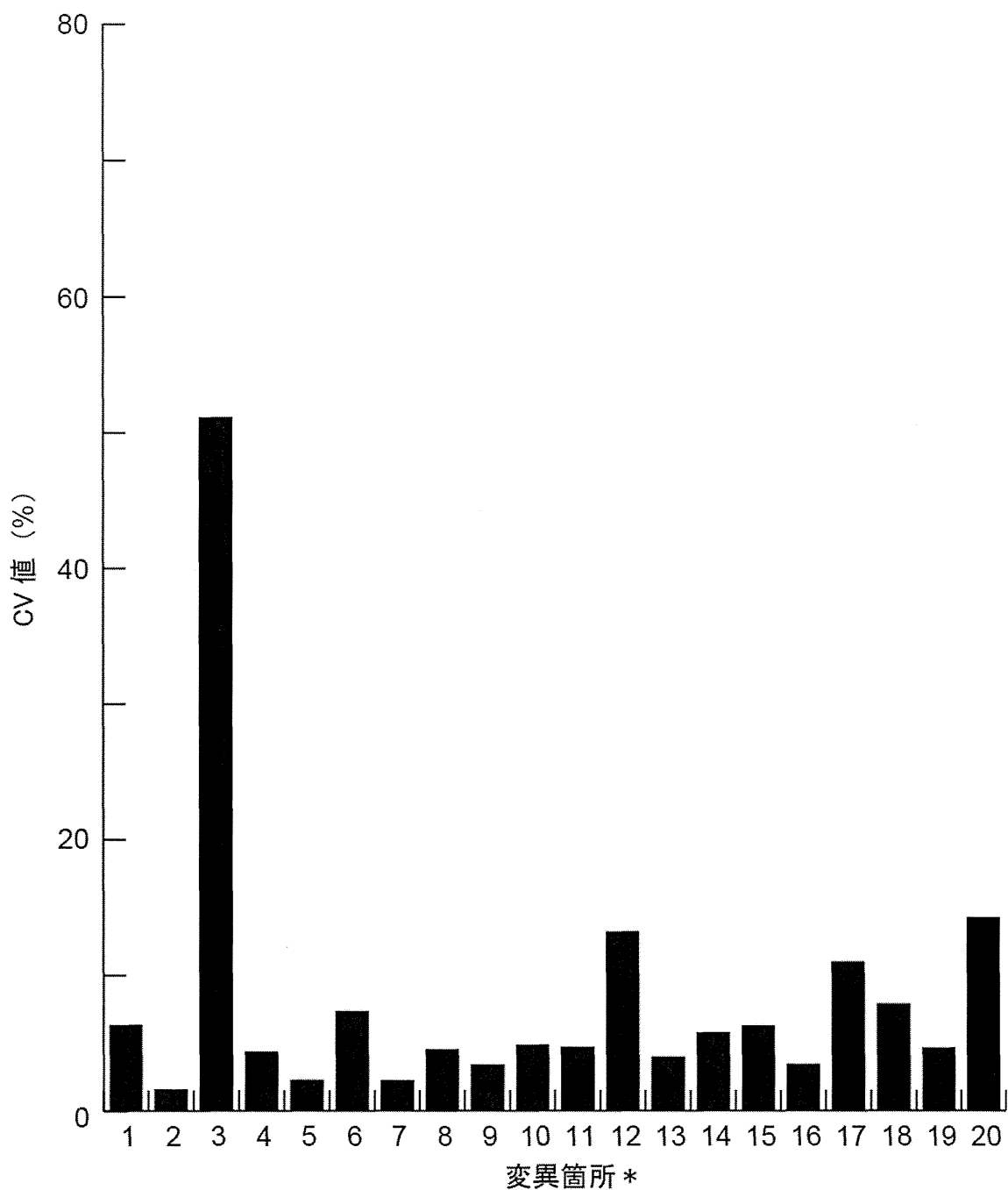


Fig. 27 全マッピング結果を合わせた時の各変異箇所における変異頻度の変動係数 (CV) 全マッピング結果をまとめた後、40Gb, 80Gb, 120Gb, 160Gb 相当のデータとなるようにダウンサンプリングし、各変異箇所におけるそれぞれのシーケンス量 (40Gb, 80Gb, 120Gb, 160Gb) に相当する塩基頻度を測定後、CV 値を算出した。*横軸の変異箇所については、Table 15 に示した。

