INGENUITY®
PATHWAY ANALYSIS

IPA Summer Release (2018): Release Notes

Table of Contents

Product Release Information

Product:	IPA
Release Number:	Summer Release 2018
Release Date:	June 30th, 2018
Customer Support:	For more information or support, visit our website or e-mail us at ts-bioinformatics@qiagen.com

About IPA

IPA® is a cloud-based software application that enables you to analyze data derived from expression and SNP microarrays, RNA-sequencing, proteomics and metabolomics experiments, and small-scale experiments (such as PCR) that generate gene lists. It also enables you to search for targeted information on genes, proteins, chemicals, diseases, and drugs, as well as build your own biological models. IPA's data analysis and experimental modeling enables you to understand the significance of your data or target of interest in relation to larger biological or chemical systems, backed by the Ingenuity® Knowledge Base, a uniquely structured repository of biological and chemical Findings, and then generate custom reports to communicate and share your insights with collaborators.

What's New in the IPA Summer Release (June 2018)

Analysis Match updates

Analysis Match enhances interpretation and drives discovery by placing your dataset in the context of thousands of IPA analyses that have been processed from public sources using Array Suite.

Powered by IPA Advanced Analytics, Analysis Match* automatically identifies the analyses of curated datasets that have significant similarities and differences, enabling you to compare results, validate interpretation and better understand causal connections between diseases, genes, and networks of upstream regulators. In this release we are *dramatically* expanding the number of available analyses.

On the day of the release, IPA is gaining over 11,000 expression datasets for a total of over 19,000 available in Analysis Match! Then, in about two weeks, an additional ~28,000 datasets from the NIH Library of Integrated Network-Based Cellular Signatures (LINCS) Program will be added for a total of ~47,000 available datasets.

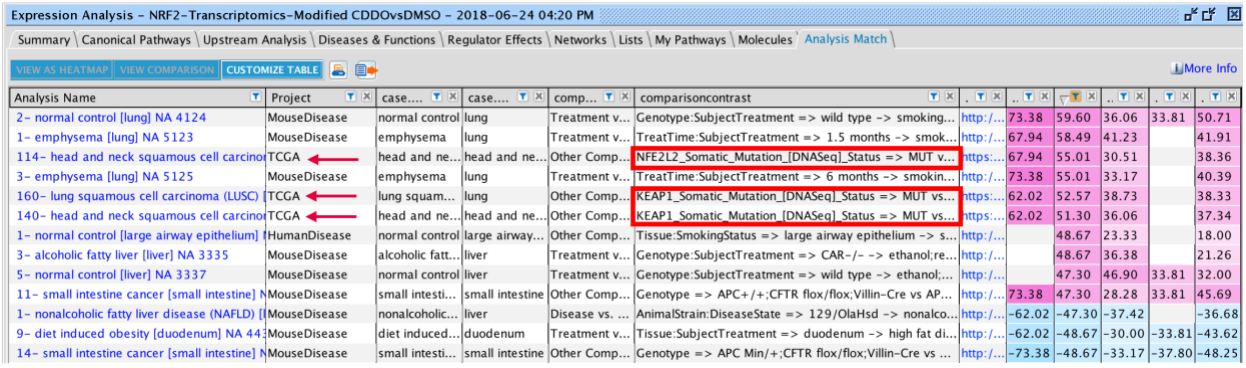
Source	Repository	Q1 2018 (March)	Q2 2018 (June/July)	Increase
DiseaseLand	HumanDisease	3761	6245	2484
	MouseDisease	2797	5746	2949
	Hematology	107	128	21
	RatDisease	123	482	359
	LINCS (mid-July)	0	28,234	28,234
OncoLand	OncoGEO	1169	1500	331
	TCGA	24	4789	4765
	MetastaticCancer	52	81	29
	Pediatrics	126	315	189

***Table 1: Number of datasets available in IPA Analysis Match in this IPA release**. All but the LINCS analyses will be available the week of July 1, 2018. LINCS analyses will be added in the middle of July 2018.*

There are three major types of updates to the Analysis Match repositories in IPA:

1. Normal growth of each repository due to curation of additional studies from GEO, SRA, etc. by OmicSoft.
2. A large expansion of TCGA due to the creation of new datasets stratified by each cancer's molecular attributes or associated patient metadata.
3. Correction of an issue where datasets of closely similar types were inadvertently combined into supersets in IPA. These datasets and their analyses have been removed and replaced with the individual (un-combined) versions. Please see page 9 for additional details.

