進化するRNA-Seq：臨床検体からシングルセル解析まで ～ウェット・ドライ解析の実験ノート

東京大学
新領域創成科学研究科

鈴木　穣

東大・柏キャンパス

Hiseq2500 x 3

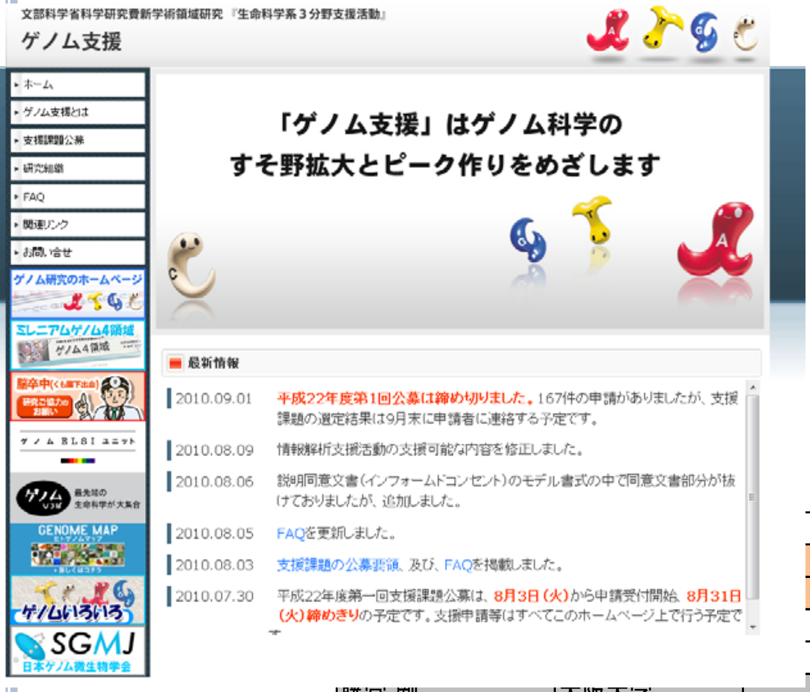
Operation:

Technicians 4

Programmers 3

ysuzuki@hgc.jp

“ゲノム支援”



文部科学省科学研究費新学術領域研究『生命科学系３分野支援活動』

ゲノム支援

「ゲノム支援」はゲノム科学の
すそ野拡大とピーク作りをめざします

最新情報

2010.09.01 平成22年度第1回公募は締め切りました。167件の申請がありましたが、支援課題の選定結果は9月末に申請者に連絡する予定です。

2010.08.09 情報解析支援活動の支援可能な内容を修正しました。

2010.08.06 説明同意文書(インフォームドコンセント)のモデル書式の中で同意文書部分が抜けておりましたが、追加しました。

2010.08.05 FAQを更新しました。

2010.08.03 支援課題の公募要領、及び、FAQを掲載しました。

2010.07.30 平成22年度第一回支援課題公募は、8月3日(火)から申請受付開始、8月31日(火)締めきりの予定です。支援申請等はすべてこのホームページ上で行う予定で

Providing NGS platform for researchers
in various research field
http://www.genome-sci.jp/

			胎児型腎臓幹細胞の成体腎での再活性化
			次世代シークエンサーを用いた生殖系列のエピゲノム修飾とトランスクリプトーム解析
		5	種内雑種を利用した対立遺伝子間の優劣に関わるDNAメチル化機構の解析
			メリステム制御の基盤を支える植物幹細胞の不等分裂の分子機構の解明
			トゲウオ科魚類における種分化の遺伝機構
藤堂 剛	大阪大学		メダカ逆遺伝学的手法を基盤とした個体・組織レベルでの損傷応答解析系の確立
太田 邦史	東京大学	8	長鎖非翻訳RNAを介したクロマチン／染色体機能の制御
武田 洋幸 (森下BS)	東京大学		組織が創るマクロでロバストなコンパートメントの成立・維持のロジック
深田 吉孝	東京大学		脳時計ニューロンにおける光シグナリングと概日リズム制御の分子解析
多羽田 哲也	東京大学		ショウジョウバエの記憶形成回路の構造および機能発現の分子基盤
三谷 啓志	東京大学		個体内における電離放射線誘発突然変異成立過程の解明
平良 眞規	東京大学		転写制御ネットワークから見る原口形成と原腸胚オーガナイザーの進化のメカニズム
國枝 武和	東京大学		極限環境耐性動物クマムシが獲得した耐性メカニズムの解明
稲田 利文	名古屋大学理学研究科		新生ポリペプチド鎖依存の翻訳アレストにおけるRACK1の機能解明
高浜 洋介	徳島大学		胸腺における自己形成と自己認識
嶋田 透	東京大学		カイコとその近縁種における寄主植物選択機構の進化
田中 知明	千葉大学		p53転写因子複合体によるクロマチン機能調節とiPSリプログラム制御機構の解明
後藤 由季子	東京大学		胎生期大脳新皮質神経幹細胞による多様な細胞の産生機構の解析
坂山 英俊	神戸大学		陸上植物の2倍体多細胞体制の起源をシャジクモ藻類の遺伝子から探る
三室 仁美	東京大学		ヘリコバクターピロリの胃粘膜感染機構と炎症惹起メカニズムの研究
國府 力	大阪大学		初期発生におけるクロマチン制御のリアルタイム解析
田中 知明	千葉大学		転写因子p53による新たな代謝調節機能と代謝環境応答のエピジェネティクス制御
福澤 秀哉	京都大学		デジタル遺伝子発現解析による微細藻類のCO2濃縮・水素発生関連遺伝子の同定と

RNA Seqの分類

タグ数をカウントするもの　(36bp Single End Read)

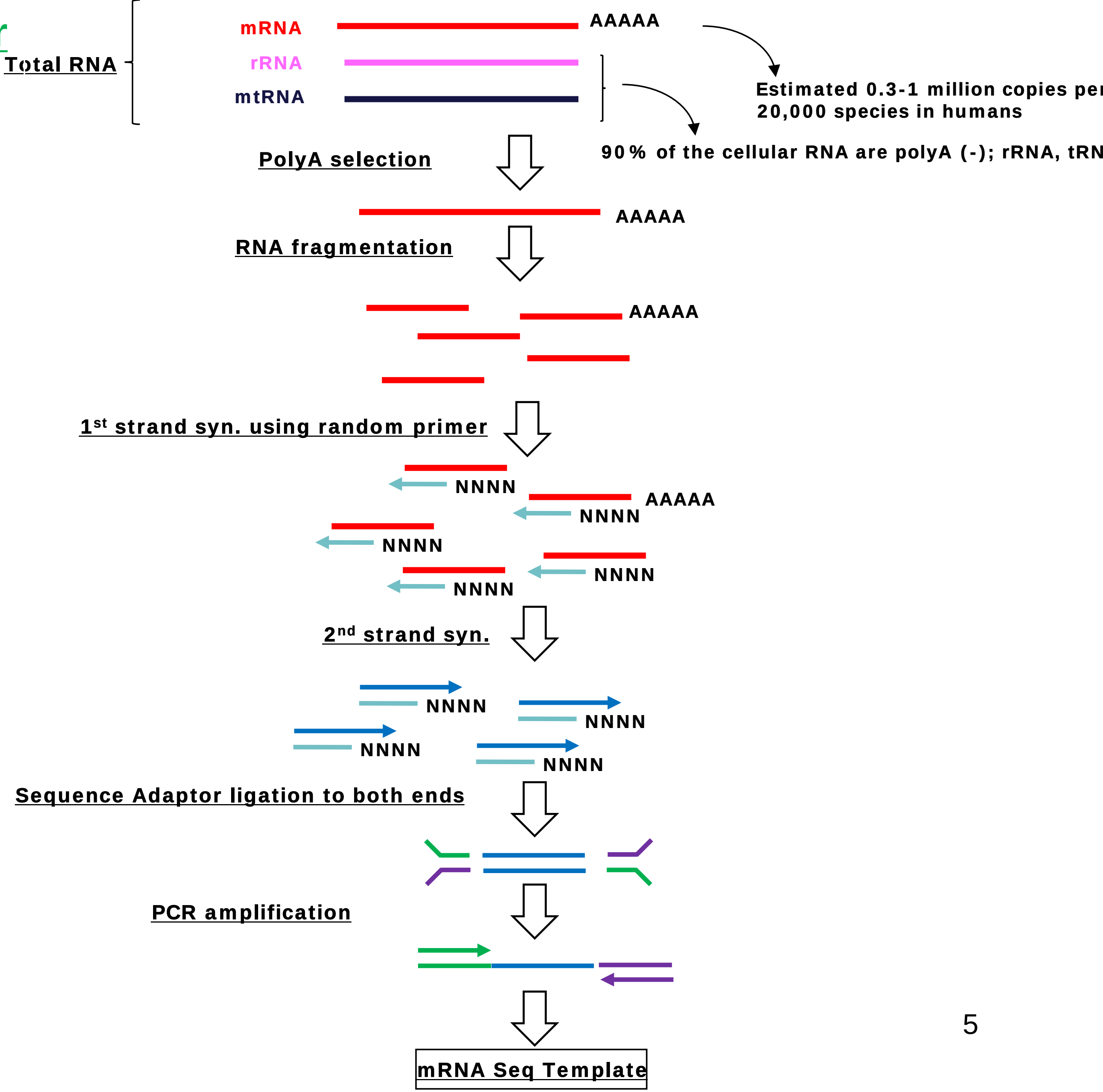
- 発現量を計測するもの
  - (mRNA) RNA Seq
  - small RNA Seq
- タンパク質との相互作用を計測するもの　RIP Seq / CLIP Seq

配列を決定するもの (>100 bp Paired End Read)

- 遺伝子アノテーションするもの　de novo アセンブリ

  mRNA Seq
- 選択的スプライシングを解析するもの

Template Prep. for RNA Seq



Total RNA

mRNA AAAAA

rRNA

mtRNA

Estimated 0.3-1 million copies per
20,000 species in humans

90% of the cellular RNA are polyA (-); rRNA, tRN

PolyA selection

AAAAA

RNA fragmentation

AAAAA

1st strand syn. using random primer

NNNN

AAAAA

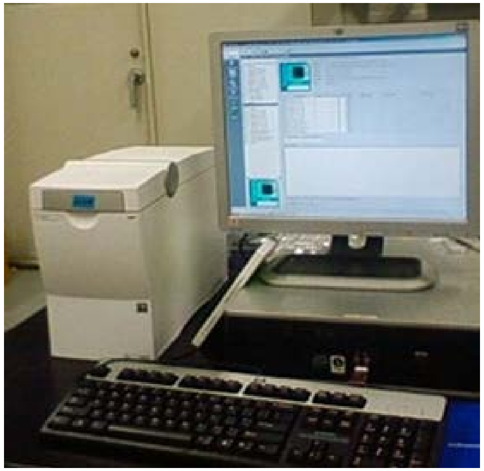
2nd strand syn.

Sequence Adaptor ligation to both ends

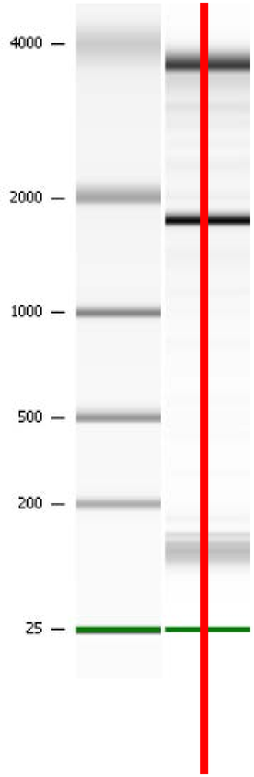
PCR amplification

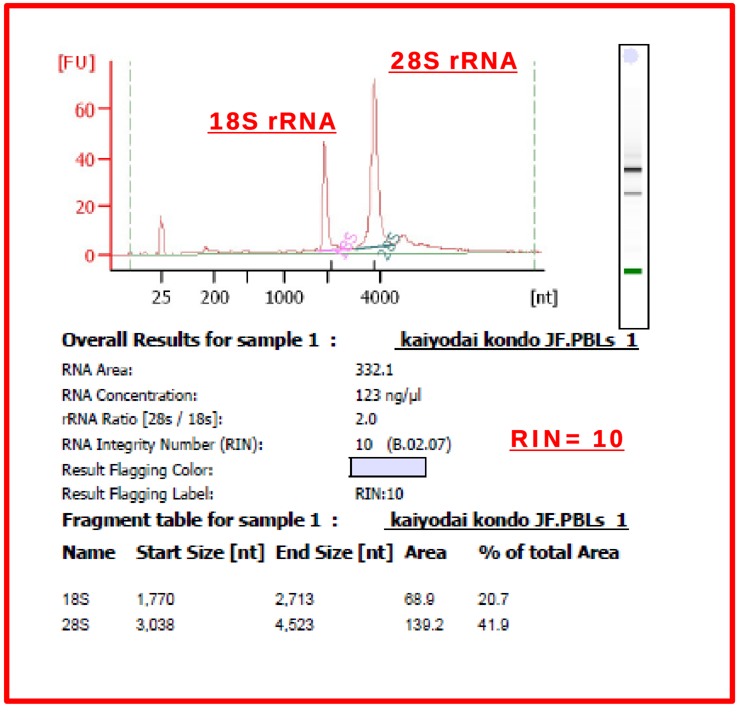
mRNA Seq Template

BioAnalyzer is essential for sample preparation



BioAnalyzer (Agilent):
Electrophoresis on microchip





Overall Results for sample 1 : kaiyodai kondo JF.PBLs 1

RNA Area:	332.1
RNA Concentration:	123 ng/µl
rRNA Ratio [28s / 18s]:	2.0
RNA Integrity Number (RIN):	10 (B.02.07)
Result Flagging Color:	▭
Result Flagging Label:	RIN:10

RIN= 10

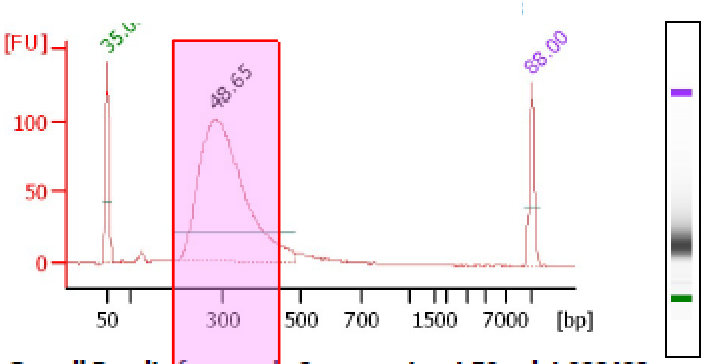
Fragment table for sample 1 : kaiyodai kondo JF.PBLs 1

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,770	2,713	68.9	20.7
28S	3,038	4,523	139.2	41.9

Dissection

Advantages in using BioAnalyzer (I)

effective material (250-450 bp)



Overall Results for sample 2 : input 50ng lot 022433

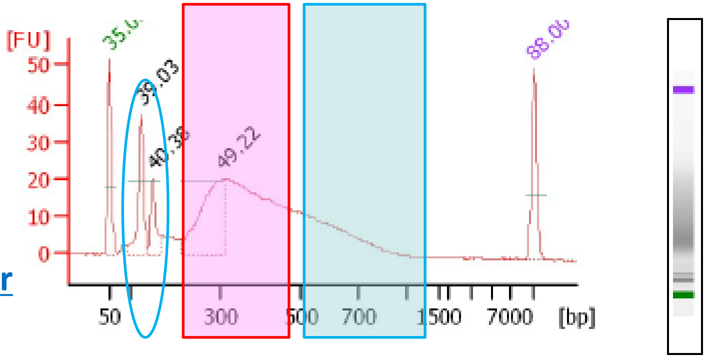
Number of peaks found: 1

Peak table for sample 2 : input 50ng lot 022433

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
2	285	48.18	256.0	

effective material (250-450 bp) non-effective material

Primer dimer



Overall Results for sample 5 : input 50ng lot023554

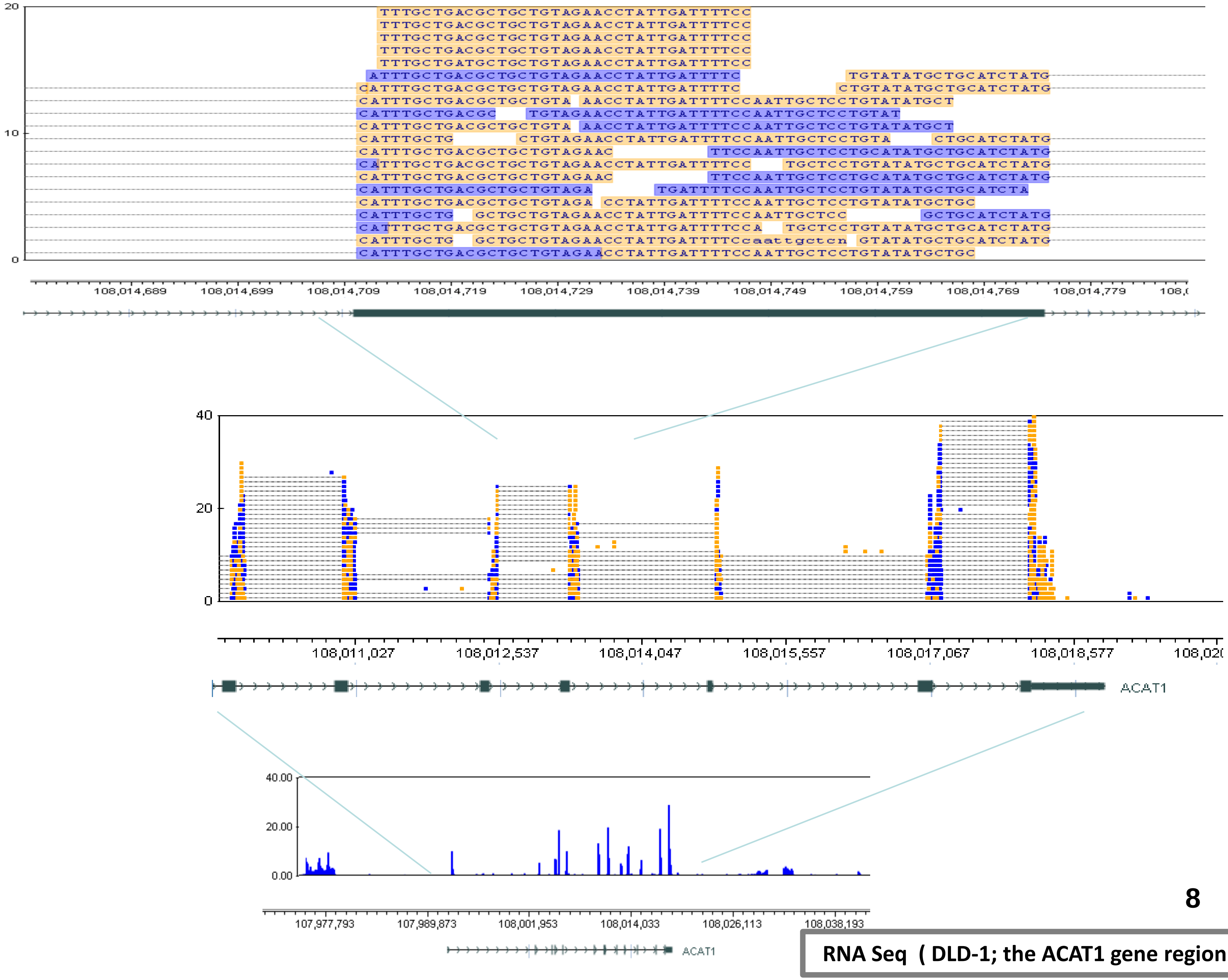
Number of peaks found: 3

Peak table for sample 5 : input 50ng lot023554

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
2	125	6.84	82.7	
3	149	3.40	34.5	
4	307	12.36	61.0	

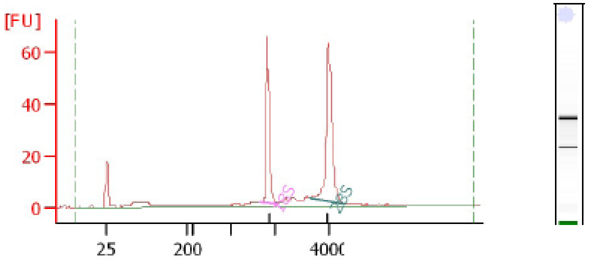
To measure effective template amount

Examples of NGS data (RNA Seq on Genome Studio Viewer)


20
10
0
TTTGCTGACGCTGCTGTAGAACCTATTGATTTTCC
TTTGCTGACGCTGCTGTAGAACCTATTGATTTTCC
TTTGCTGACGCTGCTGTAGAACCTATTGATTTTCC
TTTGCTGACGCTGCTGTAGAACCTATTGATTTTCC
TTTGCTGATGCTGCTGTAGAACCTATTGATTTTCC
ATTTGCTGACGCTGCTGTAGAACCTATTGATTTTC TGTATATGCTGCATCTATG
CATTTGCTGACGCTGCTGTAGAACCTATTGATTTTC CTGTATATGCTGCATCTATG
CATTTGCTGACGCTGCTGTA AACCTATTGATTTTCCAATTGCTCCTGTATATGCT
CATTTGCTGACGC TGTAGAACCTATTGATTTTCCAATTGCTCCTGTAT
CATTTGCTGACGCTGCTGTA AACCTATTGATTTTCCAATTGCTCCTGTATATGCT
CATTTGCTG CTGTAGAACCTATTGATTTTCCAATTGCTCCTGTA CTGCATCTATG
CATTTGCTGACGCTGCTGTAGAAC TTCCAATTGCTCCTGCATATGCTGCATCTATG
CATTTGCTGACGCTGCTGTAGAACCTATTGATTTTCC TGCTCCTGTATATGCTGCATCTATG
CATTTGCTGACGCTGCTGTAGAAC TTCCAATTGCTCCTGCATATGCTGCATCTATG
CATTTGCTGACGCTGCTGTAGA TGATTTTCCAATTGCTCCTGTATATGCTGCATCTA
CATTTGCTGACGCTGCTGTAGA CCTATTGATTTTCCAATTGCTCCTGTATATGCTGC
CATTTGCTG GCTGCTGTAGAACCTATTGATTTTCCAATTGCTCC GCTGCATCTATG
CATTTGCTGACGCTGCTGTAGAACCTATTGATTTTCCA TGCTCCTGTATATGCTGCATCTATG
CATTTGCTG GCTGCTGTAGAACCTATTGATTTTCcaattgctcn GTATATGCTGCATCTATG
CATTTGCTGACGCTGCTGTAGAACCTATTGATTTTCCAATTGCTCCTGTATATGCTGC
108,014,689
108,014,699
108,014,709
108,014,719
108,014,729
108,014,739
108,014,749
108,014,759
108,014,769
108,014,779
108,0
40
20
0
108,011,027
108,012,537
108,014,047
108,015,557
108,017,067
108,018,577
108,02
ACAT1
40.00
20.00
0.00
107,977,793
107,989,873
108,001,953
108,014,033
108,026,113
108,038,193
ACAT1

RNA Seq (DLD-1; the ACAT1 gene region)

Increasing number of templates

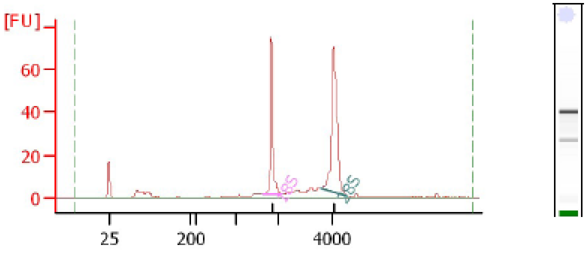


Overall Results for sample 1 : S

RNA Area:	241.1
RNA Concentration:	107 ng,
rRNA Ratio [28s / 18s]:	1.7
RNA Integrity Number (RIN):	9.8 (E
Result Flagging Color:	▭
Result Flagging Label:	RIN: 9.

Fragment table for sample 1 : S

Name	Start Size [nt]	End Size [nt]
18S	1,665	2,155
28S	3,299	4,624

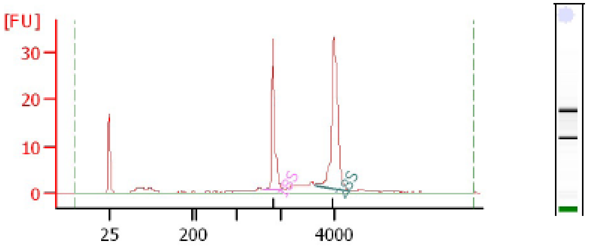


Overall Results for sample 2 : Samp

RNA Area:	264.6
RNA Concentration:	117 ng/µl
rRNA Ratio [28s / 18s]:	1.6
RNA Integrity Number (RIN):	9.8 (B.02.07
Result Flagging Color:	▭
Result Flagging Label:	RIN: 9.80

Fragment table for sample 2 : Sampl

Name	Start Size [nt]	End Size [nt]	Area
18S	1,648	2,163	64.0
28S	3,575	4,570	102.7

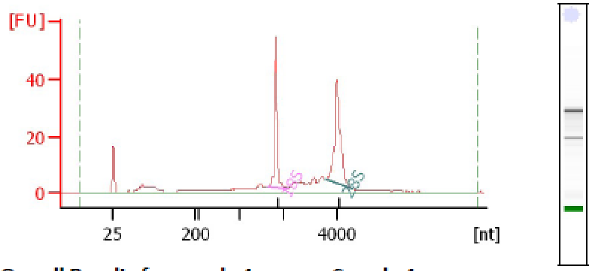


Overall Results for sample 3 : Sam

RNA Area:	139.9
RNA Concentration:	62 ng/µl
rRNA Ratio [28s / 18s]:	1.7
RNA Integrity Number (RIN):	9.8 (B.02.0
Result Flagging Color:	▭
Result Flagging Label:	RIN: 9.80

Fragment table for sample 3 : Sampl

Name	Start Size [nt]	End Size [nt]	Are
18S	1,631	2,134	31.5
28S	3,291	4,567	54.8



Overall Results for sample 4 : Sample 4

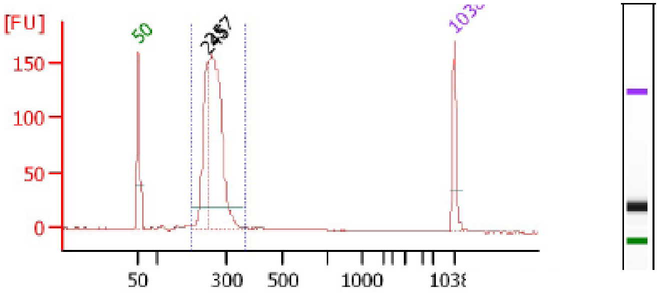
RNA Area:	230.9
RNA Concentration:	103 ng/µl
rRNA Ratio [28s / 18s]:	1.3
RNA Integrity Number (RIN):	9 (B.02.07)
Result Flagging Color:	▭
Result Flagging Label:	RIN:9

Fragment table for sample 4 : Sample 4

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,655	2,206	45.4	19.7
28S	3,576	4,589	60.0	26.0

Such as time-course RNA Seq analysis

For fair comparison of multiple data points



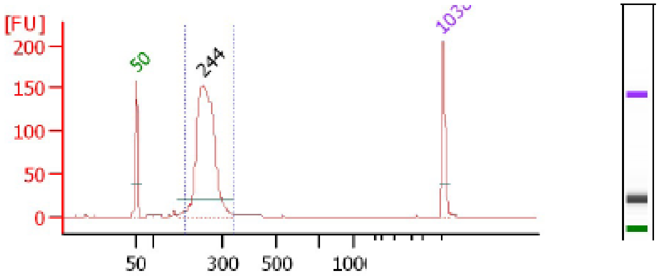
Overall Results for sample 5 : **Sample !**

Number of peaks found: 2

Area 1: 523.8

Region table for sample 5 : **Sample 5**

From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distributi CV [%]
201	367	523.8	95	263	8.8



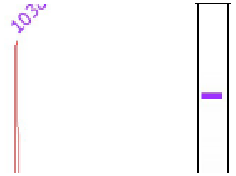
Overall Results for sample 6 :

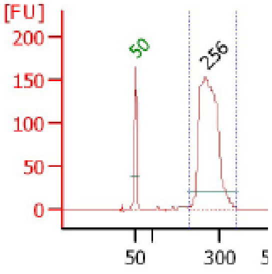
Number of peaks found: 1

Area 1: 491.

Region table for sample 6 : **Sa**

From [bp]	To [bp]	Area	% of Total	Average Size [bp]
191	343	491.7	93	255





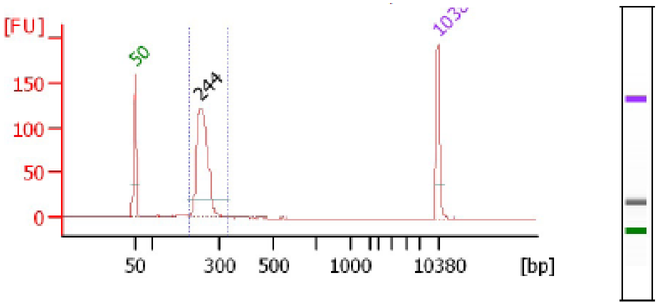
Overall Results for sample

Number of peaks found:

Area 1:

Region table for sample 7

From [bp]	To [bp]	Area	% of Total
205	361	525.2	94



Overall Results for sample 8 : **Sample 8**

Number of peaks found: 1

Area 1: 253.4

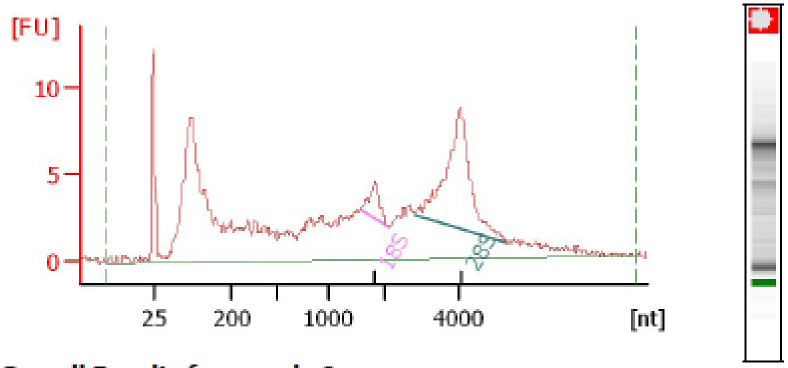
Region table for sample 8 : **Sample 8**

From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Co lor
207	334	253.4	92	252	6.4	15.60	🟦

Uniform sample prep is essential

Occasionally, “irregular samples” should be also handled

Total RNA from operation material



Overall Results for sample 8 :

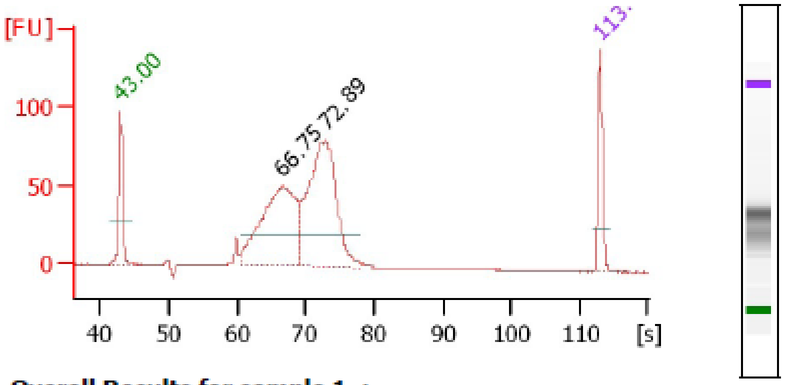
RNA Area:	248.0
RNA Concentration:	81 ng/µl
rRNA Ratio [28s / 18s]:	6.7
RNA Integrity Number (RIN):	N/A (B.02.07)
Result Flagging Color:	▨
Result Flagging Label:	RIN N/A

Fragment table for sample 8 :

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,608	2,134	4.9	2.0
28S	2,852	5,337	32.8	13.2

RIN N/A; but this is still RNA!

“irregular” template



Overall Results for sample 1 :

Number of peaks found: 2

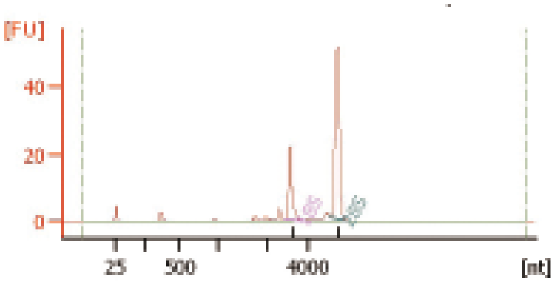
Peak table for sample 1

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀ 15	4.20	424.2	Lower Marker
2	210	10.95	78.9	
3	264	12.49	71.7	
4	▶ 1,500	2.10	2.1	Upper Marker

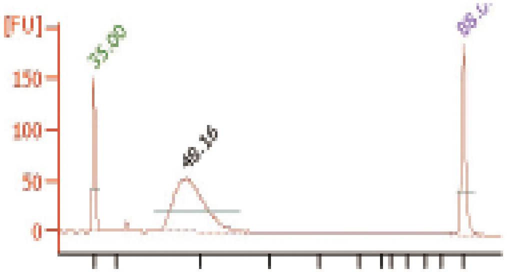
トマトのトランスクリプトーム解析 （成熟葉、老化葉）

試料調整とシークエンス

組織からのRNAの抽出
（1 μg total RNA)



シークエンスライブラリーの作成 (450ng library)



シークエンスと配列解析
(0.2ng library)

GAIIx；36-base single-end read: 1 lane

microTomゲノムへのマッピング

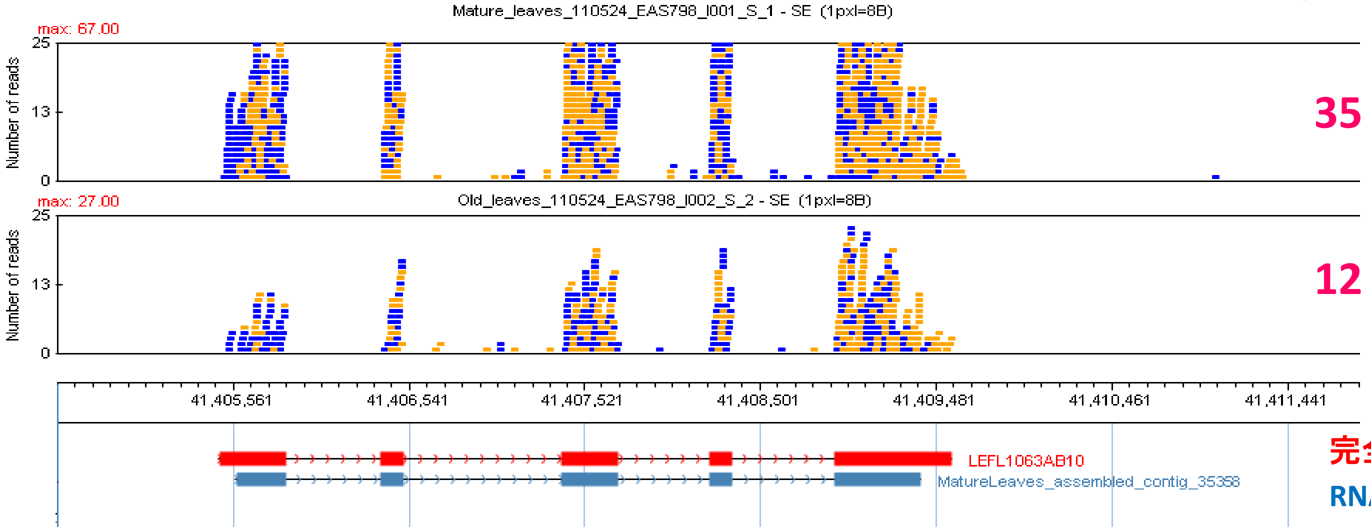
microTom完全長cDNAへのマッピング

De novo assemble (AbySS)

Sequence Summary

Tissue	# reads (36bp)	# Assembled contigs 500bp< / 1k < / 1.5k<	%Matched with cDNA 500bp< / 1k < / 1.5k<	%Matched with tBLASTX < 1e-50 500bp< / 1k < / 1.5k<
mature leaves	29,923,071	7,165/ 2,304/834	4,648/1,456/467	6,866/ 2,280/828
old leaves	28,711,676	6,118/1,890/653	4,001/1,199/361	5,869/1,871/649

完全長cDNAへの発現情報の付加



Expression level

35 rpkm

12 rpkm

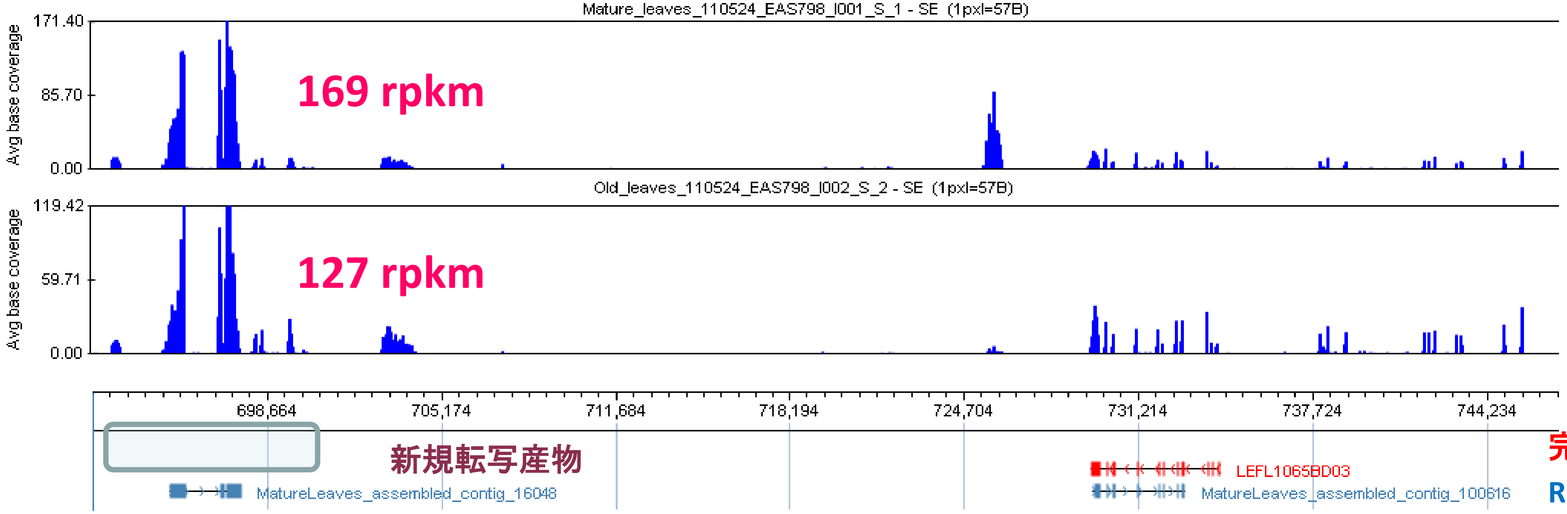
完全長cDNA

RNA Seq assembled contig

(rpkm: read per million tags per kb mRNA)

新規転写産物の発見

Expression level


Mature_leaves_110524_EAS798_l001_S_1 - SE (1pxl=57B)
171.40
85.70
0.00
Avg base coverage
Old_leaves_110524_EAS798_l002_S_2 - SE (1pxl=57B)
119.42
59.71
0.00
Avg base coverage
698,664
705,174
711,684
718,194
724,704
731,214
737,724
744,234
LEFL1065BD03
MatureLeaves_assembled_contig_16048
MatureLeaves_assembled_contig_100616