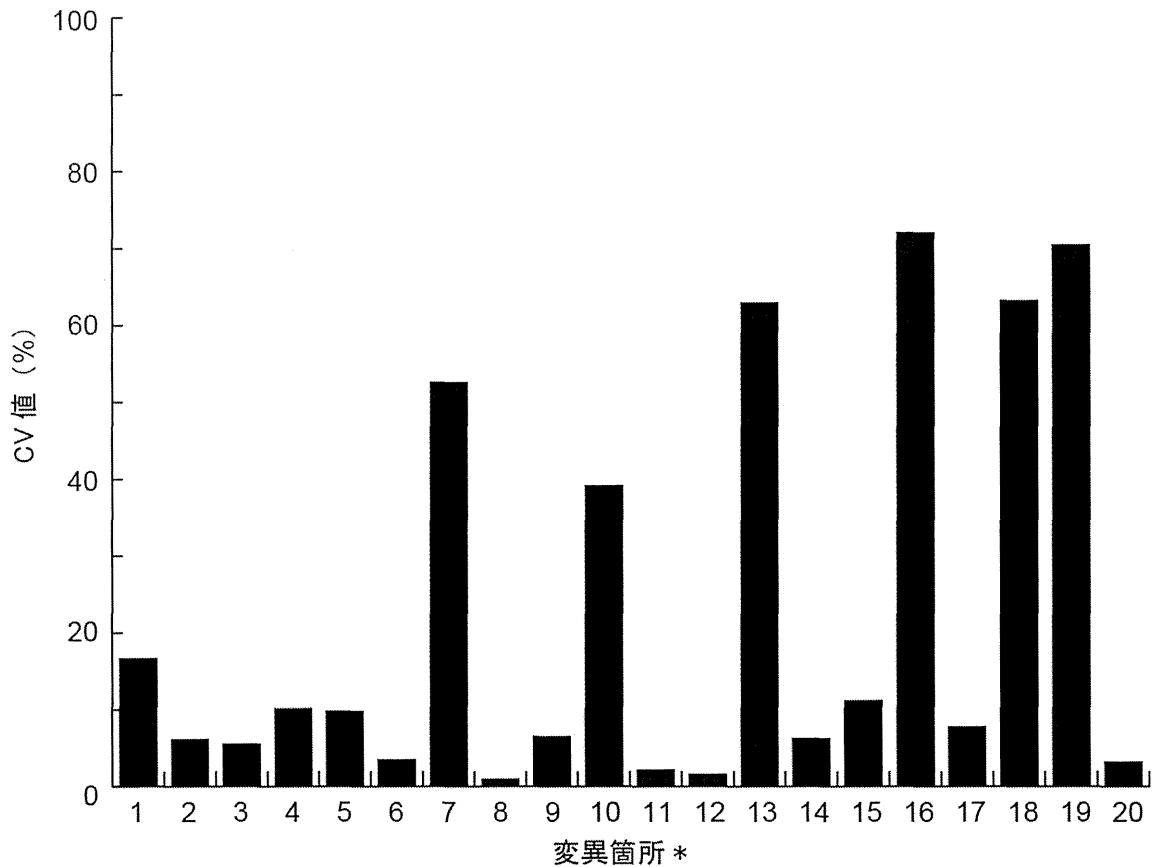


Fig. 28 全マッピング結果を合わせた時の各変異箇所における変異頻度の公表値に対する変動係数 (CV) (公表値に対する実測値のパラツキ) 全マッピング結果をまとめた後, 40Gb, 80Gb, 120Gb, 160Gb 相当のデータとなるようにダウンサンプリングし, 各変異箇所におけるそれぞれのシーケンス量 (40Gb, 80Gb, 120Gb, 160Gb) に相当する塩基頻度を測定後, CV 値を算出した. さらに, 公表値を含めた CV 値を算出し, 実測値および公表値の 2 群間の CV 値を求めた. *横軸の変異箇所については, Table 15 に示した.