Analysis Match can help highlight and elucidate potential underlying molecular mechanisms. Figure 1 shows an example Analysis Match result in IPA that highlights a few of the new TCGA datasets. The "query" Core Analysis in this case represents the transcriptome of kidney tissue from mice treated with the NRF2 (NFE2L2) activating chemical CDDO-Me ratio'ed to DMSO-treated controls. This analysis matches several from TCGA where the donor patients had somatic mutations in either NFE2L2 or its binding partner KEAP1. The matches are driven by a similar pattern of upstream regulator activation and inhibition in these conditions. IPA was able to match the hallmarks of NFE2L2 activation even though the matching analyses are from different diseases, tissues and organisms. Here, the query analysis is from the kidneys of chemically-treated mice and the matches are from (non-kidney) squamous cell cancers from human patients with somatic mutations in key NRF2 pathway genes. The activation of NFE2L2 common to these biological conditions is shown in more detail in the heat map in Figure 2 below.



Analysis Name	Project	case....	case....	comp...	comparisoncontrast						
2- normal control [lung] NA 4124	MouseDisease	normal control	lung	Treatment v...	Genotype:SubjectTreatment => wild type -> smoking...	http:/...	73.38	59.60	36.06	33.81	50.71
1- emphysema [lung] NA 5123	MouseDisease	emphysema	lung	Treatment v...	TreatTime:SubjectTreatment => 1.5 months -> smok...	http:/...	67.94	58.49	41.23		41.91
114- head and neck squamous cell carcinor	TCGA	head and ne...	head and ne...	Other Comp...	NFE2L2_Somatic_Mutation_[DNASeq]_Status => MUT v...	https:...	67.94	55.01	30.51		38.36
3- emphysema [lung] NA 5125	MouseDisease	emphysema	lung	Treatment v...	TreatTime:SubjectTreatment => 6 months -> smokin...	http:/...	73.38	55.01	33.17		40.39
160- lung squamous cell carcinoma (LUSC) [	TCGA	lung squam...	lung	Other Comp...	KEAP1_Somatic_Mutation_[DNASeq]_Status => MUT vs...	https:...	62.02	52.57	38.73		38.33
140- head and neck squamous cell carcinor	TCGA	head and ne...	head and ne...	Other Comp...	KEAP1_Somatic_Mutation_[DNASeq]_Status => MUT vs...	https:...	62.02	51.30	36.06		37.34
1- normal control [large airway epithelium] [	HumanDisease	normal control	large airway...	Other Comp...	Tissue:SmokingStatus => large airway epithelium -> s...	http:/...		48.67	23.33		18.00
3- alcoholic fatty liver [liver] NA 3335	MouseDisease	alcoholic fatt...	liver	Treatment v...	Genotype:SubjectTreatment => CAR-/- -> ethanol;re...	http:/...		48.67	36.38		21.26
5- normal control [liver] NA 3337	MouseDisease	normal control	liver	Treatment v...	Genotype:SubjectTreatment => wild type -> ethanol;...	http:/...		47.30	46.90	33.81	32.00
11- small intestine cancer [small intestine] N	MouseDisease	small intesti...	small intestine	Other Comp...	Genotype => APC+/+;CFTR flox/flox;Villin-Cre vs AP...	http:/...	73.38	47.30	28.28	33.81	45.69
1- nonalcoholic fatty liver disease (NAFLD) [	MouseDisease	nonalcoholic...	liver	Disease vs. ...	AnimalStrain:DiseaseState => 129/OlaHsd -> nonalco...	http:/...	-62.02	-47.30	-37.42		-36.68
9- diet induced obesity [duodenum] NA 44	MouseDisease	diet induced...	duodenum	Treatment v...	Tissue:SubjectTreatment => duodenum -> high fat di...	http:/...	-62.02	-48.67	-30.00	-33.81	-43.62
14- small intestine cancer [small intestine] N	MouseDisease	small intesti...	small intestine	Other Comp...	Genotype => APC Min/+;CFTR flox/flox;Villin-Cre vs ...	http:/...	-73.38	-48.67	-33.17	-37.80	-48.25