169 rpkm

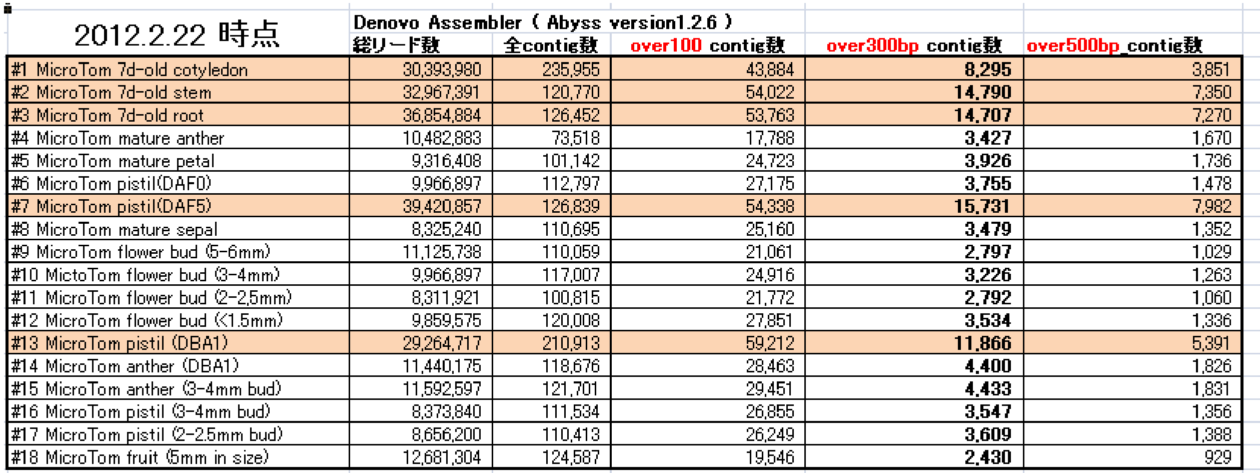
127 rpkm

新規転写産物

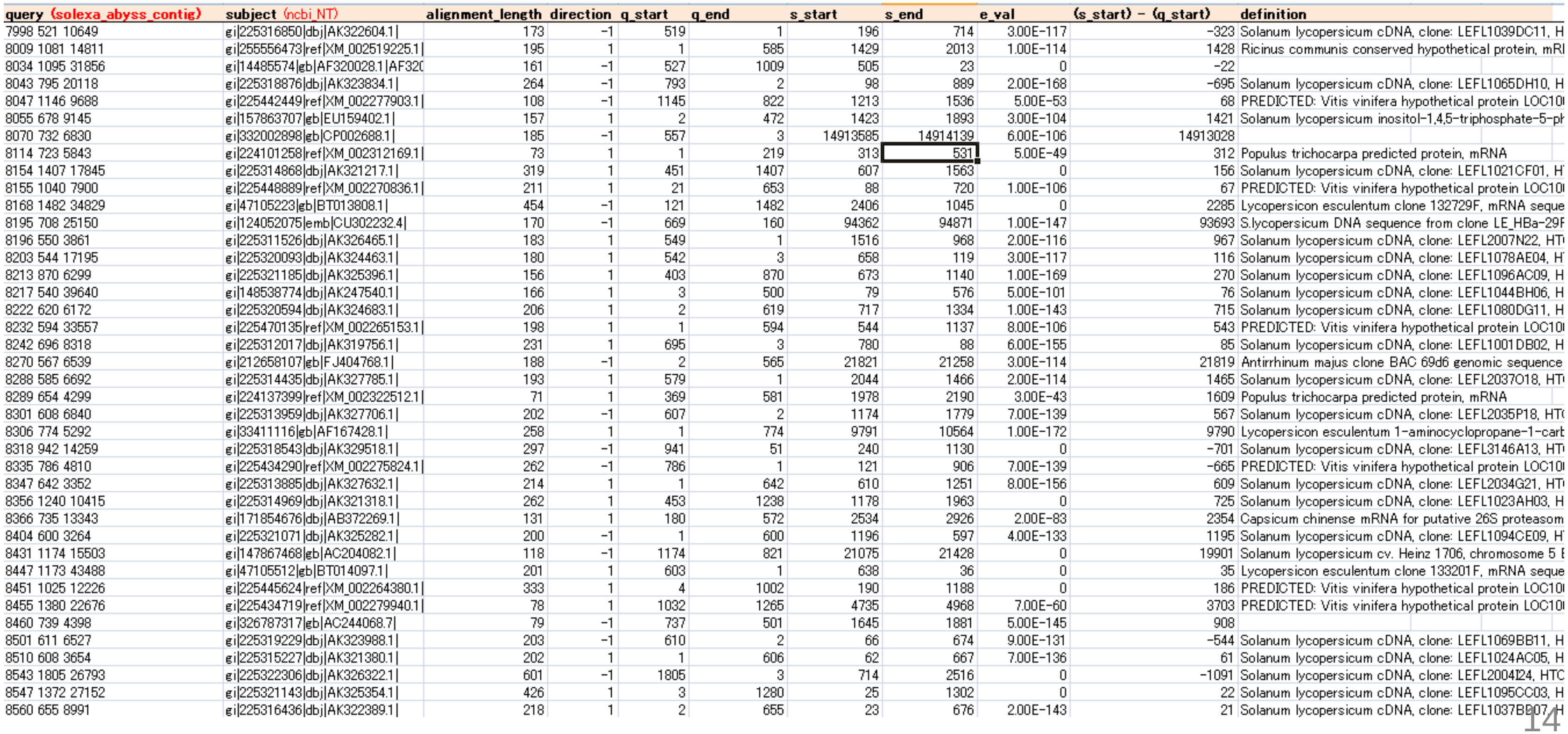
完全長cDNA

RNA Seq assembled contig

De novo assembly of microTom transcripts and their annotations



2012.2.22 時点	Denovo Assembler (Abyss version1.2.6)				
	総リード数	全contig数	over100 contig数	over300bp contig数	over500bp_contig数
#1 MicroTom 7d-old cotyledon	30,393,980	235,955	43,884	**8,295**	3,851
#2 MicroTom 7d-old stem	32,967,391	120,770	54,022	**14,790**	7,350
#3 MicroTom 7d-old root	36,854,884	126,452	53,763	**14,707**	7,270
#4 MicroTom mature anther	10,482,883	73,518	17,788	**3,427**	1,670
#5 MicroTom mature petal	9,316,408	101,142	24,723	**3,926**	1,736
#6 MicroTom pistil(DAF0)	9,966,897	112,797	27,175	**3,755**	1,478
#7 MicroTom pistil(DAF5)	39,420,857	126,839	54,338	**15,731**	7,982
#8 MicroTom mature sepal	8,325,240	110,695	25,160	**3,479**	1,352
#9 MicroTom flower bud (5-6mm)	11,125,738	110,059	21,061	**2,797**	1,029
#10 MictoTom flower bud (3-4mm)	9,966,897	117,007	24,916	**3,226**	1,263
#11 MicroTom flower bud (2-2.5mm)	8,311,921	100,815	21,772	**2,792**	1,060
#12 MicroTom flower bud (<1.5mm)	9,859,575	120,008	27,851	**3,534**	1,336
#13 MicroTom pistil (DBA1)	29,264,717	210,913	59,212	**11,866**	5,391
#14 MicroTom anther (DBA1)	11,440,175	118,676	28,463	**4,400**	1,826
#15 MicroTom anther (3-4mm bud)	11,592,597	121,701	29,451	**4,433**	1,831
#16 MicroTom pistil (3-4mm bud)	8,373,840	111,534	26,855	**3,547**	1,356
#17 MicroTom pistil (2-2.5mm bud)	8,656,200	110,413	26,249	**3,609**	1,388
#18 MicroTom fruit (5mm in size)	12,681,304	124,587	19,546	**2,430**	929



query (solexa_abyss_contig)	subject (ncbi_NT)	alignment_length	direction	q_start	q_end	s_start	s_end	e_val	(s_start) - (q_start)	definition
7998 521 10649	gi|225316850|dbj|AK322604.1|	173	-1	519	1	196	714	3.00E-117	-323	Solanum lycopersicum cDNA, clone: LEFL1039DC11, H
8009 1081 14811	gi|255556473|ref|XM_002519225.1|	195	1	1	585	1429	2013	1.00E-114	1428	Ricinus communis conserved hypothetical protein, mR
8034 1095 31856	gi|14485574|gb|AF320028.1|AF32(	161	-1	527	1009	505	23	0	-22	
8043 795 20118	gi|225318876|dbj|AK323834.1|	264	-1	793	2	98	889	2.00E-168	-695	Solanum lycopersicum cDNA, clone: LEFL1065DH10, H
8047 1146 9688	gi|225442449|ref|XM_002277903.1|	108	-1	1145	822	1213	1536	5.00E-53	68	PREDICTED: Vitis vinifera hypothetical protein LOC10
8055 678 9145	gi|157863707|gb|EU159402.1|	157	1	2	472	1423	1893	3.00E-104	1421	Solanum lycopersicum inositol-1,4,5-triphosphate-5-ph
8070 732 6830	gi|332002898|gb|CP002688.1|	185	-1	557	3	14913585	14914139	6.00E-106	14913028	
8114 723 5843	gi|224101258|ref|XM_002312169.1|	73	1	1	219	313	531	5.00E-49	312	Populus trichocarpa predicted protein, mRNA
8154 1407 17845	gi|225314868|dbj|AK321217.1|	319	1	451	1407	607	1563	0	156	Solanum lycopersicum cDNA, clone: LEFL1021CF01, H
8155 1040 7900	gi|225448889|ref|XM_002270836.1|	211	1	21	653	88	720	1.00E-106	67	PREDICTED: Vitis vinifera hypothetical protein LOC10
8168 1482 34829	gi|47105223|gb|BT013808.1|	454	-1	121	1482	2406	1045	0	2285	Lycopersicon esculentum clone 132729F, mRNA seque
8195 708 25150	gi|124052075|emb|CU302232.4|	170	-1	669	160	94362	94871	1.00E-147	93693	S.lycopersicum DNA sequence from clone LE_HBa-29F
8196 550 3861	gi|225311526|dbj|AK326465.1|	183	1	549	1	1516	968	2.00E-116	967	Solanum lycopersicum cDNA, clone: LEFL2007N22, HT
8203 544 17195	gi|225320093|dbj|AK324463.1|	180	1	542	3	658	119	3.00E-117	116	Solanum lycopersicum cDNA, clone: LEFL1078AE04, H
8213 870 6299	gi|225321185|dbj|AK325396.1|	156	1	403	870	673	1140	1.00E-169	270	Solanum lycopersicum cDNA, clone: LEFL1096AC09, H
8217 540 39640	gi|148538774|dbj|AK247540.1|	166	1	3	500	79	576	5.00E-101	76	Solanum lycopersicum cDNA, clone: LEFL1044BH06, H
8222 620 6172	gi|225320594|dbj|AK324683.1|	206	1	2	619	717	1334	1.00E-143	715	Solanum lycopersicum cDNA, clone: LEFL1080DG11, H
8232 594 33557	gi|225470135|ref|XM_002265153.1|	198	1	1	594	544	1137	8.00E-106	543	PREDICTED: Vitis vinifera hypothetical protein LOC10
8242 696 8318	gi|225312017|dbj|AK319756.1|	231	1	695	3	780	88	6.00E-155	85	Solanum lycopersicum cDNA, clone: LEFL1001DB02, H
8270 567 6539	gi|212658107|gb|FJ404768.1|	188	-1	2	565	21821	21258	3.00E-114	21819	Antirrhinum majus clone BAC 69d6 genomic sequence
8288 585 6692	gi|225314435|dbj|AK327785.1|	193	1	579	1	2044	1466	2.00E-114	1465	Solanum lycopersicum cDNA, clone: LEFL2037O18, HT
8289 654 4299	gi|224137399|ref|XM_002322512.1|	71	1	369	581	1978	2190	3.00E-43	1609	Populus trichocarpa predicted protein, mRNA
8301 608 6840	gi|225313959|dbj|AK327706.1|	202	-1	607	2	1174	1779	7.00E-139	567	Solanum lycopersicum cDNA, clone: LEFL2035P18, HT
8306 774 5292	gi|33411116|gb|AF167428.1|	258	1	1	774	9791	10564	1.00E-172	9790	Lycopersicon esculentum 1-aminocyclopropane-1-carb
8318 942 14259	gi|225318543|dbj|AK329518.1|	297	-1	941	51	240	1130	0	-701	Solanum lycopersicum cDNA, clone: LEFL3146A13, HT
8335 786 4810	gi|225434290|ref|XM_002275824.1|	262	-1	786	1	121	906	7.00E-139	-665	PREDICTED: Vitis vinifera hypothetical protein LOC10
8347 642 3352	gi|225313885|dbj|AK327632.1|	214	1	1	642	610	1251	8.00E-156	609	Solanum lycopersicum cDNA, clone: LEFL2034G21, HT
8356 1240 10415	gi|225314969|dbj|AK321318.1|	262	1	453	1238	1178	1963	0	725	Solanum lycopersicum cDNA, clone: LEFL1023AH03, H
8366 735 13343	gi|171854676|dbj|AB372269.1|	131	1	180	572	2534	2926	2.00E-83	2354	Capsicum chinense mRNA for putative 26S proteasom
8404 600 3264	gi|225321071|dbj|AK325282.1|	200	-1	1	600	1196	597	4.00E-133	1195	Solanum lycopersicum cDNA, clone: LEFL1094CE09, H
8431 1174 15503	gi|147867468|gb|AC204082.1|	118	-1	1174	821	21075	21428	0	19901	Solanum lycopersicum cv. Heinz 1706, chromosome 5 E
8447 1173 43488	gi|47105512|gb|BT014097.1|	201	1	603	1	638	36	0	35	Lycopersicon esculentum clone 133201F, mRNA seque
8451 1025 12226	gi|225445624|ref|XM_002264380.1|	333	1	4	1002	190	1188	0	186	PREDICTED: Vitis vinifera hypothetical protein LOC10
8455 1380 22676	gi|225434719|ref|XM_002279940.1|	78	1	1032	1265	4735	4968	7.00E-60	3703	PREDICTED: Vitis vinifera hypothetical protein LOC10
8460 739 4398	gi|326787317|gb|AC244068.7|	79	-1	737	501	1645	1881	5.00E-145	908	
8501 611 6527	gi|225319229|dbj|AK323988.1|	203	-1	610	2	66	674	9.00E-131	-544	Solanum lycopersicum cDNA, clone: LEFL1069BB11, H
8510 608 3654	gi|225315227|dbj|AK321380.1|	202	1	1	606	62	667	7.00E-136	61	Solanum lycopersicum cDNA, clone: LEFL1024AC05, H
8543 1805 26793	gi|225322306|dbj|AK326322.1|	601	-1	1805	3	714	2516	0	-1091	Solanum lycopersicum cDNA, clone: LEFL2004I24, HTC
8547 1372 27152	gi|225321143|dbj|AK325354.1|	426	1	3	1280	25	1302	0	22	Solanum lycopersicum cDNA, clone: LEFL1095CC03, H
8560 655 8991	gi|225316436|dbj|AK322389.1|	218	1	2	655	23	676	2.00E-143	21	Solanum lycopersicum cDNA, clone: LEFL1037BD07, H

ある魚類のdenovo

● data process

Solexa Read 76PE

(Pass Filtered , remove the read including N)

↓

AbySS (version 1.2.6)

↓

> 500bp contig 抽出

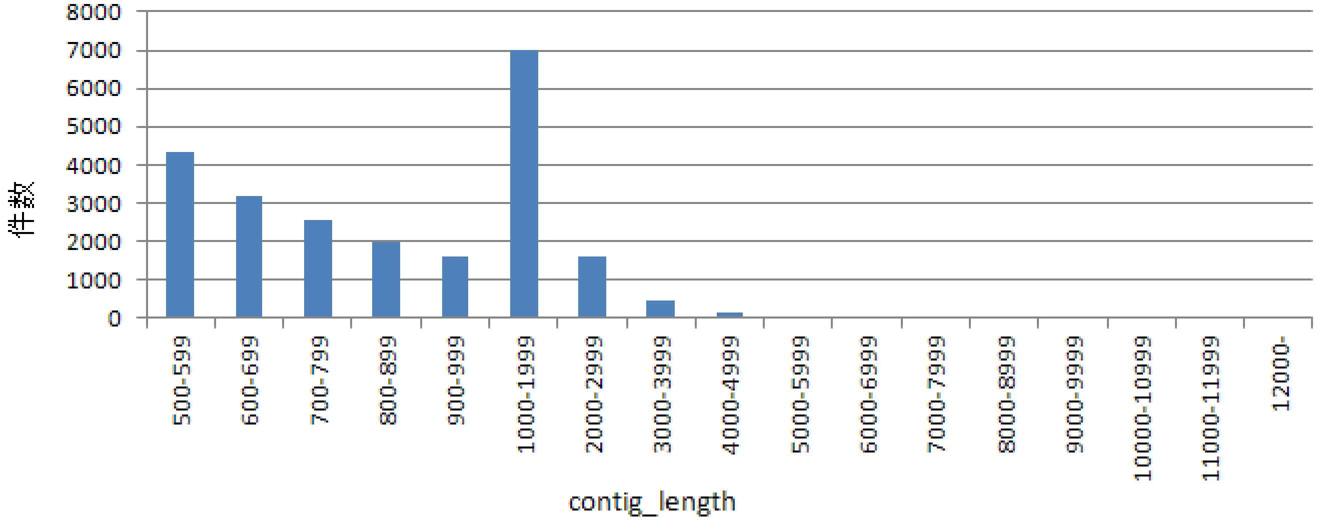
↓ ↘

tBlastX (Query:contig , DB: NT)

ELAND (Ref:contig)

●assemble result

Sample	# Reads (76bp)	# Assembled contigs 500bp< Average contig length	#Matched with tBLASTX < 1e-50 500bp<
JDPBLs-1	46,771,912	23,045 （Average 1,141bp)	11,549

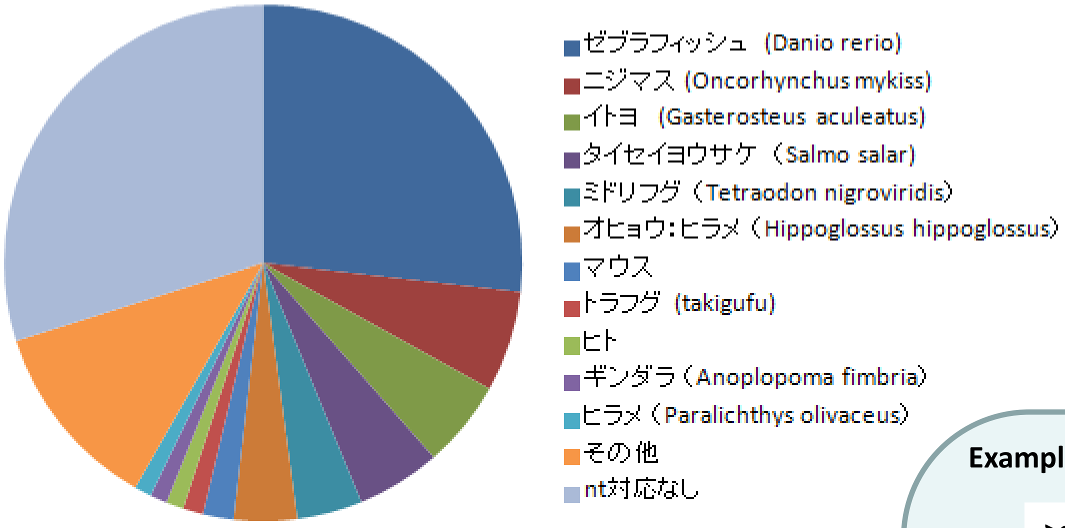


contig_length	件数
500-599	4323
600-699	3190
700-799	2561
800-899	1959
900-999	1599
1000-1999	6992
2000-2999	1633
3000-3999	487
4000-4999	163
5000-5999	75
6000-6999	30
7000-7999	18
8000-8999	6
9000-9999	6
10000-10999	1
11000-11999	
12000-	2
total	23045

近藤研との共同研究

ある魚類のdenovo

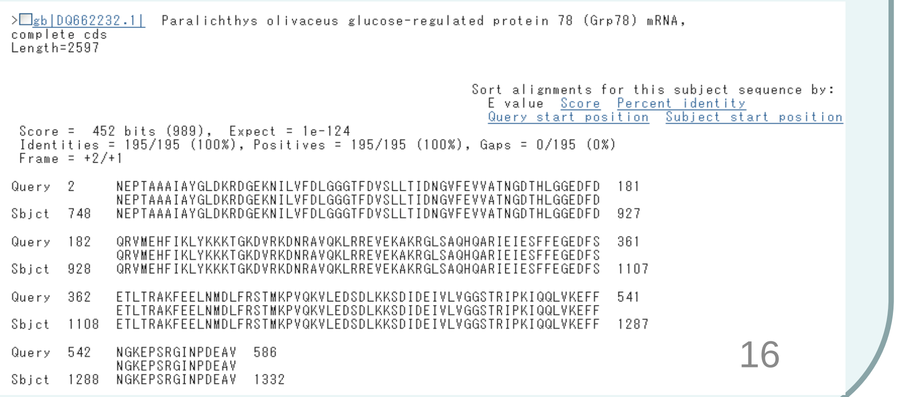
●tblastx assembled contig to NT



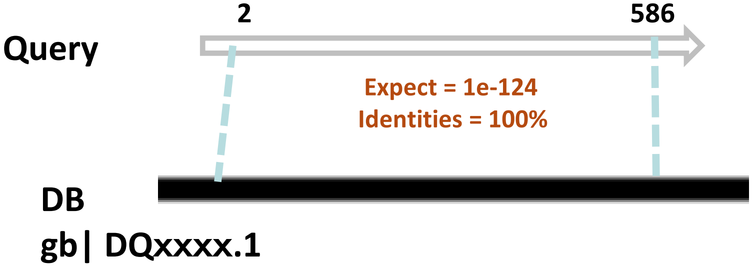
tblastx結果　内訳	
ゼブラフィッシュ（*Danio rerio*）	27%
ニジマス（*Oncorhynchus mykiss*）	6%
イトヨ（*Gasterosteus aculeatus*）	5%
タイセイヨウサケ（*Salmo salar*）	5%
ミドリフグ（*Tetraodon nigroviridis*）	4%
オヒョウ:ヒラメ（*Hippoglossus hippoglossus*）	4%
マウス	2%
トラフグ（*takigufu*）	1%
ヒト	1%
ギンダラ（*Anoplopoma fimbria*）	1%
ヒラメ（Paralichthys olivaceus）	1%
その他	12%
nt対応なし	30%

Example: xxx Assembled contig : Query length 588bp

```
>contig_102559 588 97855
CAATGAGCCAACTGCTGCTGCCATTGCTTATGGTCTGGACAAGAGAGATGGCGAGAAGAACATTCTTGT
GTTCGATCTGGGTGGCGGCACCTTCGATGTCTCCCTCTTGACCATCGACAATGGTGTGTTTGAAGTGGTG
GCCACCAACGGTGACACTCACCTGGGAGGTGAGGACTTCGACCAGCGCGTCATGGAGCACTTCATCAAG
CTGTACAAGAAGAAAACTGGCAAAGATGTGCGCAAAGACAACCGTGCTGTGCAGAAGCTGCGTCGTGA
GGTTGAGAAGGCAAAGAGGGGGCTGTCCGCCCAGCACCAGGCCCGCATTGAGATCGAGTCCTTCTTTGA
GGGAGAAGACTTCTCTGAGACTCTGACCCGTGCCAAGTTTGAAGAGCTGAACATGGACCTGTTCCGTTCC
ACCATGAAGCCTGTGCAGAAGGTGCTGGAAGATTCCGACCTGAAGAAATCTGACATCGATGAGATTGTC
CTGGTTGGAGGCTCCACCCGTATCCCCAAAATTCAGCAGCTGGTGAAGGAGTTCTTCAATGGCAAGGAGC
CATCTAGGGGCATCAACCCTGATGAGGCTGTGGC
```



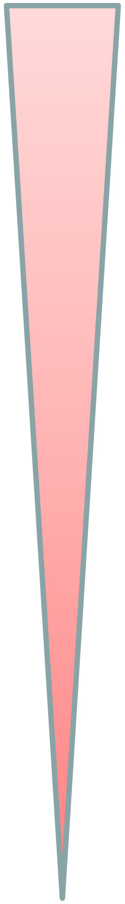
```
>□gb|DQ662232.1|  Paralichthys olivaceus glucose-regulated protein 78 (Grp78) mRNA,
complete cds
Length=2597


                                        Sort alignments for this subject sequence by:
                                          E value  Score  Percent identity
                                          Query start position  Subject start position
 Score =  452 bits (989),  Expect = 1e-124
 Identities = 195/195 (100%), Positives = 195/195 (100%), Gaps = 0/195 (0%)
 Frame = +2/+1

Query  2     NEPTAAAIAYGLDKRDGEKNILVFDLGGGTFDVSLLTIDNGVFEVVATNGDTHLGGEDFD  181
             NEPTAAAIAYGLDKRDGEKNILVFDLGGGTFDVSLLTIDNGVFEVVATNGDTHLGGEDFD
Sbjct  748   NEPTAAAIAYGLDKRDGEKNILVFDLGGGTFDVSLLTIDNGVFEVVATNGDTHLGGEDFD  927

Query  182   QRVMEHFIKLYKKKTGKDVRKDNRAVQKLRREVEKAKRGLSAQHQARIEIESFFEGEDFS  361
             QRVMEHFIKLYKKKTGKDVRKDNRAVQKLRREVEKAKRGLSAQHQARIEIESFFEGEDFS
Sbjct  928   QRVMEHFIKLYKKKTGKDVRKDNRAVQKLRREVEKAKRGLSAQHQARIEIESFFEGEDFS  1107

Query  362   ETLTRAKFEELNMDLFRSTMKPVQKVLEDSDLKKSDIDEIVLVGGSTRIPKIQQLVKEFF  541
             ETLTRAKFEELNMDLFRSTMKPVQKVLEDSDLKKSDIDEIVLVGGSTRIPKIQQLVKEFF
Sbjct  1108  ETLTRAKFEELNMDLFRSTMKPVQKVLEDSDLKKSDIDEIVLVGGSTRIPKIQQLVKEFF  1287

Query  542   NGKEPSRGINPDEAV  586
             NGKEPSRGINPDEAV
Sbjct  1288  NGKEPSRGINPDEAV  1332
```



contig領域 タグ集計		
tag	ppm	rpkm
2035	94	159.86

鋳型調整

出発材料量



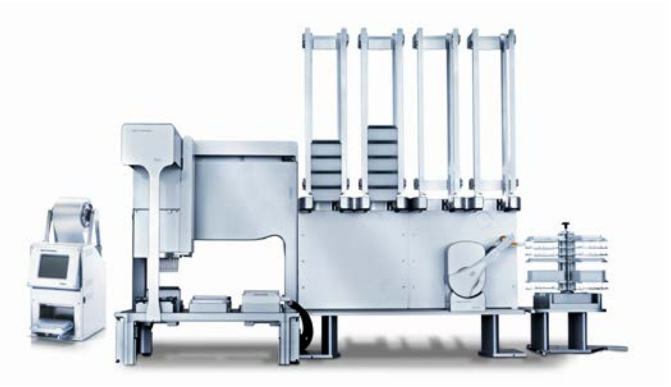
>200ng

>10ng

100-1000細胞

1細胞

Illumina/Agilent RNA Seq



QIAGEN RepliG

Clontech Smarter

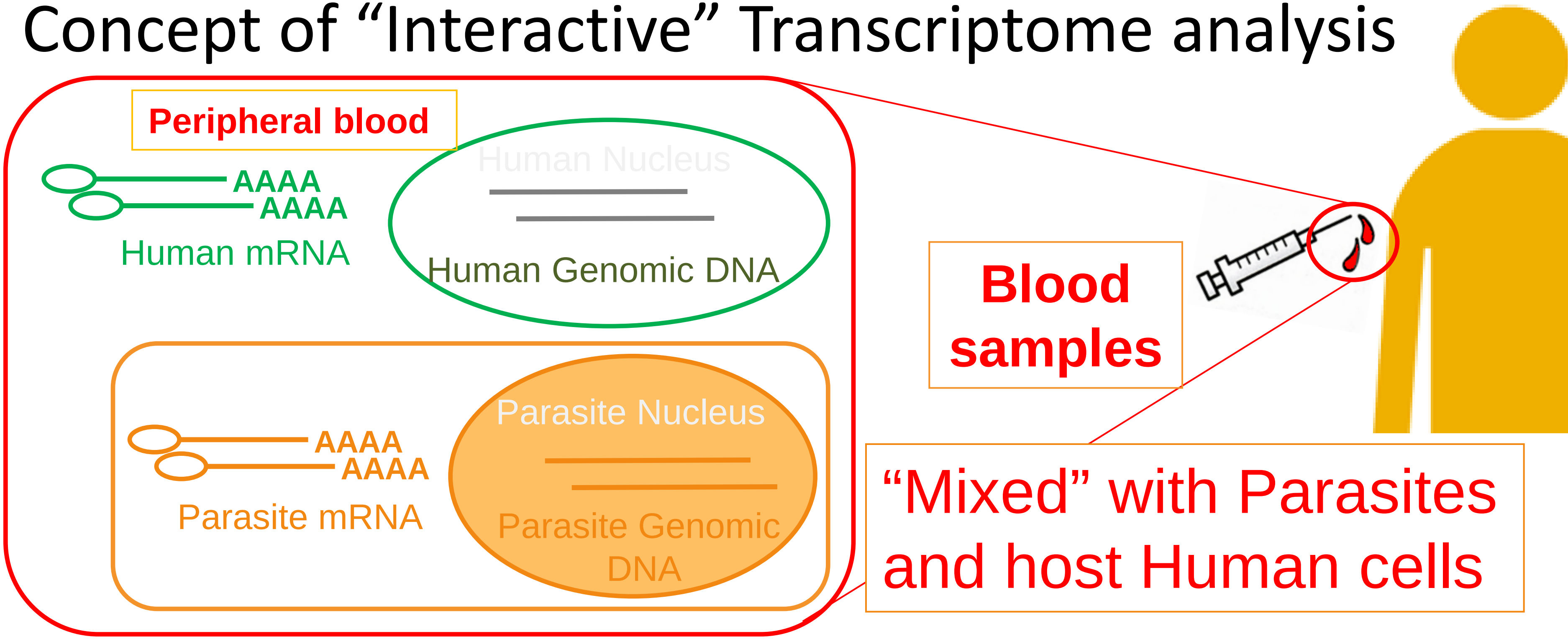
情報解析

鈴木絢子・鈴木穣　実験医学増刊　印刷中

用途	ソフトウェア	URL	概要
マッピング	BWA	http://bio-bwa.sourceforge.net/	ショートリードをゲノムにマッピングする（Li H. and Durbin R. 2009 *Bioinformatics*）。
	Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	ショートリードを少ないメモリで参照配列に高速にアライメントする（Langmead and Steven L Salzberg. 2012 *Nat Methods*）。
	TopHat2	http://tophat.cbcb.umd.edu/	スプライスジャンクションを考慮したマッピングをおこなう（Kim et al. 2013 *Genome Biol*）。
遺伝子発現解析	Cufflinks	http://cufflinks.cbcb.umd.edu/	異なるスプライスバリアントごとの発現量の計算や新規転写産物のアセンブルを行う（Trapnell et al. 2010 *Nat Biotechnol*）。
	Cuffdiff	同上	Cufflinksのコマンドの一つ。群間の発現量やスプライスパターンの差異を検出する（Trapnell et al. 2013 Nat Biotechnol）
	DEseq	http://bioconductor.org/packages/release/bioc/html/DESeq.html	群間のRNA Seqタグ数や発現量の差を統計的に抽出する（Anders and Huber. 2010 *Genome Biol*）。
融合遺伝子探索	TopHat-fusion	http://tophat.cbcb.umd.edu/fusion_index.html	TopHat2ベースで、シングルまたはペアエンドリードから融合遺伝子を抽出する（Kim and Salzberg. 2011 *Genome Biol*）。
	deFuse	http://compbio.bccrc.ca/software/defuse/	ペアエンドのRNA Seqリードから、融合部位を抽出する（McPherson et al. 2011 *PLoS Comput Biol*）。
	SOAPfuse	http://soap.genomics.org.cn/soapfuse.html	ペアエンドのRNA Seqリードから、融合部位を抽出する（Jia et al. 2013 *Genome Biol*）。
アセンブル	Trans-Abyss	http://www.bcgsc.ca/platform/bioinfo/software/trans-abyss	トランスクリプトームde novoアセンブラ（Robertson et al. 2010 *Nat Methods*）。
	Trinity	http://trinityrnaseq.sourceforge.net/	ショートリード向けのトランスクリプトームアセンブラ。必要なメモリ量は大きい（Grabherr et al. 2011 *Nat Biotechnol*）。
可視化ツール	UCSC Genome Browser	http://genome.ucsc.edu/cgi-bin/hgGateway	データをアップロードして表示することができる（Kent et al. 2002 *Genome Res*）。
	IGV	https://www.broadinstitute.org/igv/home	BAM、BEDファイルなどを簡単に可視化でき、操作性が高い（Robinson et al. 2011 *Nat Biotechnol*）。

Yamagishi et al Genome Res (2014)

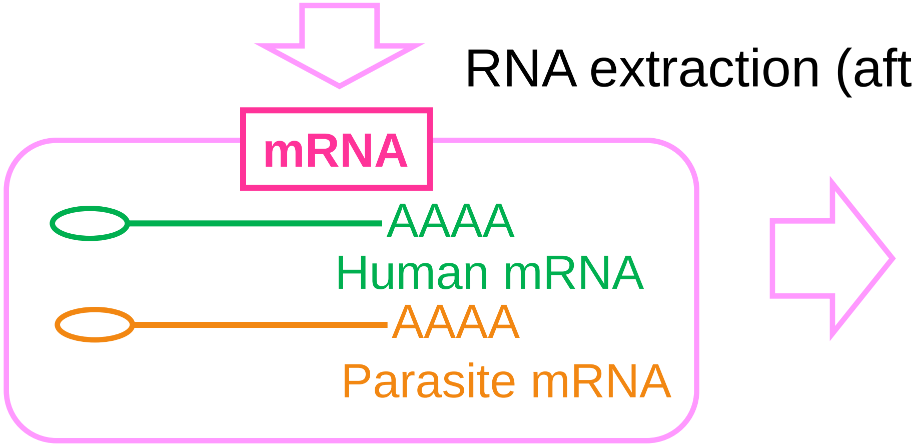
Concept of "Interactive" Transcriptome analysis



Blood samples

"Mixed" with Parasites and host Human cells

RNA extraction (after shipping to Japan)



RNA Seq

After generating sequence tags, species were separated by mapping tags to the respective genomes

To avoid delicate material handling in fields

To monitor human gene expressions simultaneously

Read Statistics (malaria patients)

	Human	***P. falciparum***
Number of samples	116 (24 from Manado, 92 from Bitung)	
Total number of mapped reads	3,016,323,916 (25M reads on average)	
Number of mapped reads	2,794,371,292	244,767,495
Average frequency of parasite reads	10.2%	

新技術：方法論の多様化

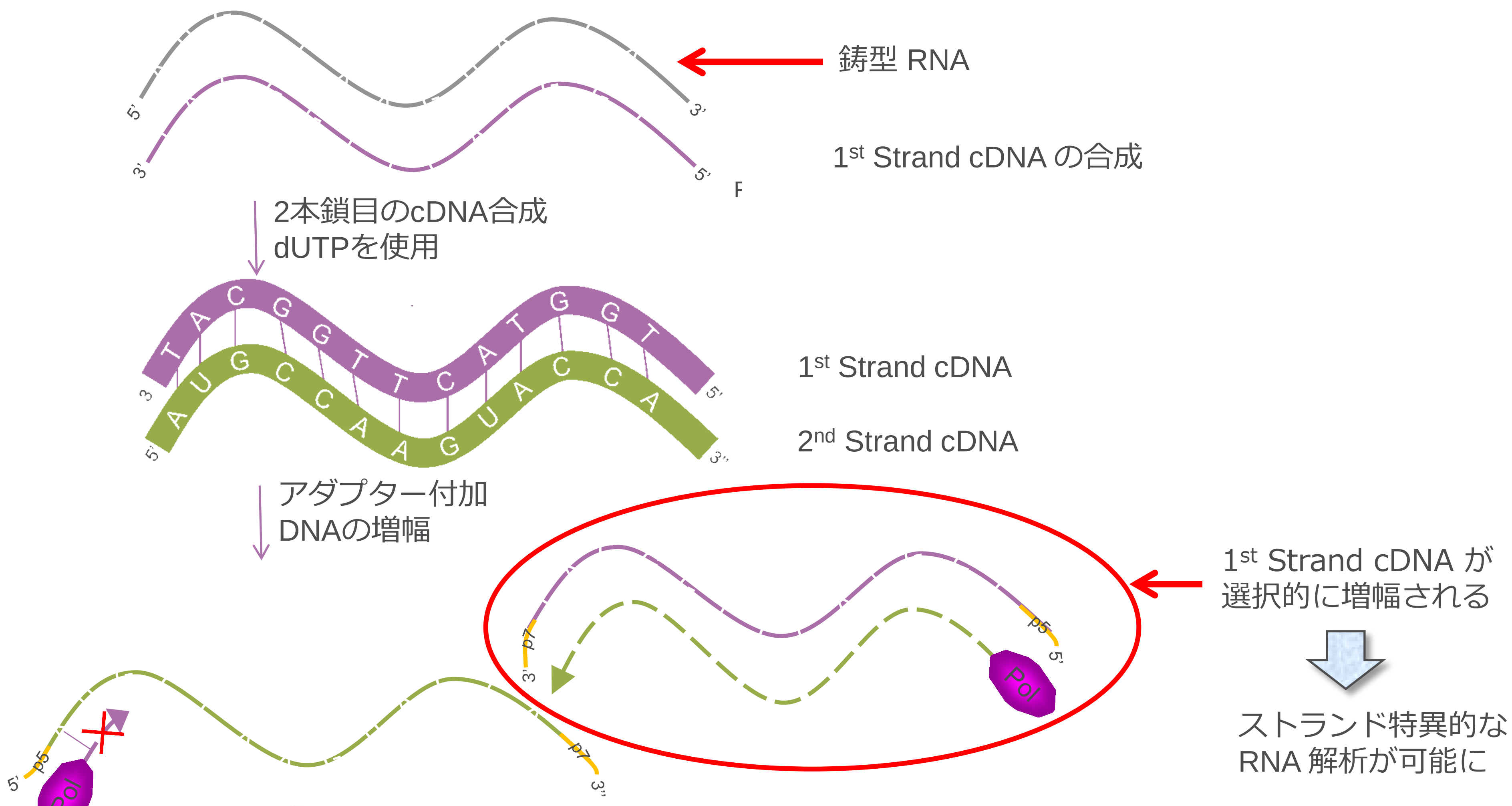
illumina®

TruSeq® Stranded mRNA Sample Preparation Guide

アジレント SureSelect
Strand-Specific RNA ライブラリ調製
イルミナマルチプレックスシーケンス対応

Whole-Transcriptome ライブラリ調製プロトコル

2本鎖目のcDNA合成時にdUTPを使用することでこの鎖が増幅されず、ストランド情報を維持



鋳型 RNA

1st Strand cDNA の合成

2本鎖目のcDNA合成
dUTPを使用

1st Strand cDNA

2nd Strand cDNA

アダプター付加
DNAの増幅

1st Strand cDNA が
選択的に増幅される

ストランド特異的な
RNA 解析が可能に

ポイント

- デオキシウラシル (dUTP) を鋳型に使えないDNAポリメラーゼで PCR
- dUTP を使った 2nd Strand cDNA は増幅されず、1st Strand cDNA のみが増幅される

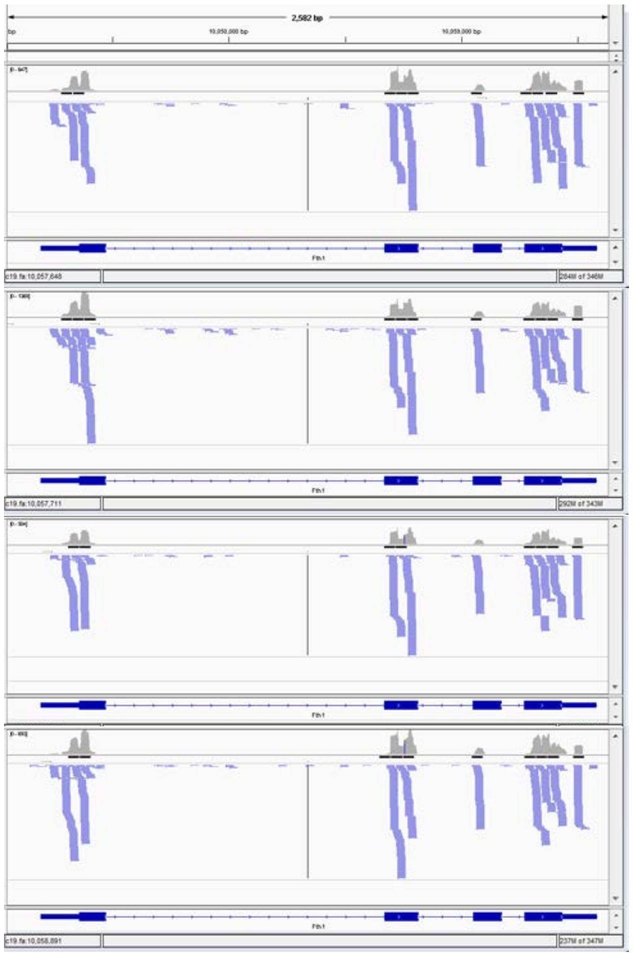
Agilent

D0

D4

D8

N9



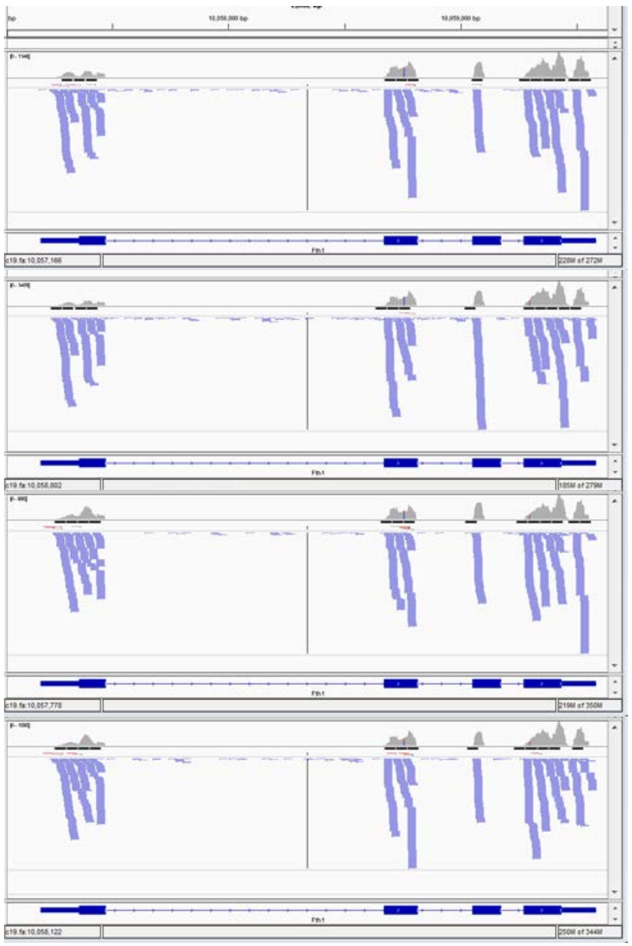
Illumina

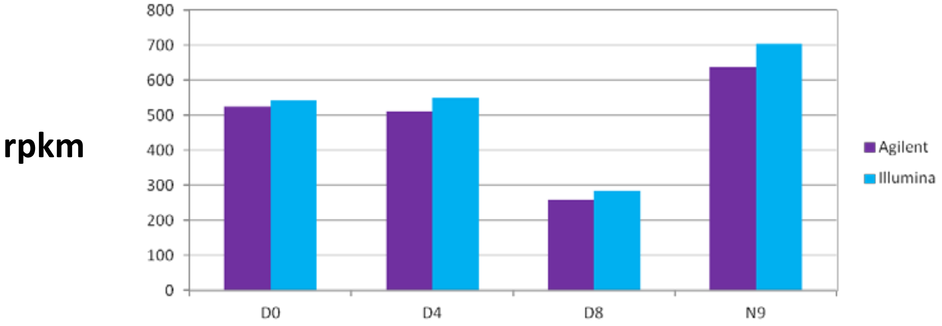
D0

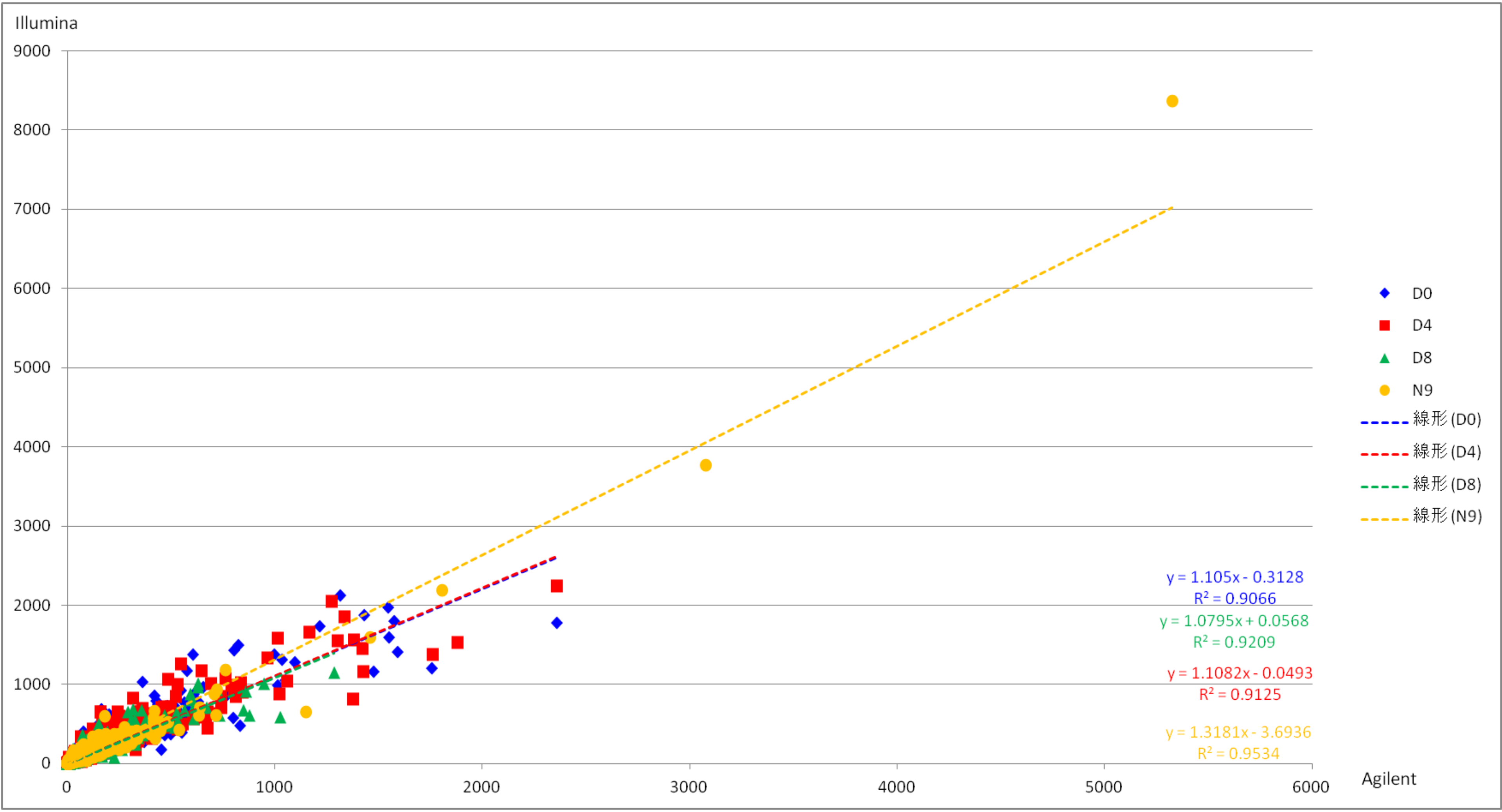
D4

D8

N9

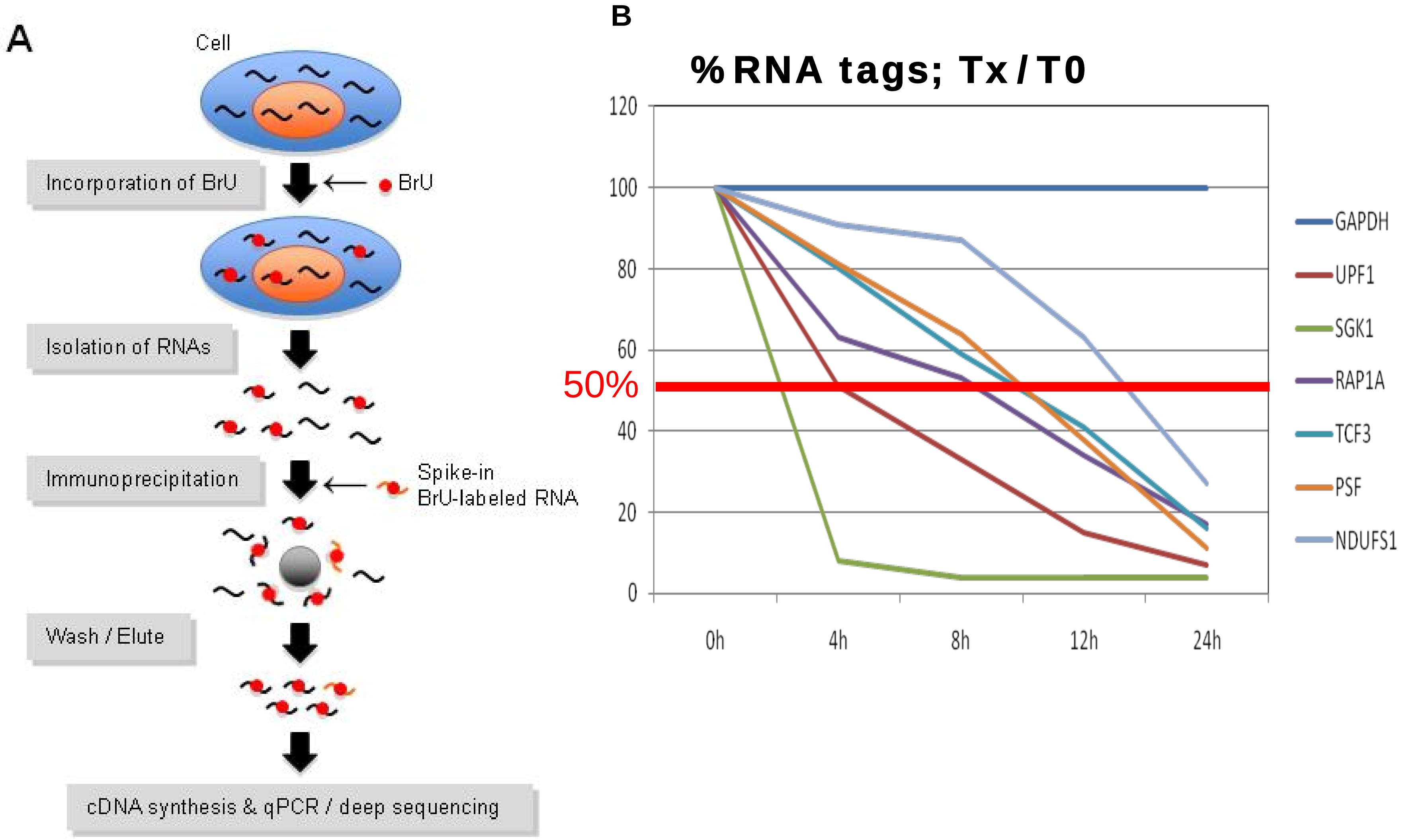





Illumina
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
6000
Agilent
◆ D0
■ D4
▲ D8
● N9
線形 (D0)
線形 (D4)
線形 (D8)
線形 (N9)
y = 1.105x - 0.3128
R² = 0.9066
y = 1.0795x + 0.0568
R² = 0.9209
y = 1.1082x - 0.0493
R² = 0.9125
y = 1.3181x - 3.6936
R² = 0.9534