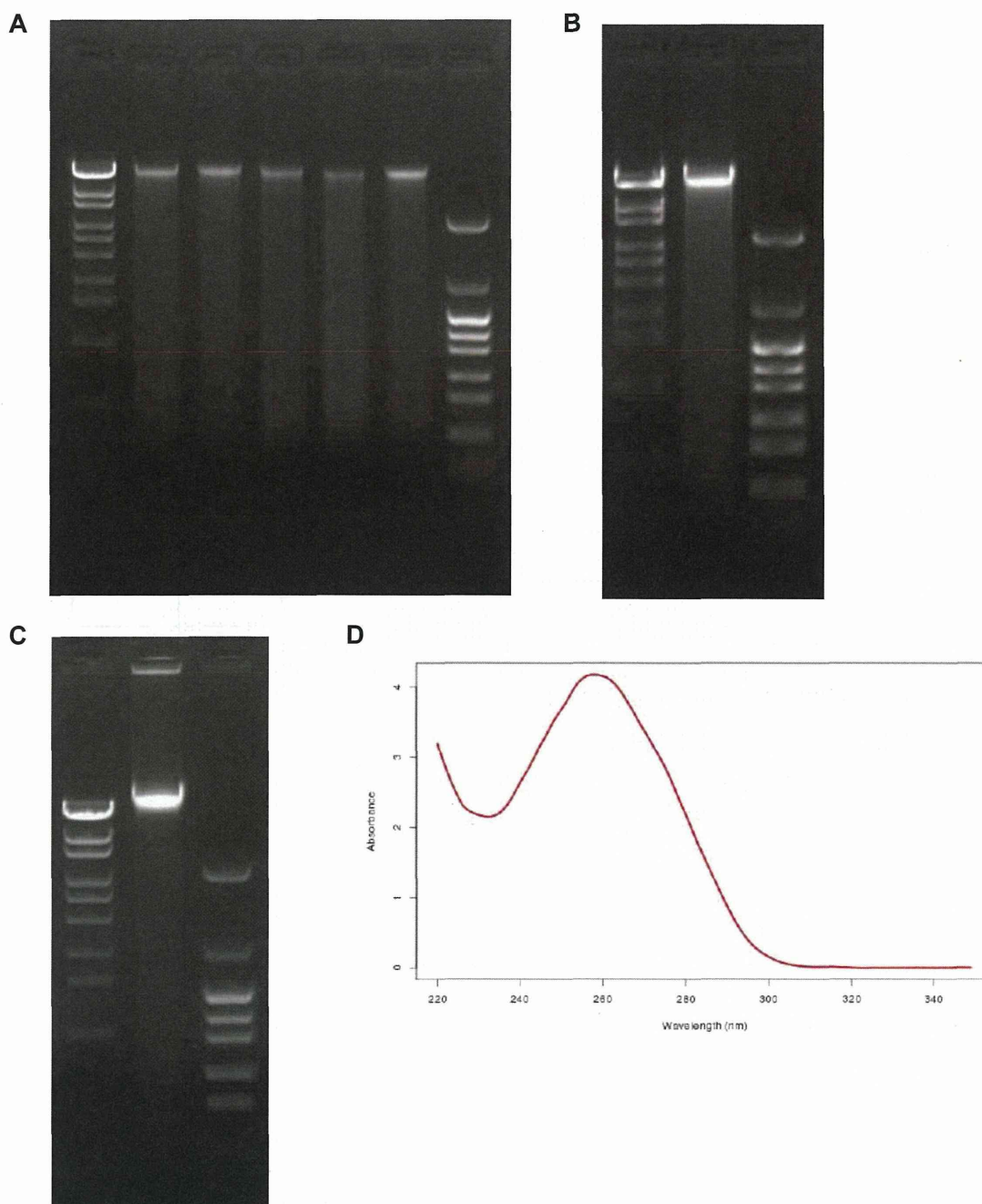


Fig. 29 ゲノム DNA 抽出結果

(A, B) 6人の健常日本人男性(JOM001, JOM002, JOM003, JOM004, JOM005, JOM006)の精液 250 μL から抽出したゲノム DNA の一部を電気泳動した。収量は左側の検体からそれぞれ, 3.2 μg , 1.1 μg , 2.5 μg , 7.7 μg , 3.1 μg , 4.0 μg であった。(C) 左から 5 番目の個人(JOM005) について, さらに 1750 μL の精液から 73.0 μg のゲノム DNA を得た。サイズマーカーは λ -EcoT14I および pHY を用いた。(D) C に示したサンプルに対して吸光定量を行い, 品質の確認を行った。

Table 16 SNP タイピング 2,294,794 サイトの結果

ID	Homo (Ref/Ref)	Homo (Alt/Alt)	Hetero
Sample 1	1,599,066	305,672	385,367
Sample 2	1,599,828	307,376	384,500
Sample 3	1,600,578	307,857	383,589
Sample 4	1,597,215	307,531	387,059
Sample 5	1,600,864	308,225	381,883

Table 17

ID	SNPs1	SNPs2	SNPs3	SNPs4
Sample 1	17	41	4	6
Sample 2	19	37	12	1
Sample 3	24	39	9	2