Figure 1: The Core Analysis of the expression profile of kidney from mice treated with the NFE2L2 activator CDDO-Me. *This analysis matches several TCGA datasets that have either NFE2L2 or KEAP1 somatic mutations. The mouse kidney expression data was obtained from PMID 26422507.*

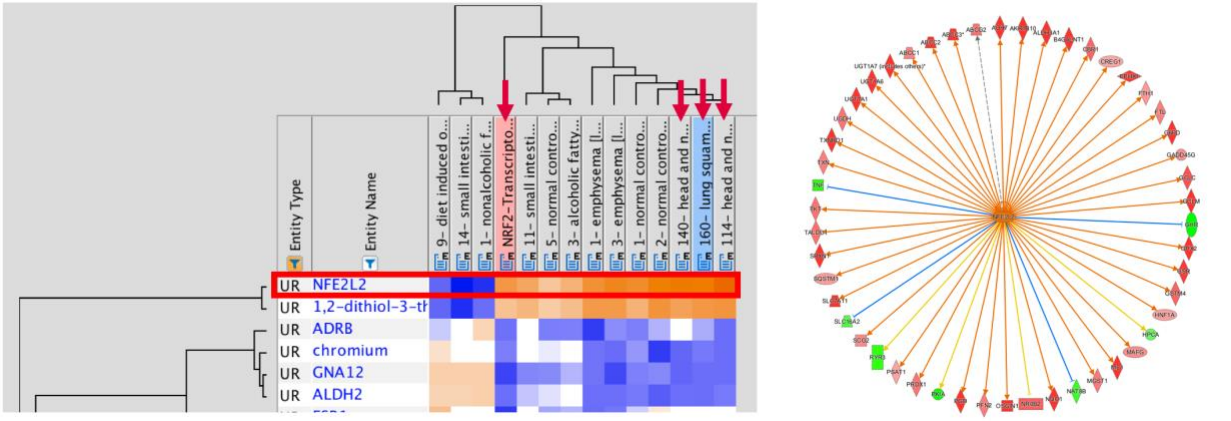


Figure 2: Heatmap from Analysis Match results showing several rows of the Upstream Regulator signature from kidneys of CDDO-Me treated mice compared to matching analyses*. The heatmap shows the signature z-scores versus several matching analyses from the expanded TCGA and MouseDisease repositories. The row corresponding to the Upstream Regulator NFE2L2 is shown in a red box in this screenshot. The pink column header indicates the "query" analysis and the blue column (#160) represents lung (LUSC) cancer with KEAP1 somatic mutations. The network predicting NFE2L2 activation in the KEAP1 mutant analysis is shown at the right where NFE2L2 is displayed as activated (orange) in the center, and its known targets from that*

dataset are arranged around it. Red nodes represent up-regulated genes and green nodes indicate genes that are down-regulated in KEAP1-mutated squamous cell cancers relative to KEAP1 wild type cancer tissue of the same type. Two analyses from TCGA shown with the red arrows (#140 and 114 above) have activating somatic mutations in NFE2L2 itself.

The collection of driver mutations chosen for each cancer type for TCGA were based on QIAGEN QIAseq Targeted DNA Panels as described here. A minimum of two case and two control samples are required to create a dataset.

Table 2 shows the count of datasets created for each cancer type.

Cancer type	**Available single-gene mutation datasets**
acute myeloid leukemia (LAML)	45
adrenocortical carcinoma (ACC)	63
bile duct cancer (CHOL)	48
bladder carcinoma (BLCA)	262
brain glioma (LGG)	184
breast invasive carcinoma (BRCA)	274
cervical squamous cell carcinoma (CESC); endocervical adenocarcinoma	166
chromophobe renal cell carcinoma	32
colon adenocarcinoma (COAD)	282
diffuse large B-cell lymphoma (DLBC)	91
endometrial carcinoma (UCEC)	259
esophageal carcinoma (ESCA)	194
gastric adenocarcinoma (STAD)	261
glioblastoma (GBM)	140
head and neck squamous cell carcinoma (HNSC)	244
hepatocellular carcinoma (LIHC)	192
lung adenocarcinoma (LUAD)	273
lung squamous cell carcinoma (LUSC)	274
malignant mesothelioma (MESO)	25
ovarian serous cystadenocarcinoma (OV)	108
pancreas adenocarcinoma (PAAD)	161
papillary renal cell carcinoma (KIRP)	113
pheochromocytoma and paraganglioma (PCPG)	23
prostate adenocarcinoma (PRAD)	141
rectum adenocarcinoma (READ)	187
renal clear cell carcinoma (KIRC)	178
sarcoma (SARC)	86
skin melanoma (SKCM)	266
testicular germ cell tumor (TGCT)	55
thymoma (THYM)	18
thyroid carcinoma (THCA)	73
uterine carcinosarcoma (UCS)	56
uveal melanoma (UVM)	15

