

PrimePCR™ Pathway Analysis: Pathway Curation and Real-Time PCR Panel Design Strategy

Svetlana Bureeva,¹ Alexey Guryanov,¹ Mary Ferrero,² Viresh Patel,² Megan Clarey,² and Sam Ropp²
¹Thomson Reuters, Carlsbad, CA 92008; ²Bio-Rad Laboratories, Inc., Hercules, CA 94547

Introduction

Pathway analysis has become an important application in life science research and drug development. Investigating how individual targets communicate and respond within complex molecular networks has helped elucidate how biological processes function at a cellular level. Academic institutions use network analysis to advance our fundamental understanding of molecular and cellular biology while pharmaceutical companies now apply pathway analysis to the development of next generation therapeutics. With the advent of genomics and personalized medicine, pathway-based analysis will become an even more valuable tool in aiding our understanding and ability to navigate signaling and disease networks. As the body of literature regarding these networks continues to rapidly increase, challenges such as assay design and therapeutics development can utilize this knowledge to increase the chance of success in research and development.

Bio-Rad Laboratories has partnered with GeneGo, a Thomson Reuters business and leading provider of systems biology tools, to offer a complete pathway solution for real-time PCR-based gene expression analysis. As part of the collaboration, Thomson Reuters provided interactive pathway maps for 260 canonical pathways. Each pathway belongs to one or more general biological categories such as Cancer. The pathway maps illustrate protein interactions and regulation to provide a comprehensive picture of signaling and disease processes (Figure 1). The curated pathways were also used to design panels of real-time PCR assays specifically tailored to the top-ranked gene targets for differential gene expression analysis. Each gene target within a pathway was assigned a score based on frequency of differential expression and popularity in scientific research, both historically and currently. The resulting scores were used to rank and select which assays were included on the corresponding real-time PCR pathway panel.

Thomson Reuters was selected to devise the pathway gene list and rankings based on their superior knowledge base that is widely used for data mining and pathway analysis applications (for example, biomarker discovery, drug target identification). In a recent independent publication by New York University School of Medicine comparing

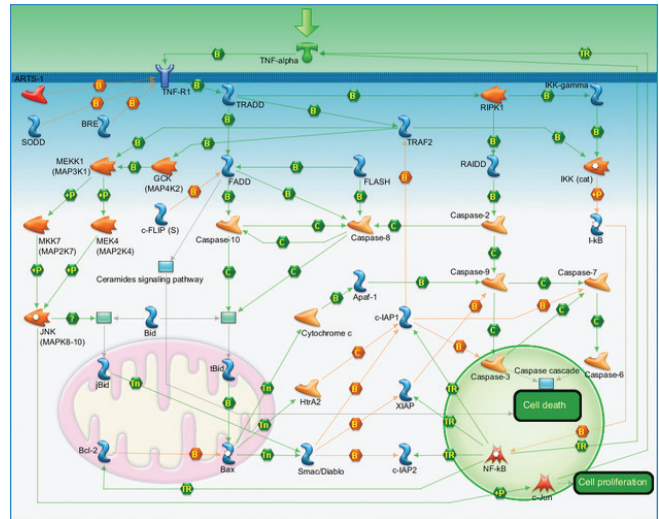


Fig. 1. Apoptosis and survival — TNFR1 signaling pathway map. This pathway map is one example of the 260 canonical pathway maps provided by Thomson Reuters. Pathway maps illustrate protein interactions and regulation to provide a comprehensive picture of signaling and disease processes. On the Bio-Rad website, each target on the map can be selected to view related gene and protein information.

ten public and commercial pathway information providers, Thomson Reuters flagship software, MetaCore, was found to be the only knowledge base enabling statistically significant overlap (84% cases) with experimental data when studying target genes of transcription factors (Shmelkov et al. 2011). The highest quality of protein interaction data in MetaCore is supported by 100% manual curation provided by a team of experts with extensive training. Additional benefits include the directionality of interactions showing the signaling path and the effect of interactions (for example, activation or inhibition), the mechanism of interactions (for example, phosphorylation) and a comprehensive ontology of protein groups and complexes. Taken together, this pathway plate design strategy allows for the most focused and tailored approach to gene expression profiling with the specificity, sensitivity, and reproducibility of real-time PCR.

Target Ranking

Gene assays present on pathway panels have been prioritized based on three main criteria: how often a gene changes expression level in transcriptome studies, how much attention was paid to this gene in the overall scientific research, and how interesting the scientific community found this gene in the last 2 years.