Tani et al Genome Res (2012)

“BRIC” Analysis for determining mRNA half-life (Akimitsu lab)



Cell

Incorporation of BrU

BrU

Isolation of RNAs

Immunoprecipitation

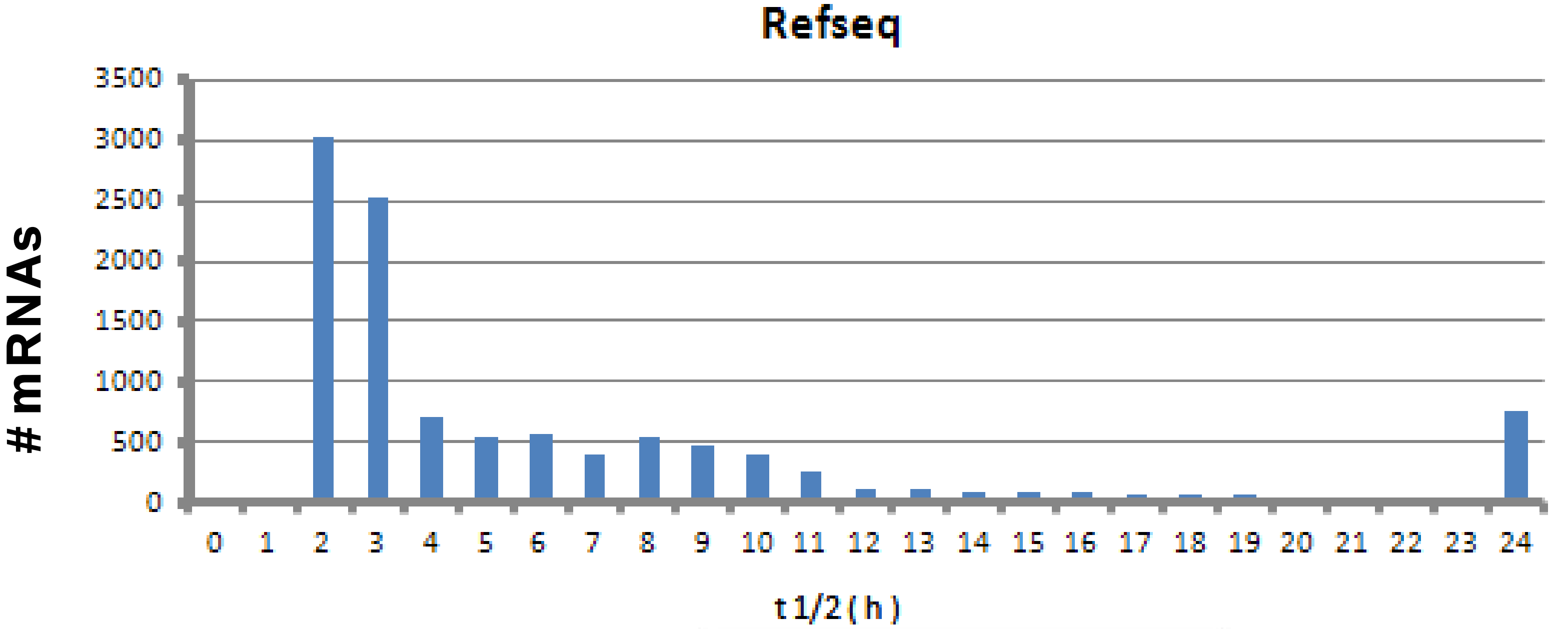
Spike-in BrU-labeled RNA

Wash / Elute

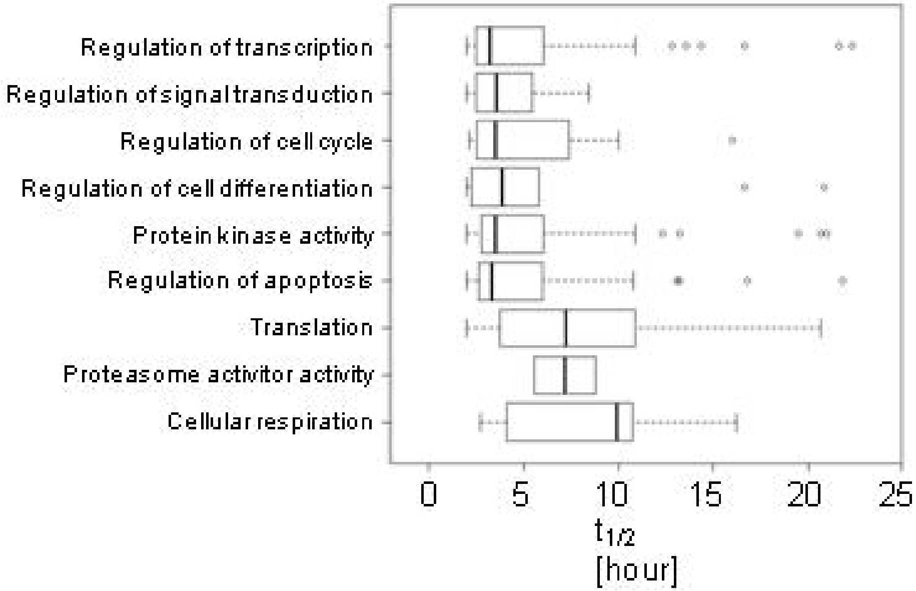
cDNA synthesis & qPCR / deep sequencing

BRIC can monitor the T1/2 for each RNA

BRIC revealed Half-lives of mRNAs in a genome-wide manner

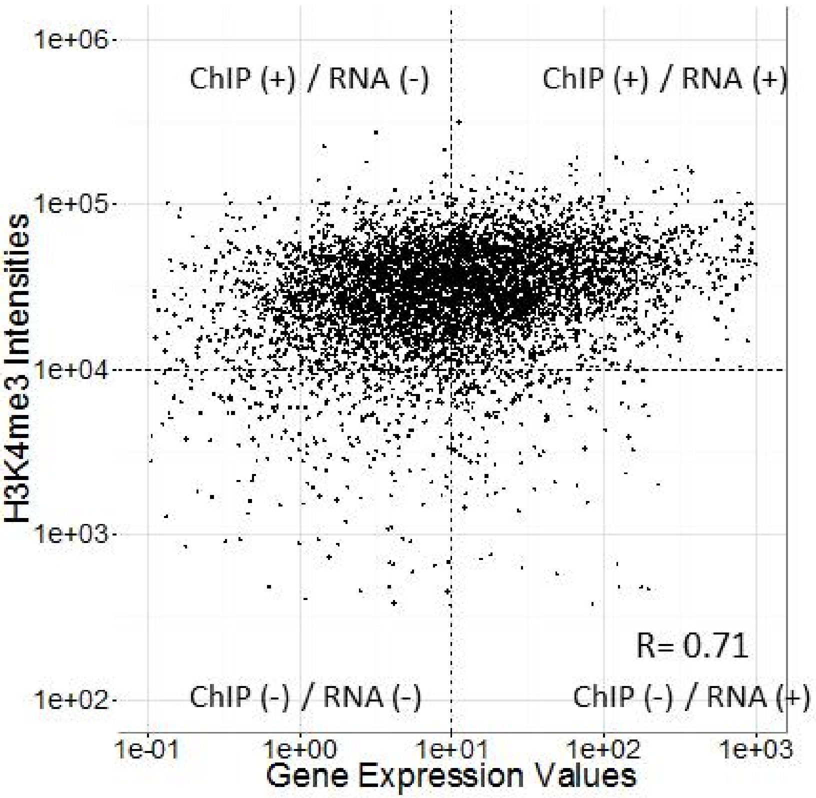


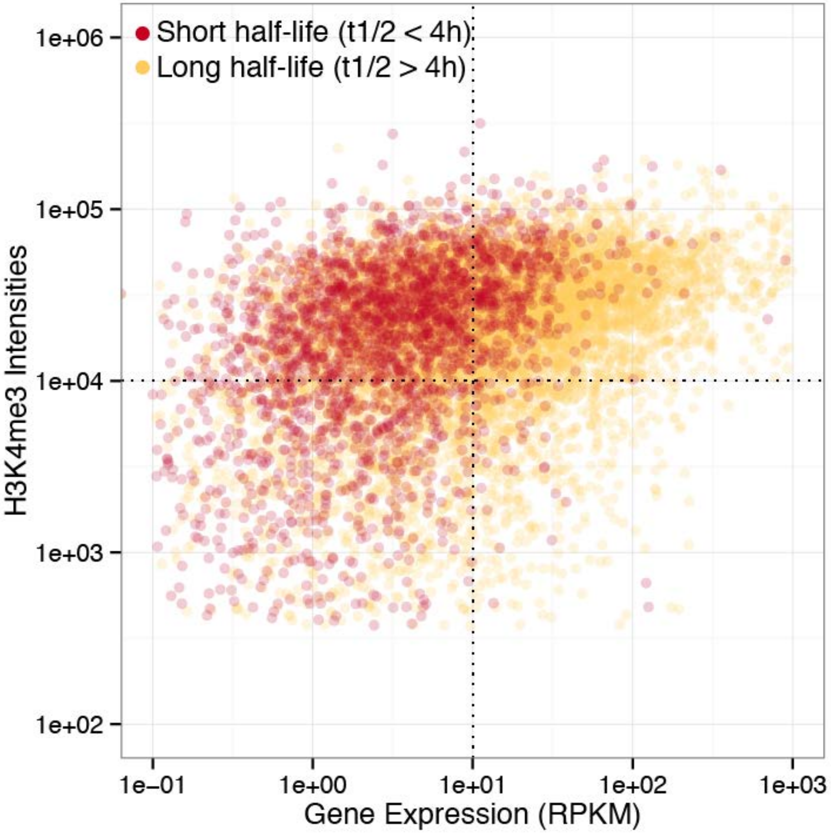
GO term analysis



RNAs related to "regulations" are enriched in short-lived RNAs

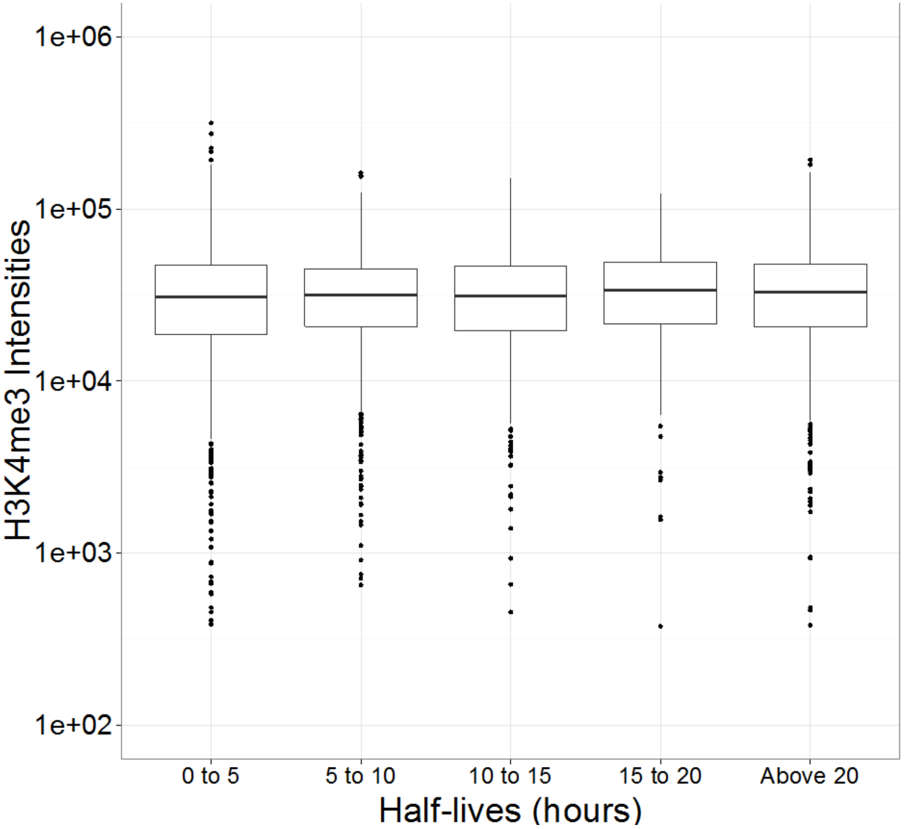
mRNAs of short half-lives are enriched in the population of ChIP+/RNA-

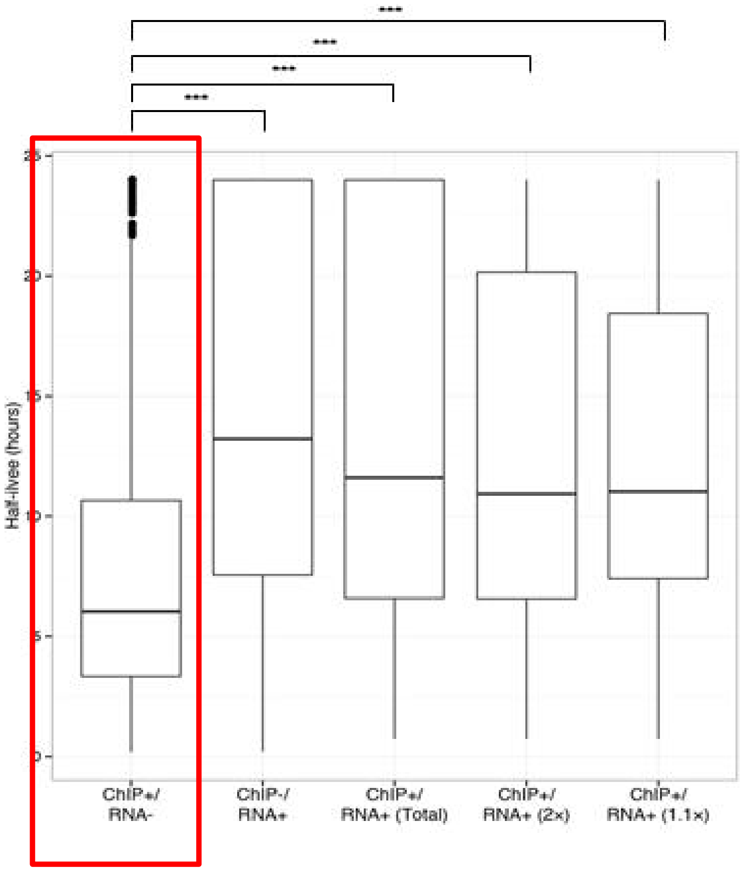




Maekawa et al submitted

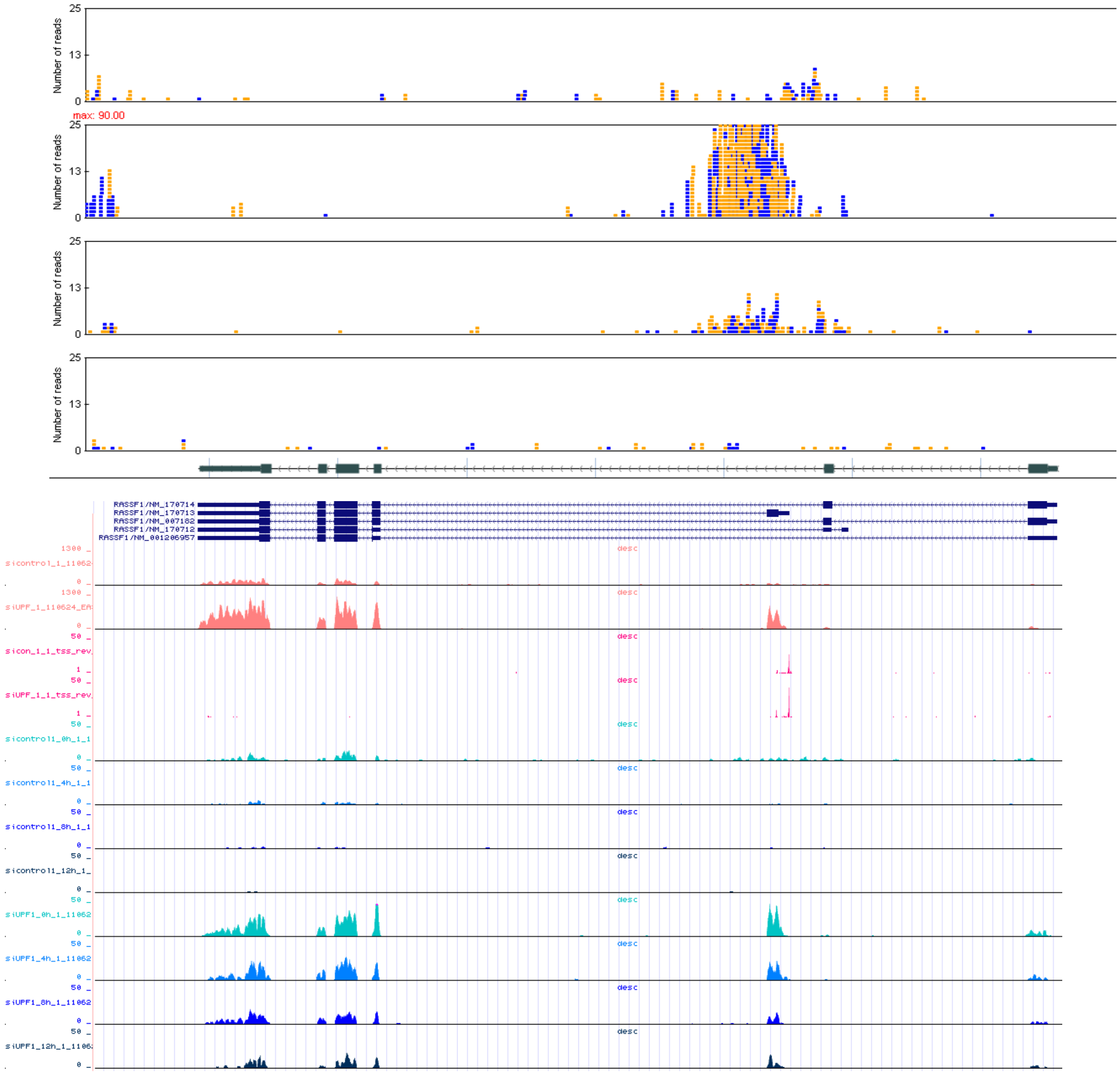
half-lives of mRNAs are controlled independently from transcriptional initiation





ChIP+/RNA-

RefSeq: NM_001206957.1
Description: Homo sapiens Ras association (RalGDS/AF-6) domain family member 1 (RASSF1), transcript variant H, mRNA.
Position: chr3:50367217-50378367
Strand: - Gene Symbol: RASSF1



ChIP-seq

Pol II

K4m3

Ac

input

Total RNA

sicontrol : 79.41 ppm

siUPF : 314.02ppm

TSS-seq

sicontrol : 16.3 ppm

siUPF : 26.8 ppm

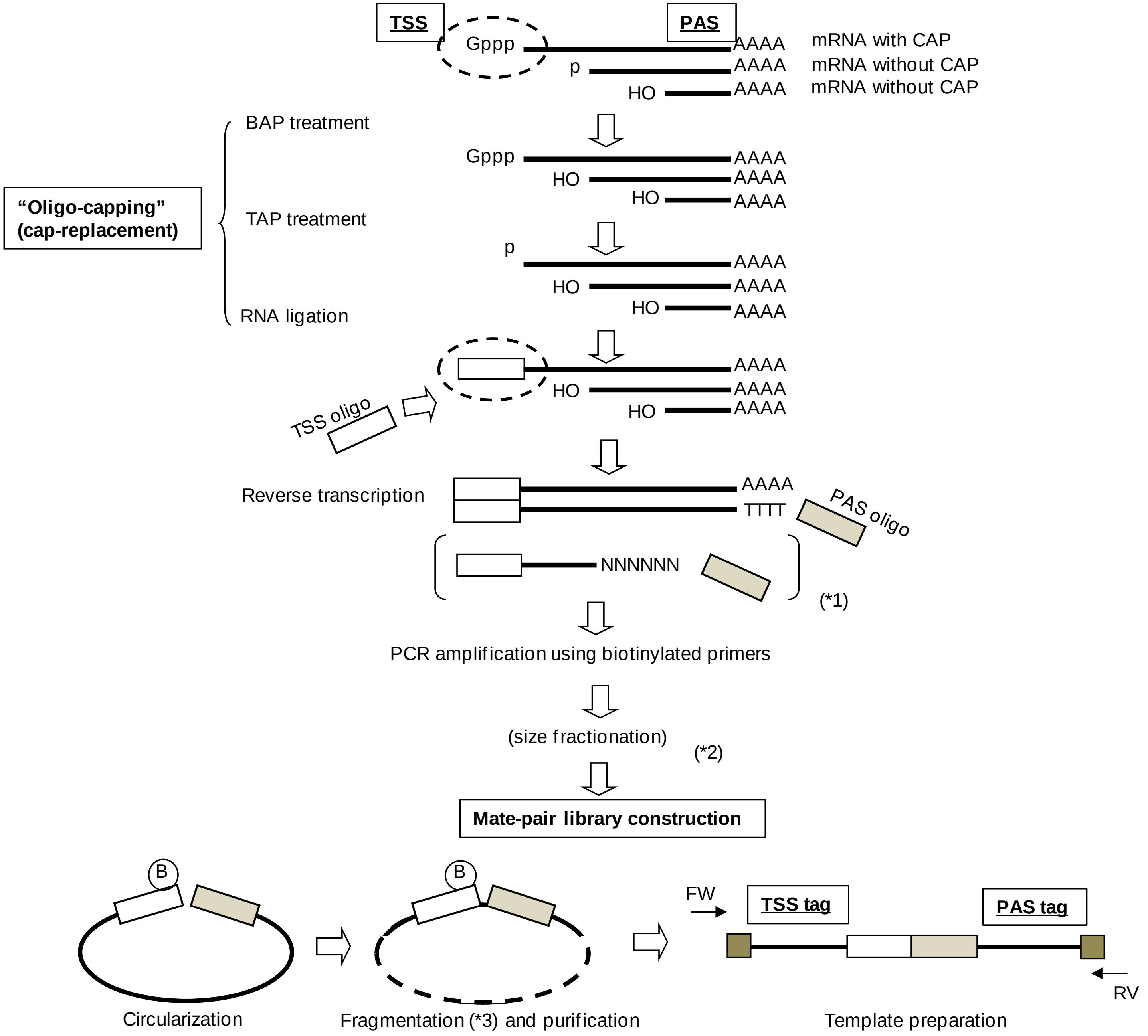
BRIC-seq

siCont t1/2 : 0.68h

BRIC-seq

siUPF t1/2 : 7.51h

Matsumoto et al NAR in press



TSS

PAS

Gppp AAAA mRNA with CAP

p AAAA mRNA without CAP

HO AAAA mRNA without CAP

"Oligo-capping" (cap-replacement)

BAP treatment

Gppp AAAA

HO AAAA

HO AAAA

TAP treatment

p AAAA

HO AAAA

HO AAAA

RNA ligation

TSS oligo

AAAA

HO AAAA

HO AAAA

Reverse transcription

AAAA

TTTT

PAS oligo

NNNNNN

(*1)

PCR amplification using biotinylated primers

(size fractionation)

(*2)

Mate-pair library construction

B

Circularization

B

Fragmentation (*3) and purification

FW

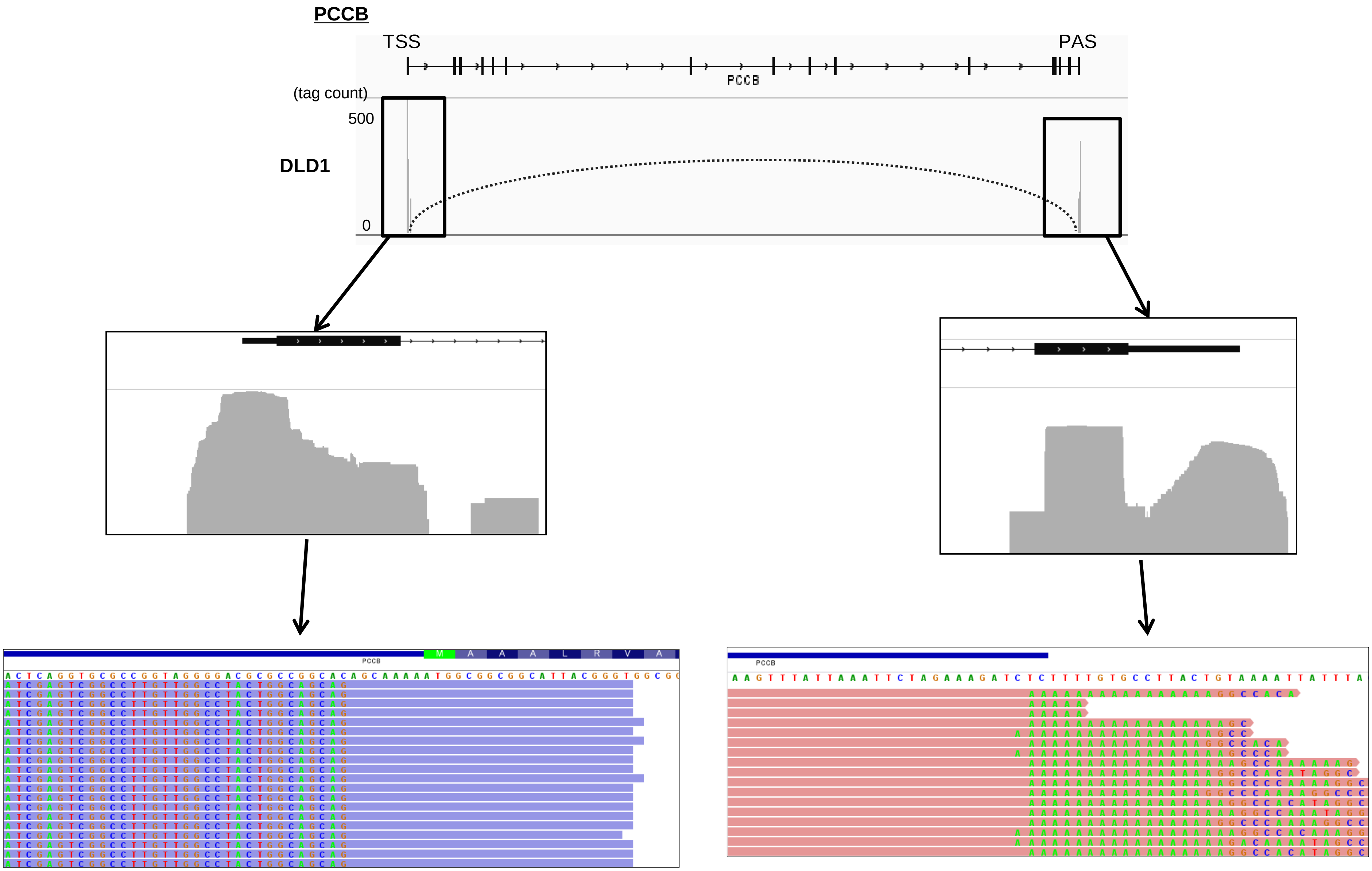
TSS tag

PAS tag

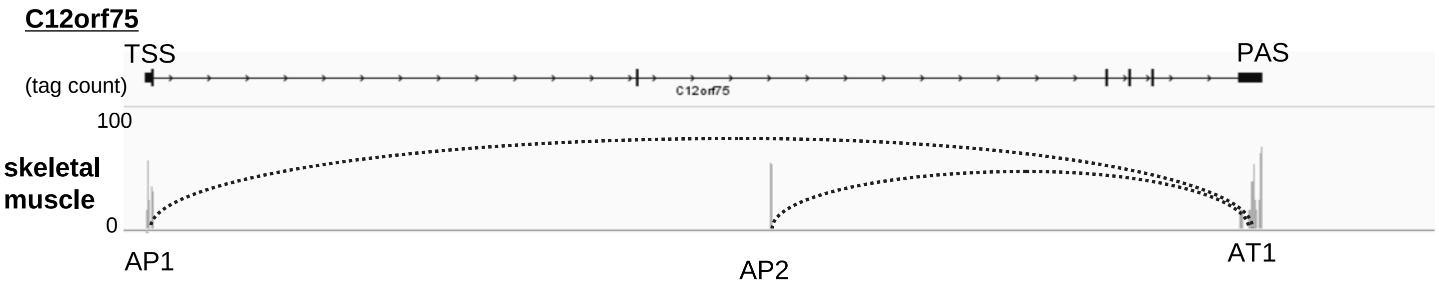
RV

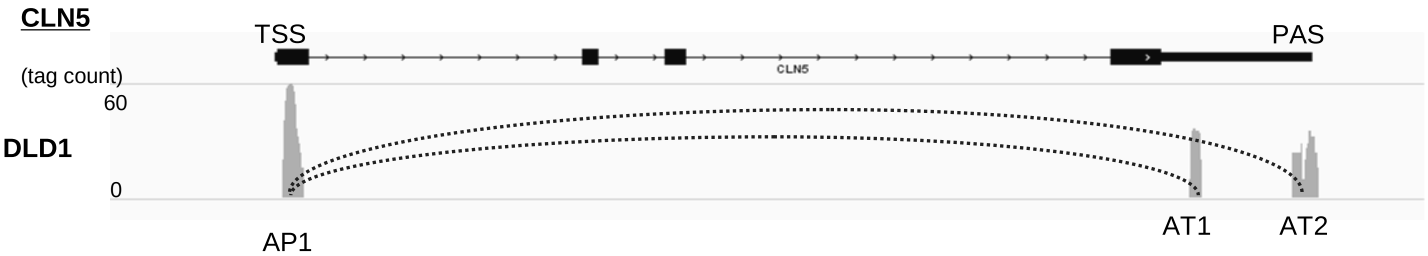
Template preparation

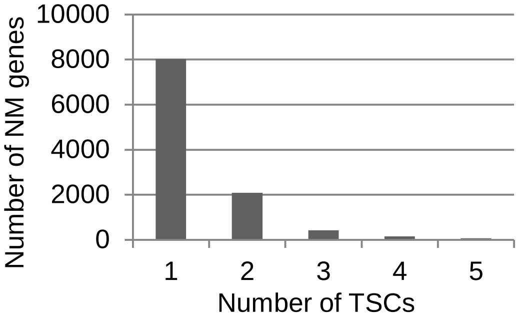
Mate Pair library can detect TSS/TTS simultaneously

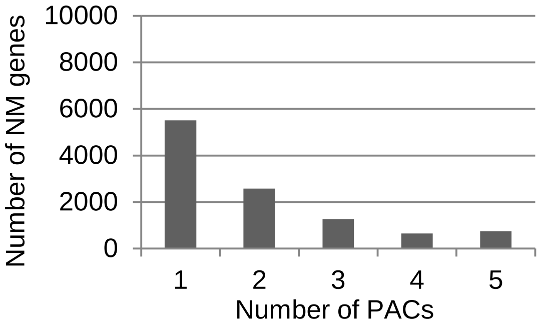

PCCB
TSS
PAS
PCCB
(tag count)
500
DLD1
0
PCCB
M A A A L R V A
A C T C A G G T G C G C C G G T A G G G G A C G C G C C G G C A C A G C A A A A A T G G C G G C G G C A T T A C G G G T G G C G G
A T C G A G T C G G C C T T G T T G G C C T A C T G G C A G C A G
PCCB
A A G T T T A T T A A A T T C T A G A A A G A T C T C T T T T G T G C C T T A C T G T A A A A T T A T T T A
A A A A A A A A A A A A A A A A G G C C A C A

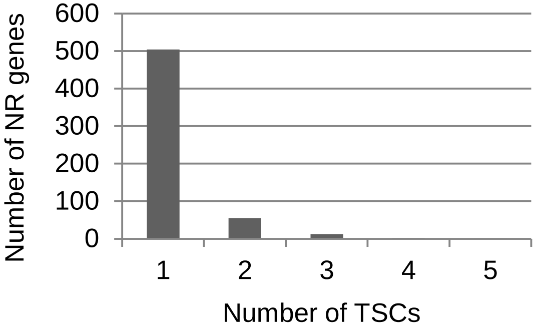
Alternative TSS/TTS and their relations

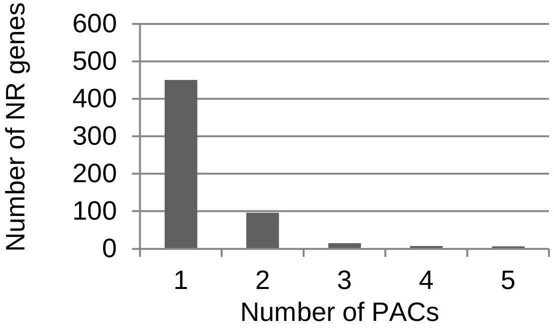






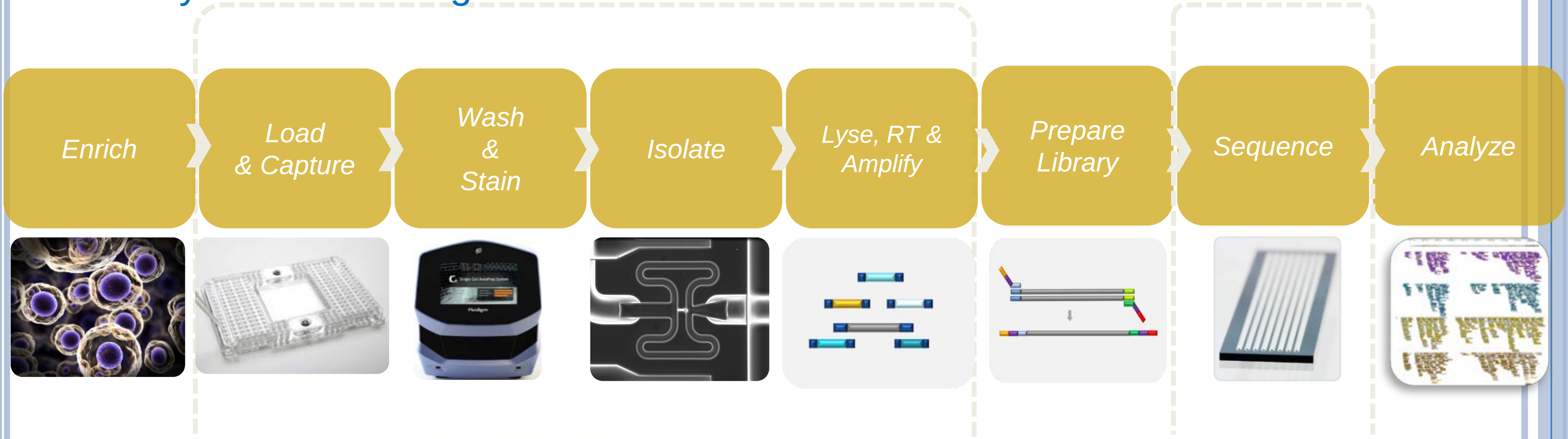






Semi-Automated Single-cell RNA Seq analysis

“C1 System” of Fluidigm



Enrich

Load & Capture

Wash & Stain

Isolate

Lyse, RT & Amplify

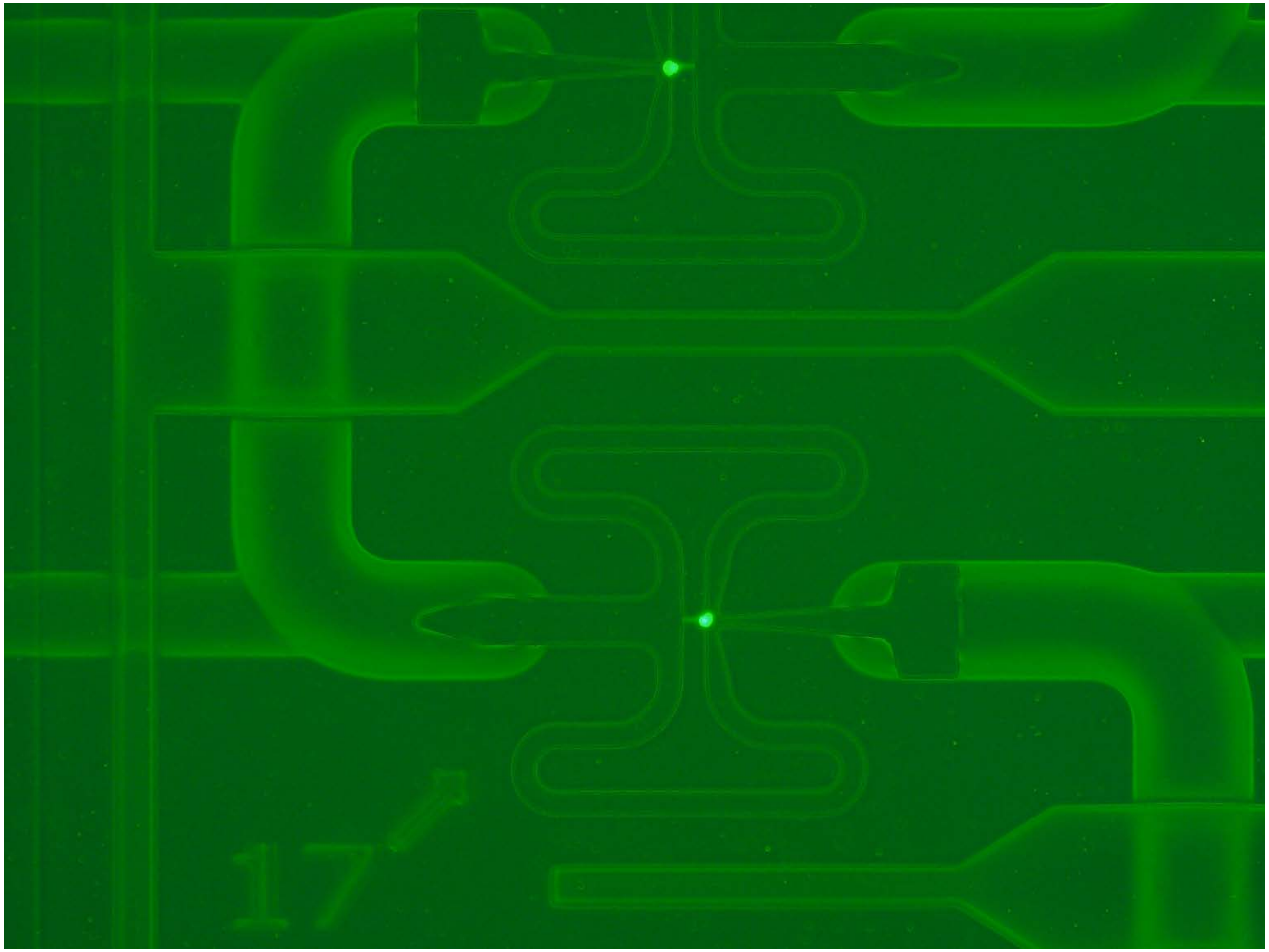
Prepare Library

Sequence

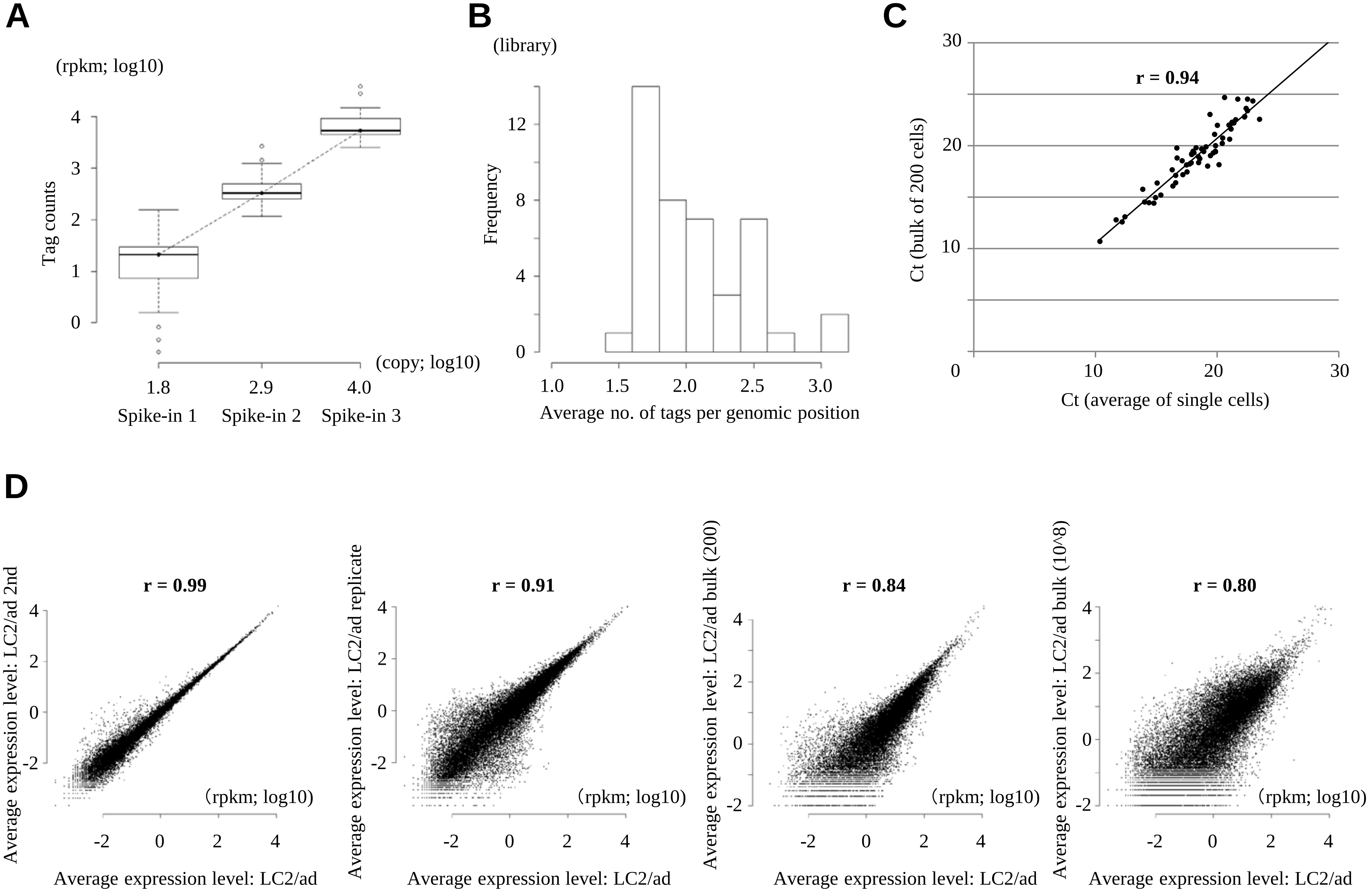
Analyze

C$_1$ Single-Cell Auto Prep System

Any Illumina System



成功率： 80% (Fluidigm)-> 60-70% (デモでの経験）



Suzuki et al submitted

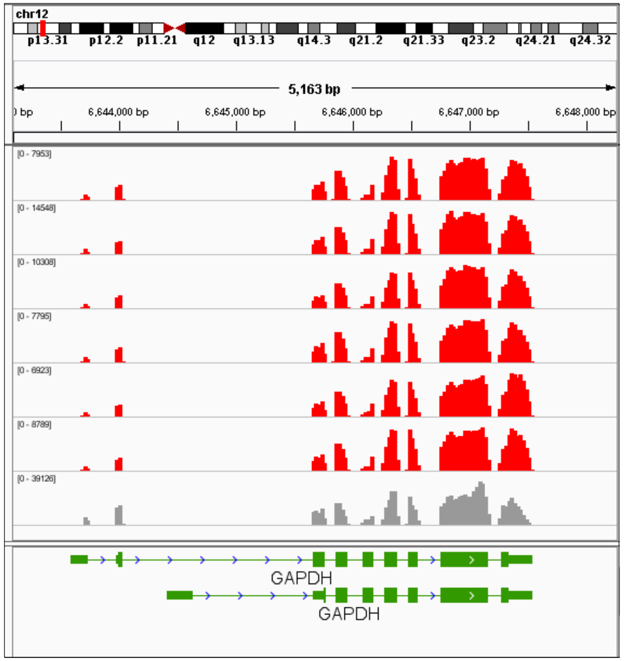
Distinct splice patterns in different single-cells