Table 2. Count of datasets created for each cancer. Each dataset corresponds to that cancer type stratified by a somatic mutation in a different gene.

*You may not have Analysis Match active on your license today, but please consult with your local QIAGEN customer solutions manager for additional details on how to get access.

Other Application Improvements

Analysis Match Improvements

- The word “test” has been removed from the names of the analyses, and the value in the “ComparisonIndex” metadata field has been appended. For example, “test4- normal control [hippocampus] lithium” will now be called “4- normal control [hippocampus] lithium 98”. The prefix of “test” equates to the term “observation” in IPA (a dataset of “experimental vs. control”), but inadvertently implied that the datasets were perhaps a test and not the final version.
- A “Weblink” metadata field has been added to each Analysis Match analysis to link to its public record. This makes it easy to look up details about a study, including any publications deriving from it as well as sample data. This is shown in Figure 3 below:

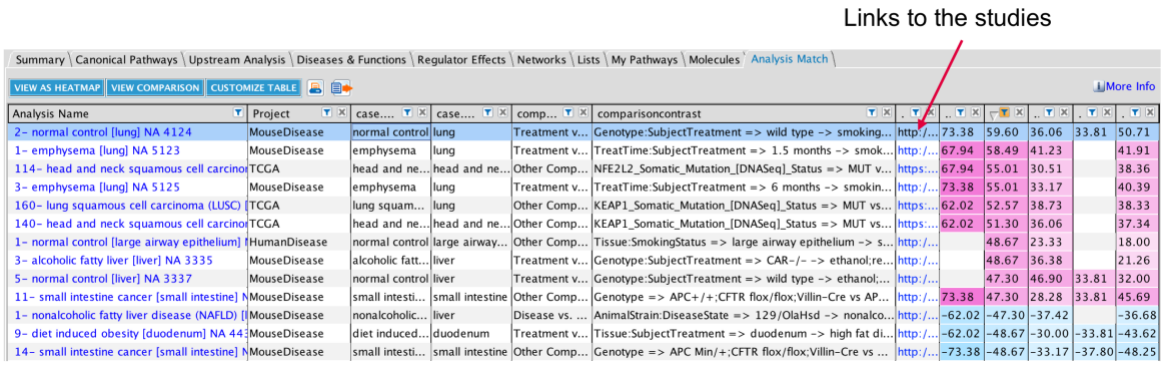


Figure 3: Analysis Match tab showing the new column with hyperlinks to the studies. The link in the row highlighted in blue above is https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65124

Preview of IPA speed enhancements

The speed of opening a Core Analysis or a Comparison Analysis will be improved in a special upcoming release of IPA in the middle of August 2018. But, you can try a Beta version of the improvements today! Right-click on the Core Analysis or Comparison Analysis file in the Project Manager and choose “Open Faster (Beta)” to see the improvement (as shown in Figure 4 below). We are releasing this option as a Beta preview until its final release in mid-August because a few features of the analysis are not yet fully implemented in this mode. However, the results you view are the same as when using the standard way of opening an analysis, but are returned much more quickly.

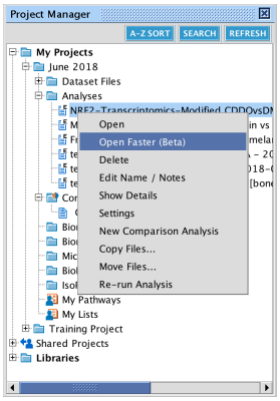


Figure 4. Right-click on a Core Analysis or Comparison Analysis and choose "Open Faster (Beta)". This will launch the analysis more quickly and will become the only method to open the analyses in a mid-August special release of IPA.