In order to evaluate how often a particular gene changes expression level or is relevant to transcriptome studies (“expression score”), we calculated three parameters covering:

- Number of diseases for which this gene is a biomarker changing expression level (data source: MetaCore)
- Number of samples in which this gene changes expression with a fold change of 1.2 and P value ≤ 0.05 normalized on sample count (data source: Thomson Reuters OMICs data repository)
- Number of incoming interactions with mechanism “influence on expression” and “transcription regulation” (data source: MetaCore)

In order to estimate how much attention was paid to a gene in the overall scientific research (“all years score”), we calculated seven parameters covering the following:

- Number of diseases, for which this gene is a biomarker (data sources: MetaCore and Integrity Biomarkers Module)
- Number of articles that have been published in biomarker research related to this gene (data sources: MetaCore and Integrity Biomarkers Module)
- Number of articles that have been published to study protein interactions of this gene, protein groups, and complexes including this gene (data source: MetaCore). The amount was evaluated separately for human and mouse genes
- Number of pathway maps on which this gene is present (data source: MetaCore)

In order to evaluate how interesting this gene was for the scientific community in the last 2 years (“2010–2011 score”), we calculated nine parameters covering the following:

- Number of articles that have been published in biomarker research related to this gene in 2010–2011: total count and total count normalized on number of articles published in all years (data sources: MetaCore and Integrity Biomarkers Module)
- Number of articles that have been published to study protein interactions of this gene and protein groups and complexes including this gene in 2010–2011: total count and total count normalized on number of articles published in all years (data source: MetaCore)
- Number of customer visits of corresponding gene pages in MetaCore in 2010–2011

At the next step, the calculated parameters were processed to remove outliers and normalize the values from 0 to 1. Normalized parameters were summed up to calculate three scores for each protein: expression, all years, and 2010–2011 scores. Results for each score were normalized again and included in the final score with equal weights. The final scores for each human and mouse protein were used to prioritize pathway map participants and protein neighbors on interaction networks.

To validate our algorithm, we ran several tests to explore how well our scores correlate with number of publications registered by article repositories. For example, our scores show that the TWIST1 protein is one of the proteins that has a higher 2010–2011 score than all years score meaning that the interest of the scientific community in this protein has grown in the last 2 years (Figure 2). The number of articles published on the protein according to the Web of Knowledge, an article database of Thomson Reuters, confirms this observation.

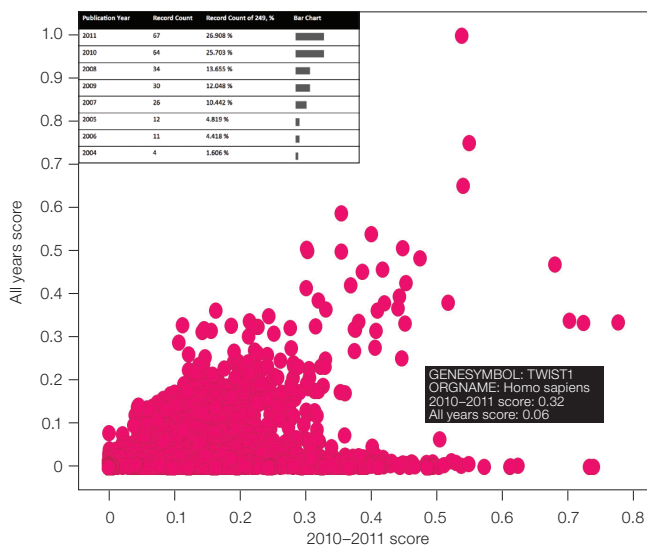


Fig. 2. A distribution of scores of interest for proteins in all years and 2010–2011 years. A higher 2010–2011 score for TWIST1 is confirmed by the increased number of publications related to this protein in the last 2 years.

PrimePCR Pathway and Collection Panel Design

In addition to the 260 pathway maps, Thomson Reuters provided the corresponding list of genes that code for the protein targets present in each pathway and the general biological category. Using the target ranking procedure described above, the top-ranked gene targets from each pathway and category were selected for inclusion on a pre-designed real-time PCR plate, referred to as a “panel.” In addition to the 260 pathway-specific panels, collection panels were designed for each of the 57 general biological categories. For example, a pathway panel has been designed for “Apoptosis and Survival — NGF Signaling Pathway” using the top-ranked genes within that canonical pathway. In addition, an “Apoptosis and Survival” collection panel was designed using the highest-ranked gene targets from the cumulative list of genes across the 17 pathways that report to the general “Apoptosis and Survival” category. For both types of panels, PCR plate designs were selected to accommodate the largest number of gene targets while maximizing the number of samples that will fit on a given 96- or 384-well plate format. This panel design strategy allows for a pathway-focused analysis or a more general survey of gene targets across a given biological process (Figure 3).