Sample 4	13	23	7	1
Sample 5	25	30	3	2

SNPs1: 全体と比較して日本人に稀な SNP 群

SNPs2: 欧州人と比較して日本人に稀な SNP 群

SNPs3: 東アジア人と比較して日本人に稀な SNP 群

SNPs4: 東アジア人には稀だが、日本人に多い SNP 群

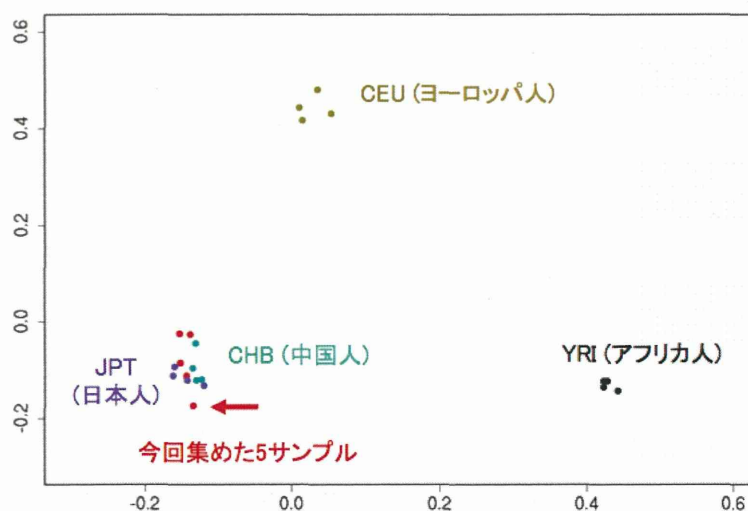


Fig. 30 SNP タイピングのデータをもとに行った主成分分析

1000 人ゲノムから選んだ 16 サンプルと、今回 SNP タイピングを行った 5 サンプル(赤色)との比較. アセンブリ対象の Sample 5 を赤矢印で示した.

Table 18 Illumina HiSeq から得られた塩基数 (Mb: mega base)

Library type	Insert size	Mb	Mb (Q > 15)	Read length
Paired end	260 bp	269,140	261,298	150 nt × 2
Paired end	360 bp	137,934	132,559	125 nt × 2
Paired end	660 bp	161,552	154,689	150 nt × 2
Mate pair	2 kb	102,230	97,729	150 nt × 2
Mate pair	5 kb	45,245	43,738	150 nt × 2
Mate pair	9 kb	46,134	44,512	150 nt × 2
Total	N/a	762,235	601,966	N/a