Content Updates

~50,000 new findings (bringing the total to greater than 6.4 million findings), including:

~32,000 new Expert findings
~1,300 new ExpertAssist findings
~12,000 new mutation-to-disease findings from ClinVar
~2,500 new cancer mutation disease association findings from COSMIC
~1,100 new disease-to-target findings from ClinicalTrials.gov
~1,200 new drug-to-disease findings from ClinicalTrials.gov
~4,000 new protein-protein interactions from the BioGRID database
~141 new protein-protein interactions from the IntAct database

Noteworthy content additions:

- More than 700 new chemicals (mostly metabolites) have been added in this release.
- Findings for ~75 previously unannotated lncRNAs have been added, and the literature coverage for already annotated lncRNAs has increased.
- Mapping of Ensembl IDs has been improved significantly for less-well annotated genes and transcripts. For example, successful mapping of lncRNAs from Ensembl is improved.

Version or date of third party databases

Source name	Latest version number, download date, or Knowledge Base import date
An Open Access Database of Genome-wide Association Results	Feb-2009
APPRIS	human GRCh38 and rs108v24, mouse GRCm38 and rs106v24 18-Oct-2017
BIND	Dec-2008
BioGRID	Version 3.4.160, 15-May-2018
Chemical Carcinogenesis Research Information System (CCRIS)	Oct-2006
Clinical Trials	20-Mar-2018
ClinVar	06-Apr-2018
Cognia	11-Oct-2007
Conserved Domain Database	Version 3.16, 28-Mar-2017
COSMIC	Version 84, 13-Feb-2018
DIP	14-Oct-2008
DrugBank	27-Apr-2016
Entrez Gene	08-Jan-2018
Gene Ontology (GO)	15-May-2018
GVK Biosciences	31-Mar-2017
Hazardous Substances Data Bank (HSDB)	Apr-2007
Human Metabolome Database (HMDB)	Jun-2009
Human Phenotype Ontology (HPO)	01-May-2018
HumanCyc	16.0, Feb-2012
IntAct	15-May-2018
Interactome studies	2008
MIPS	Apr-2008
miRBase	V21, Jun-2014
miRecords	08-Apr-2011
Mouse Genome Database (MGD)	18-May-2018
Obesity Gene Map Database	Dec-2006
Online Mendelian Inheritance in Man (OMIM)	26-May-2017
Orphanet	Nov-2016
PubChem	16-May-2018
TarBase	18-Mar-2011
TargetScan Human	7.1, Jun-2016

Identifier Mapping

KB version: Rohan
IPA version: Q2 2018

Identifier source versions:

Affymetrix: na36 (1/08/2018)
Agilent: various (1/08/2018)
dbSNP: b151 (chicken, dog)

	b150 (human) b149 (drosophila, rat) b148 (bovine) b146 (mouse) b142 (zebrafish) b138 (Arabidopsis, nematode) (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi)
EntrezGene:	01/08/2018
Ensembl:	Ensembl 91 (12/2017) (http://uswest.ensembl.org/info/website/archives/index.html)
GenBank:	NCBI-GenBank Flat File Release 222 (12/2017) (ftp://ftp.ncbi.nih.gov/genbank/release.notes/)
HGNC:	01/08/2018
HomoloGene:	HomoloGene Release 68 (04/09/2014) (http://www.ncbi.nlm.nih.gov/homologene/statistics/)
Illumina:	10/25/2010
MiRBase:	version 21 (June 2014)
RefSeq:	Release 85 (11/12/2017) (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/release-notes/)
UCSC hg18:	May 2009
UCSC hg19:	June 2013
UCSC hg38:	July 2016
Unigene:	Human(#236), Mouse(#194), Rat(#195), Bovine (#100), Chicken (#46), Dog (#25), Fruitfly (#75), Nematode (#52), Zebrafish(#126), Arabidopsis(#74) (ftp://ftp.ncbi.nih.gov/repository/UniGene/, log files)
Uniprot:	UniProt release 2017_11, 12/20/2017 (ftp://ftp.uniprot.org/pub/databases/uniprot/previous_releases/)