For pathway-specific panels where the number of gene targets on the pathway didn't completely fill the maximum potential wells on the PCR plate, would-be empty wells were filled with the highest-ranked targets from the associated general category for the corresponding pathway map. For collection panels where multiple 96-well plates could be designed, the highest-ranking targets were combined on "tier 1" plates, and lesser-ranked targets were used to complete tier 2, 3, and 4 plates, respectively. This tiered system allows researchers the option to study well-characterized targets known to be of interest for a given biological process, or to pursue less well-characterized candidates. For example, if the "Apoptosis and Survival — NGF Signaling Pathway" panel needed additional gene targets, the highest-ranking remaining targets from the "Apoptosis and Survival" collection would be used. For all panels, target assays were placed into the plate alphabetically to allow easy target identification in each pathway. In addition to target assays, a PrimePCR panel includes three reference genes and five experimental controls for each sample.

Conclusions

Pathway analysis provides an efficient and focused approach for studying a large number of related gene targets in a single experiment. The pathway curation and ranking strategy employed by Thomson Reuters and Bio-Rad ensures that the gene assays present on each real-time PCR pathway and collection panel are the most relevant for gene expression profiling based on differential expression studies and the frequency with which gene targets appear in the peer-reviewed literature. All PrimePCR assays were designed following strict guidelines on maximum transcript coverage, have minimal overlap with known single nucleotide polymorphisms, and span large introns where possible. In addition, they have all been fully validated in the lab, passing stringent quality controls. As a result, PrimePCR panels set a new quality standard for commercially available predesigned qPCR pathway panels.

Reference

Shmelkov E et al. (2011). Assessing quality and completeness of human transcriptional regulatory pathways on a genome-wide scale. *Biol Direct* 6, 15.

Integrity and Web of Knowledge are service marks and MetaCore is a trademark of Thomson Reuters.

Practice of the polymerase chain reaction (PCR) may require a license.

A. 24 genes with 3 reference genes and controls x 3 samples

	1	2	3	4	5	6	7	8	9	10	11	12	
A	1	9	17	Ref 1	1	9	17	Ref 1	1	9	17	Ref 1	A
B	2	10	18	Ref 2	2	10	18	Ref 2	2	10	18	Ref 2	B
C	3	11	19	Ref 3	3	11	19	Ref 3	3	11	19	Ref 3	C
D	4	12	20	gDNA	4	12	20	gDNA	4	12	20	gDNA	D
E	5	13	21	PCR	5	13	21	PCR	5	13	21	PCR	E
F	6	14	22	RQ 1	6	14	22	RQ 1	6	14	22	RQ 1	F
G	7	15	23	RQ 2	7	15	23	RQ 2	7	15	23	RQ 2	G
H	8	16	24	RT	8	16	24	RT	8	16	24	RT	H
	1	2	3	4	5	6	7	8	9	10	11	12	

Fig. 3. PrimePCR pathway and collection panel templates. The top-ranked gene targets from each pathway and category were selected for inclusion on PrimePCR panels. PCR plate templates were selected to accommodate the largest number of gene targets while maximizing the number of experimental samples that will fit on a given 96- or 384-well plate format. Experimental samples are color-coded, alternating between gray and tan. Control assays are represented on the plate templates as follows: genomic DNA contamination control assay (gDNA), positive PCR control assay (PCR), RNA quality assay (RQ1 and RQ2), and reverse transcription control assay (RT). **A–B**, examples of 96-well plate templates; **C–E**, examples of 384-well plate templates.