Table 19 トリミング後の塩基数

Library type	Insert size	Trimmed Mb	Mb (Q > 15)	Actual size
Paired end	260 bp	259,596	257,686	240 bp
Paired end	360 bp	135,509	131,901	330 bp
Paired end	660 bp	154,125	152,814	640 bp
Mate pair	2 kb	59,506	58,999	1.6 kb
Mate pair	5 kb	27,084	26,920	4.6 kb
Mate pair	9 kb	27,584	27,410	8.6 kb
Total	N/a	663,404	655,730	N/a

Table 20 ALLPATHS-LG によるゲノムサイズ等の推定

一倍体ゲノムサイズ *	2,741,890,408 bp
G+C 含量	40.2%
反復配列の割合 (K = 25)	27.0%

* GRCh37/hg19 の参考値: 3,137,161,264 bp

Table 21 ALLPATHS-LG による de novo アセンブリの結果

contig の最短長	1000 bp
全 contig 数	485,318
全 scaffold 数	333,852
総 contig 長	2,254,772,438 bp
総 scaffold 長 (gap を含む)	2,535,926,213 bp
N50 contig	9.3 kb
N50 scaffold	39 kb
N50 scaffold (gap を含む)	45 kb

Table 22 PacBio RS II から得られたデータ

PacBio RS II	ライブラリ S1	ライブラリ S2	S1 + S2
塩基数	9,146 Mb	35,694 Mb	44,840 Mb
リード数	757,979	3,673,817	4,431,796
平均リード長	12,066 bp	9716 bp	N/a
最長リード長	49,865 bp	53,506 bp	53,506 bp

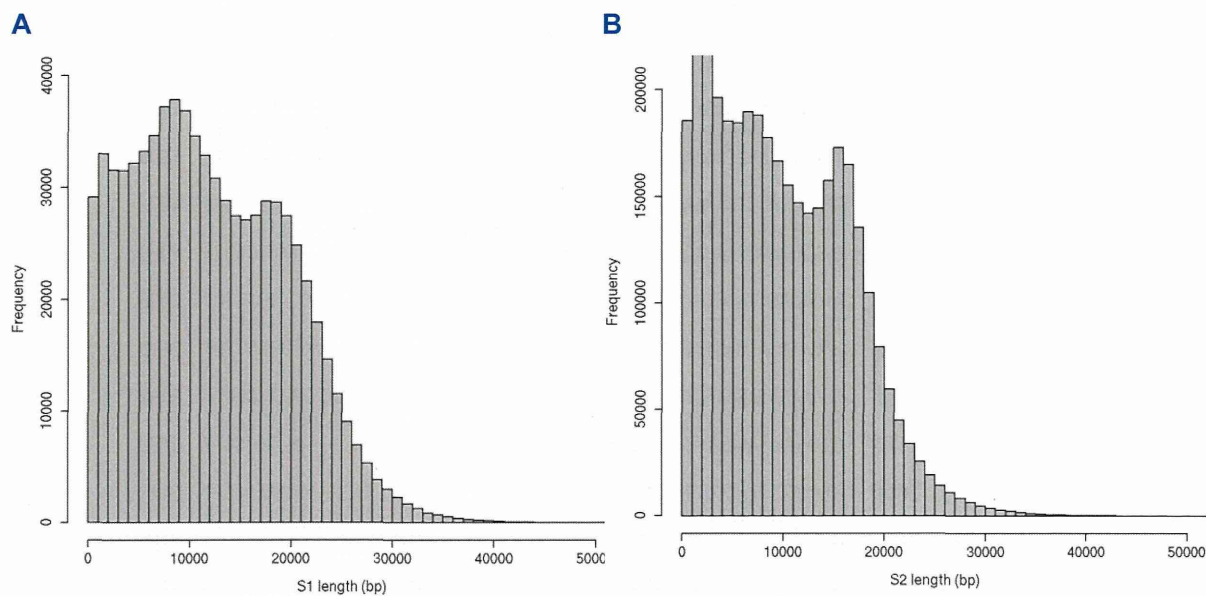


Fig. 31 PacBio RS II から得られたロングリードの長さの分布

(A) ライブラリ S1 から得られたリード長の分布. (B) ライブラリ S2 から得られたリード長の分布. S2 から得られた配列データの方が多いため、縦軸のスケールが一致していないことに注意.