GAPDH

Number of cells

1

10^7

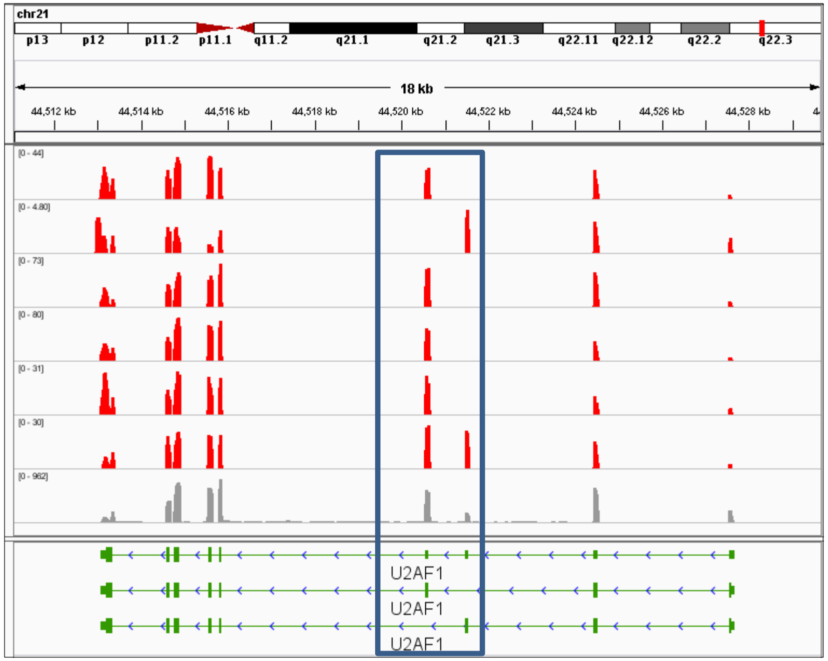


U2AF1

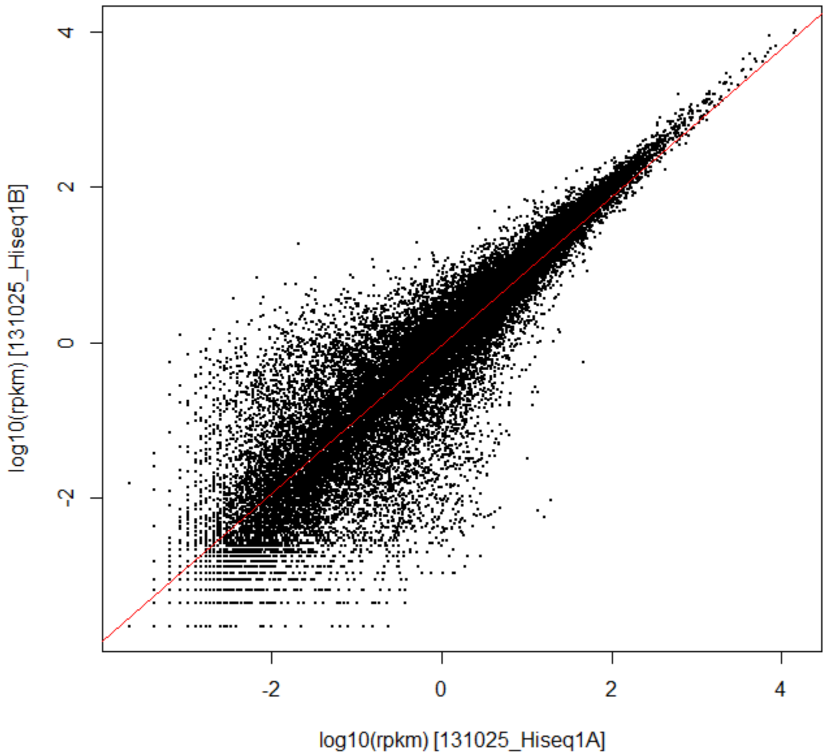
Number of cells

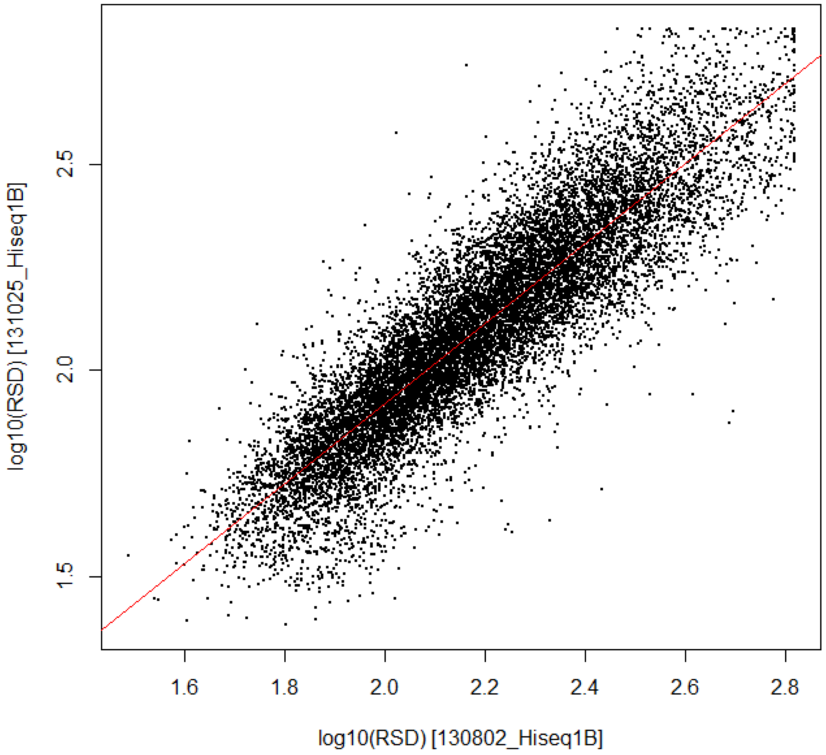
1

10^7



相関係数
1 回目 (C1_LC2AD : 131025_HISEQ1A) VS
2 回目 (LC2AD_2ND : 131025_HISEQ1B)





log10(rpkm)

y = 0.95409x + -0.03752
R = 0.9140295

LC2ad vs LC2ad_2nd

y = 0.97418x + -0.02766
R = 0.8898153

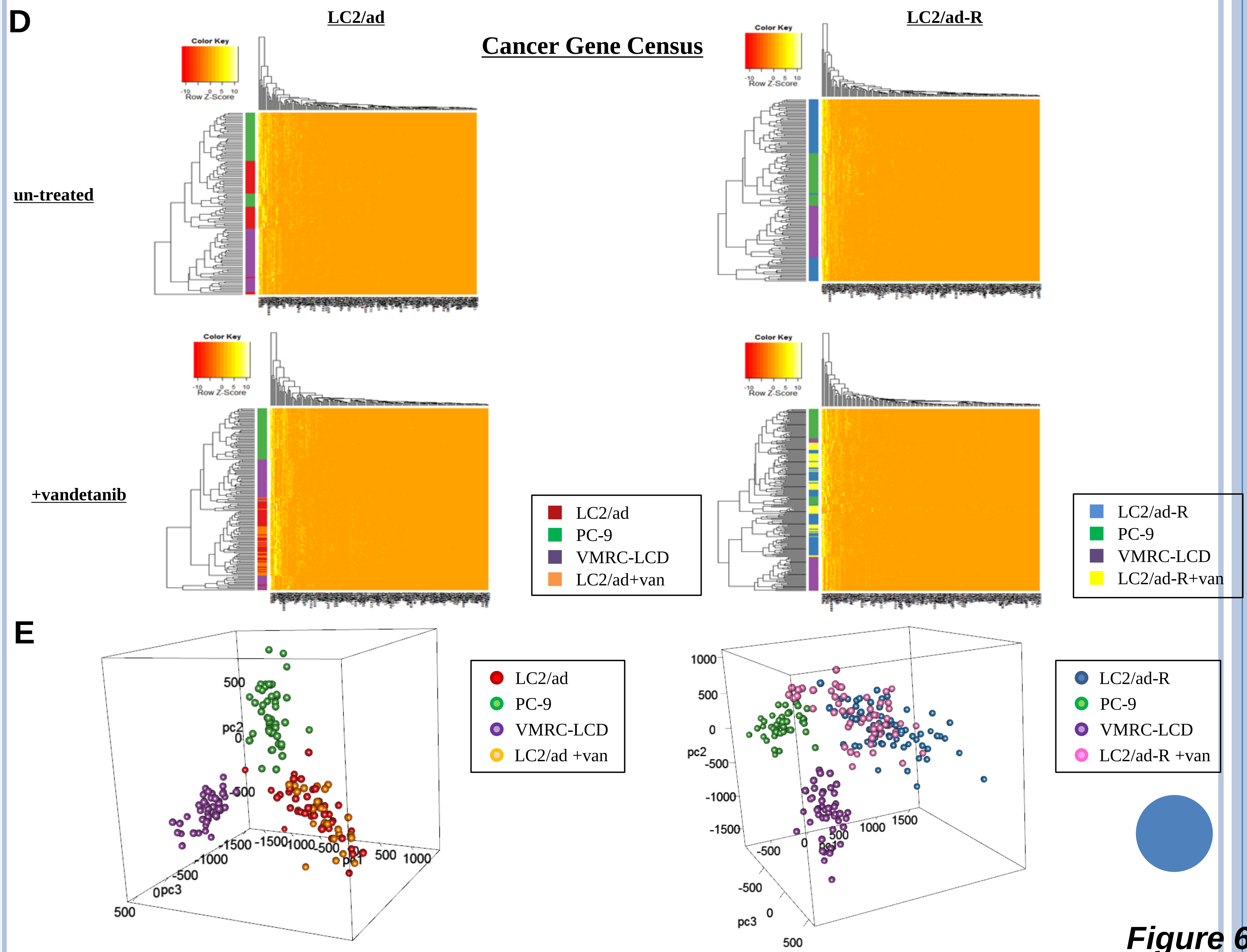
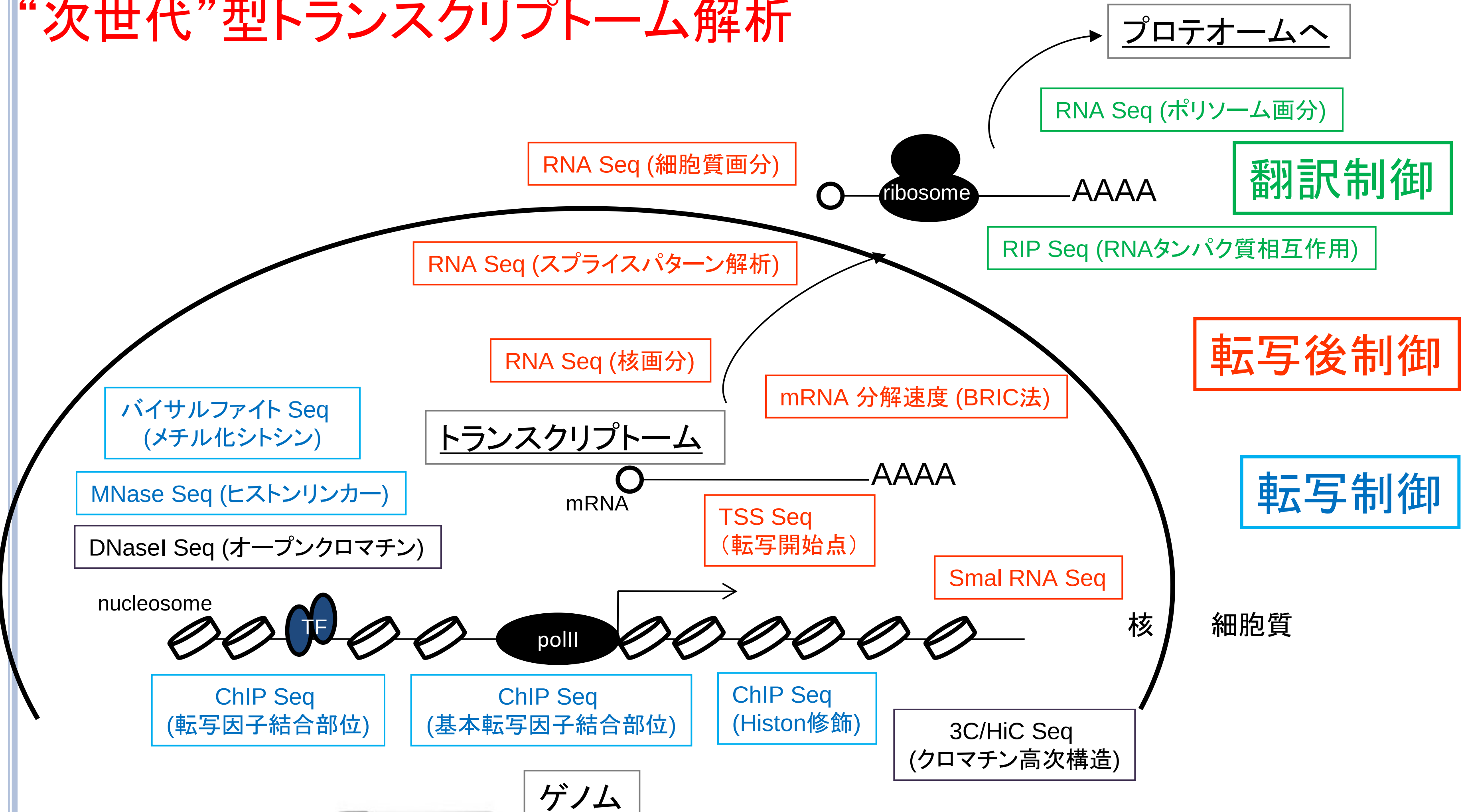

D
LC2/ad
Cancer Gene Census
LC2/ad-R
un-treated
+vandetanib
Color Key
-10 0 5 10
Row Z-Score
LC2/ad
PC-9
VMRC-LCD
LC2/ad+van
LC2/ad-R
PC-9
VMRC-LCD
LC2/ad-R+van
E
500
pc2
0
-500
-1500 -1500 -1000 -500 0 500 1000
pc1
-1000
-500
0 pc3
500
LC2/ad
PC-9
VMRC-LCD
LC2/ad +van
1000
500
0
pc2
-500
-1000
-1500
-500 0 500 1000 1500
pc1
-500
0
pc3
500
LC2/ad-R
PC-9
VMRC-LCD
LC2/ad-R +van

Figure 6

“次世代”型トランスクリプトーム解析



プロテオームへ

RNA Seq (ポリソーム画分)

RNA Seq (細胞質画分)

ribosome

AAAA

翻訳制御

RIP Seq (RNAタンパク質相互作用)

RNA Seq (スプライスパターン解析)

転写後制御

RNA Seq (核画分)

mRNA 分解速度 (BRIC法)

バイサルファイト Seq
(メチル化シトシン)

トランスクリプトーム

MNase Seq (ヒストンリンカー)

mRNA

AAAA

転写制御

TSS Seq
(転写開始点)

DNaseI Seq (オープンクロマチン)

Smal RNA Seq

nucleosome

TF

polII

核

細胞質

ChIP Seq
(転写因子結合部位)

ChIP Seq
(基本転写因子結合部位)

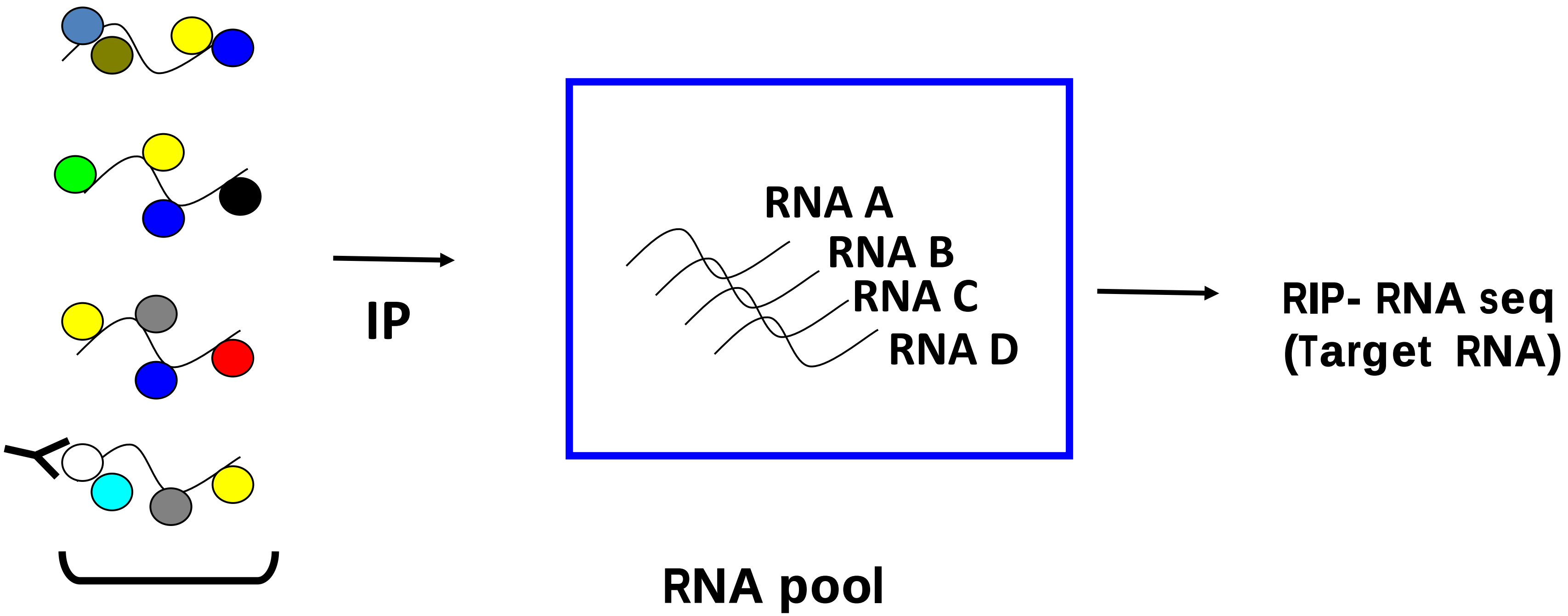
ChIP Seq
(Histon修飾)

3C/HiC Seq
(クロマチン高次構造)

ゲノム

共通検出器としての次世代シークエンサー

Schematic diagram of RIP(RNA immunoprecipitation) -Seq



IP

RNA A

RNA B

RNA C

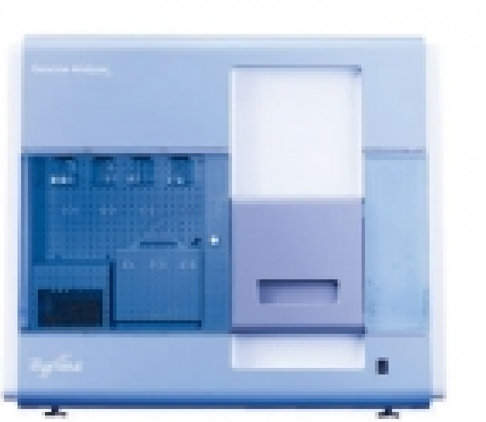
RNA D

RNA pool

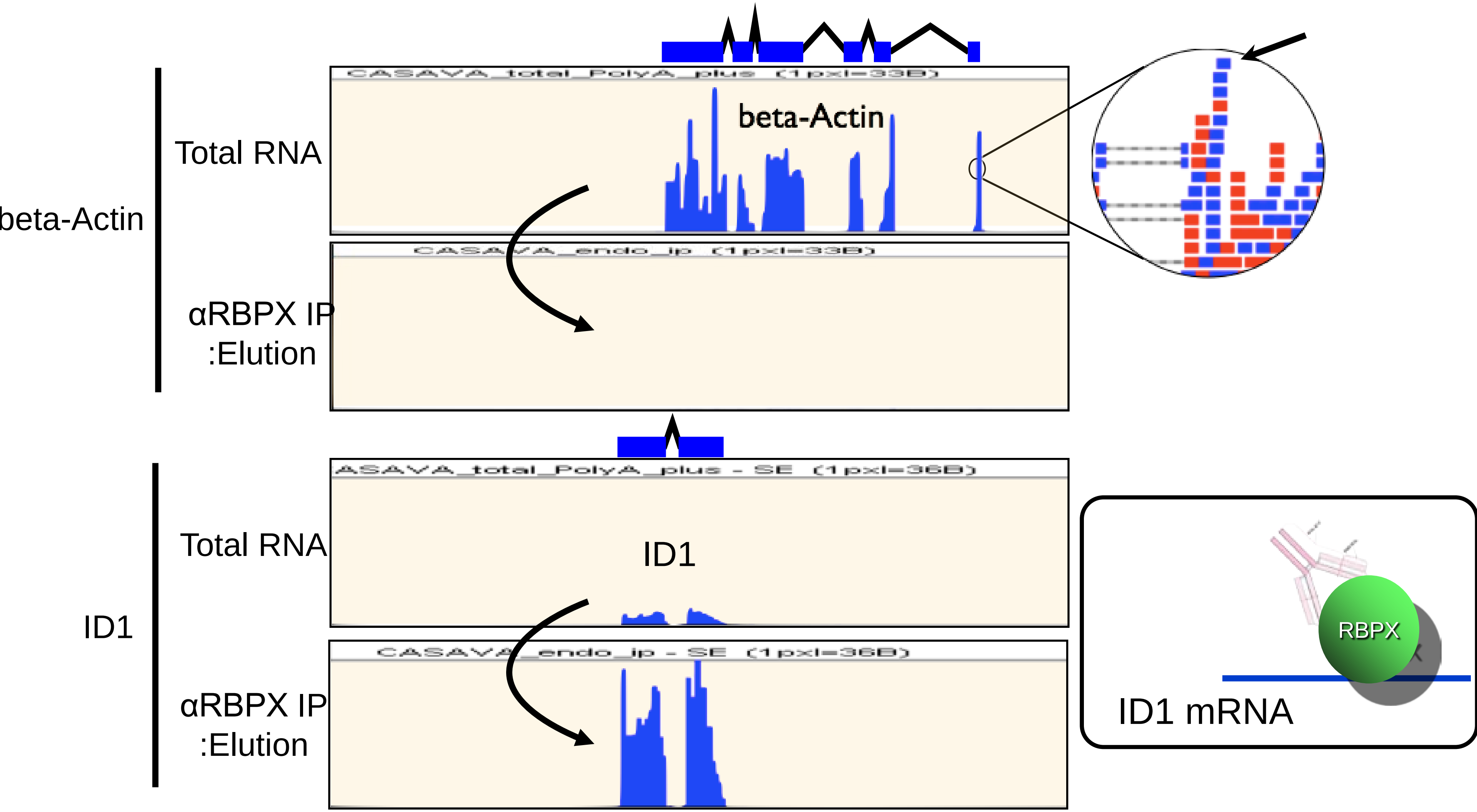
RIP- RNA seq
(Target RNA)

● RNA binding protein

∿ Target RNA



Identification of RNA binding protein target mRNAs



beta-Actin

Total RNA

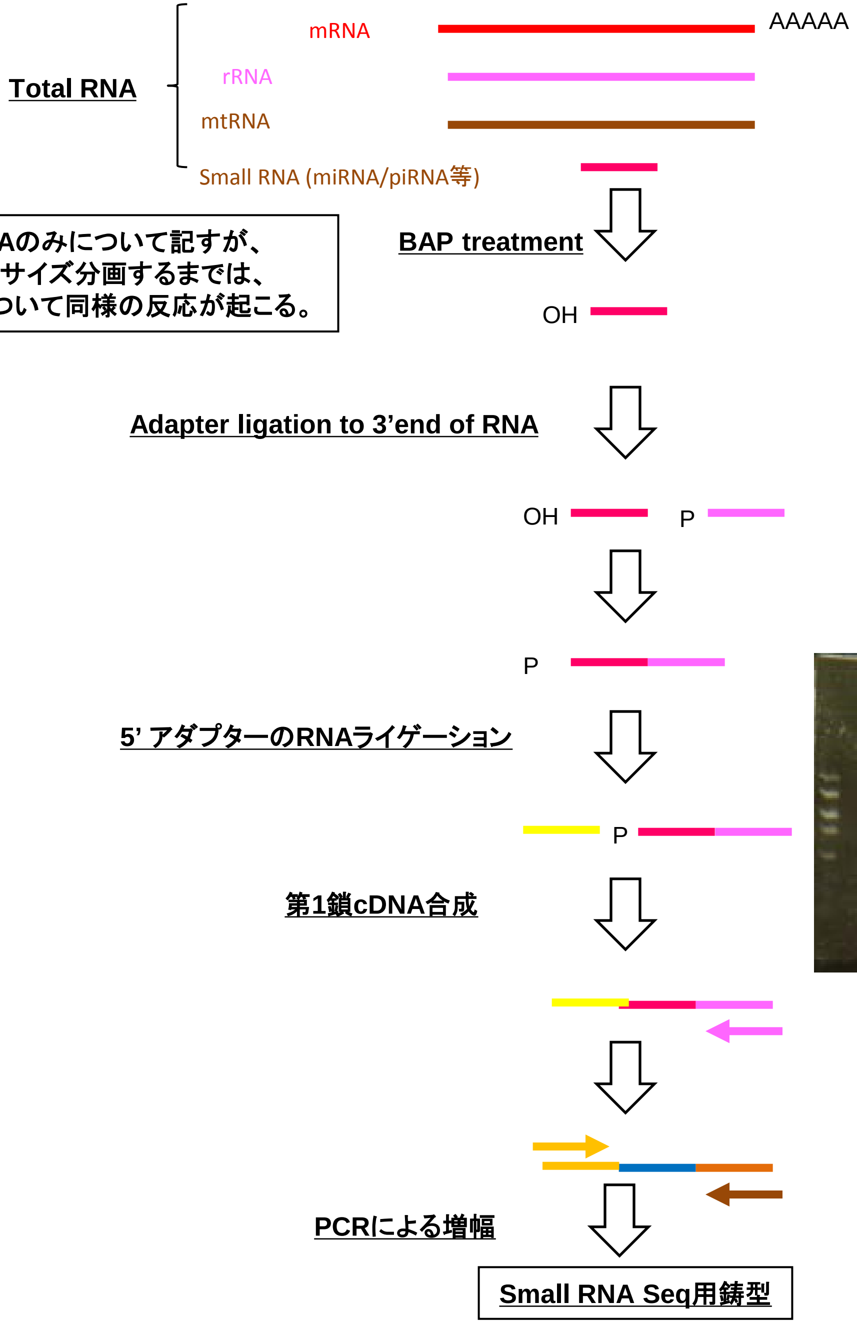
αRBPX IP
:Elution

ID1

Total RNA

αRBPX IP
:Elution

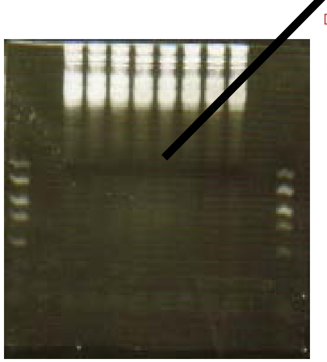
ID1 mRNA

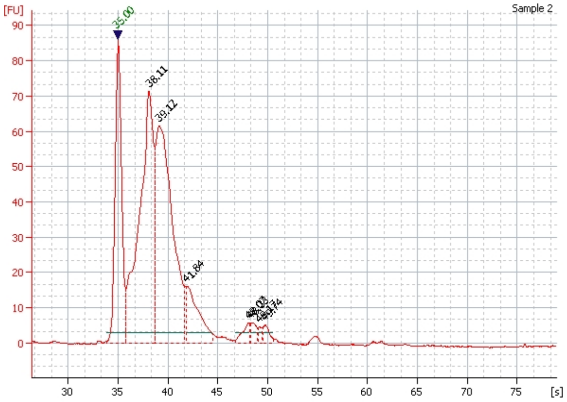


※図はsmall RNAのみについて記すが、
最後のステップでサイズ分画するまでは、
すべてのRNAについて同様の反応が起こる。

	Takara Protocol	Illumina protocol　(v1.5)
Total RNA input	100ug	1ug
Size selection	Needed	Not needed

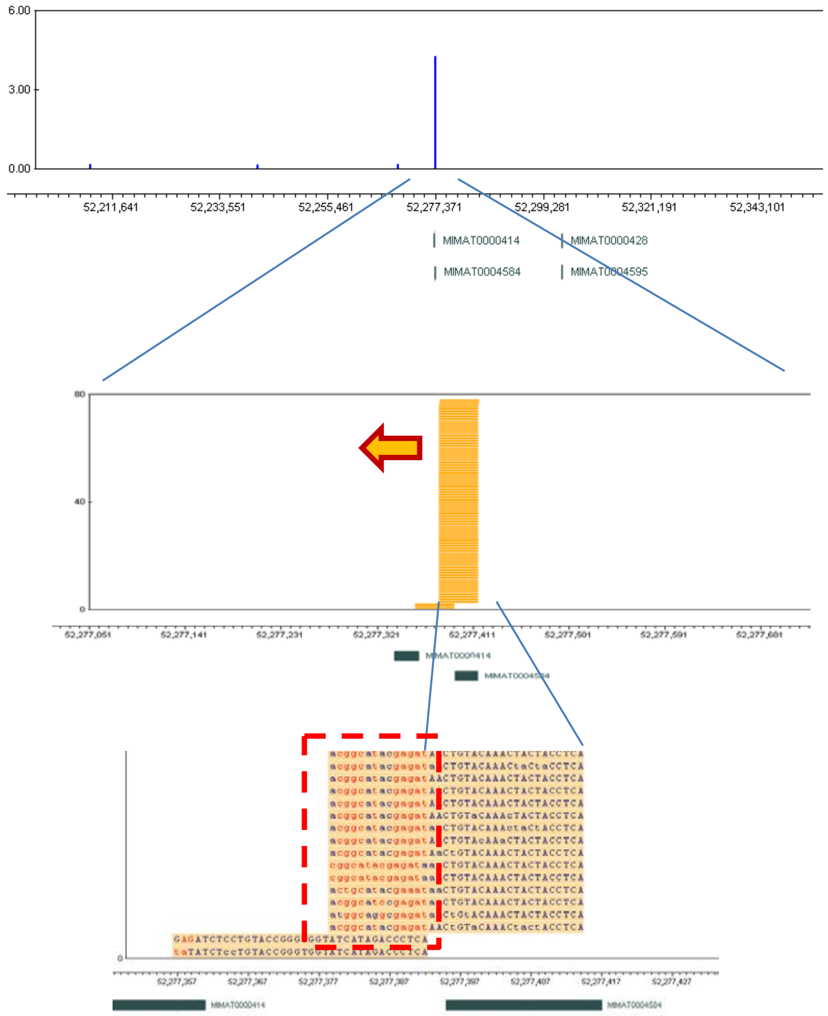
約18 nt~30 nt分画の
Small RNAを単離



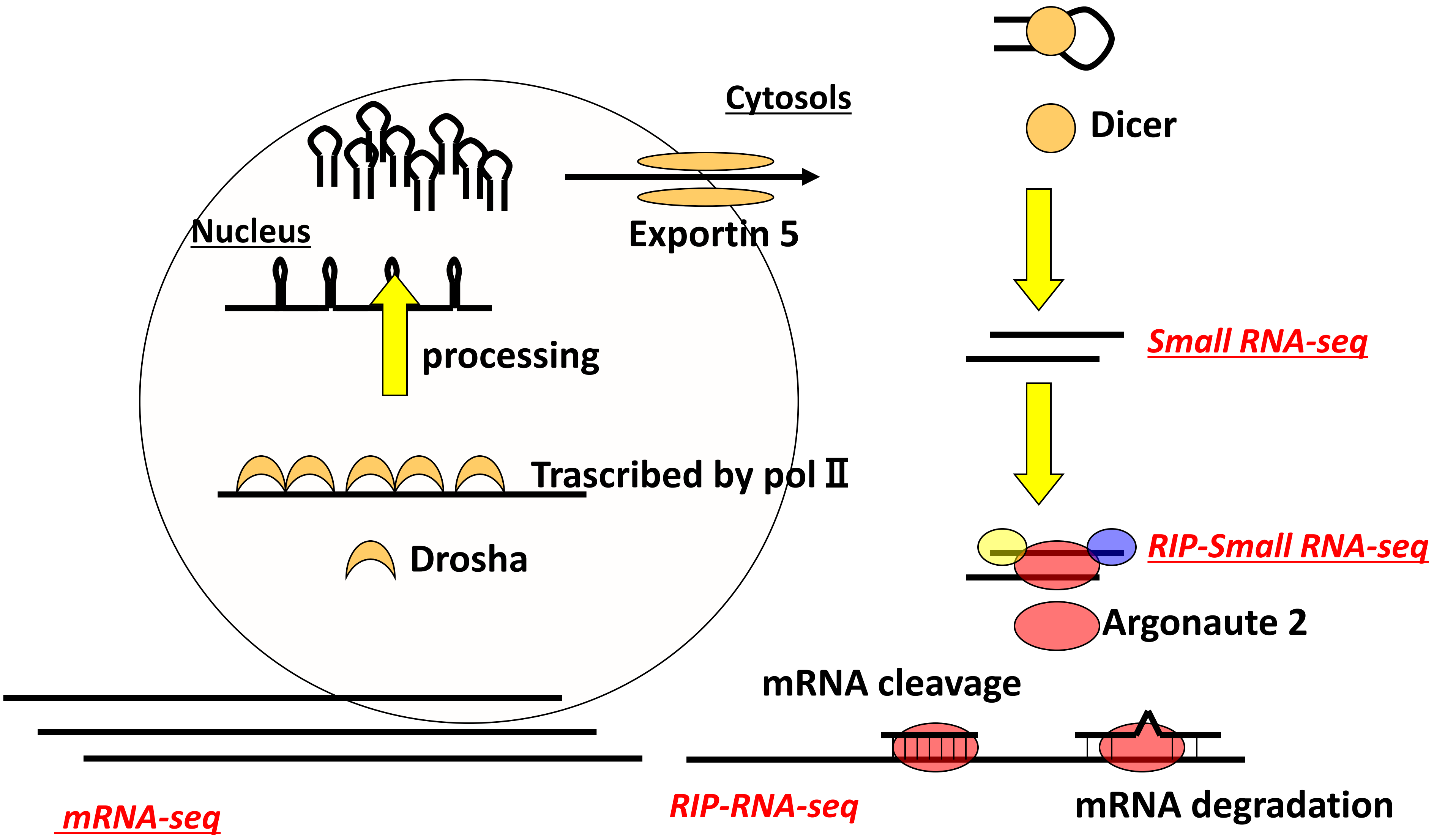


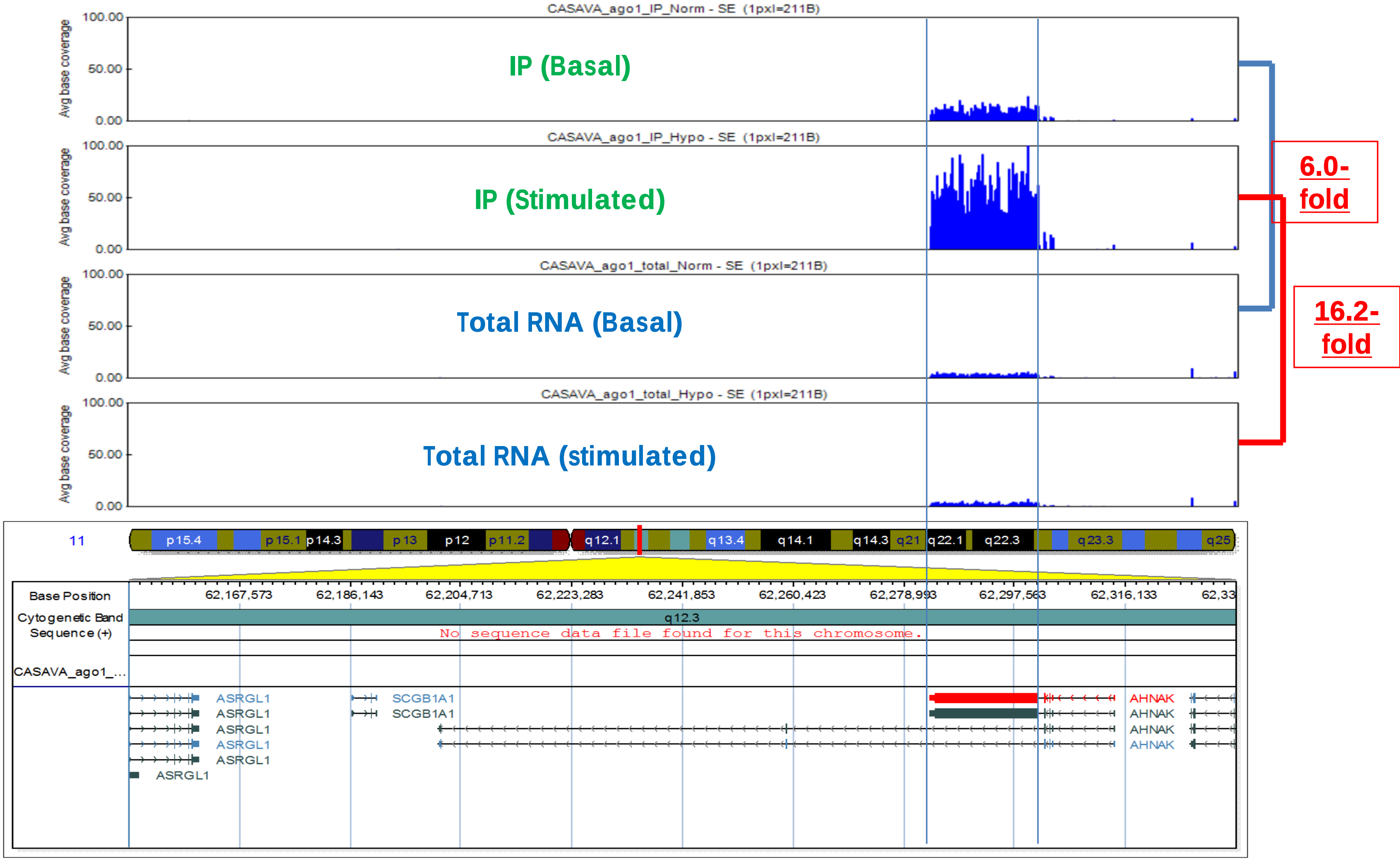
Kanematsu et al Gene 2012

small RNA Seq (DLD-1; the MIMAT0004584 gene region)



Schematic diagram of biogenesis of microRNAs and post-transcriptional silencing of target mRNA





IP (Basal)

IP (Stimulated)

Total RNA (Basal)

Total RNA (stimulated)

6.0-fold

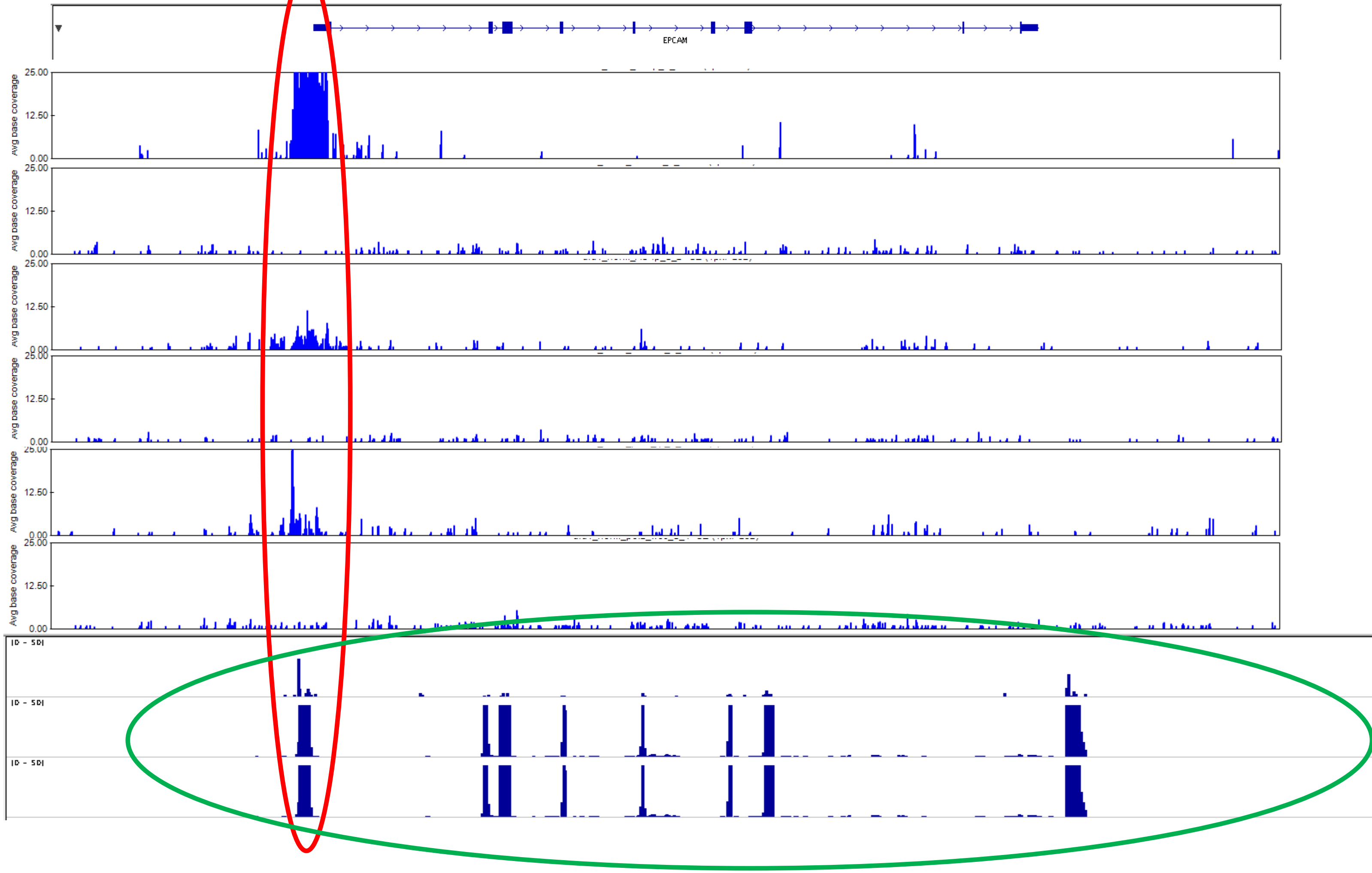
16.2-fold

次世代シークエンスデータの統合的解析

DLD-1 cell (colon caner)

転写制御の網羅的理解へ

Chr2: 47,443,347 - 47,477,133 (NM_002354)



Annotated mRNA

DLD-1_H3K4me3 (IP)

DLD-1_H3K4me3 (background)

DLD-1_H3Ac (IP)

DLD-1_H3Ac (background)

DLD-1_pol II (IP)

DLD-1_pol II (background)

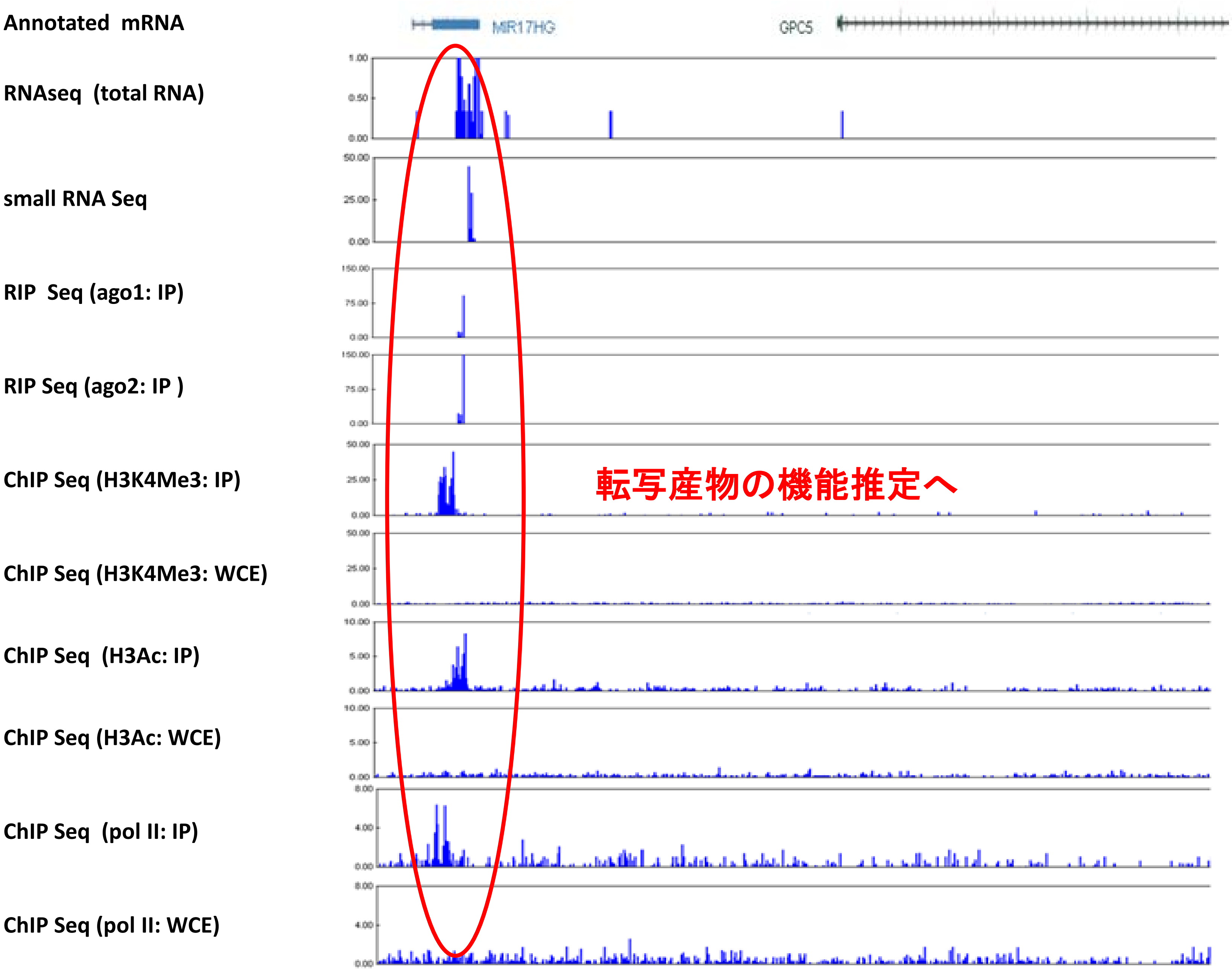
DLD-1_TSSseq

DLD-1_RNAseq

DLD-1_Polysome

mRNA動態の網羅的理解へ

B The MIR17HG_gene region (DLD-1 cells)



Annotated mRNA

RNAseq (total RNA)

small RNA Seq

RIP Seq (ago1: IP)

RIP Seq (ago2: IP)

ChIP Seq (H3K4Me3: IP)

ChIP Seq (H3K4Me3: WCE)

ChIP Seq (H3Ac: IP)

ChIP Seq (H3Ac: WCE)

ChIP Seq (pol II: IP)