B. 88 genes with 3 reference genes and controls x 1 sample

	1	2	3	4	5	6	7	8	9	10	11	12	
A	1	9	17	25	33	41	49	57	65	73	81	Ref 1	A
B	2	10	18	26	34	42	50	58	66	74	82	Ref 2	B
C	3	11	19	27	35	43	51	59	67	75	83	Ref 3	C
D	4	12	20	28	36	44	52	60	68	76	84	gDNA	D
E	5	13	21	29	37	45	53	61	69	77	85	PCR	E
F	6	14	22	30	38	46	54	62	70	78	86	RQ 1	F
G	7	15	23	31	39	47	55	63	71	79	87	RQ 2	G
H	8	16	24	32	40	48	56	64	72	80	88	RT	H
	1	2	3	4	5	6	7	8	9	10	11	12	

C. 88 genes with 3 reference genes and controls x 4 samples

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
A	1	9	17	25	33	41	49	57	65	73	81	Ref 1	1	9	17	25	33	41	49	57	65	73	81	Ref 1	A
B	2	10	18	26	34	42	50	58	66	74	82	Ref 2	2	10	18	26	34	42	50	58	66	74	82	Ref 2	B
C	3	11	19	27	35	43	51	59	67	75	83	Ref 3	3	11	19	27	35	43	51	59	67	75	83	Ref 3	C
D	4	12	20	28	36	44	52	60	68	76	84	gDNA	4	12	20	28	36	44	52	60	68	76	84	gDNA	D
E	5	13	21	29	37	45	53	61	69	77	85	PCR	5	13	21	29	37	45	53	61	69	77	85	PCR	E
F	6	14	22	30	38	46	54	62	70	78	86	RQ 1	6	14	22	30	38	46	54	62	70	78	86	RQ 1	F
G	7	15	23	31	39	47	55	63	71	79	87	RQ 2	7	15	23	31	39	47	55	63	71	79	87	RQ 2	G
H	8	16	24	32	40	48	56	64	72	80	88	RT	8	16	24	32	40	48	56	64	72	80	88	RT	H
I	1	9	17	25	33	41	49	57	65	73	81	Ref 1	1	9	17	25	33	41	49	57	65	73	81	Ref 1	I
J	2	10	18	26	34	42	50	58	66	74	82	Ref 2	2	10	18	26	34	42	50	58	66	74	82	Ref 2	J
K	3	11	19	27	35	43	51	59	67	75	83	Ref 3	3	11	19	27	35	43	51	59	67	75	83	Ref 3	K
L	4	12	20	28	36	44	52	60	68	76	84	gDNA	4	12	20	28	36	44	52	60	68	76	84	gDNA	L
M	5	13	21	29	37	45	53	61	69	77	85	PCR	5	13	21	29	37	45	53	61	69	77	85	PCR	M
N	6	14	22	30	38	46	54	62	70	78	86	RQ 1	6	14	22	30	38	46	54	62	70	78	86	RQ 1	N
O	7	15	23	31	39	47	55	63	71	79	87	RQ 2	7	15	23	31	39	47	55	63	71	79	87	RQ 2	O
P	8	16	24	32	40	48	56	64	72	80	88	RT	8	16	24	32	40	48	56	64	72	80	88	RT	P
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	

D. 120 genes with 3 reference genes and controls x 3 samples

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
A	1	17	33	49	65	81	97	113	1	17	33	49	65	81	97	113	1	17	33	49	65	81	97	113	A
B	2	18	34	50	66	82	98	114	2	18	34	50	66	82	98	114	2	18	34	50	66	82	98	114	B
C	3	19	35	51	67	83	99	115	3	19	35	51	67	83	99	115	3	19	35	51	67	83	99	115	C
D	4	20	36	52	68	84	100	116	4	20	36	52	68	84	100	116	4	20	36	52	68	84	100	116	D
E	5	21	37	53	69	85	101	117	5	21	37	53	69	85	101	117	5	21	37	53	69	85	101	117	E
F	6	22	38	54	70	86	102	118	6	22	38	54	70	86	102	118	6	22	38	54	70	86	102	118	F
G	7	23	39	55	71	87	103	119	7	23	39	55	71	87	103	119	7	23	39	55	71	87	103	119	G
H	8	24	40	56	72	88	104	120	8	24	40	56	72	88	104	120	8	24	40	56	72	88	104	120	H
I	9	25	41	57	73	89	105	Ref 1	9	25	41	57	73	89	105	Ref 1	9	25	41	57	73	89	105	Ref 1	I
J	10	26	42	58	74	90	106	Ref 2	10	26	42	58	74	90	106	Ref 2	10	26	42	58	74	90	106	Ref 2	J
K	11	27	43	59	75	91	107	Ref 3	11	27	43	59	75	91	107	Ref 3	11	27	43	59	75	91	107	Ref 3	K
L	12	28	44	60	76	92	108	gDNA	12	28	44	60	76	92	108	gDNA	12	28	44	60	76	92	108	gDNA	L
M	13	29	45	61	77	93	109	PCR	13	29	45	61	77	93	109	PCR	13	29	45	61	77	93	109	PCR	M
N	14	30	46	62	78	94	110	RQ 1	14	30	46	62	78	94	110	RQ 1	14	30	46	62	78	94	110	RQ 1	N
O	15	31	47	63	79	95	111	RQ 2	15	31	47	63	79	95	111	RQ 2	15	31	47	63	79	95	111	RQ 2	O
P	16	32	48	64	80	96	112	RT	16	32	48	64	80	96	112	RT	16	32	48	64	80	96	112	RT	P
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	

