



# Risk-Based Decision Making in Discovery using Quantigene Plex Assay

Eva Adriaensen (Open Analytics), Bie Verbist (Janssen R&D),  
Rytis Bagdziunas (Open Analytics), Olivier Thas (Universiteit Gent)  
October 4, 2018 – NCS Paris

# Outline

## 1. QuantiGene Plex assay

## 2. Compound prioritization

- Dose response modelling
- Principal component analysis
- Ranking

## 3. Risk-based compound prioritization

- Sampling
- Risk-based ranking

# QuantiGene Plex assay

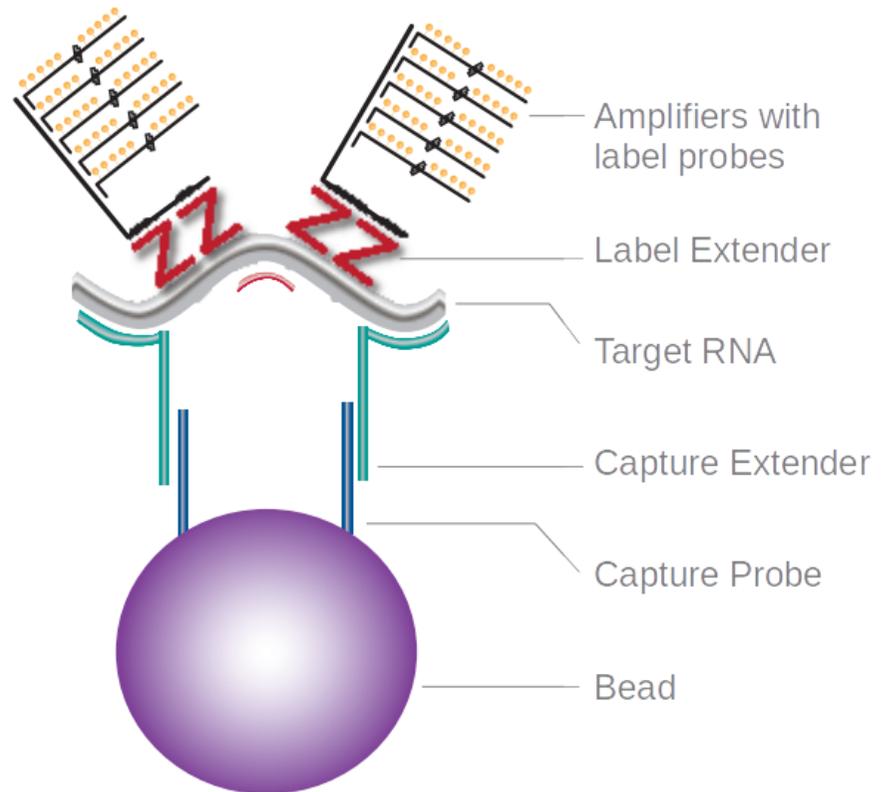
Multiplexed gene expression  
detection and quantitation