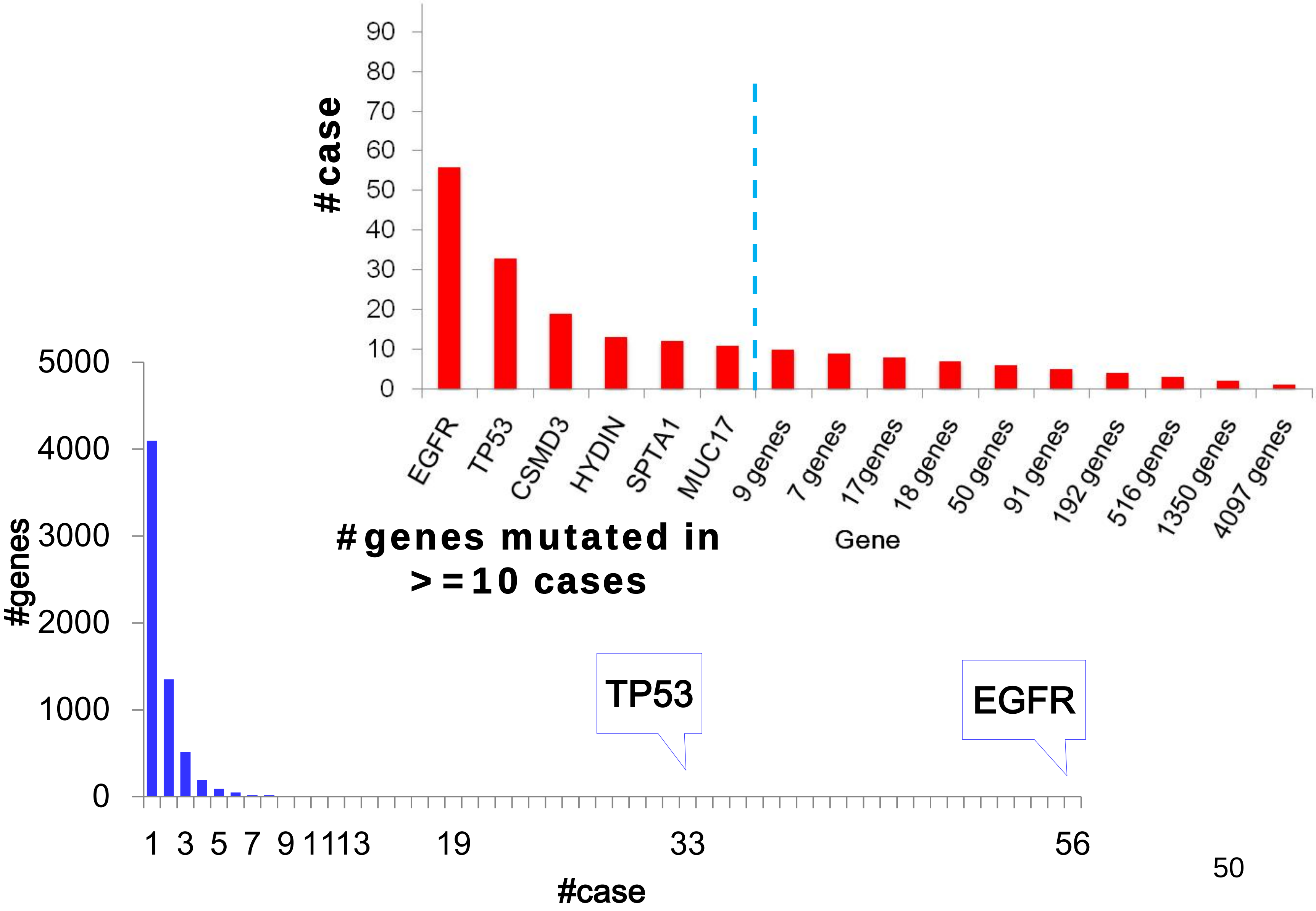
ChIP Seq (pol II: WCE)

転写産物の機能推定へ

肺腺がん細胞株のカタログ化
（と多階層オミクス解析のモデル）

Suzuki et al PLoS ONE 2013

Mutataion patterns of lung adenocarcinoma in 97 Japanese patients

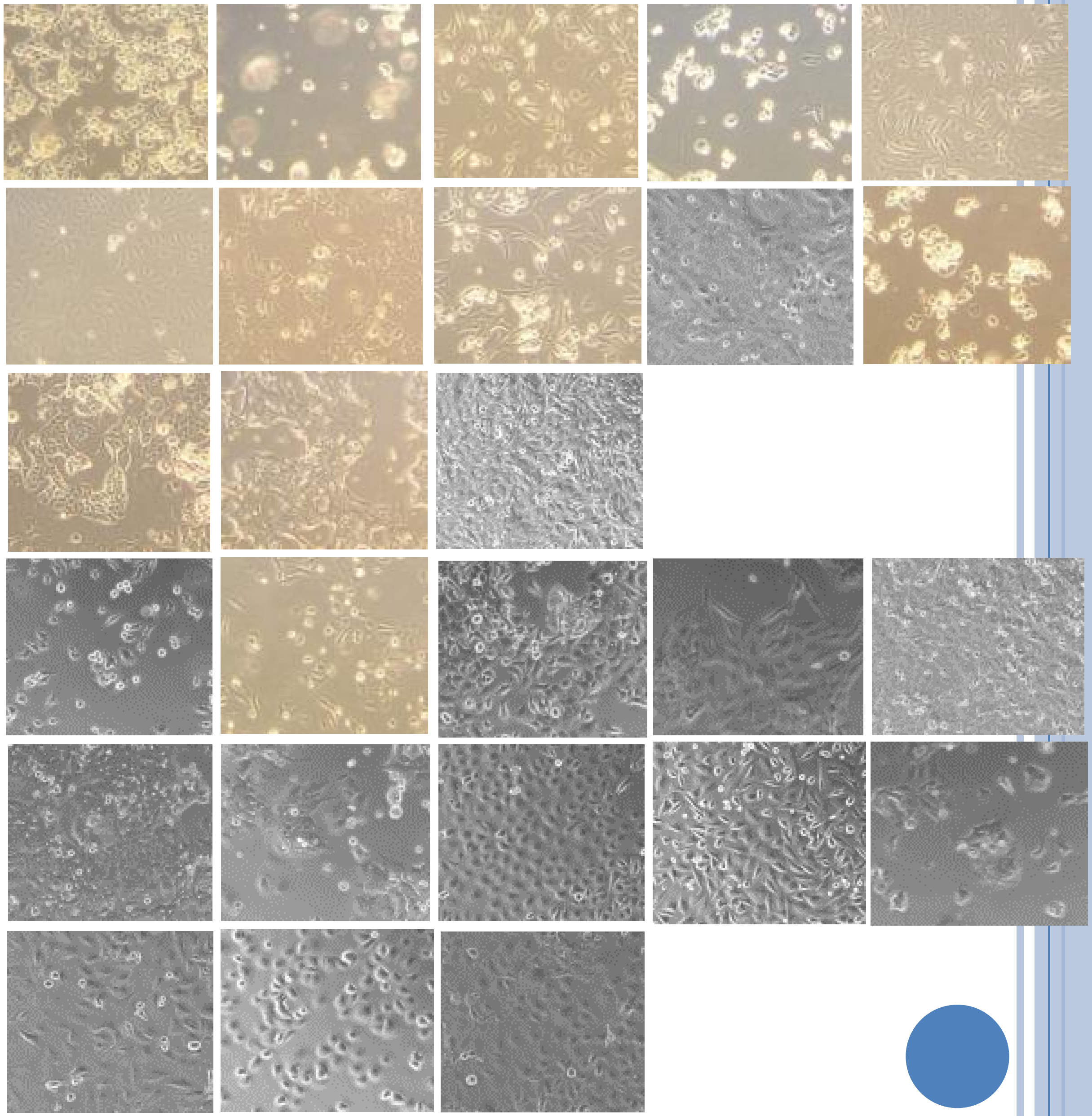


Suzuki et al submitted

Materials

26 lung adenocarcinoma cell lines

name	origin
PC-3	Japanese
PC-7	Japanese
PC-9	Japanese
PC-14	Japanese
RERF-LC-Ad1	Japanese
RERF-LC-Ad2	Japanese
RERF-LC-KJ	Japanese
RERF-LC-MS	Japanese
RERF-LC-OK	Japanese
VMRC-LCD	Japanese
ABC-1	Japanese
LC2/ad	Japanese
II-18	Japanese
A427	Caucasian
A549	Caucasian
H322	Caucasian
H2228	Unknown
H1299	Caucasian
H1437	Caucasian
H1648	Black
H1650	Caucasian
H1703	Caucasian
H1819	Caucasian
H1975	Unknown
H2126	Caucasian
H2347	Caucasian



All cell lines were provided from Dr. Tsuchihara and Dr. Kohno in National Cancer Center.

Genome

Whole-genome sequencing:

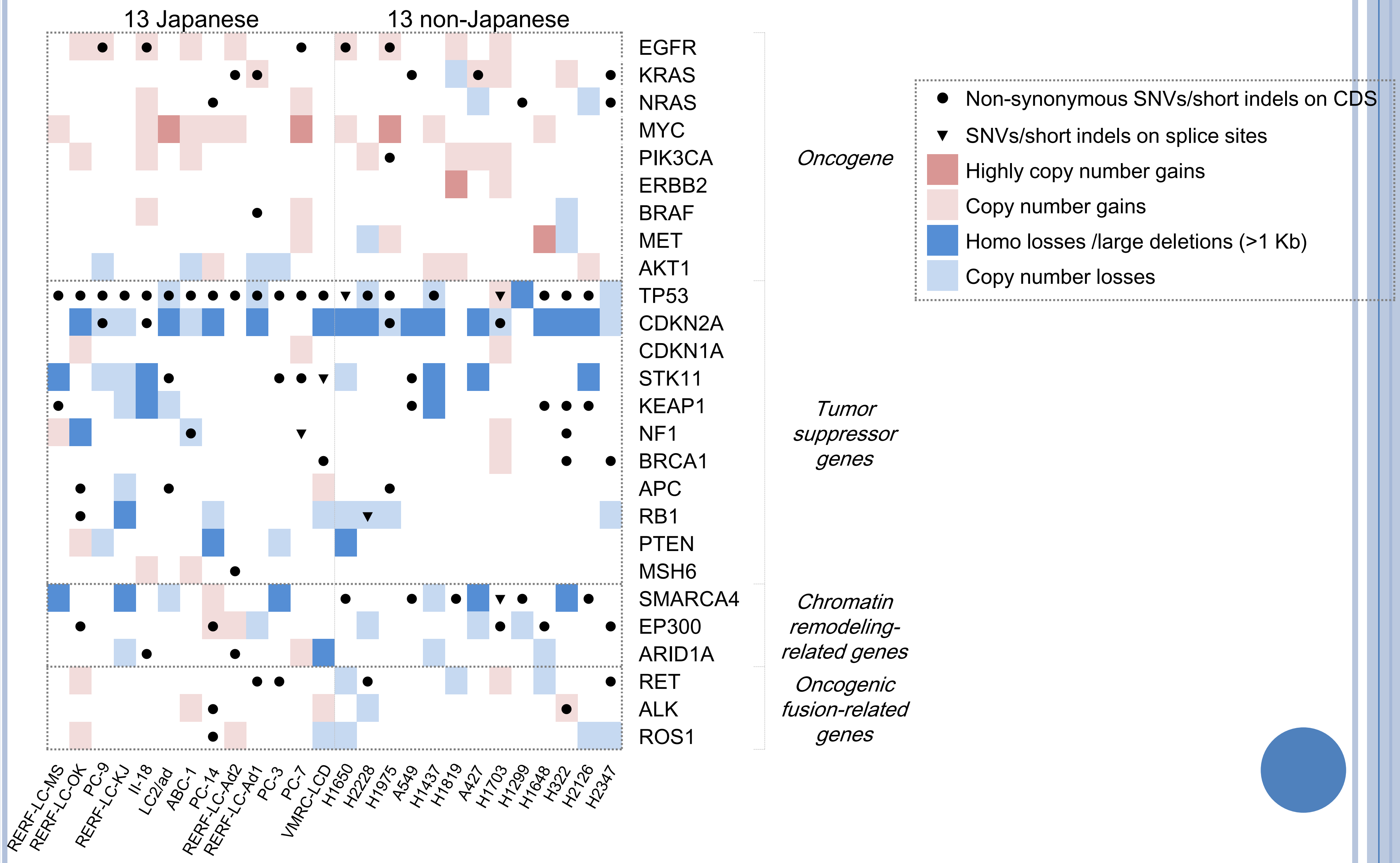
- ✓ Single nucleotide variants (SNVs), Insertion/deletions (indels)
- ✓ Copy number aberrations (CNAs)
- ✓ Chromosome rearrangements

Summary of SNVs/indels

	Total number of positions (Avg. of 26 cell lines)	
	SNVs	Short indels
Total	12,732,271 (3,302,407)	1,916,622 (453,821)
Germline	10,010,429 (3,177,173)	1,597,810 (429,846)
Somatic candidates	2,721,842 (125,234)	318,812 (23,975)
Genic[*]	892,941 (39,695)	118,268 (8,516)
Upstream (-500 from TSS)	11,796 (551)	2,049 (159)
UTRs	24,902 (1,086)	13 (0.8)
CDS	16,354 (687)	573 (37)
Synonymous	4,505 (188)	***
Non-synonymous	11,849 (499)	***
Splice sites[†]	346 (14)	39 (3)
Intronic and others	839,543 (37,357)	115,594 (8,315)
Intergenic	1,828,901 (85,539)	200,544 (15,459)

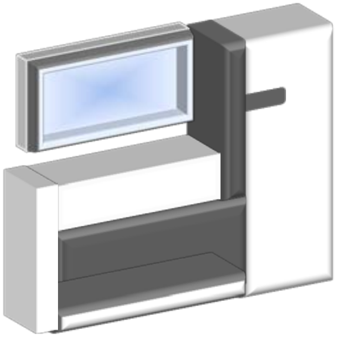
Genome

Genomic mutation status in 26 cancer-related genes



Ding et al. *Nature* 2008; Blanco et al. *Hum Mutat* 2009; Imielinski et al. *Cell* 2012

Sequencing data



Whole-genome sequencing

Sequencing: illumina HiSeq2000/2500; 101PE

mRNA-Seq

Sequencing: illumina HiSeq2000/2500; 101PE

Bisulfite sequencing

Capture: Agilent SureSelect Methyl-Seq Target Enrichment System (84 Mb)

Sequencing: illumina HiSeq2000/2500; 101PE

ChIP-Seq for histone modifications and RNA Polymerase II

Sequencing: HiSeq2000/2500; 36SE

IP	Marker
H3K4me3	Active
H3K4/9ac	Active
Pol II	Active
H3K36me3	Active (elongation)
H3K9me3	Silent, Heterochromatin
H3K27me3	Silent
H3K4me1	Active, Enhancer
H3K27ac	Active, Enhancer

Comprehensive catalogues of genome, transcriptome and epigenome in 26 lung adenocarcinoma cell lines

Small-molecule inhibitors to chromatin-associated factors

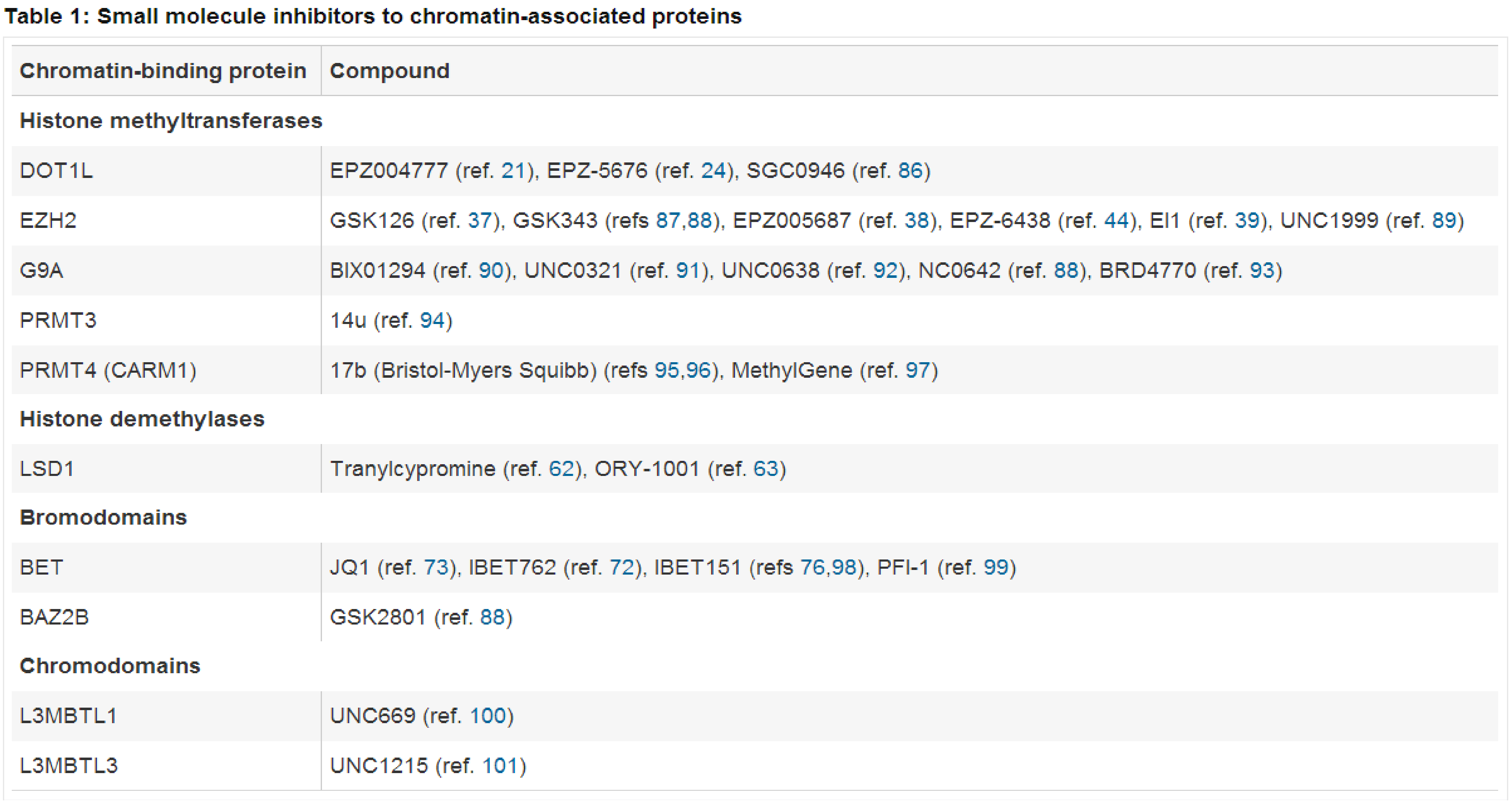


Table 1: Small molecule inhibitors to chromatin-associated proteins

Chromatin-binding protein	Compound
Histone methyltransferases	
DOT1L	EPZ004777 (ref. 21), EPZ-5676 (ref. 24), SGC0946 (ref. 86)
EZH2	GSK126 (ref. 37), GSK343 (refs 87,88), EPZ005687 (ref. 38), EPZ-6438 (ref. 44), EI1 (ref. 39), UNC1999 (ref. 89)
G9A	BIX01294 (ref. 90), UNC0321 (ref. 91), UNC0638 (ref. 92), NC0642 (ref. 88), BRD4770 (ref. 93)
PRMT3	14u (ref. 94)
PRMT4 (CARM1)	17b (Bristol-Myers Squibb) (refs 95,96), MethylGene (ref. 97)
Histone demethylases	
LSD1	Tranylcypromine (ref. 62), ORY-1001 (ref. 63)
Bromodomains	
BET	JQ1 (ref. 73), IBET762 (ref. 72), IBET151 (refs 76,98), PFI-1 (ref. 99)
BAZ2B	GSK2801 (ref. 88)
Chromodomains	
L3MBTL1	UNC669 (ref. 100)
L3MBTL3	UNC1215 (ref. 101)

Helin & Dhanak. 2013 *Nature*
Chromatin proteins and modifications as drug targets

JQ1:
a small-molecule bromodomain inhibitor

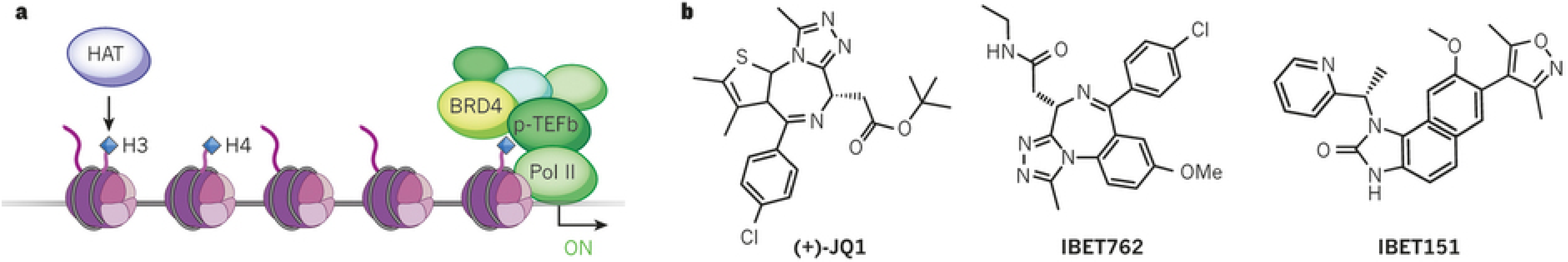
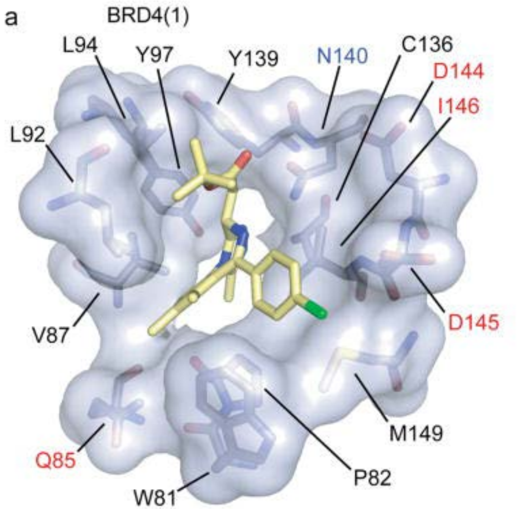


Fig. 4 Bromodomain proteins and their inhibitors.

Helin & Dhanak. 2013 *Nature*
Chromatin proteins and modifications as drug targets



Filippakopoulos et al. 2010 *Nature*
Selective inhibition of BET bromodomains

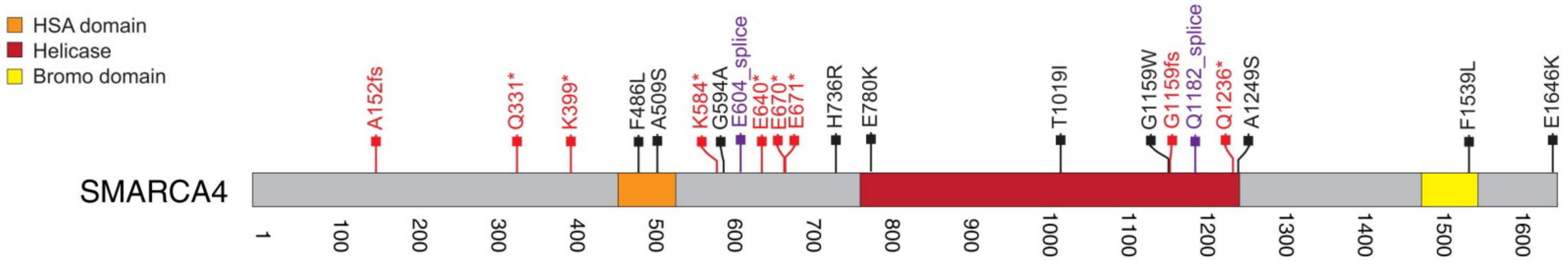
Fig. 3a The acetyl-lysine binding pocket of BRD4(1) is shown as a semi-transparent surface with contact residues labelled and depicted in stick representation. Carbon atoms in (+)-JQ1 are coloured yellow to distinguish them from protein residues. Distinguishing surface residues are shown in red; the family conserved asparagine is shown in blue.

Genomic aberrations in chromatin remodeling-related genes

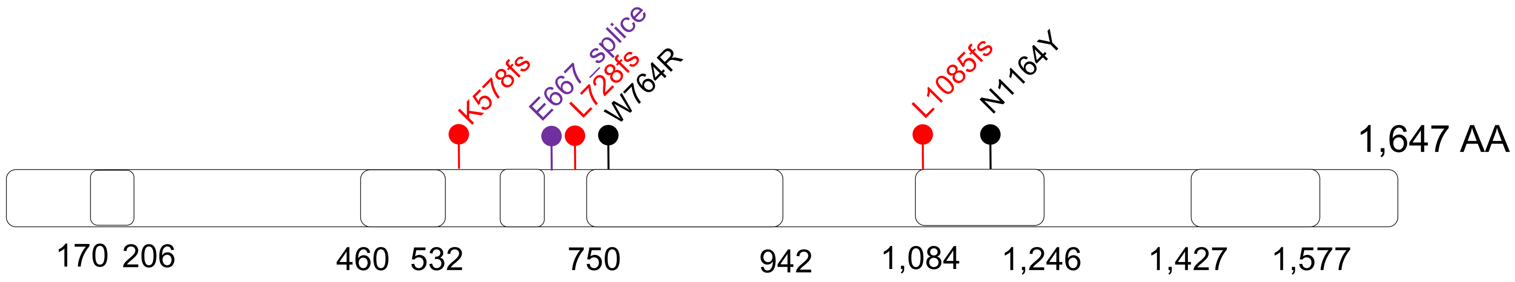
SMARCA4 (BRG1)

SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4

183 lung adenocarcinoma (Imielinski et al. 2012 *Cell*, Figure S3c)



26 lung adenocaricnoma cell lines

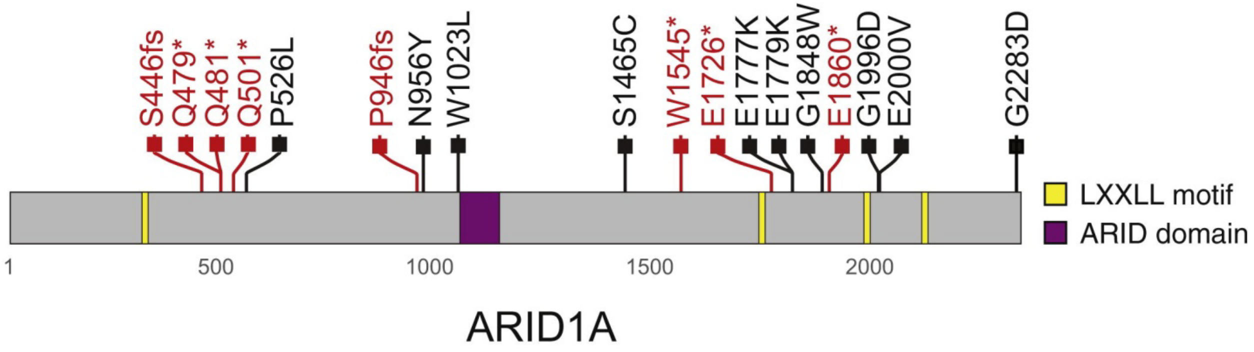


+ large deletions (>1 kb) in five cell lines

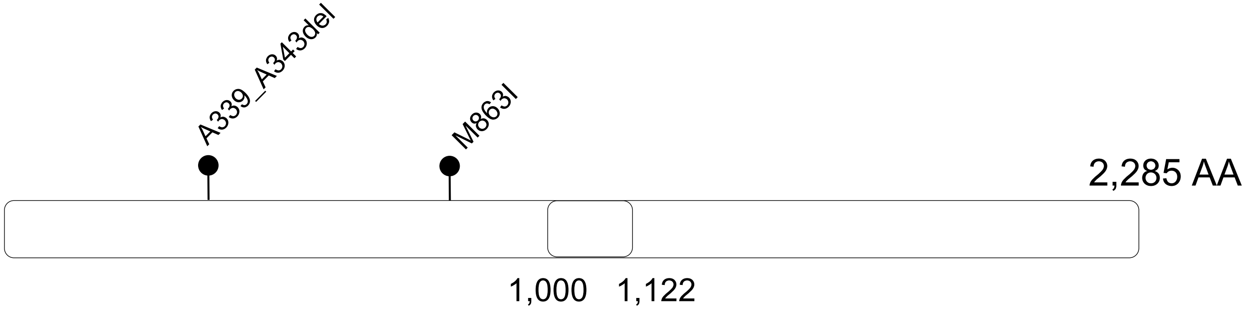
Genomic aberrations in chromatin remodeling-related genes

ARID1A (BAF250) AT rich interactive domain 1A (SWI-like)

183 lung adenocarcinoma (Imielinski et al. 2012 *Cell*, Figure 3c)



26 lung adenocaricnoma cell lines

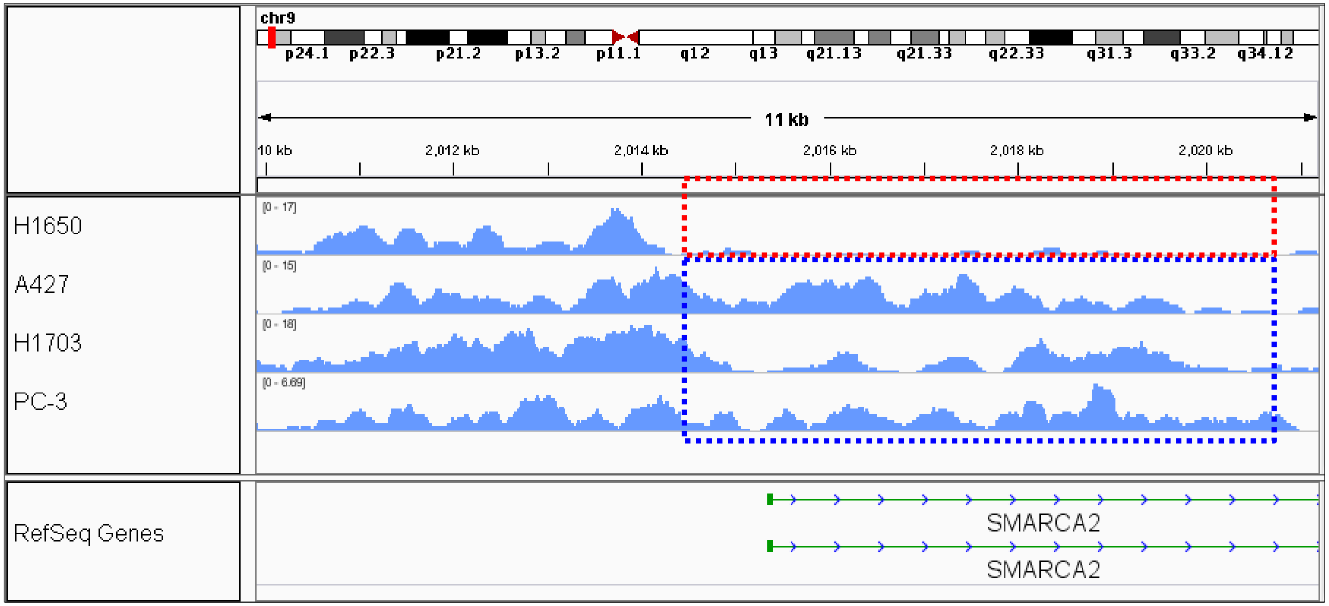


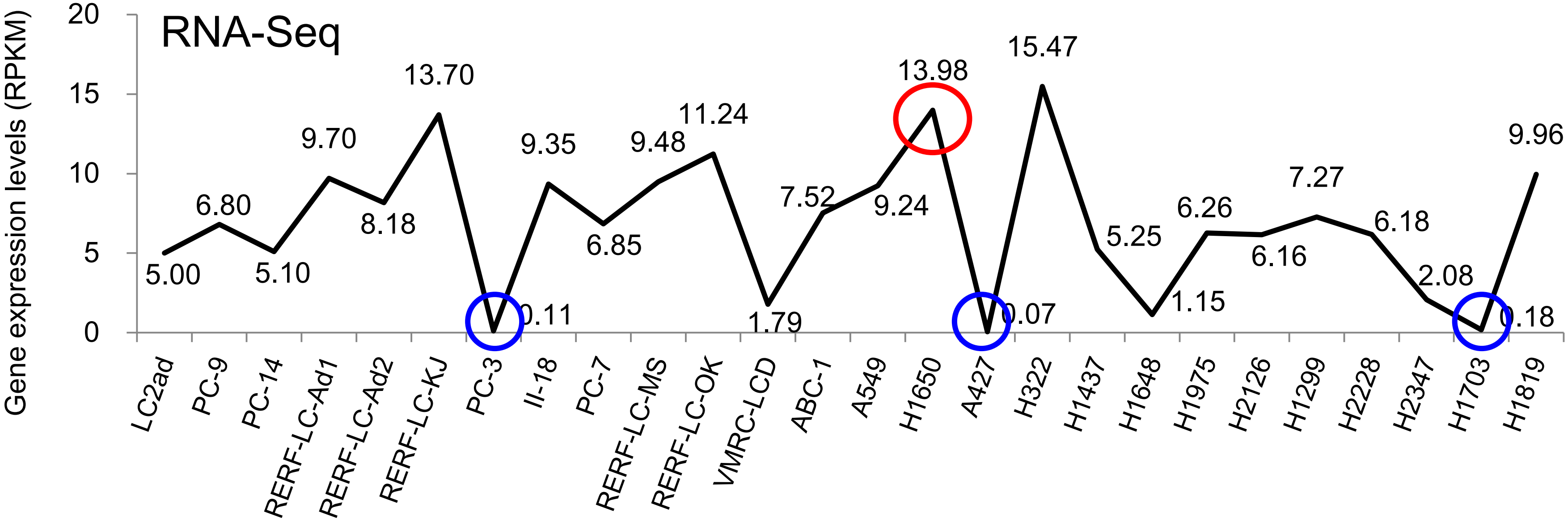
+ large deletions (>1 kb) in one cell line

Epigenomic aberrations in chromatin remodeling-related genes

SMARCA2 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2

ChIP-Seq H3K27me3 (transcriptional repressive mark)



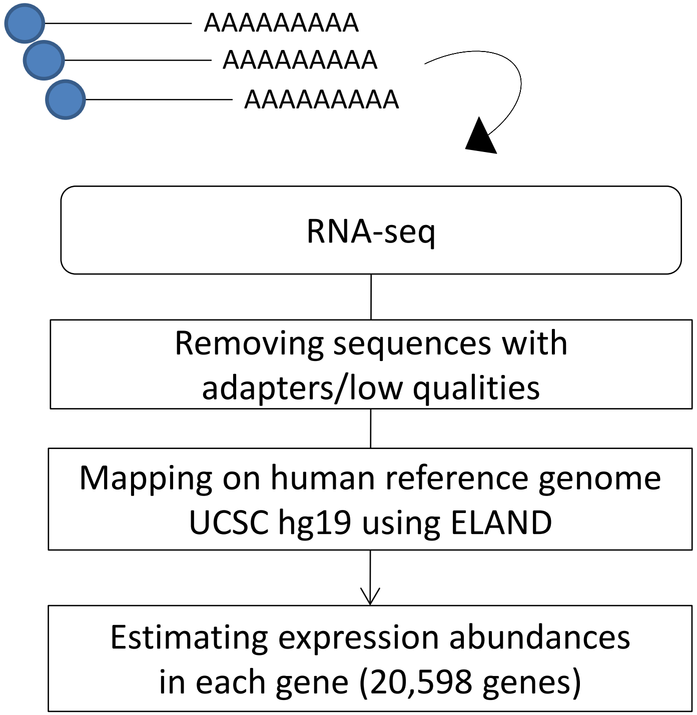


Transcriptome

RNA-seq:

- ✓ Gene expression profiles
- ✓ Fusion transcripts

Gene expression profiles from RNA-seq



RNA-seq

Removing sequences with adapters/low qualities

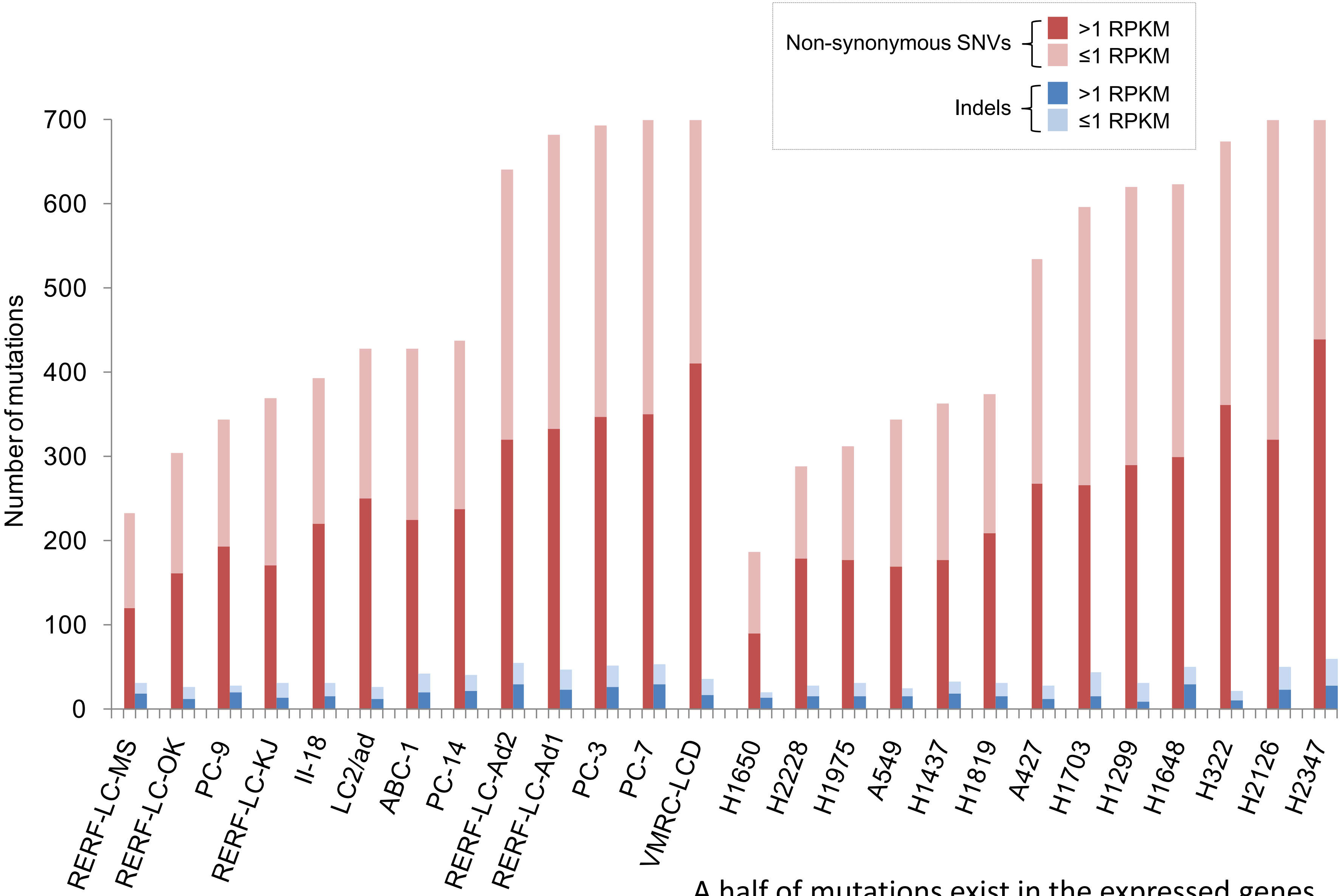
Mapping on human reference genome UCSC hg19 using ELAND

Estimating expression abundances in each gene (20,598 genes)

	Used sequences (Read1)	Num of genes	
		>1 RPKM	>5 RPKM
PC-3	49,914,547	12,205	9,240
PC-7	50,925,975	12,129	9,009
PC-9	34,167,521	12,817	9,532
PC-14	53,977,381	12,169	9,037
RERF-LC-Ad1	56,406,046	12,298	9,206
RERF-LC-Ad2	45,580,359	12,392	8,804
RERF-LC-KJ	60,803,665	12,054	8,938
RERF-LC-MS	52,715,099	13,045	9,090
RERF-LC-OK	33,086,988	12,309	8,954
VMRC-LCD	45,944,953	12,502	8,711
ABC-1	37,993,504	11,715	8,384
LC2/ad	43,665,988	12,366	9,206
II-18	63,869,445	11,955	9,038
A549	20,440,396	12,155	8,998
A427	41,895,881	11,866	9,011
H322	54,487,583	12,457	9,351
H2228	56,465,940	12,409	9,106
H1299	51,120,991	11,735	8,958
H1437	49,890,034	12,275	8,921
H1648	38,908,100	12,604	9,317
H1650	26,635,691	12,716	9,595
H1703	87,705,180	11,736	8,695
H1819	75,262,673	12,494	9,185
H1975	36,195,247	12,715	9,634
H2126	46,862,796	12,143	9,016
H2347	50,325,156	12,278	9,030

Genome Transcriptome

Genomic mutations on CDS and gene expression

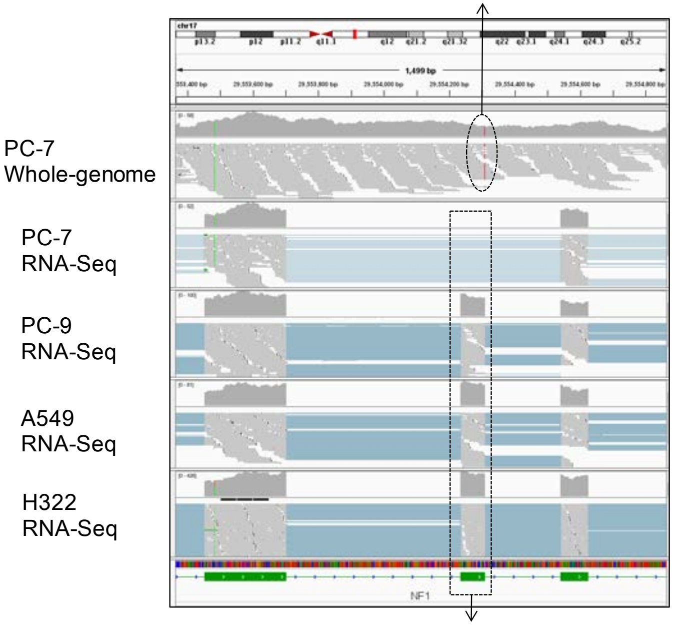


A half of mutations exist in the expressed genes.

Aberrant splicing patterns in tumor-suppressor genes

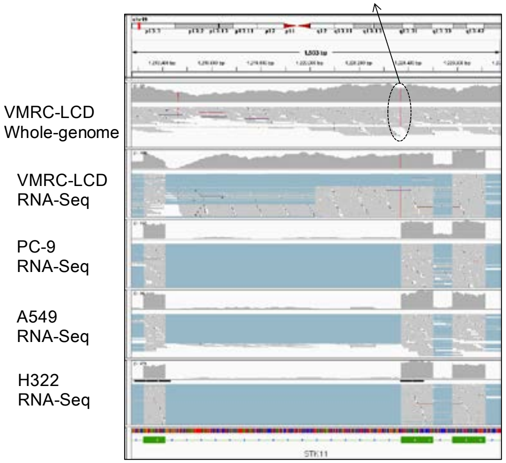
Cell line	Symbol	Mutation
PC-7	NF1	Intron 19, donor, GT>TT
VMRC-LCD	STK11	Intron 3, acceptor, AG>AT
H2228	RB1	Intron 6, acceptor, AG>A
H1650	TP53	Intron 6, acceptor, AG>GG
H1703	TP53	Intron 8, donor, GT>TT

splice site SNVs
19th intron, donor, GT>TT

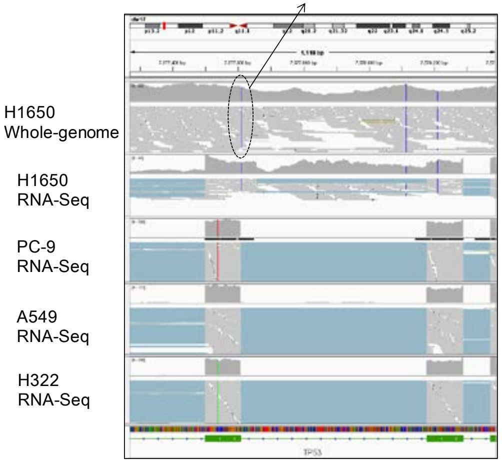


19th exon of NF1

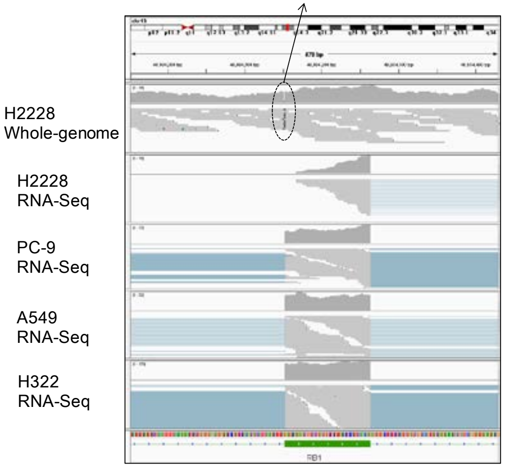
splice site SNVs
3th intron, acceptor, AG>AT



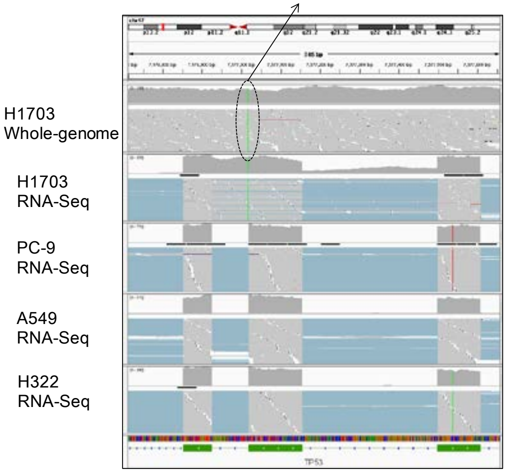
splice site SNVs
6th intron, acceptor, AG>GG



splice site indels
6th intron, acceptor, AG>A

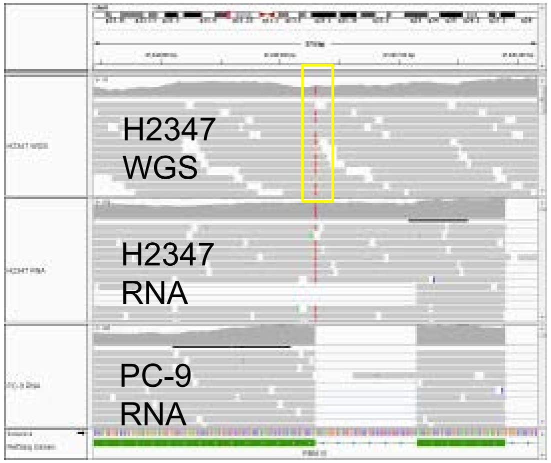


splice site SNVs
8th intron, donor, GT>TT



Examples of aberrant splicing patterns

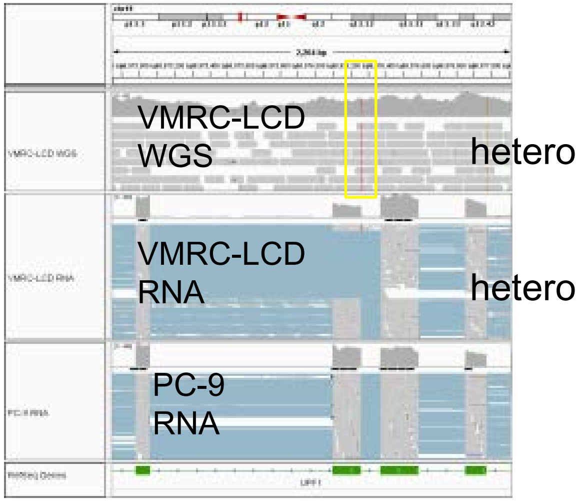
RBM10 RNA binding motif protein 10



H2347; Intron 20, donor, GT>TT;
Intron read-through (p.V785_splice)

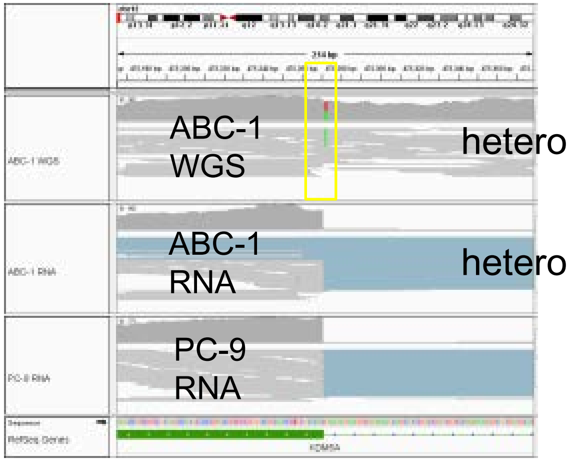
RBM10 was reported as a frequently mutated gene in lung adenocarcinoma (Imielinski et al. 2012 *Cell*).