E. 376 genes with 3 reference genes and controls x 1 sample

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
A	1	17	33	49	65	81	97	113	129	145	161	177	193	209	225	241	257	273	289	305	321	337	353	369	A
B	2	18	34	50	66	82	98	114	130	146	162	178	194	210	226	242	258	274	290	306	322	338	354	370	B
C	3	19	35	51	67	83	99	115	131	147	163	179	195	211	227	243	259	275	291	307	323	339	355	371	C
D	4	20	36	52	68	84	100	116	132	148	164	180	196	212	228	244	260	276	292	308	324	340	356	372	D
E	5	21	37	53	69	85	101	117	133	149	165	181	197	213	229	245	261	277	293	309	325	341	357	373	E
F	6	22	38	54	70	86	102	118	134	150	166	182	198	214	230	246	262	278	294	310	326	342	358	374	F
G	7	23	39	55	71	87	103	119	135	151	167	183	199	215	231	247	263	279	295	311	327	343	359	375	G
H	8	24	40	56	72	88	104	120	136	152	168	184	200	216	232	248	264	280	296	312	328	344	360	376	H
I	9	25	41	57	73	89	105	121	137	153	169	185	201	217	233	249	265	281	297	313	329	345	361	Ref 1	I
J	10	26	42	58	74	90	106	122	138	154	170	186	202	218	234	250	266	282	298	314	330	346	362	Ref 2	J
K	11	27	43	59	75	91	107	123	139	155	171	187	203	219	235	251	267	283	299	315	331	347	363	Ref 3	K
L	12	28	44	60	76	92	108	124	140	156	172	188	204	220	236	252	268	284	300	316	332	348	364	gDNA	L
M	13	29	45	61	77	93	109	125	141	157	173	189	205	221	237	253	269	285	301	317	333	349	365	PCR	M
N	14	30	46	62	78	94	110	126	142	158	174	190	206	222	238	254	270	286	302	318	334	350	366	RQ 1	N
O	15	31	47	63	79	95	111	127	143	159	175	191	207	223	239	255	271	287	303	319	335	351	367	RQ 2	O
P	16	32	48	64	80	96	112	128	144	160	176	192	208	224	240	256	272	288	304	320	336	352	368	RT	P
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	



**Bio-Rad
Laboratories, Inc.**

Life Science
Group

Web site www.bio-rad.com **USA** 800 424 6723 **Australia** 61 2 9914 2800 **Austria** 01 877 89 01 **Belgium** 09 385 55 11 **Brazil** 55 11 5044 5699
Canada 905 364 3435 **China** 86 21 6169 8500 **Czech Republic** 420 241 430 532 **Denmark** 44 52 10 00 **Finland** 09 804 22 00
France 01 47 95 69 65 **Germany** 089 31 884 0 **Greece** 30 210 9532 220 **Hong Kong** 852 2789 3300 **Hungary** 36 1 459 6100 **India** 91 124 4029300
Israel 03 963 6050 **Italy** 39 02 216091 **Japan** 03 6361 7000 **Korea** 82 2 3473 4460 **Mexico** 52 555 488 7670 **The Netherlands** 0318 540666
New Zealand 64 9 415 2280 **Norway** 23 38 41 30 **Poland** 48 22 331 99 99 **Portugal** 351 21 472 7700 **Russia** 7 495 721 14 04
Singapore 65 6415 3188 **South Africa** 27 861 246 723 **Spain** 34 91 590 5200 **Sweden** 08 555 12700 **Switzerland** 061 717 95 55
Taiwan 886 2 2578 7189 **Thailand** 800 88 22 88 **United Kingdom** 020 8328 2000