Gene model source versions:

Hg38/mm38:	
Ensembl Human:	Homo_sapiens.GRCh38.91.gtf.gz
Ensembl Mouse:	Mus_musculus.GRCm38.91.gtf.gz
RefSeq human:	GCF_000001405.37_GRCh38.p11
RefSeq mouse:	GCF_000001635.26_GRCm38.p6

Hg19/mm10 from UCSC:	
Ensembl Human:	April 7th 2014
Ensembl Mouse:	April 7th 2014
RefSeq Human:	June 30th 2014
RefSeq Mouse:	June 30th 2014

Issues addressed in this release:

1. An issue was discovered in a subset of datasets created for Analysis Match. Datasets with the same combination of this metadata: testN, case.DiseaseState, case.Tissue, case.Treatment, for example: “test1, diabetic nephropathy, kidney glomeruli, NA” were merged into “super” datasets that represents an averaging of a set of similar datasets. This issue has been fixed by removing the merged datasets and replacing them with the individual (unmerged) datasets. Analysis Match results are not static, so the matches will automatically update when you next open an analysis.

2. The Illumina array Infinium MethylationEPIC v1.0 B4 is now supported. A technical issue had prevented it from appearing in IPA in the Q1 release.

3. Findings about antibodies that target CD3 and CD28 had been incorrectly modeled as inhibitory when they are activating. These findings have been corrected.

4. An issue has been resolved where on some Windows computers, characters in row labels had been truncated when exporting Canonical Pathway heat maps from Comparison Analyses at high resolution.

Minimum Software Requirements:

PC

OS	Windows 10 Windows 8 Windows 7 Windows XP SP2 or later
Browser	IE6 or later Firefox 5 or later Chrome 10 or later Safari 5.0.5 or later
Java (JRE) §	JRE 1.7.0_xx JRE 1.8.0_xx

Mac

OS	macOS High Sierra macOS Sierra OS 10.11 El Capitan OS 10.10 Yosemite
Browser	Firefox 33 or later Safari 7 or later Chrome 10 or later
Java (JRE) §	JRE 1.8.0_xx JRE 1.70_xx

Minimum Hardware Requirements:

- PC – 1.25GHz, 2GB RAM (for lightweight usage of IPA)*
- PC – 2GHz, 4GB RAM (Recommended)
- Mac – 1.25GHz, 2GB RAM (for lightweight usage of IPA)*
- Mac – 2GHz, 4 GB RAM (Recommended)

Minimum Screen Resolution of 1024×768

*Lightweight usage of IPA includes search, build/overlay operations and small dataset upload and analysis creation. For larger analyses and comparison analyses, IPA requires more memory.

For Advanced Analytics (Causal Network Analysis, BioProfiler, IsoProfiler, Phosphorylation Analysis, and Relationship Export) and Analysis Match:
Core™ i5 processor or equivalent running at 2 GHz or higher with 64-bit OS and Java, and at least 3 GB RAM free for Java. Screen resolution of at least 1280×800.

Notes:

- §We recommend that you install the IPA client on your computer with this installer: https://analysis.ingenuity.com/pa/installer/select. The installed IPA client still requires you to have internet access to launch but does *not require* you to install Java (a JRE) or to launch IPA from a web browser.

- Alternatively, you can launch IPA using Web Start, which requires an up-to-date version of Java installed on your computer and is launched from a browser at this URL: https://analysis.ingenuity.com. Help on installing and/or launching IPA can be found at the following links:
  - Mac: http://qiagen.force.com/KnowledgeBase/articles/Basic_Technical_Q_A/Running-IPA-on-Mac
  - Windows: http://qiagen.force.com/KnowledgeBase/articles/Basic_Technical_Q_A/Running-IPA-on-Windows

- Due to Microsoft's discontinuation of support for the Windows XP operating system in April 2014, the Ingenuity Apps Team at QIAGEN has transitioned to legacy support for Ingenuity Pathway Analysis (IPA) on XP. IPA will continue to function on the XP operating system and we anticipate no issues. However, we recommend that you upgrade to a newer operating system as soon as possible. Despite this reduced support for XP, the IPA team will continue to investigate any reported defects and attempt to offer workarounds for any identified issues if they arise.