UPF1 UPF1 regulator of nonsense transcripts homolog (yeast)



VMRC-LCD; Intron 21, donor, GT>TT;
Exon skipping

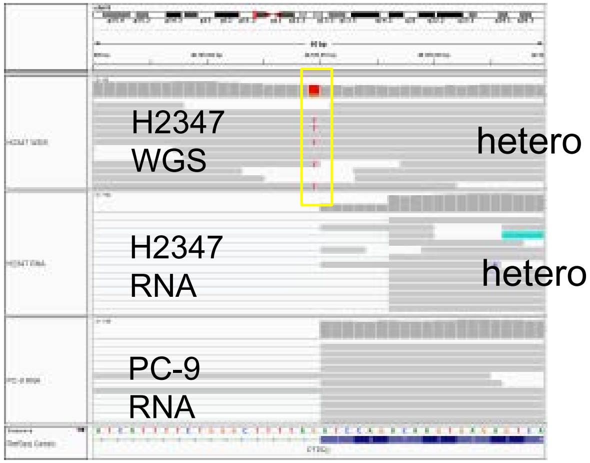
KDM5A lysine (K)-specific demethylase 5A



ABC-1; Intron 3, acceptor, AG>TG;
Exon skipping

PTPRJ protein tyrosine phosphatase, receptor type, J

PTPRJ-C11orf54 fusion was detected in H322 cell line.

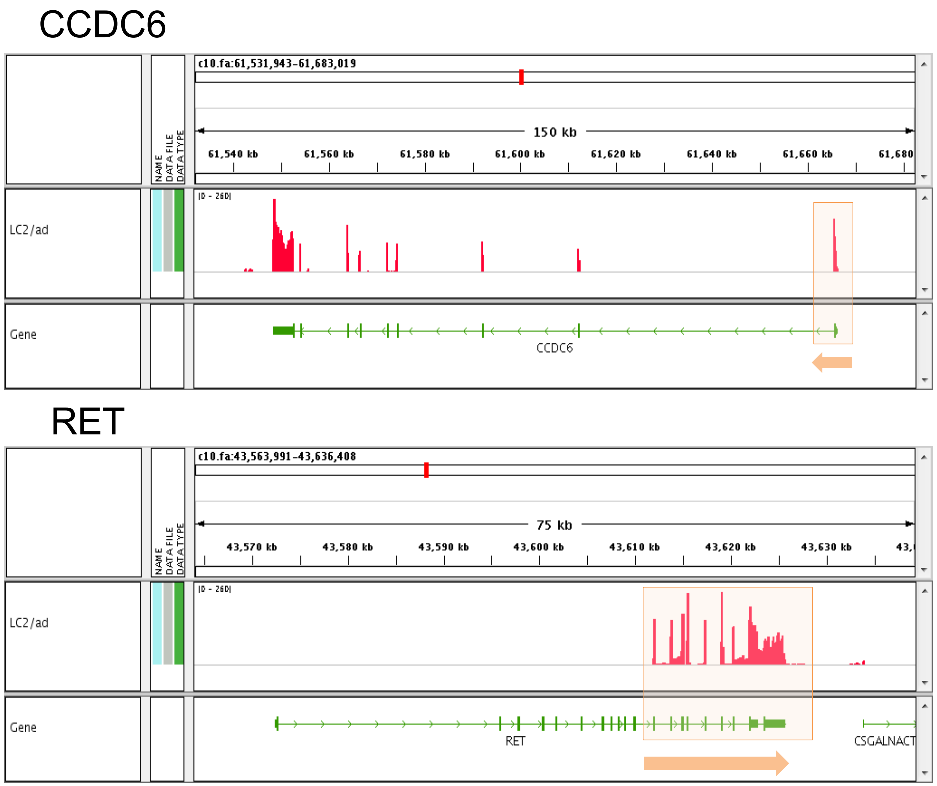


H2347; Intron 22, acceptor, AG>AT;
Deletion (p.I1187_Q1188del)

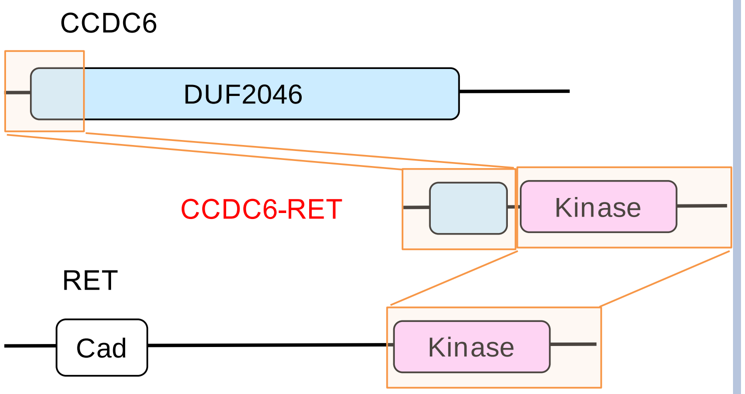
Known oncogenic fusion transcripts

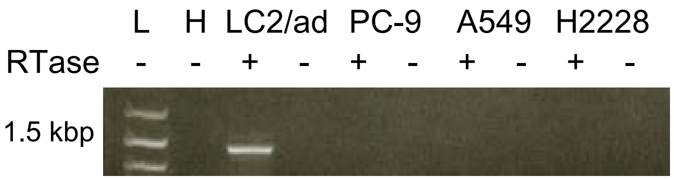
CCDC6-RET fusion in LC2/ad

Cell line	Fusion	Chrom	Strand	Coordinates		Spanning reads	Spanning pairs	Spanning pairs where one end spans a fusion
				On the left	On the right			
LC2/ad	CCDC6-RET	chr10-chr10	rf	61,665,879	43,612,031	184	27	98


CCDC6

RET



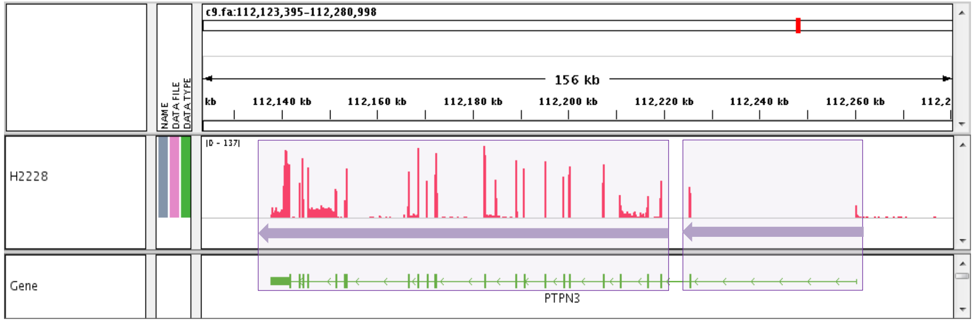


L: Ladder, H: H_2O

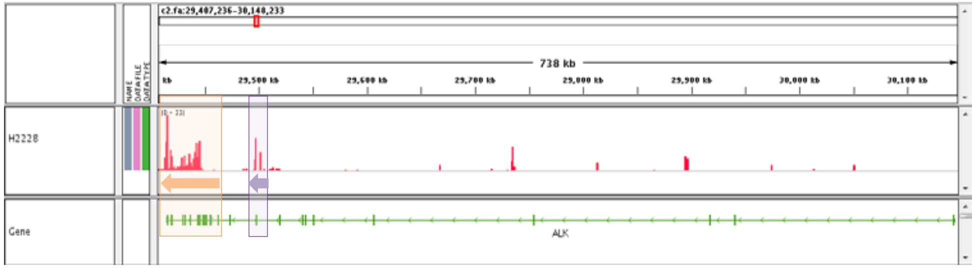
From the RNA-seq data, known driver fusion transcripts such as CCDC6-RET in LC2/ad were identified (Matsubara et al. 2012; Takeuchi et al. 2012; Suzuki et al. 2013).

ALK-related fusions (ALK-PTPN3, EML4-ALK) in H2228

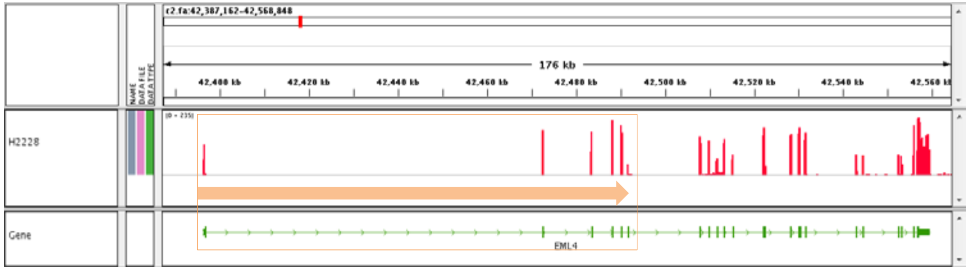
PTPN3

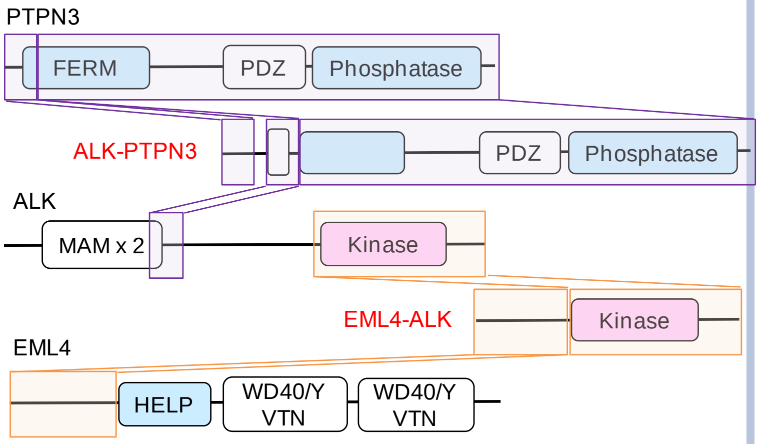


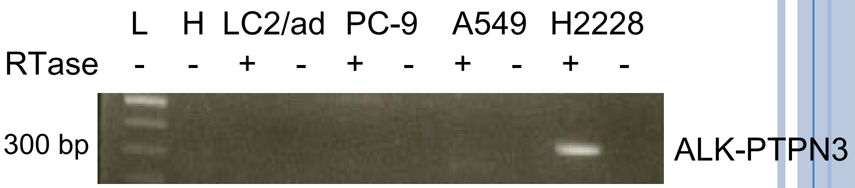
ALK

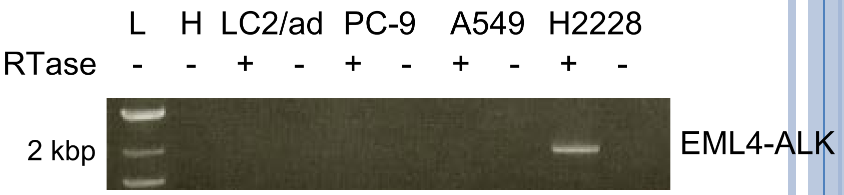


EML4









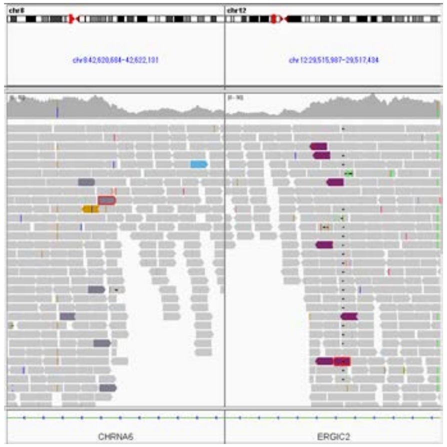
L: Ladder, H: H_2O

From the RNA-seq analysis, ALK-PTPN3 fusion was detected in H2228 cell line as reported in the previous study (Jung et al. *Genes Chromosomes Cancer* 2012). EML4-ALK was also previously reported and detected by RT-PCR but not detected by the computational analysis.

Novel fusion transcripts

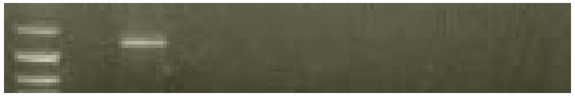
ERGIC2-CHRNA6 in H1437

WGS

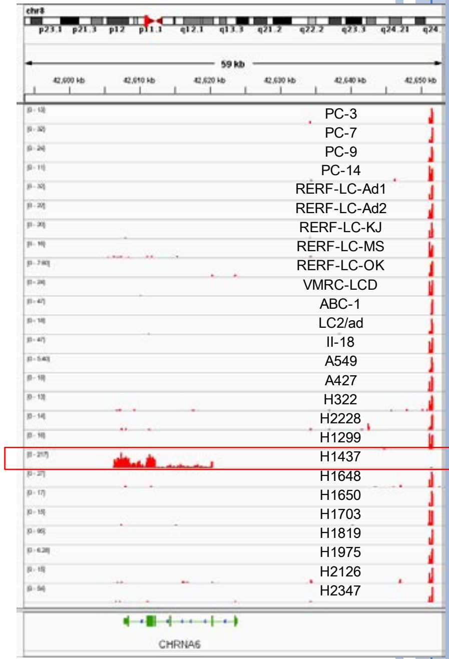


ERGIC2 ERGIC and golgi 2
CHRNA6 cholinergic receptor, nicotinic, alpha 6 (neuronal)

	L	H	H1437		PC-9		A549		H322	
RTase	-	-	+	-	+	-	+	-	+	-

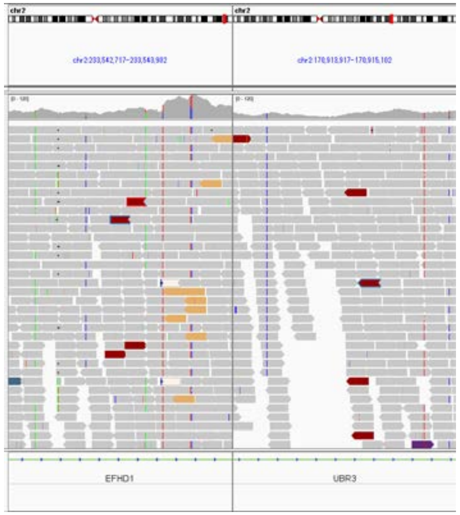


L: Ladder, H: H_2O



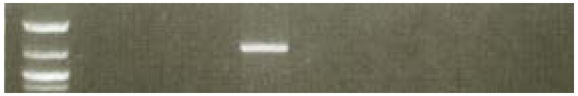
EFHD1-UBR3 in PC-9

WGS



EFHD1 EF-hand domain family, member D1
UBR3 ubiquitin protein ligase E3 component n-recognin 3 (putative)

	L	H	LC2/ad		PC-9		A549		H322	
RTase	-	-	+	-	+	-	+	-	+	-

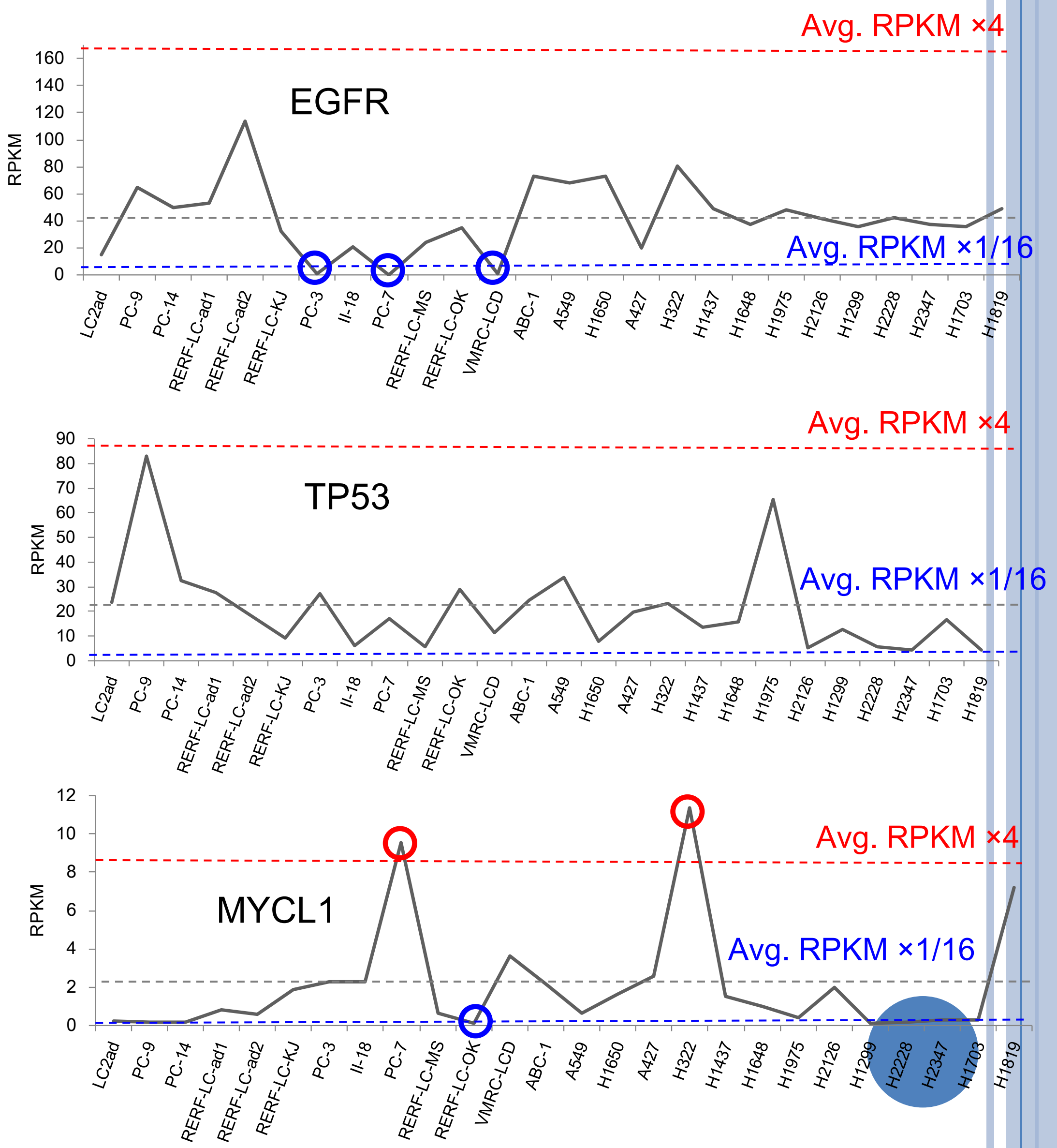


L: Ladder, H: H_2O

実際にfunctionalかどうかはわからない。

Differentially expressed genes in 26 cell lines

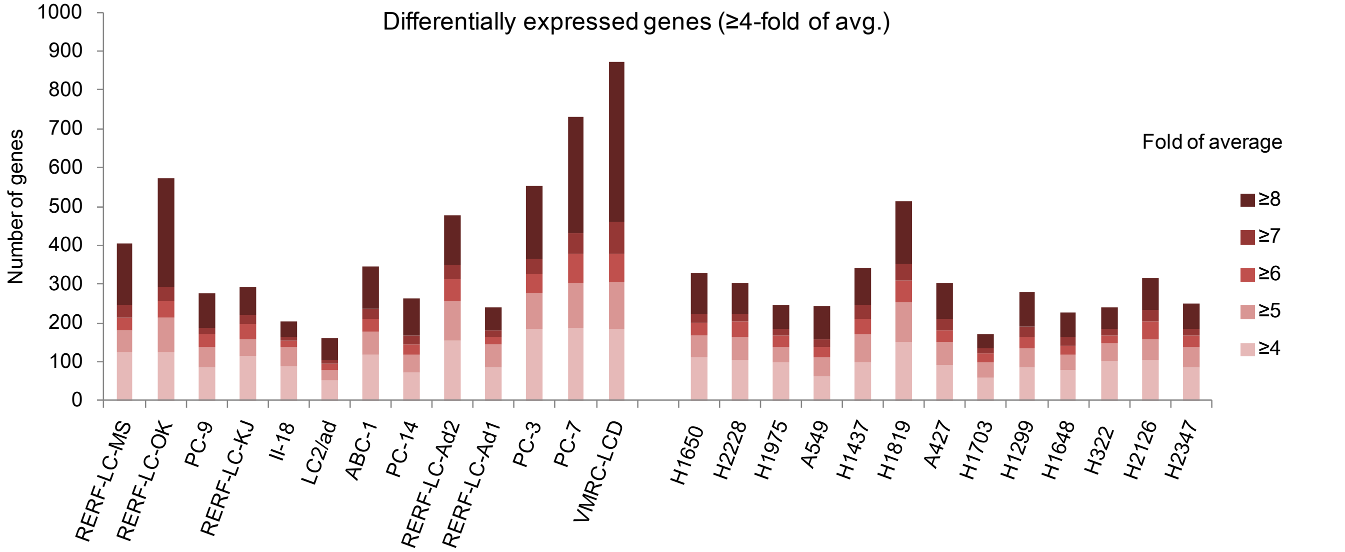
	Num of genes*	
	High expression (>4 fold of avg.)	Low expression (<1/16 fold of avg.)
PC-3	554	2,323
PC-7	731	2,700
PC-9	277	1,504
PC-14	264	2,019
RERF-LC-Ad1	240	1,661
RERF-LC-Ad2	477	1,583
RERF-LC-KJ	293	2,178
RERF-LC-MS	403	918
RERF-LC-OK	573	2,109
VMRC-LCD	871	1,818
ABC-1	346	2,636
LC2/ad	160	1,527
II-18	203	2,478
A549	242	1,968
A427	304	2,869
H322	241	1,828
H2228	304	1,663
H1299	279	2,775
H1437	341	2,007
H1648	226	1,389
H1650	328	1,511
H1703	170	2,697
H1819	512	1,626
H1975	248	1,587
H2126	315	2,033
H2347	251	1,739

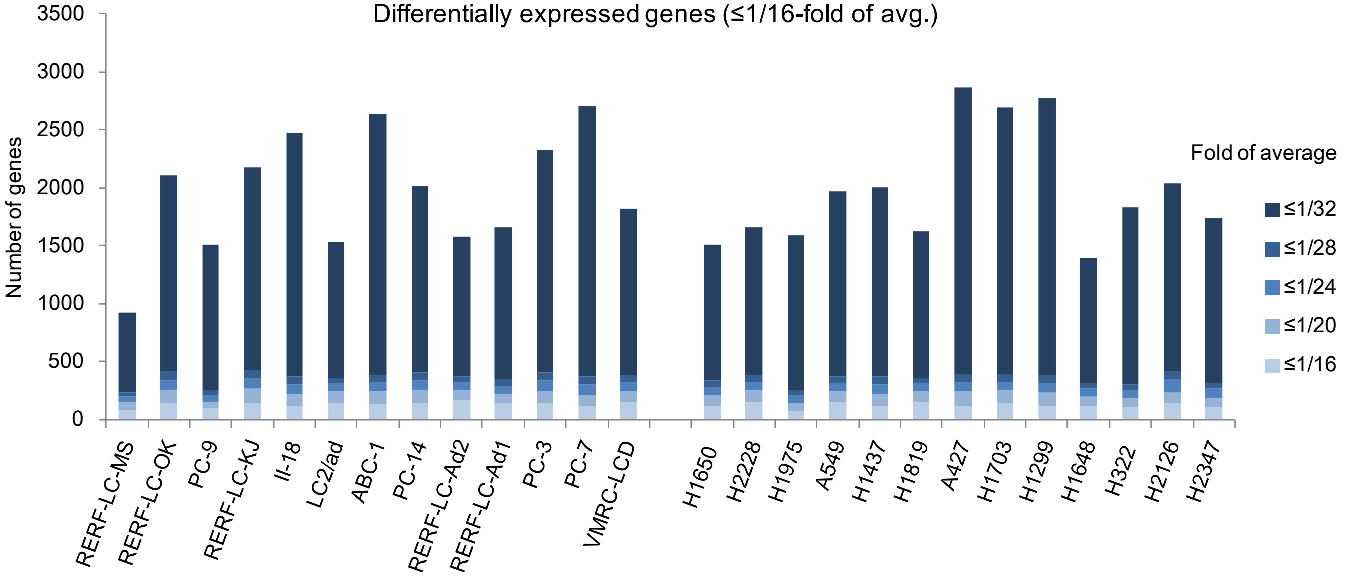


*Total 16,573 genes were used in this analysis: Avg. RPKM > 0, ≥1 cell lines with >1 RPKM

Transcriptome

Differentially expressed genes in 26 cell lines

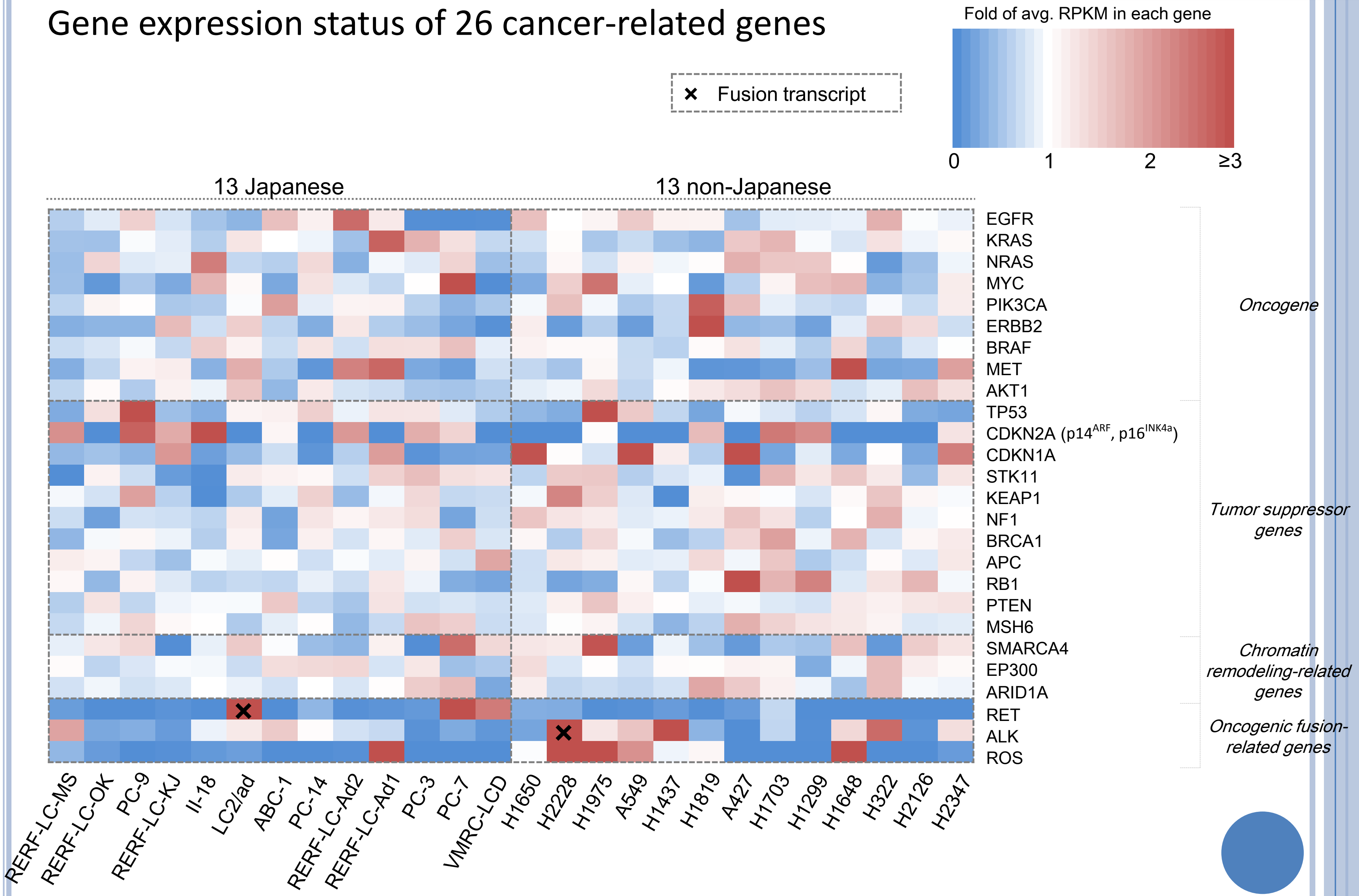




細胞株によってdifferentially expressed geneの数に差がある。

Transcriptome

Gene expression status of 26 cancer-related genes



細胞株によって発現量に差がある遺伝子がどのような制御を受けているか？
→エピゲノム解析へ

Epigenome①

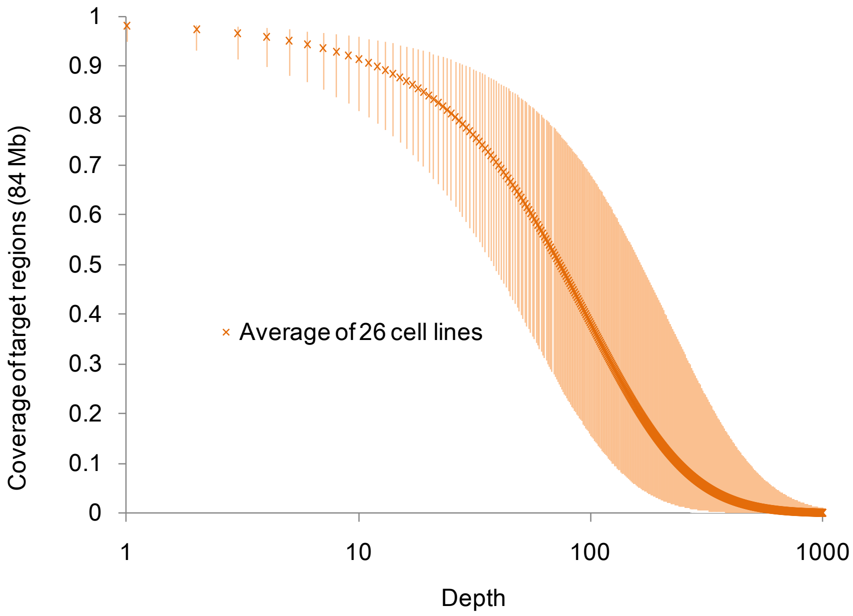
Target captured-bisulfite sequencing:

- ✓ DNA methylation profiles in regulatory regions

Target captured-bisulfite sequencing

Approximately 100 million mapped reads (50 million pairs) were obtained in each cell line.

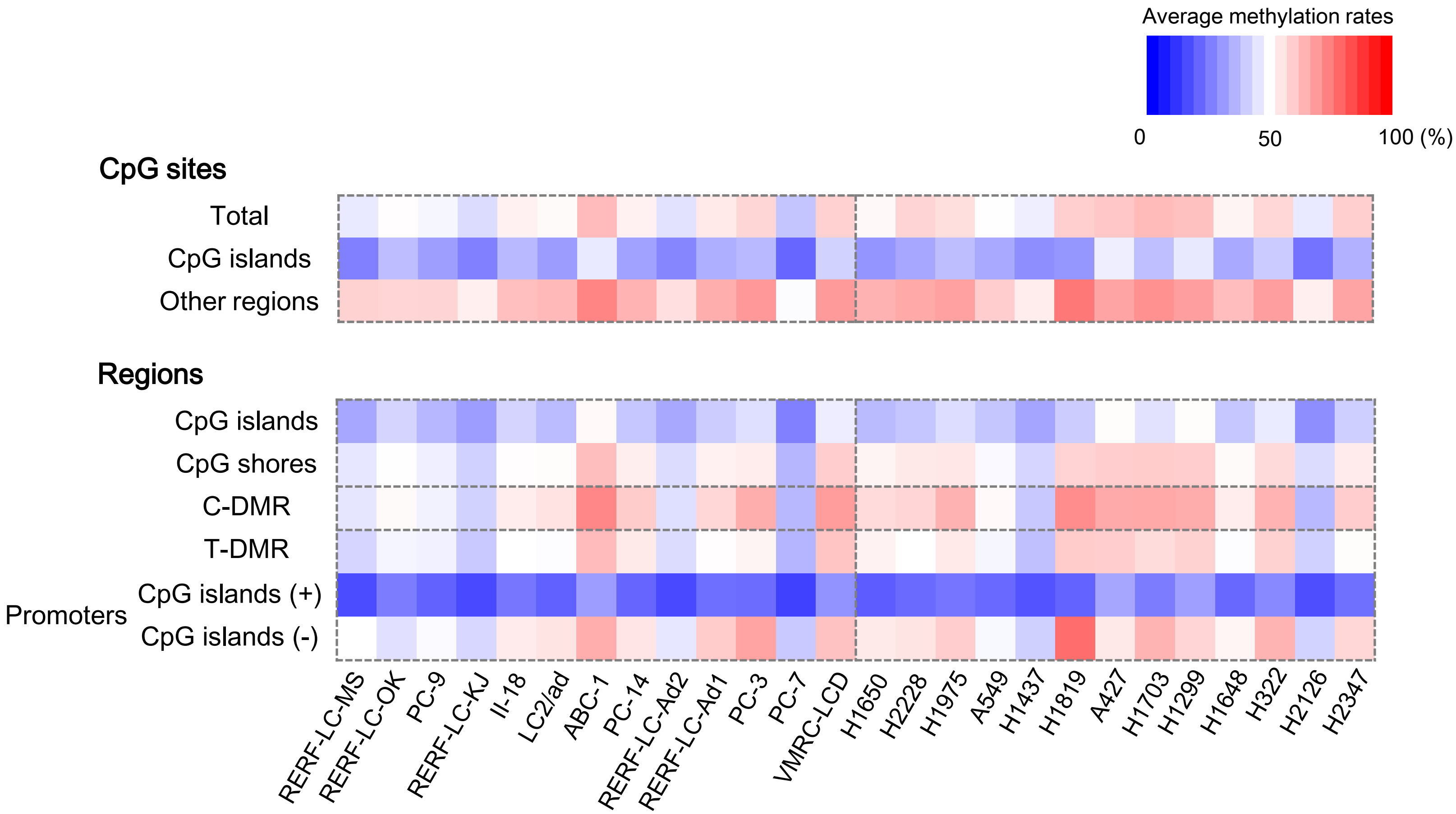
Average depth: 109.7
x10 coverage: 91%
(Total length of the bait regions: 84Mb)



	Mapped sequences (R1+R2)	Depth (avg)	Coverage (x10)	Conversion rate (x5)	CpG sites (>x5)
PC-3	157,902,653	161.4	0.93	0.99	3,673,159
PC-7	109,919,011	110.9	0.93	0.99	3,418,929
PC-9	87,012,056	89.6	0.90	0.99	3,231,320
PC-14	204,216,479	210.3	0.96	0.99	4,064,068
RERF-LC-Ad1	87,043,746	89.1	0.90	0.99	3,264,395
RERF-LC-Ad2	78,300,691	83.0	0.92	0.99	3,448,211
RERF-LC-KJ	72,844,738	74.9	0.88	0.99	3,068,971
RERF-LC-MS	102,938,936	109.0	0.94	0.99	3,598,662
RERF-LC-OK	161,552,507	165.0	0.95	0.99	3,758,532
VMRC-LCD	84,681,570	89.5	0.91	0.99	3,136,774
LC2/ad	112,097,386	116.0	0.93	0.99	3,548,548
ABC-1	93,158,547	93.1	0.93	0.99	3,493,903
II-18	99,682,438	165.0	0.91	0.99	3,327,001
A549	87,966,180	91.0	0.91	0.99	3,324,364
A427	53,499,542	54.3	0.81	0.99	2,614,641
H322	153,896,186	165.8	0.95	0.99	4,161,775
H2228	122,705,759	81.6	0.90	0.99	4,815,543
H1299	118,923,875	82.2	0.91	0.99	4,533,930
H1437	98,311,209	63.1	0.88	0.99	4,382,225
H1648	102,033,841	104.4	0.91	0.99	3,357,747
H1650	105,694,196	109.4	0.93	0.99	3,460,378
H1703	127,897,486	81.6	0.91	0.99	5,513,896
H1819	220,008,485	223.4	0.95	0.99	4,085,231
H1975	79,688,628	81.7	0.91	0.99	3,274,116
H2126	124,651,437	80.2	0.90	0.99	4,991,289
H2347	115,973,241	76.1	0.89	0.99	4,661,415

Depths and coverage were calculated using BEDTools (Quinlan AR and Hall IM. 2010 *Bioinformatics*).
Conversion rate: (TA+TT+TC) / (CA+CT+CC+TA+TT+TC).

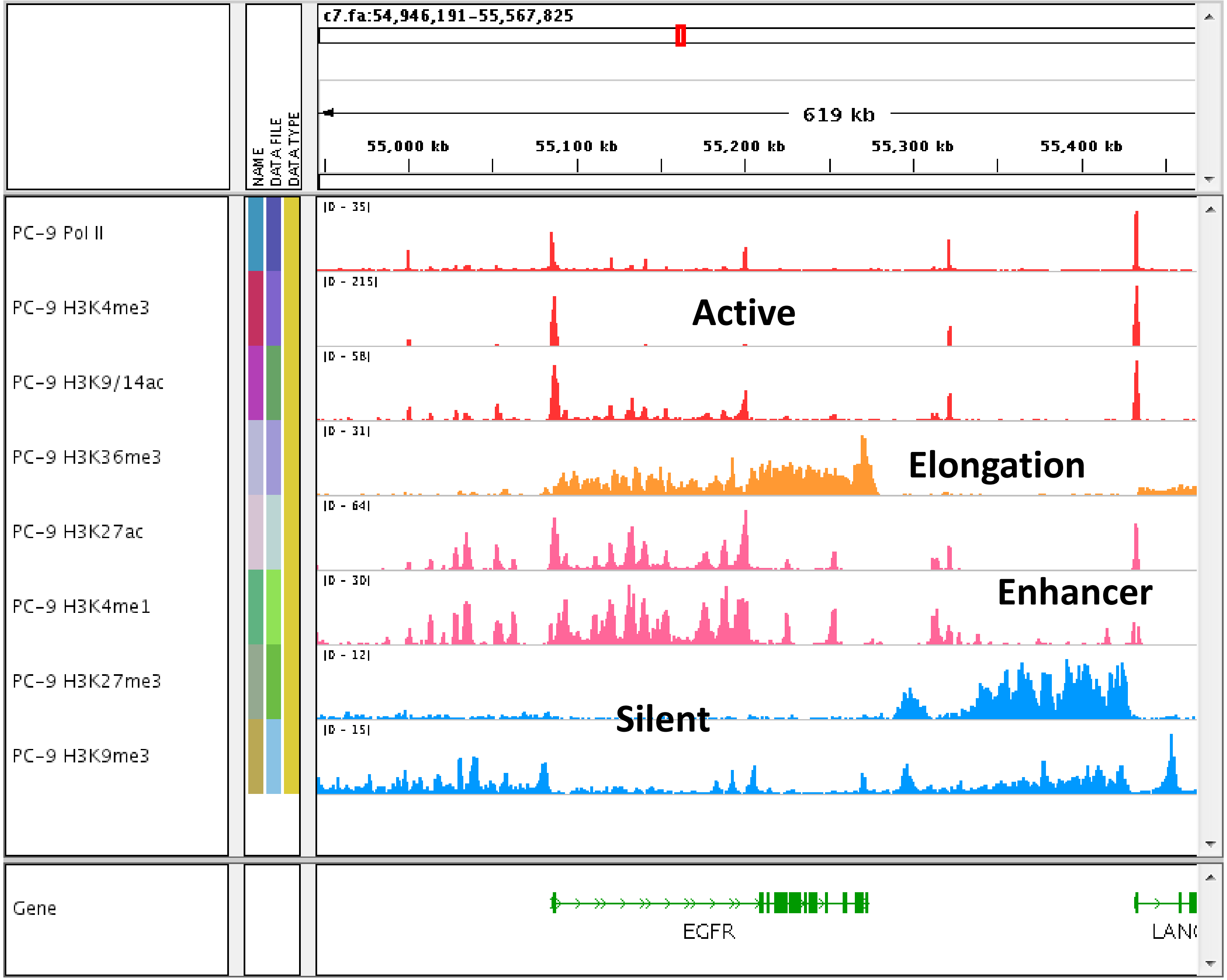
Average methylation rates in each cell line



CpG islandsは、低メチル化。
CpG islands以外のCpG siteのメチル化率は、cell lineによって異なり、variationがある。

Epigenome

Histone modification & RNA Polymerase II binding status
PC-9



Active

Elongation

Enhancer

Silent

ChIP-seq

Mapped sequences (avg. of 26 cell lines)

WCE	H3K4me3	H3K9/14ac	Pol II	H3K36me3	H3K4me1	H3K27ac	H3K27me3	H3K9me3
19,100,553	26,140,455	19,596,187	26,056,772	24,264,604	25,900,257	25,690,276	21,584,812	21,155,573

MACS2 peaks (avg. of 26 cell lines)

	H3K4me3	H3K9/14ac	Pol II	H3K36me3	H3K4me1	H3K27ac	H3K27me3	H3K9me3
narrow peaks	21,209	34,374	15,715	107,708	108,882	61,061	53,587	39,559
narrow & broad peaks	16,208	23,753	13,997	47,710	75,854	38,297	42,163	51,760

Replicates

H1975 H3K4me3

rep#1: 130705_Hiseq3A

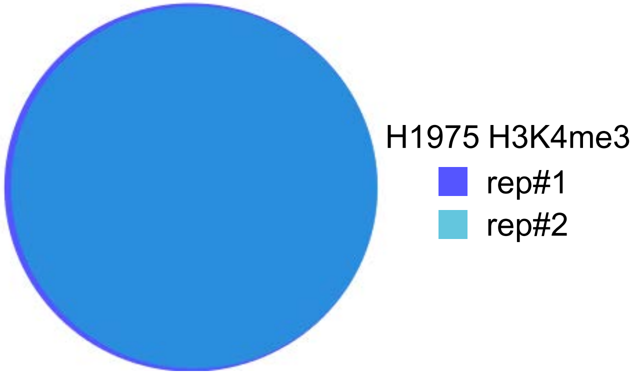
rep#2: 130625_Hiseq3A

control (WCE): 130625_Hiseq3A

Number of genes overlapping* with MACS2 peaks

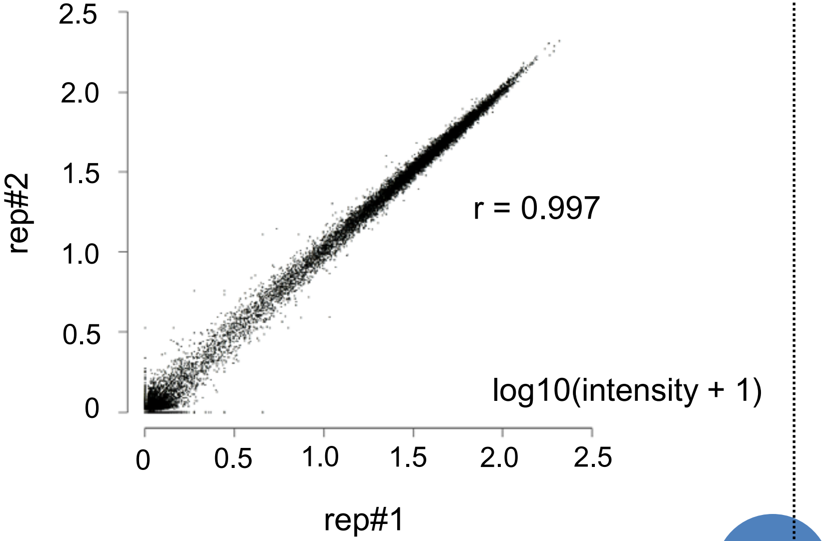
	rep#1	rep#2
H1975 H3K4me3	12,104	11,708

11,703 (96.6%)



Signal intensities

(intensity) = (IP PPM*)/(WCE PPM*)



*±1.5 Kb from TSS

r: Pearson correlation coefficient

Comparison with ENCODE data

ENCODE DCC (Data Coordination Center)

A549 H3K4me3

Our dataset: 120531_SangiB
Our dataset control (WCE): 120626_SangiA
ENCODE rep#1, rep#2: wgEncodeEH001905 (DCC Acc)
ENCODE control (standard control): wgEncodeEH001904

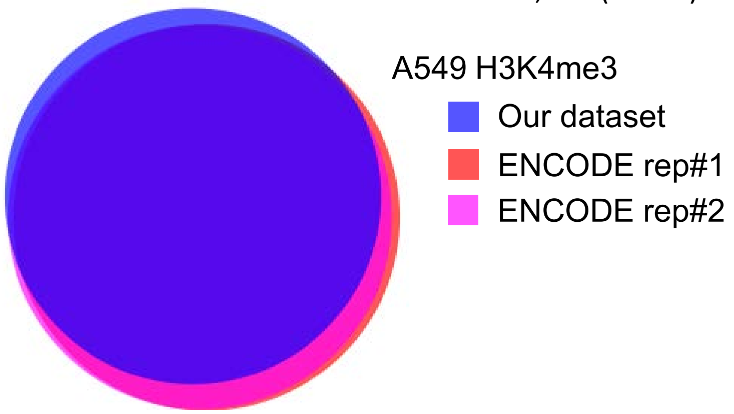
Number of genes overlapping* with narrow peaks

	Our dataset	ENCODE rep#1	ENCODE rep#2
A549 H3K4me3	11,898	13,424	13,375

11,820 (87.5%)

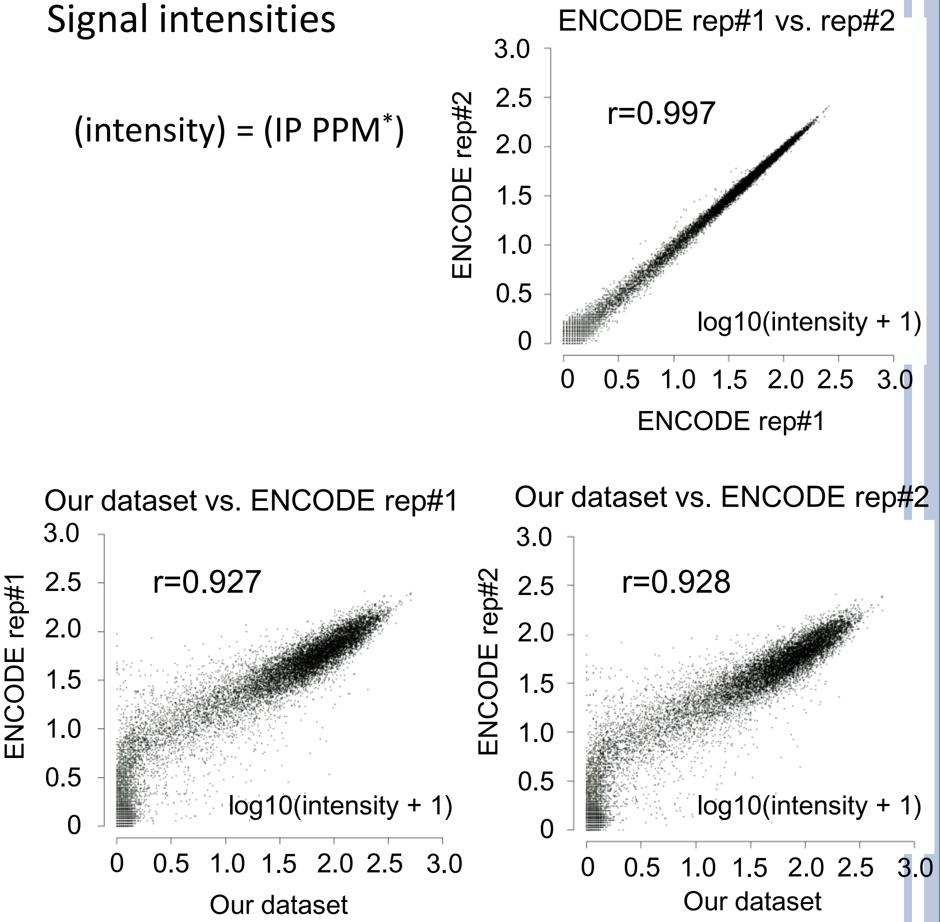
11,807 (87.7%)

13,262 (98.0%)



Signal intensities

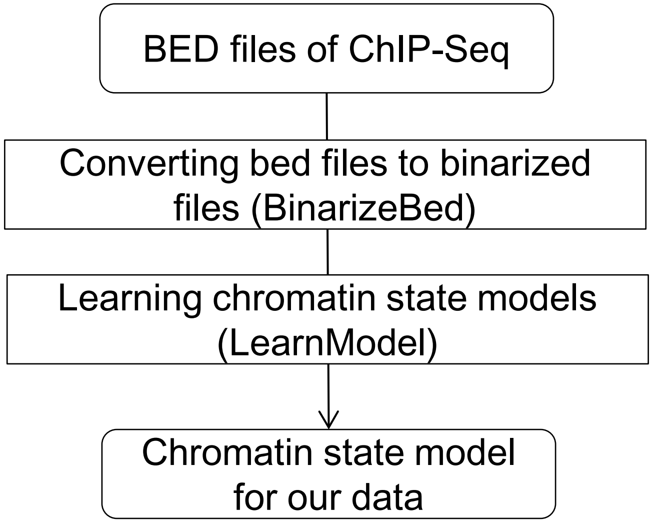
(intensity) = (IP PPM*)



*±1.5 Kb from TSS

ChromHMM

Using ChromHMM, chromatin states were detected and characterized from ChIP-Seq data of the eight chromatin marks.

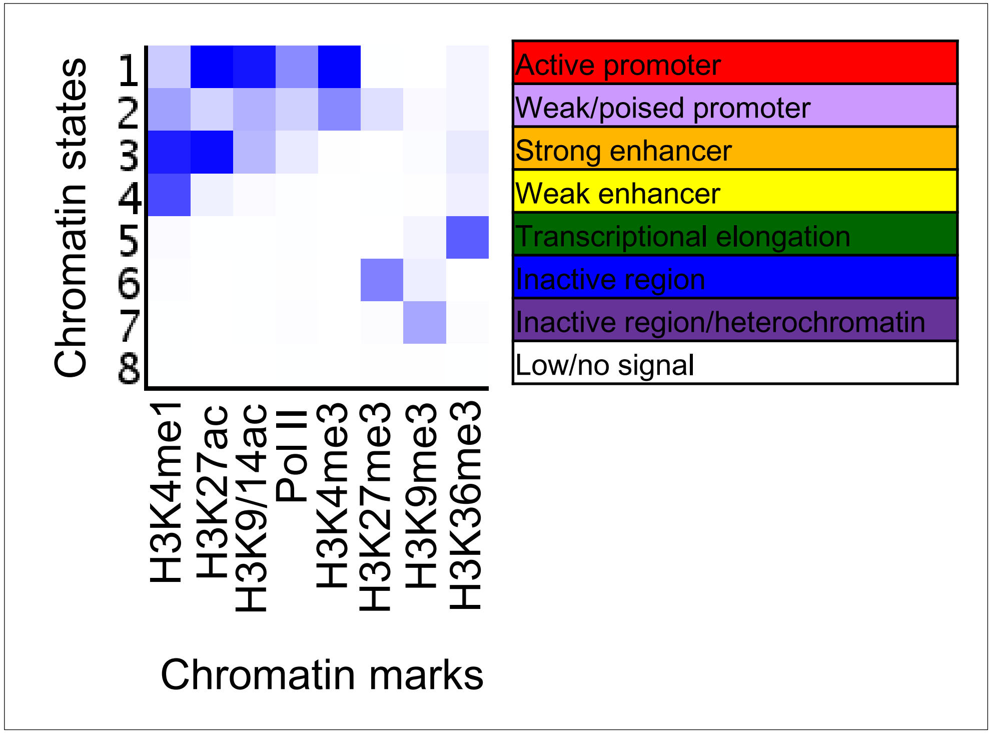


BED files of ChIP-Seq

Converting bed files to binarized files (BinarizeBed)

Learning chromatin state models (LearnModel)

Chromatin state model for our data

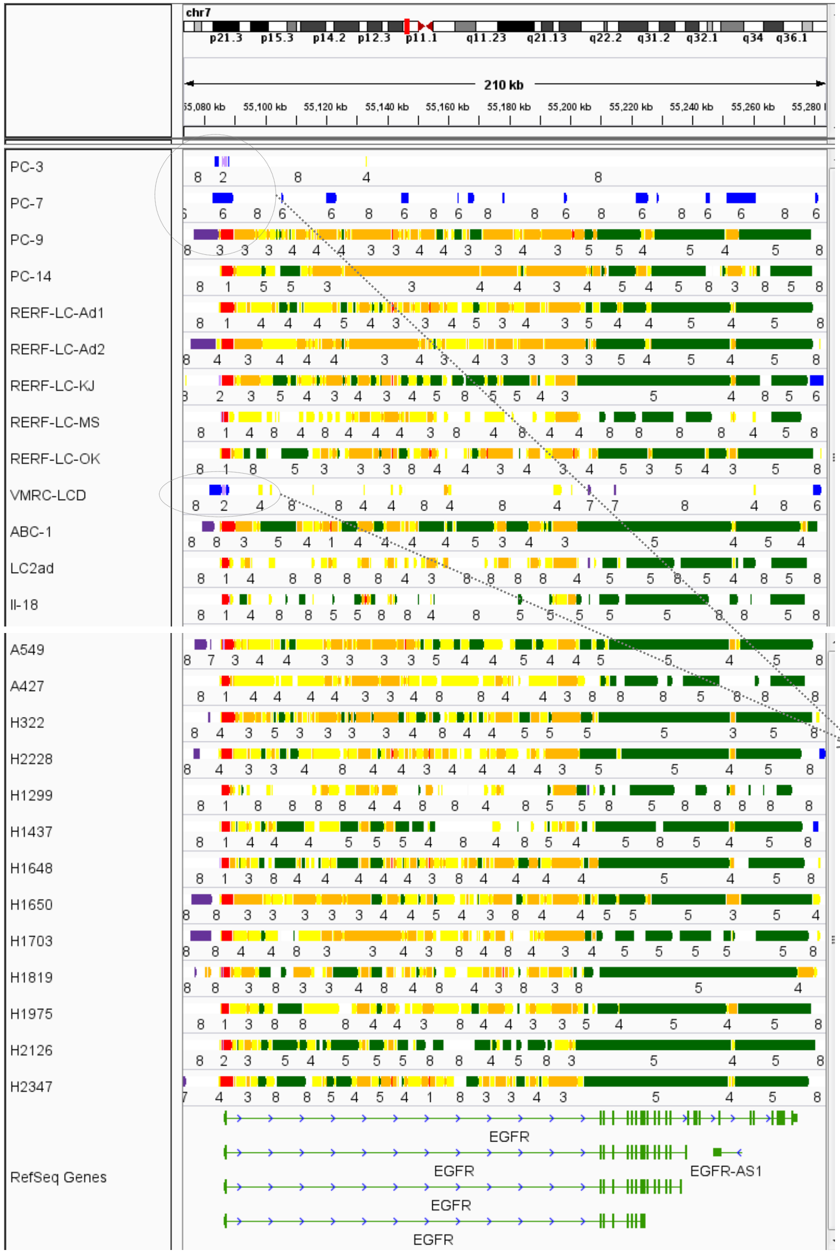


We learned and analyzed eight chromatin states.

ChromHMM: a program for the learning chromatin states using a multivariate Hidden Markov model

Ernst et al. 2011 *Nature*
Ernst and Kellis. 2012 *Nat methods*

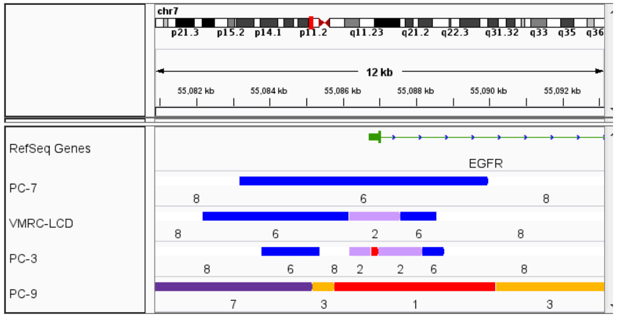
ChromHMM on IGV (EGFR)



Candidate state annotation

1	Active promoter
2	Weak/poised promoter
3	Strong enhancer
4	Weak enhancer
5	Transcriptional elongation
6	Inactive region
7	Inactive region/heterochromatin
8	Low/no signal

Chromatin states around TSS of EGFR

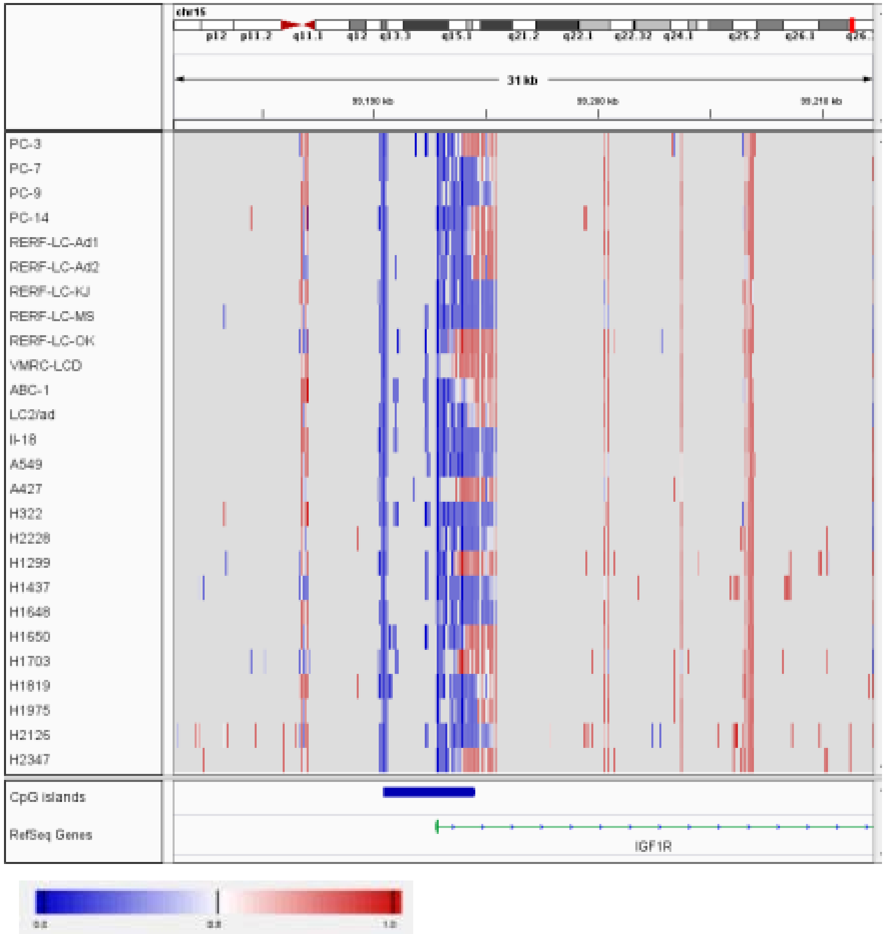

chr7
p21.3 p15.2 p14.1 p11.2 q11.23 q21.2 q22.3 q31.32 q33 q35 q36
12 kb
55,082 kb 55,084 kb 55,086 kb 55,088 kb 55,090 kb 55,092 kb
RefSeq Genes
EGFR
PC-7
8 6 8
VMRC-LCD
8 6 2 6 8
PC-3
8 6 8 2 2 6 8
PC-9
7 3 1 3

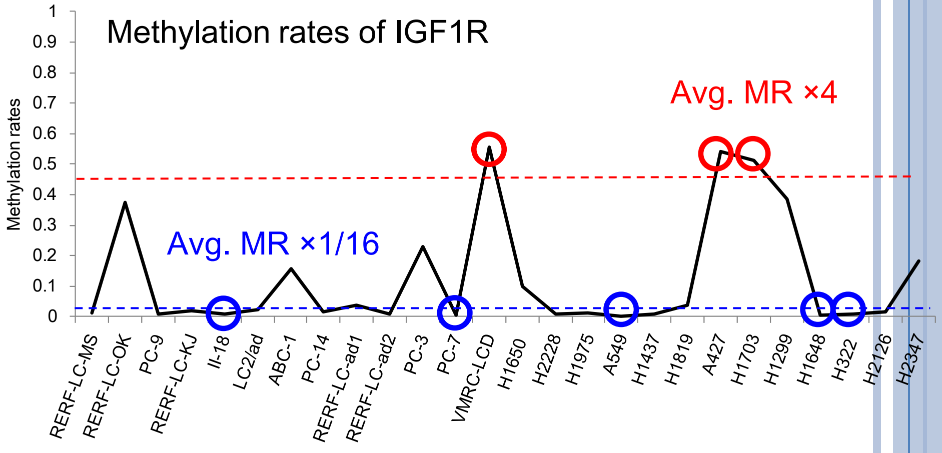
Active chromatin marks

H3K4me3	Pol II	H3K36me3
×	×	×
○	×	×
○	○	×
○	○	○

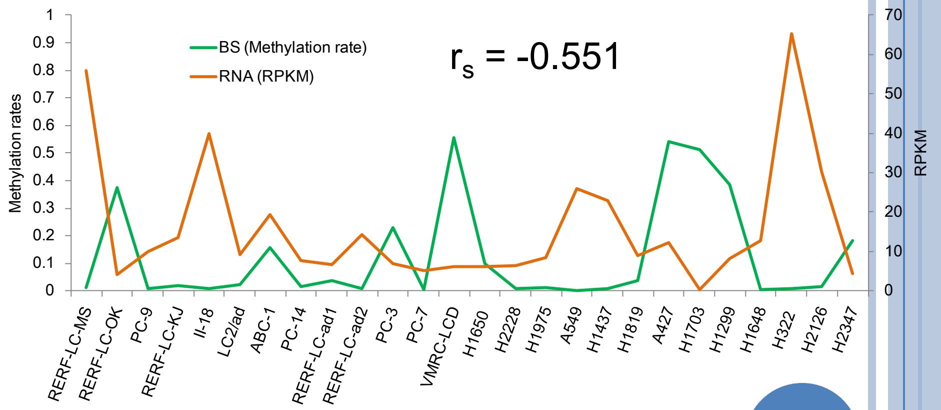
Differentially methylated genes in 26 cell lines (example)

IGF1R insulin-like growth factor 1 receptor





RPKM (RNA-Seq) vs. Methylation rates (BS-Seq)

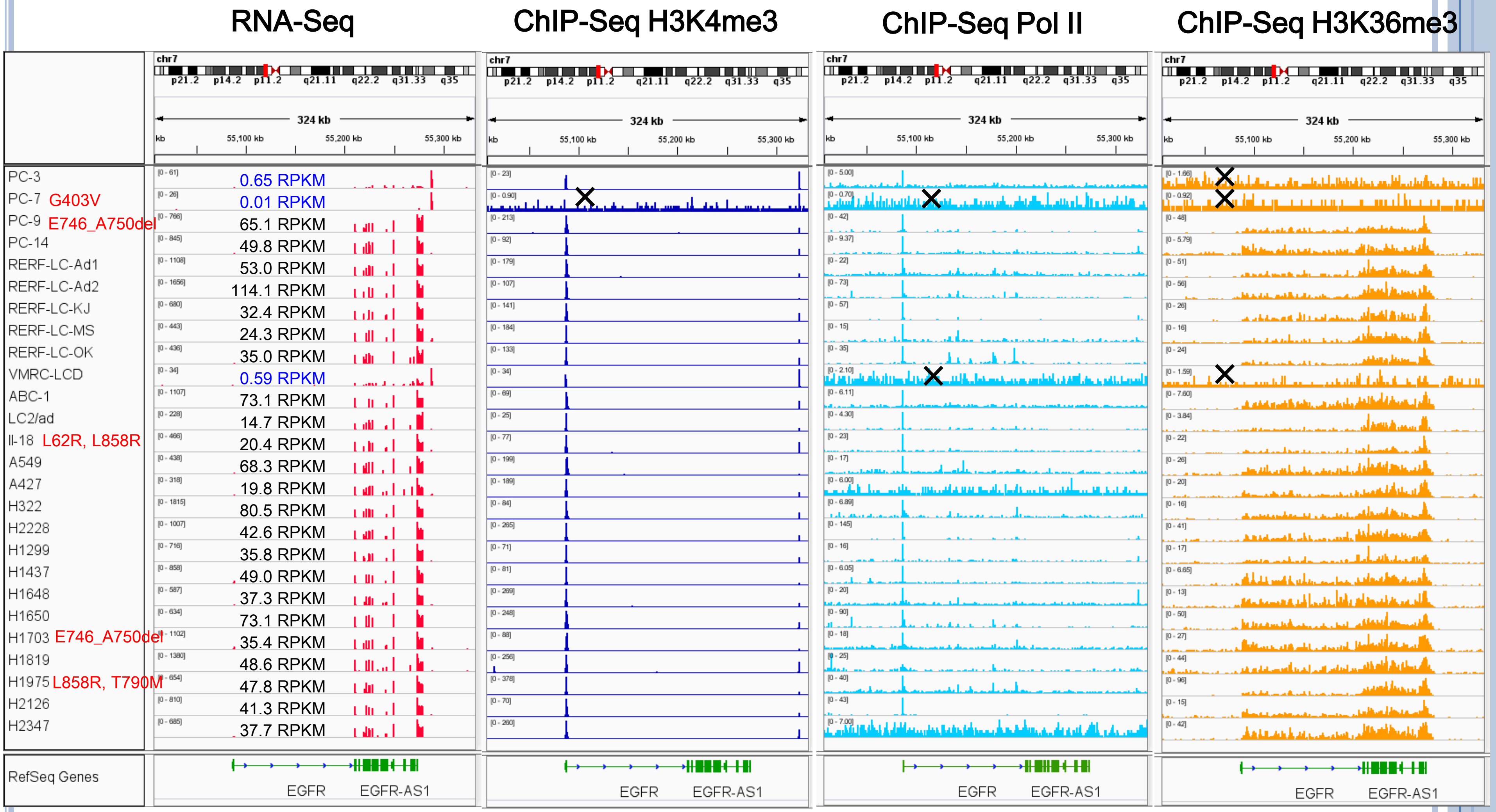


IGF1R gene was detected as one of the differentially methylated genes in the 26 cell lines. In IGF1R promoters, three cell lines are highly methylated and five cell lines show lower DNA methylation.

EGFR epidermal growth factor receptor

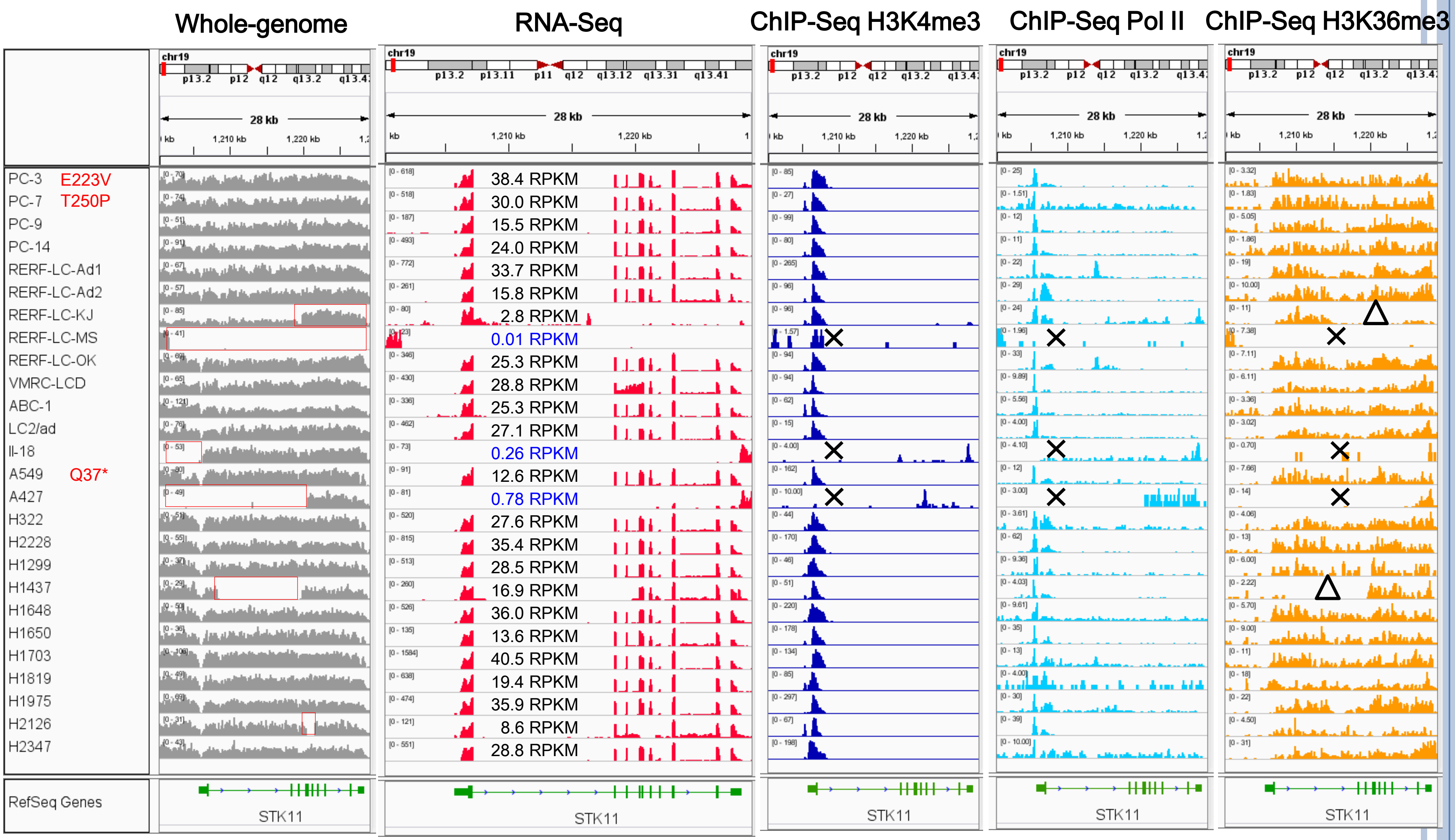
Integrated analyses

PC-7: Non-adherent cell



Cell line	H3K4me3	Pol II	H3K36me3
PC-7	×	×	×
VMRC-LCD	○	×	×
PC-3	○	△	×

STK11遺伝子についての遺伝子発現異常パターン

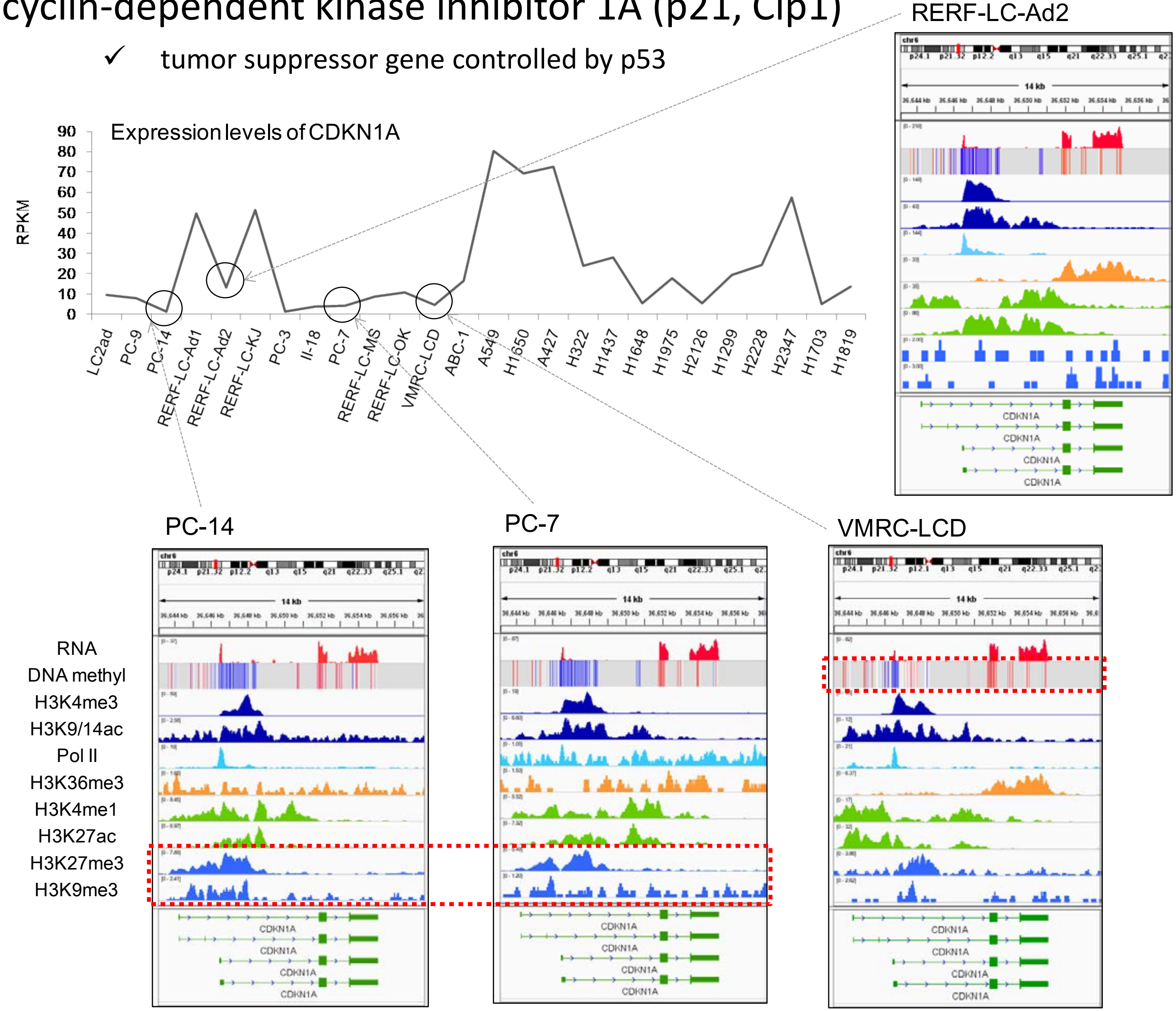


ゲノム異常　　遺伝子発現異常　　エピゲノム異常

Integrated analyses

CDKN1A cyclin-dependent kinase inhibitor 1A (p21, Cip1)

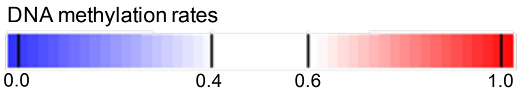
- ✓ tumor suppressor gene controlled by p53



ゲノム変異はないが、DNAメチル化やヒストンのrepressive markで発現が制御されている

Integrated analyses

CDKN2A cyclin-dependent kinase inhibitor 2A

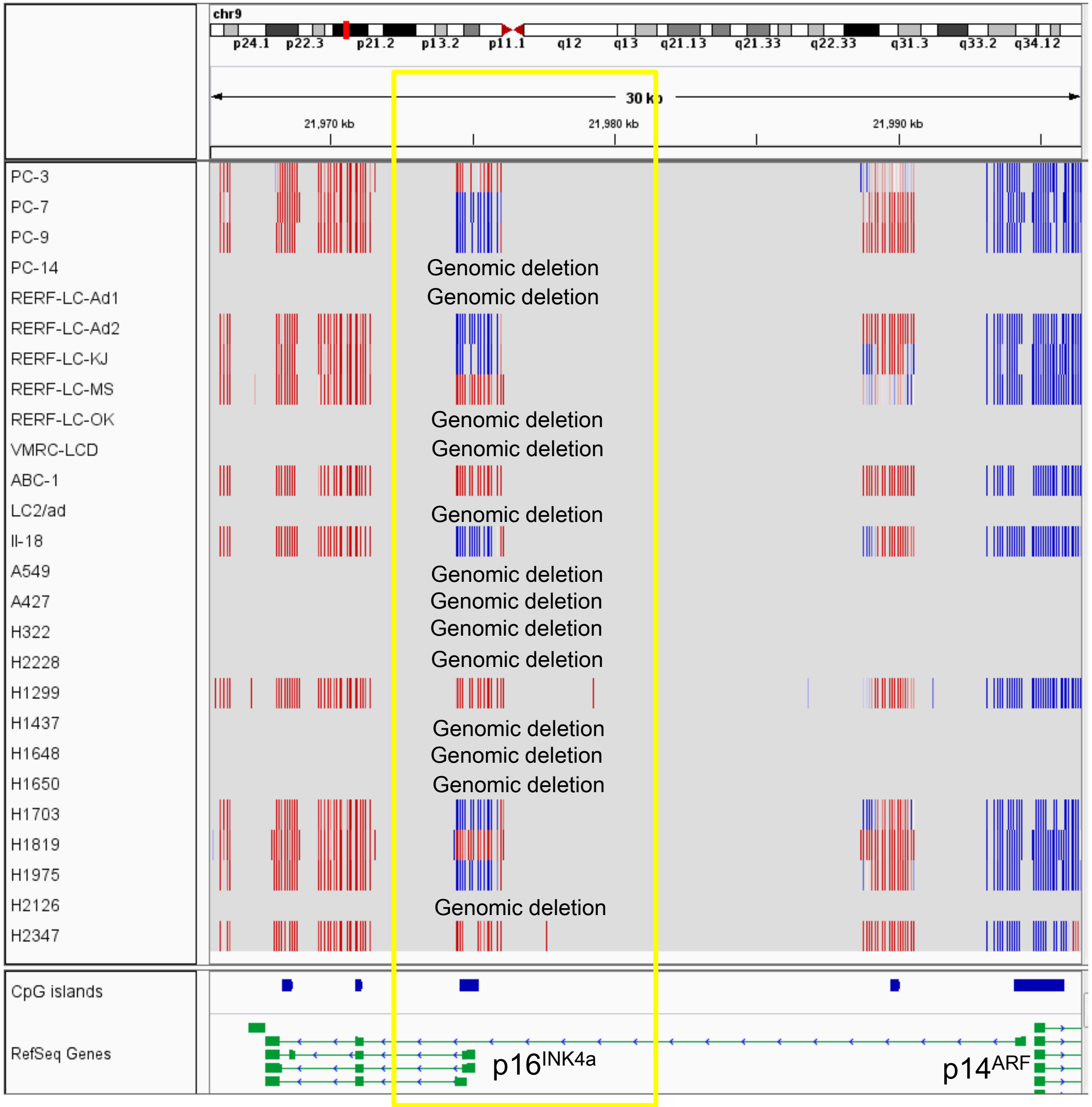


G67V (p16^{INK4a})

62-base deletion (p16^{INK4a}/p14^{ARF})

D84V (p16^{INK4a})

E69* (p16^{INK4a})



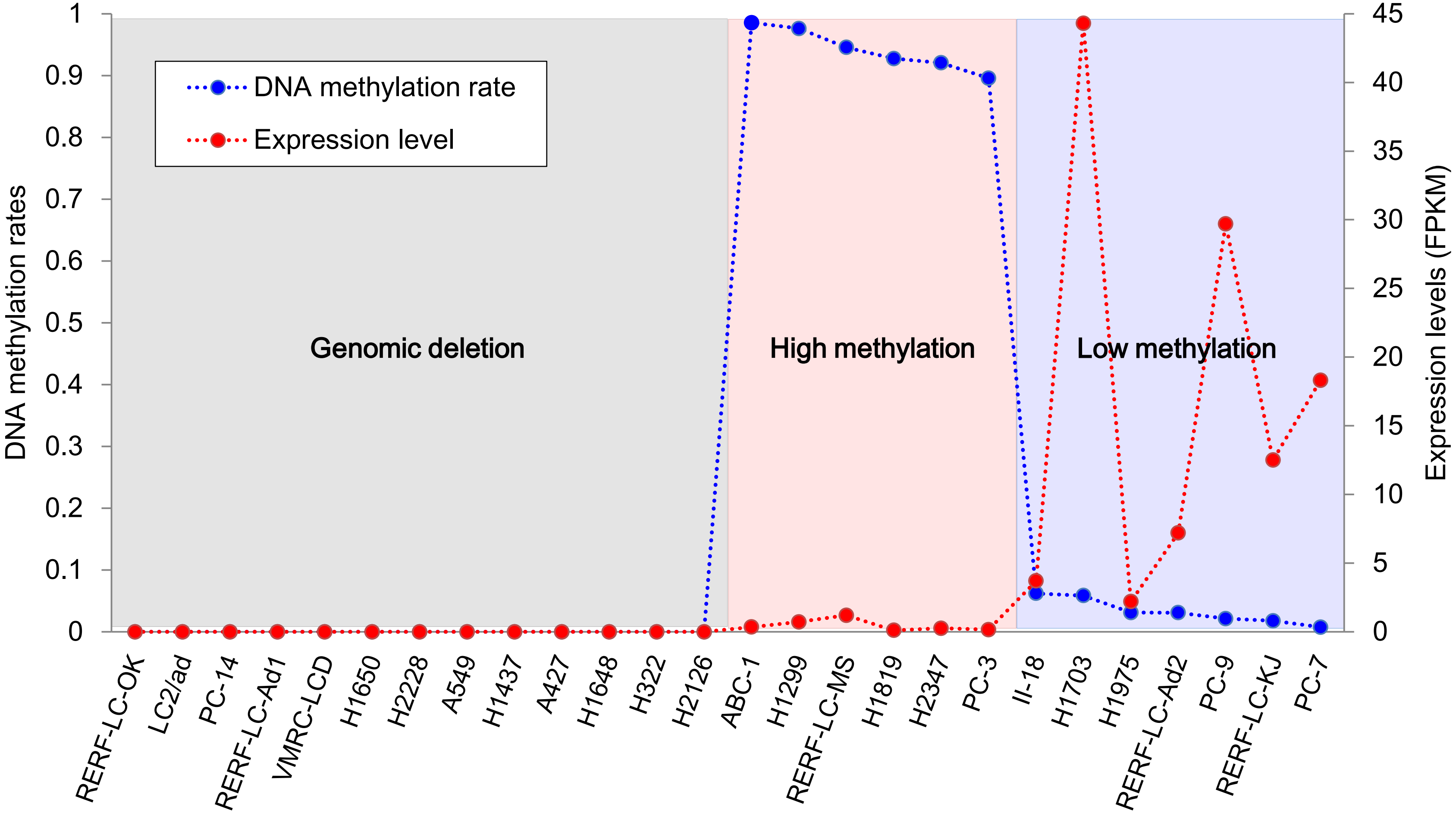
p16^{INK4a}の異常

Genomic deletion:
13 cell lines
SNVs/indels:
4 cell lines
DNA methylation:
6 cell lines

ゲノム変異とDNAメチル化が発現量に大きく寄与している

Negative correlation between DNA methylation rates and expression levels

CDKN2A ($p16^{INK4a}$)

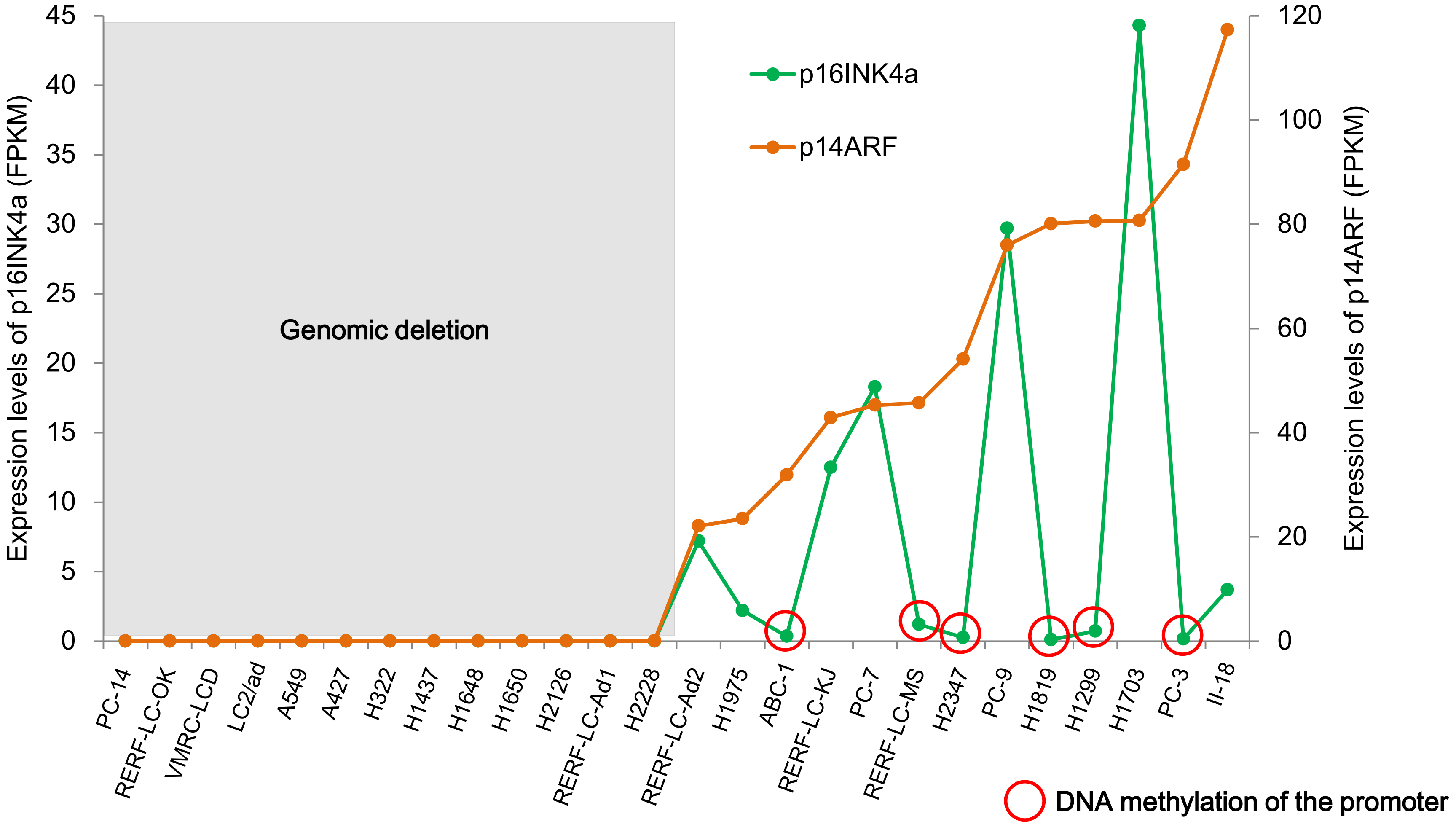


Promoter of $p16^{INK4a}$ was deleted in 13 cell lines and highly methylated in 6 cell lines. Expression levels of $p16^{INK4a}$ were down-regulated by genomic deletions or DNA methylation of the promoter.

*FPKMs of p16 and p14 were calculated using TopHat2-Cufflinks.

Integrated analyses

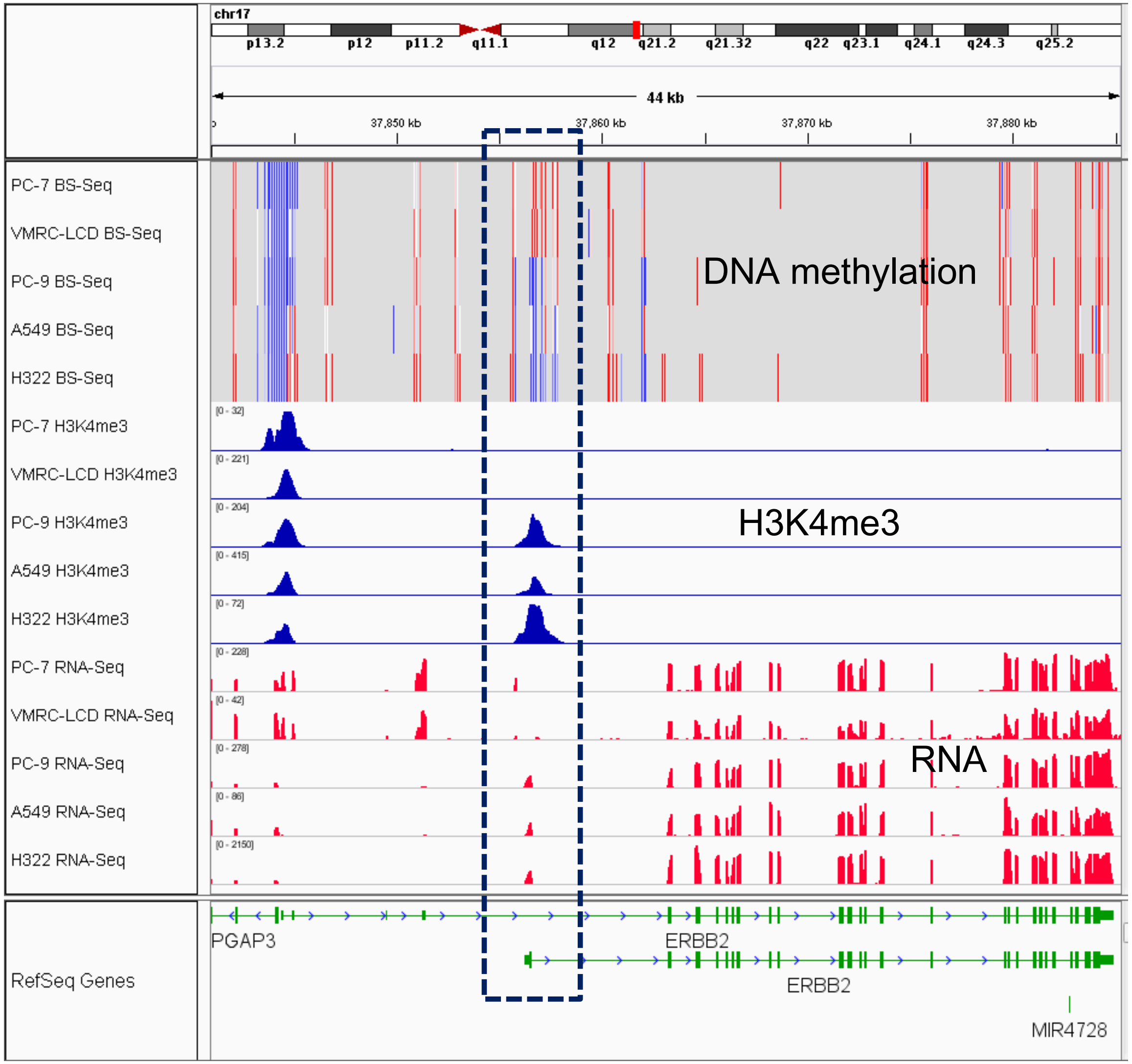
Expression levels of $p14^{ARF}$ and $p16^{INK4a}$



$p16^{INK4a}$ のプロモーターがDNAメチル化をうけていない細胞については、$p16^{INK4a}$ の発現量は$p14^{ARF}$の発現量と相関があるように見える。
ただし、H1975とII-18の$p16^{INK4a}$発現量は、低めである。
それぞれnonsense SNVsと62-base deletionをもっている ← 分解されている？
（↑ちなみにH3K4me3のintensityは高い）

Integrated analyses

ERBB2 v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2

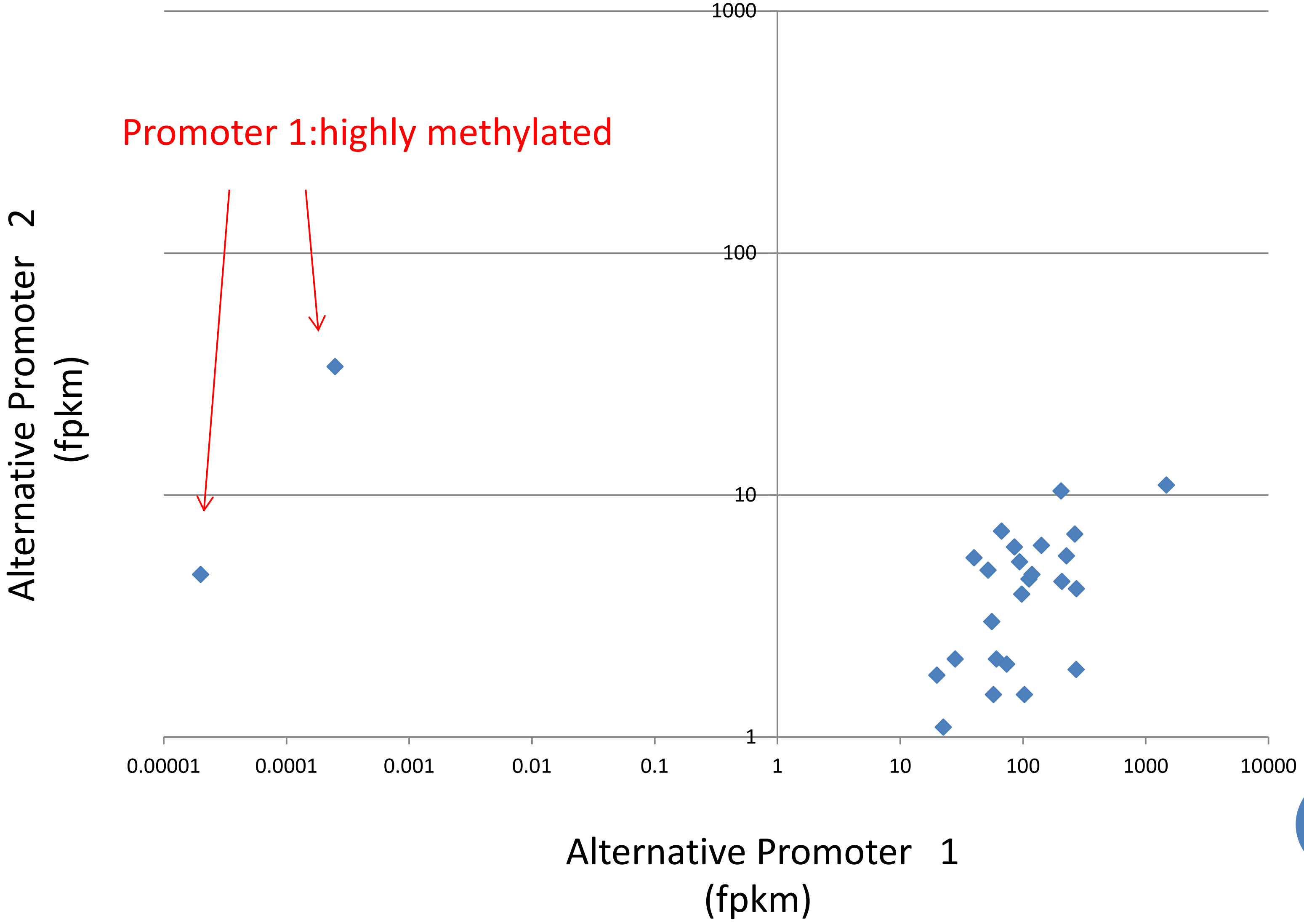


Cell line	FPKM	
	NM_004448	NM_001005862
PC-3	67.2	7.1
PC-7	0.00025	33.9
PC-9	56.0	3.0
PC-14	40.0	5.5
RERF-LC-Ad1	85.3	6.1
RERF-LC-Ad2	205.1	10.4
RERF-LC-KJ	273.1	4.1
RERF-LC-MS	52.2	4.9
RERF-LC-OK	57.7	1.5
VMRC-LCD	2.0e-5	4.7
LC2/ad	102.9	1.5
ABC-1	271.3	1.9
II-18	112.3	4.5
A549	22.5	1.1
A427	60.8	2.1
H322	265.3	6.9
H2228	19.9	1.8
H1299	28.1	2.1
H1437	94.2	5.3
H1648	141.9	6.2
H1650	207.8	4.4
H1703	73.8	2.0
H1819	1476.2	11.0
H1975	98.0	3.9
H2126	227.1	5.6
H2347	118.5	4.7

PC-7とVMRC-LCDでは、NM_04448の転写開始点付近がDNAメチル化を受けている
→NM_04448が発現していない。PC-7はNM_001005862の発現量が高め。

*FPKMs were calculated using TopHat2-Cufflinks.

Gene Expression of Alternative Promoters of the ERBB2 gene



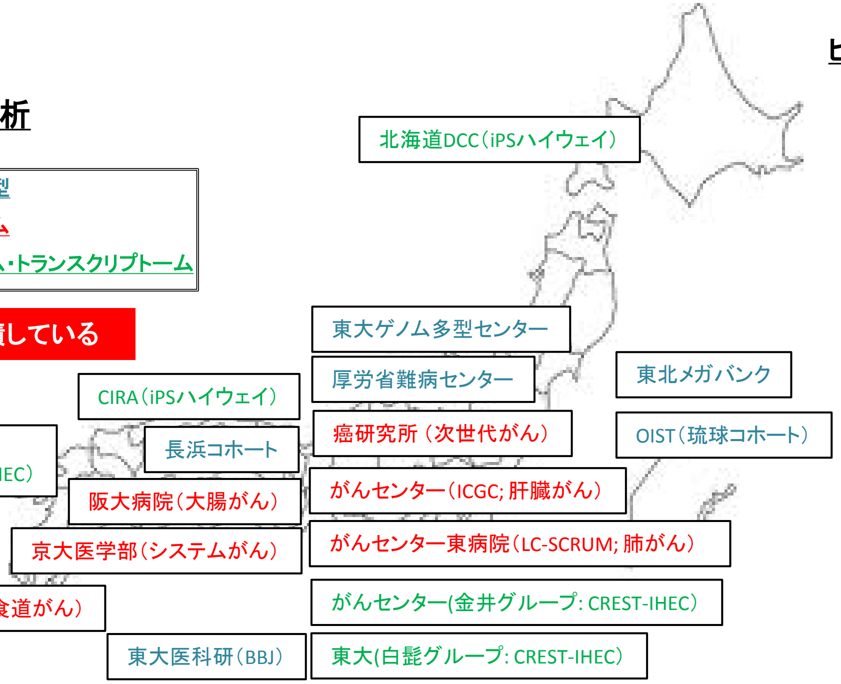
データベースへの統合

DBTSSの拡張: DB-KERO

全国に展開するヒトゲノム解析

ゲノム多型
がんゲノム
エピゲノム・トランスクリプトーム

ゲノムデータは急速に蓄積している



北海道DCC（iPSハイウェイ）

東大ゲノム多型センター

厚労省難病センター

東北メガバンク

CIRA（iPSハイウェイ）

癌研究所（次世代がん）

OIST（琉球コホート）

九大医学部
(佐々木グループ: CREST-IHEC)

長浜コホート

阪大病院（大腸がん）

がんセンター（ICGC; 肝臓がん）

京大医学部（システムがん）

がんセンター東病院（LC-SCRUM; 肺がん）

九大病院（食道がん）

がんセンター(金井グループ: CREST-IHEC)

東大医科研（BBJ）

東大(白髭グループ: CREST-IHEC)

ヒトオミクスデータ推定蓄積量

ゲノム多型(WGS/WES)：>2000人

がんゲノム(WGS/WES/Target Seq）：>1000症例

トランスクリプトーム(RNA Seq）：>1000例

エピゲノム(BS/ChIP Seq）：<100例

+培養細胞＋PDX＋モデル系：>5000例

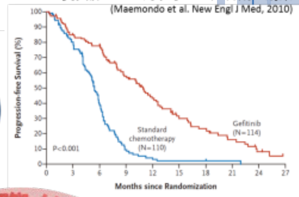
+マウス等モデル生物：???例

+個別研究者の蓄積するオミクス情報：???例

創薬スクリーニングの系に用いられるが、オミクス情報の統合が不十分

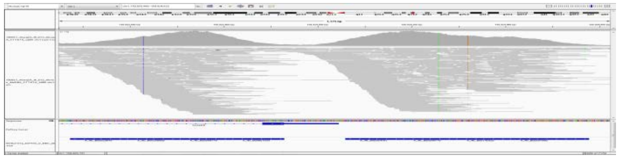
創薬スクリーニング

EGFR変異肺がんに対するEGFR阻害剤の効果
(Maemondo et al. New Engl J Med, 2010)

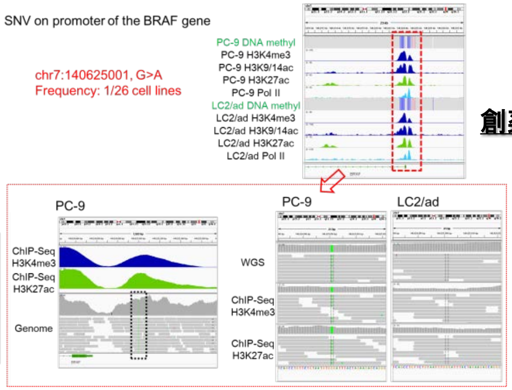


データ統合が目指すヒトゲノム臨床応用研究

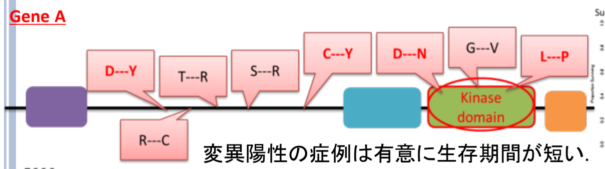
WGS/WES解析

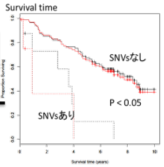


Regulatory SNVsの解析

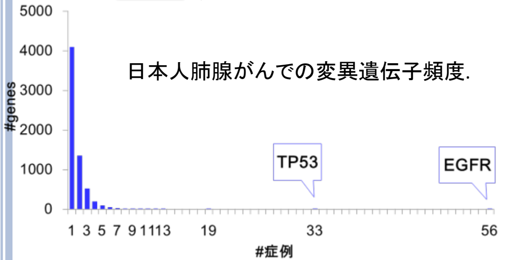


Coding SNVsの解析例





変異陽性の症例は有意に生存期間が短い.



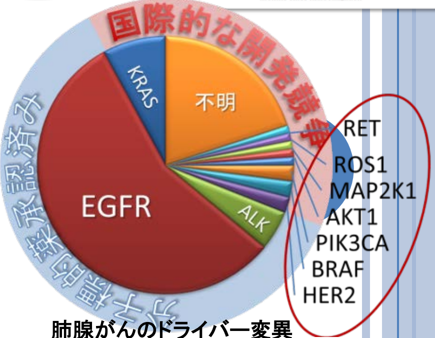
日本人肺腺がんでの変異遺伝子頻度.

・症例間で変異遺伝子が重複することは例外的な遺伝子を除いて、まれ

・Passenger変異<->Driver変異の区分が困難

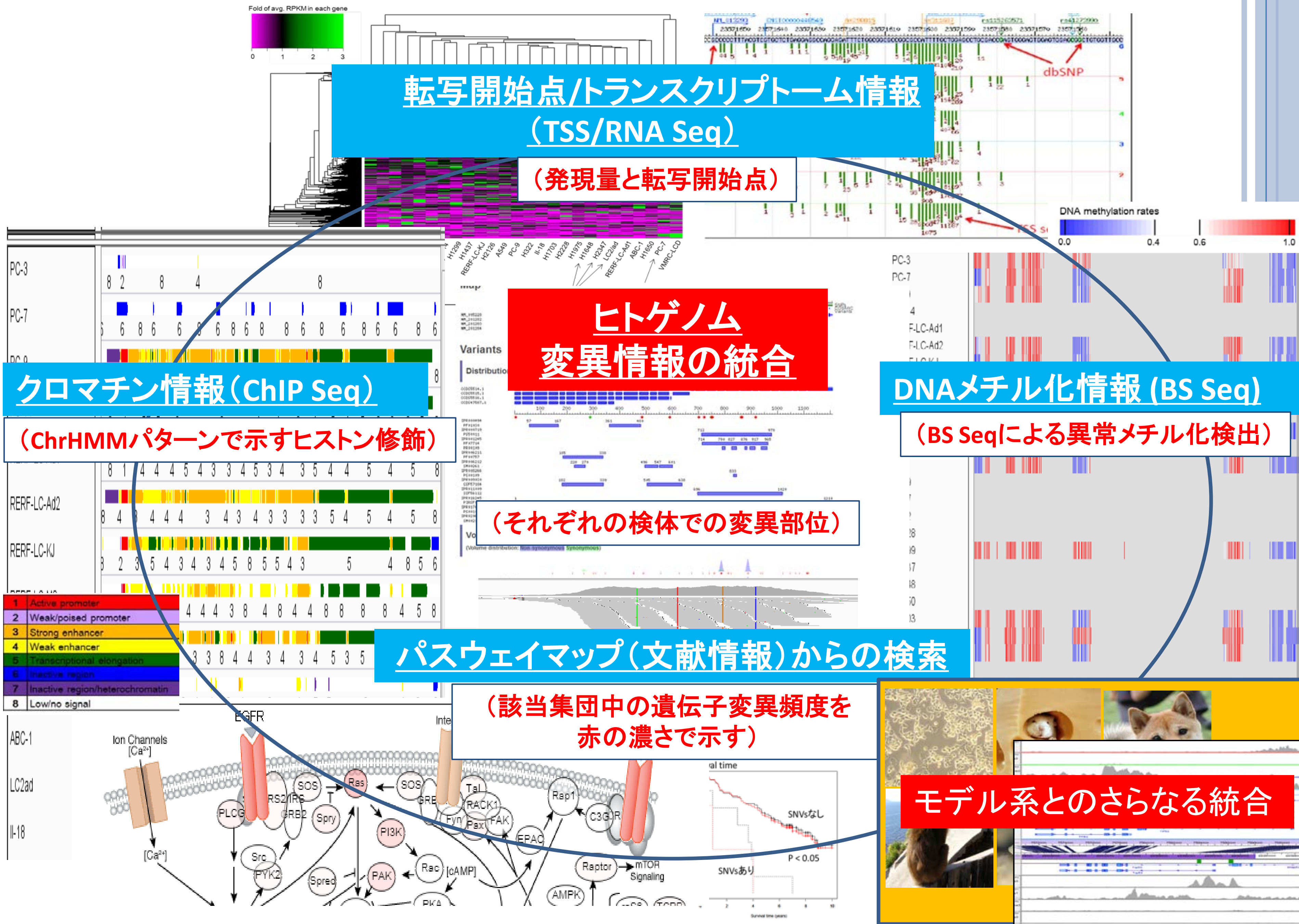
・Regulatory SNPについての情報が圧倒的に不足

創薬ゲノミクス・臨床応用へ直結しない



肺腺がんのドライバー変異

ヒト応用研究を志向したオミクス情報の統合（EGFR遺伝子を例に）



転写開始点/トランスクリプトーム情報（TSS/RNA Seq）

（発現量と転写開始点）

ヒトゲノム変異情報の統合

（それぞれの検体での変異部位）

クロマチン情報（ChIP Seq）

（ChrHMMパターンで示すヒストン修飾）

DNAメチル化情報 (BS Seq)

（BS Seqによる異常メチル化検出）

パスウェイマップ（文献情報）からの検索

（該当集団中の遺伝子変異頻度を赤の濃さで示す）

モデル系とのさらなる統合

資料2－1

SNV on promoter of BRAF

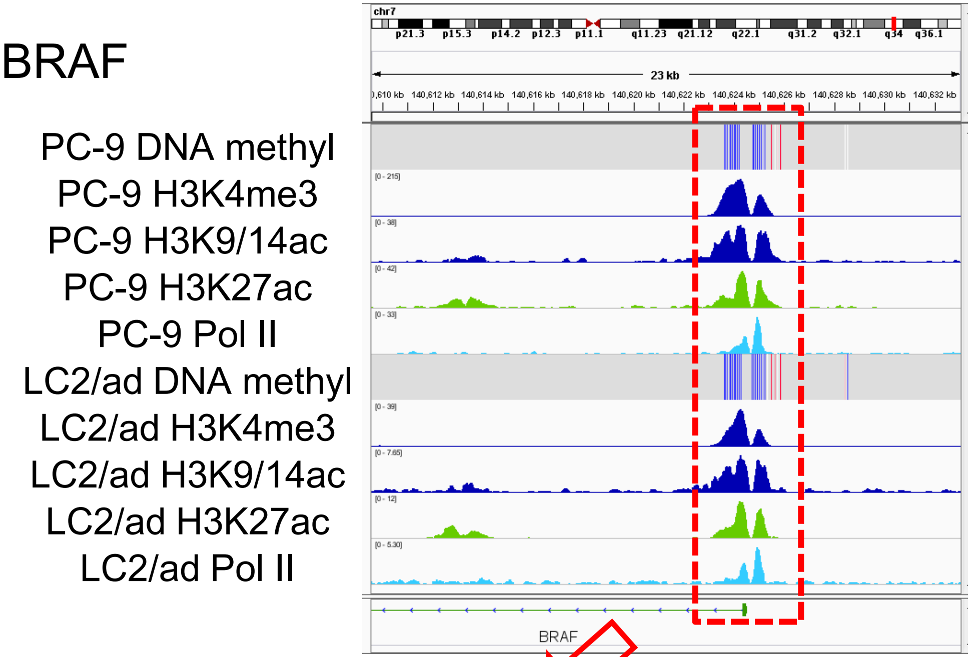
chr7:140625001, G>A
Frequency: 1/26 cell lines

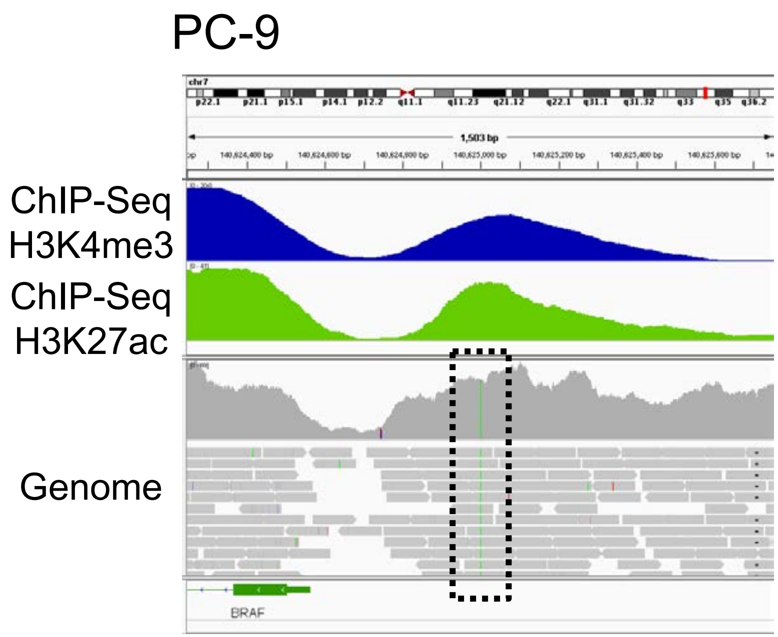
＝疾患ゲノムのその座標で“何が起きているのか”を網羅的に検索

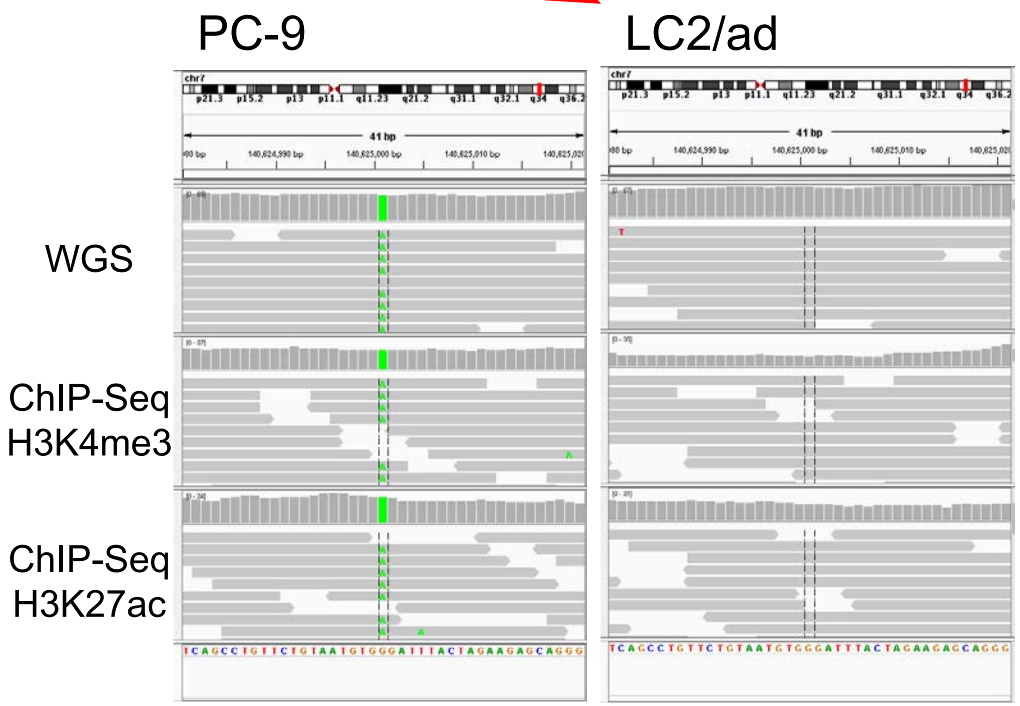
このゲノム変異はエピゲノム、トランスクリプトームに変化を与えない。

⇩

中立変異の可能性が高い？

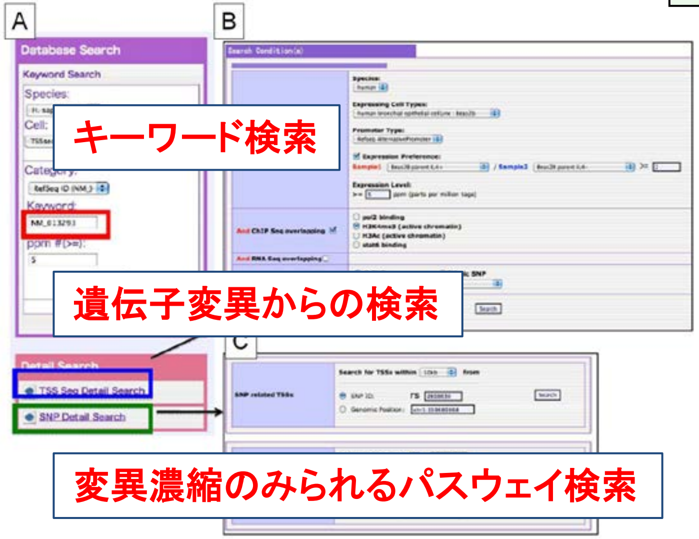






検索（テキスト検索）

（公開DB）



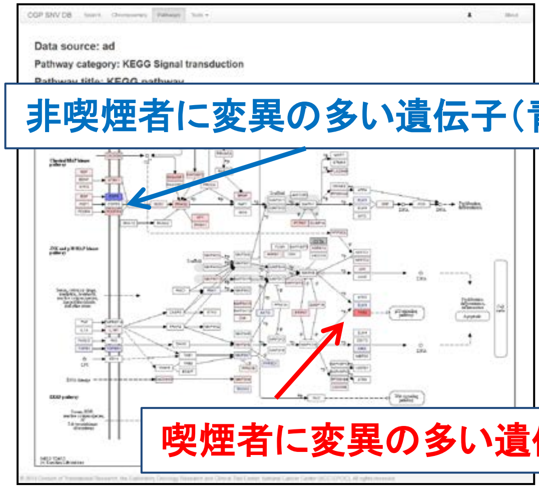
キーワード検索

遺伝子変異からの検索

変異濃縮のみられるパスウェイ検索

検索（クリッカブルマップ）

（非公開DB）

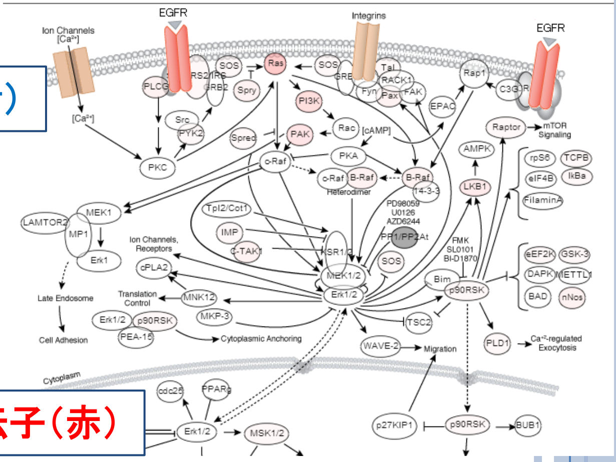


Data source: ad
Pathway category: KEGG Signal transduction
Pathway title: KEGG pathway

非喫煙者に変異の多い遺伝子（青）

喫煙者に変異の多い遺伝子（赤）

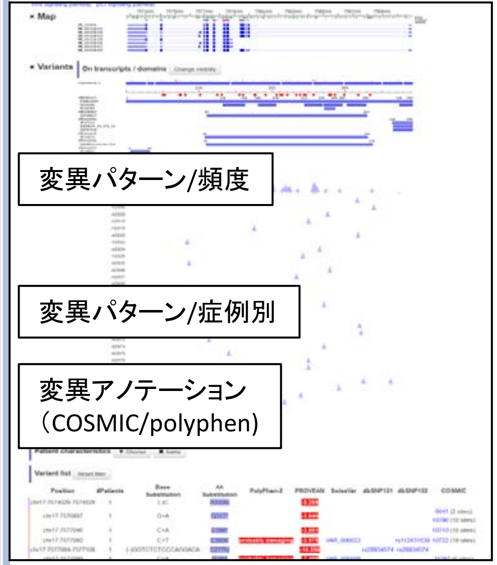
（該当集団中の遺伝子変異頻度を赤の濃さで示す）



KEGGからの自動生成

文献（ウェブ）からのマニュアル描画

結果表示（変異情報）

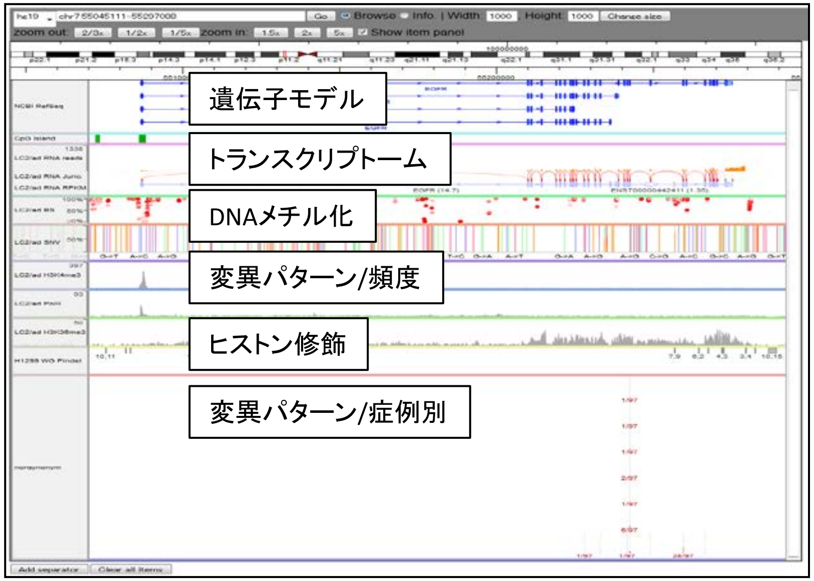


変異パターン/頻度

変異パターン/症例別

変異アノテーション（COSMIC/polyphen）

結果表示（ゲノムブラウザ）



遺伝子モデル

トランスクリプトーム

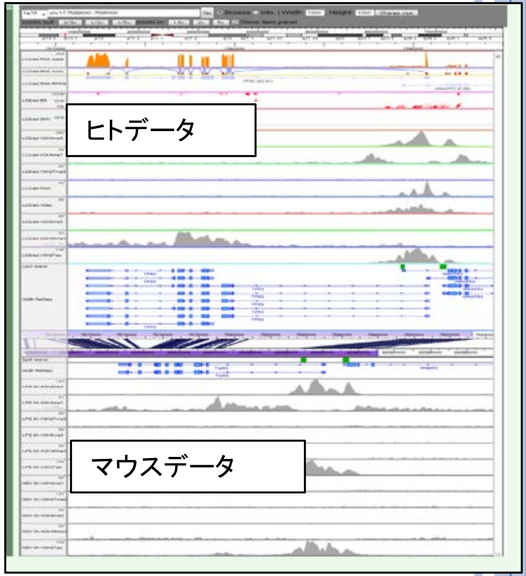
DNAメチル化

変異パターン/頻度

ヒストン修飾

変異パターン/症例別

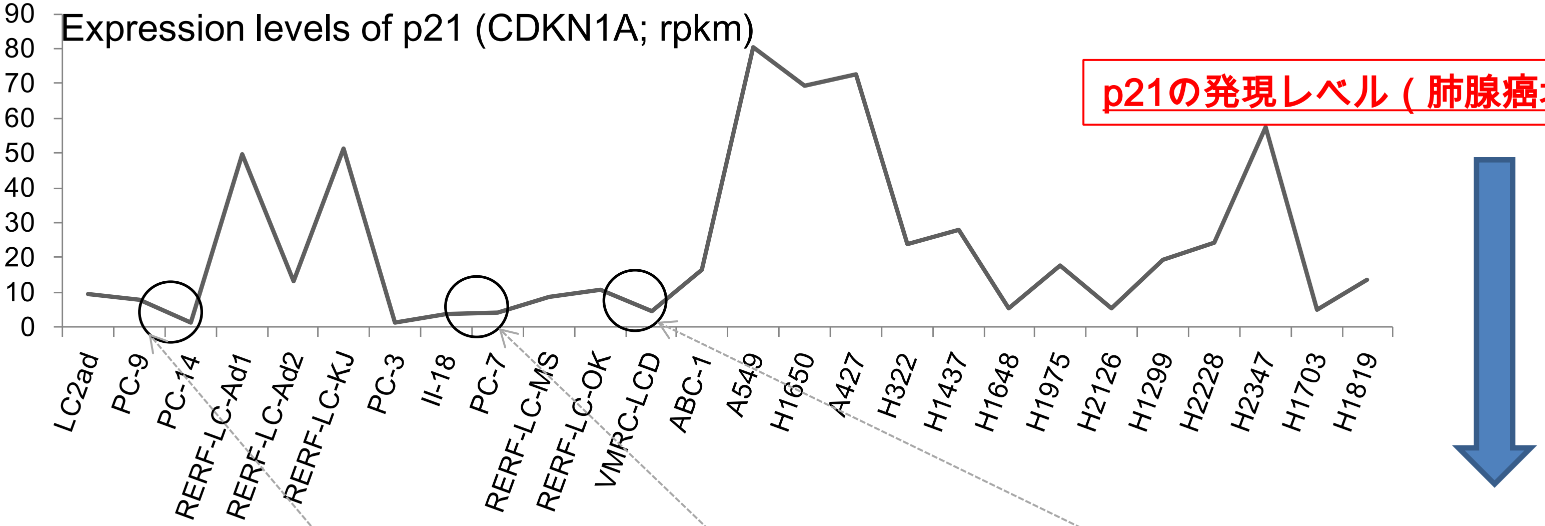
結果表示（比較ゲノム）



ヒトデータ

マウスデータ

p21遺伝子についての遺伝子発現異常パターン



p21の発現レベル（肺腺癌培養細胞26種類）

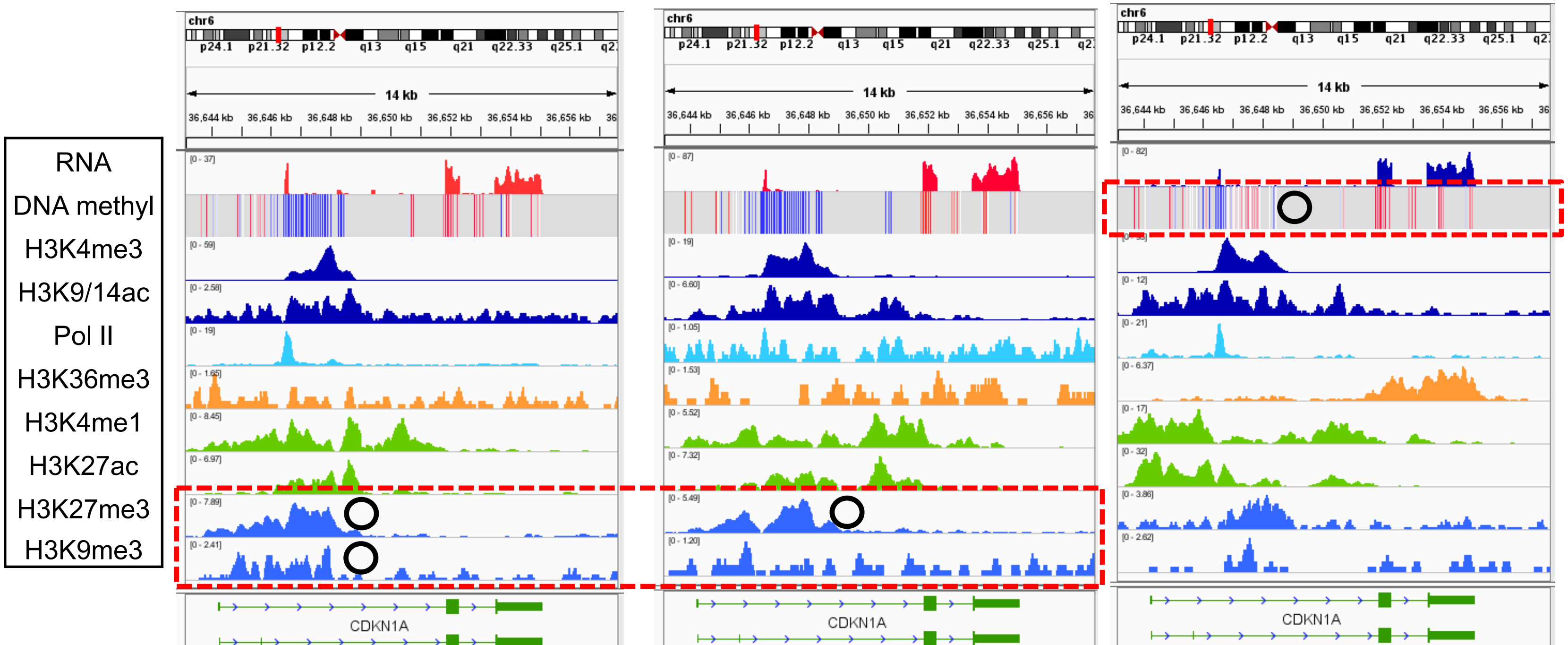
種々のヒストン修飾の影響が大きい細胞

DNAメチル化の影響が大きい細胞

PC-14

PC-7

VMRC-LCD



ヒト疾患ゲノム統合DB (DBMGS)：

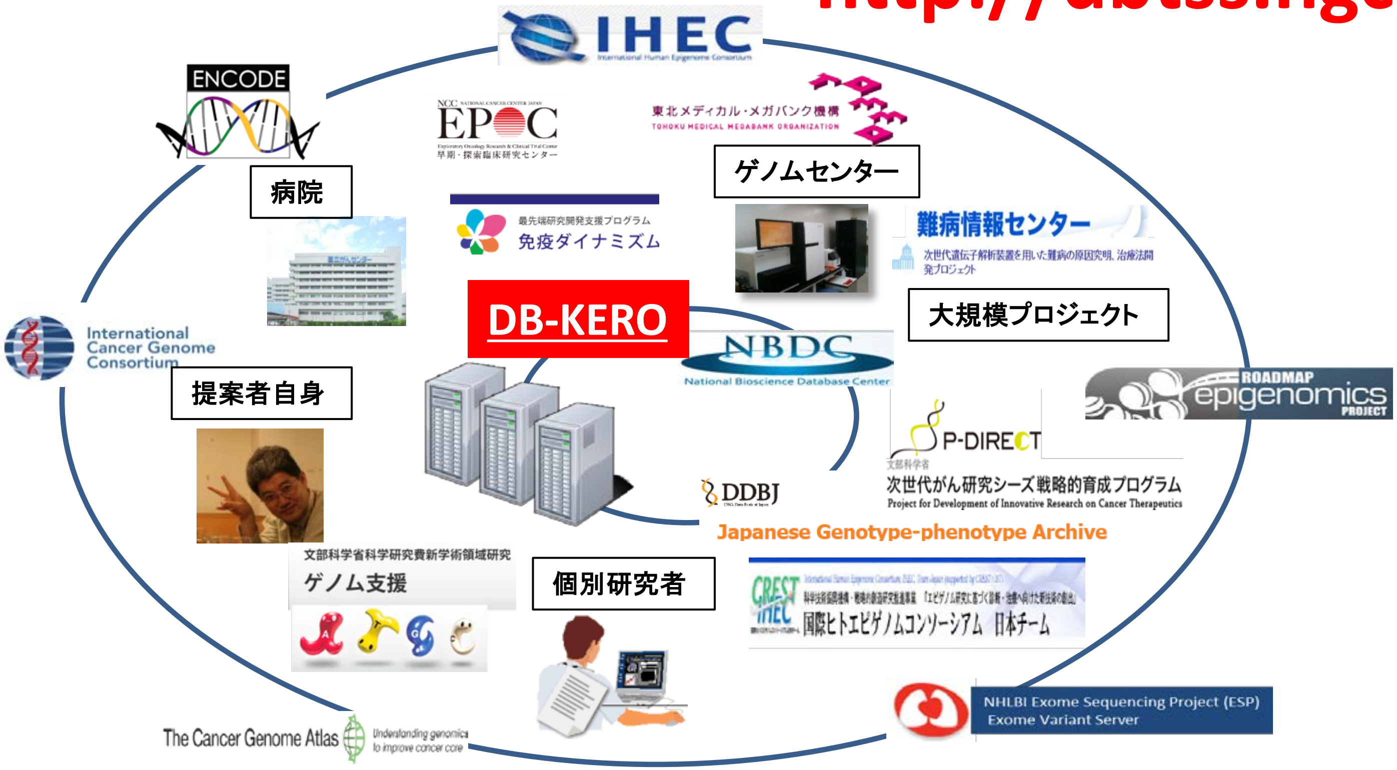
KERO(Kashiwa Encyclopedia of Regulatory Omics)

ヒトゲノム・エピゲノム・トランスクリプトームデータの統合

ヒト疾患ゲノム変異への機能的注釈

パターン検索システムの開発と実装

http://dbtss.hgc.jp/



IHEC

International Human Epigenome Consortium

ENCODE

NCC EPOC

早期・探索臨床研究センター

東北メディカル・メガバンク機構

TOHOKU MEDICAL MEGABANK ORGANIZATION

病院

ゲノムセンター

最先端研究開発支援プログラム

免疫ダイナミズム

難病情報センター

次世代遺伝子解析装置を用いた難病の原因究明、治療法開発プロジェクト

DB-KERO

大規模プロジェクト

International Cancer Genome Consortium

NBDC

National Bioscience Database Center

ROADMAP epigenomics PROJECT

提案者自身

P-DIRECT

文部科学省

次世代がん研究シーズ戦略的育成プログラム

Project for Development of Innovative Research on Cancer Therapeutics

DDBJ

Japanese Genotype-phenotype Archive

文部科学省科学研究費新学術領域研究

ゲノム支援

個別研究者

CREST IHEC

国際ヒトエピゲノムコンソーシアム 日本チーム

The Cancer Genome Atlas

Understanding genomics to improve cancer care

NHLBI Exome Sequencing Project (ESP)

Exome Variant Server

オミクスデータ統合が加速するヒトゲノム臨床応用研究

＝疾患ゲノムのその座標で“何が起きているのか”を網羅的に検索

Summary

情報提供

新機器・新技術=> 止まらない技術革新

新しいプロトコール (Stranded, MatePair, BRIC…)

シングルセル解析：フリューダイムC1システム

統合解析のモデルケース=> 遺伝子に固有のサイレンシング機構

肺腺がん培養細胞をモデルとして

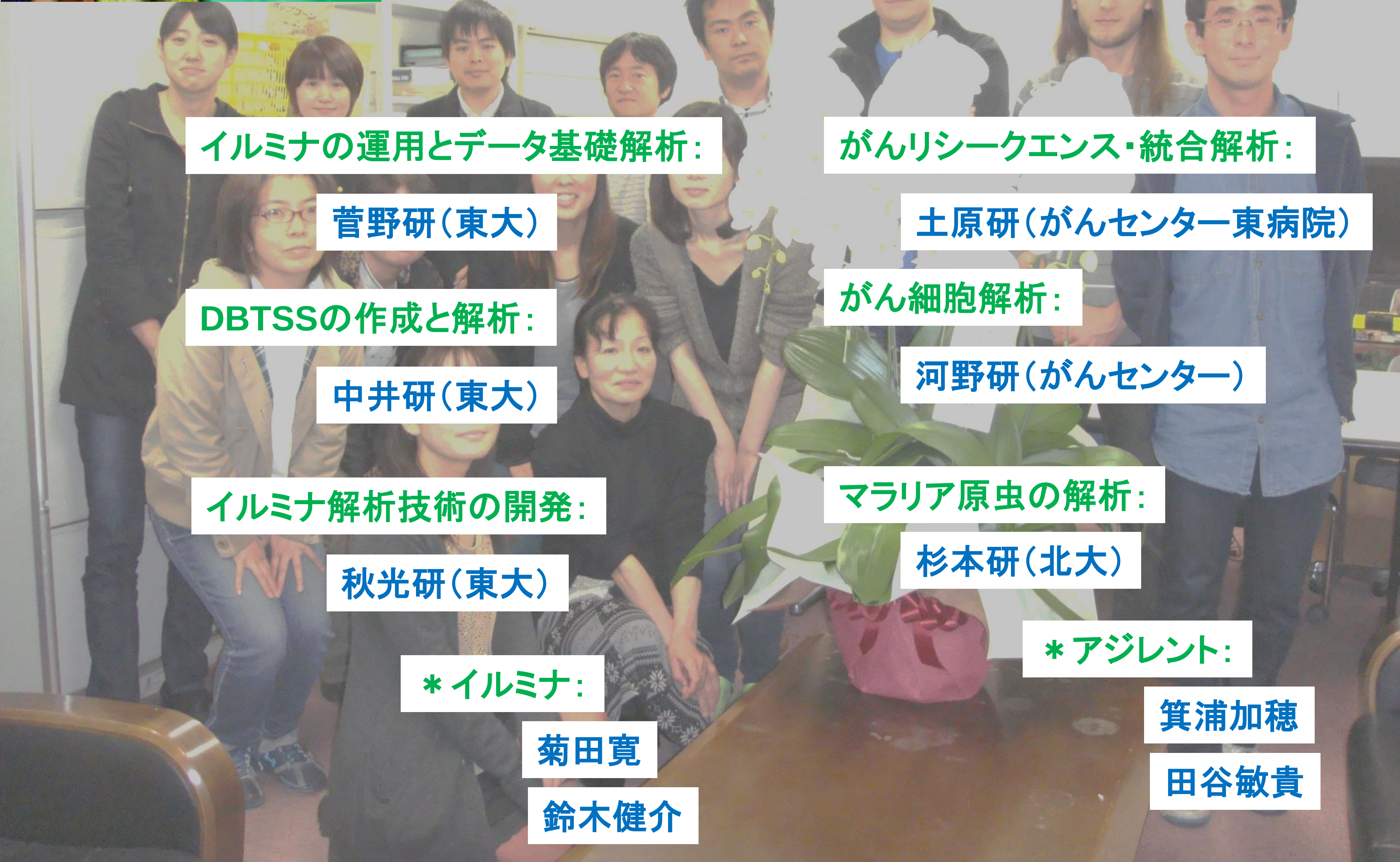
->機能解析/スクリーニングの場としての培養細胞情報の整備

情報の統合=> 情報の統合化による知識発見

多階層オミクスデータベースの構築：

->疾患ヒトゲノム変異の生物学的機能注釈を目指して

ACKNOWLEDGEMENTS



イルミナの運用とデータ基礎解析：

菅野研（東大）

DBTSSの作成と解析：

中井研（東大）

イルミナ解析技術の開発：

秋光研（東大）

＊イルミナ：

菊田寛

鈴木健介

がんリシークエンス・統合解析：

土原研（がんセンター東病院）

がん細胞解析：

河野研（がんセンター）

マラリア原虫の解析：

杉本研（北大）

＊アジレント：

箕浦加穂

田谷敏貴