## Branched DNA (bDNA) signal amplification

- Direct RNA quantitation
- No target amplification

## xMAP/Luminex magnetic bead capturing

- Measure up to 80 genes in a single well
- Disease or tox genes



# QuantiGene Plex assay

## Data normalization

1. Raw read out (MFI)



2. Background corrected



3. Normalized expression  
(housekeeping genes)



4. Fold change vs untreated

Location	Sample	GENE1	GENE2	GENE3	GENE4	GENE5	HKG1	GENE6	GENE7	HKG2	GENE8	GENE9	GENE10	GENE11	GENE12	GENE13	HKG3	GENE14	GENE15	GENE16	GENE17
01(L.A1)	Background0	8	3.5	9	7	6	5	7	8	8	11.5	11	4	11	7	6	12.5	11	9	12	6
02(L.B1)	Background0	10	5	11	11	7	8	6	6	5	13	11	8.5	10	9	6	13	9	4.5	19.5	8.5
03(L.C1)	Background0	12	7	11	15	11	6.5	8	4	6	19	12	6	12	7	7	10.5	13	11	17	3
04(L.D1)	Background0	11	7	10.5	12	13	4	9	12	4	16	11	5	17	10	8	14	11	10	13	8.5
10(L.B2)	Unknown1	703	6761	4694	644	828	68919	7762	8350	66096	7948	961	1236	1708.5	5336	20.5	59578	30	384	11771	2798
11(L.C2)	Unknown1	552	5367.5	3872	567	592.5	67361	6478.5	7220	64432	7672	785	925	1370	4845.5	13.5	58454	28	315	9540	2594.5
12(L.D2)	Unknown1	775	5943	4578	767	681	66922	7398.5	7884	66096	7910	968	1293	1628	5290	19	60576.5	27	349	11059	2985
13(L.E2)	Unknown11	7514	2154.5	2803.5	1105	211.5	68460	4111.5	5505	70663	2342	593.5	1171	1428	1501	15	63158	32	146	9341.5	937
14(L.F2)	Unknown11	9500	2911	3840	1449.5	311.5	71971	5368	7851.5	68735	2846	829	1651.5	1742.5	1998	29	64352	30	208	10947	1305.5
15(L.G2)	Unknown11	5902	1808	2344	1032	172	67335	3638.5	5052	68597.5	2109	520	981	1153	1358	21	61299	17	110	7578	832
18(L.B3)	Unknown2	747	4528	3059.5	352	417	66234	4316.5	4957	64076	4657	600	992.5	1158	3691	19	52074	19	147	8023.5	1894
19(L.C3)	Unknown2	1243	6881	4387.5	506	843	70089	6247	6714	67863	6167	860	1266	1661.5	5212	27	58683.5	33	235.5	10338	2816.5
20(L.D3)	Unknown2	1203	6899.5	3952.5	419	706	70319	5855	7267	67817	5721.5	860	1328	1514.5	4753	20.5	57765	36	198	10277.5	2807
21(L.E3)	Unknown12	5821	803	1402	486	133	63755	2152	3601	60221	981	338.5	806.5	851	858.5	15	48126	24	84	4417.5	543
22(L.F3)	Unknown12	8163	1047.5	1646	651	135.5	69171	2573	4495.5	62998	1185	452	955	957	1002	13	53852.5	16	84	5586.5	640
23(L.G3)	Unknown12	5833	745.5	1270	467	92	62493	1943	3276	56870	1025	355	631	822.5	753	18	47311.5	31.5	84	4152	467
26(L.B4)	Unknown3	1689	3809	2524.5	232	541	68988	3509	5151	65086	3521	643	871.5	1074	3170	10.5	46175	21	130	5607	1737.5
27(L.C4)	Unknown3	1379	3560	2162.5	205	443.5	64248.5	3267	4481	58523	3135	571	801.5	1066	2819	17	43284	24	125	4710	1625
28(L.D4)	Unknown3	1779	4293	2741	266	557	66027	3886	5581	64076	4105	676	1039.5	1317	3613.5	9.5	47208	19	127.5	5668	1882
29(L.E4)	Unknown13	5683	558.5	972	353	81.5	59418	1572	3051.5	51592	613.5	293	688	636.5	634.5	15.5	39382	12	76	2865.5	363
30(L.F4)	Unknown13	5565	448	963.5	335.5	92	57467	1472	2981	51993	632.5	293	598	636	591	12	39451	23	59	2849	388
31(L.G4)	Unknown13	6195	611.5	1101	405	90	63294.5	1790	3318.5	56136	640	308	667	682	694.5	18	42894	14	64.5	3282	437
34(L.B5)	Unknown4	1225.5	2482	1503.5	130	341.5	56732.5	2338	3753	51087	2246.5	423	691.5	726	2100	12.5	33874	14	93.5	3270	1097
35(L.C5)	Unknown4	1300	2609	1489	119	341	65603	2357	3615	50020	2129	426	647	742.5	2103	12.5	33300.5	19	105.5	3337	1156.5
36(L.D5)	Unknown4	1134.5	2406	1340.5	135	331	56411	2291	3370.5	49457	2220	377	590	661	2017	14	33530	13.5	74.5	3141	1097.5
37(L.E5)	Unknown14	2972	254.5	536	197.5	49.5	46198	928	1746	35274	347.5	185	346.5	317	352	12.5	26163	11	37	1632.5	249
38(L.F5)	Unknown14	3143	302	479	203	66	46795	936.5	1592	35687	327	164	381	353.5	341	15	25842	13	39.5	1603.5	217
39(L.G5)	Unknown14	2897.5	249	527.5	172	56	46038	849	1557	34092.5	325	153	304	305	359	13	25222	10	46	1554	205.5
42(L.B6)	Unknown5	689	1375	806.5	68	172	41517	1246	1950	30121.5	1073	221	366	343.5	1268.5	13.5	20242	17	41	1820.5	577.5
43(L.C6)	Unknown5	719	1256	814	76	210	41379	1299.5	2028	29548.5	1042.5	203.5	299.5	344.5	1181	14	19978	21	49	1704	593
44(L.D6)	Unknown5	748	1266	739.5	70	186	39405	1205	1896.5	28963	1009	192	306.5	357	1150	7.5	19289.5	13	43	1675	586
45(L.E6)	Unknown15	1225	83	218	80	29	24717	367	737.5	15443.5	135	65	125	124.5	105	10	11978	9	23	672	79
46(L.F6)	Unknown15	1202	101	198	92	25	22709	376.5	622.5	14565	140	68	168	120.5	129	5	11571.5	9	19	659	87.5
47(L.G6)	Unknown15	1239.5	91	184	86.5	26	22938.5	384	624	14196.5	140	60.5	159.5	119	135	4.5	11212.5	16	22	659	77
50(L.B7)	Unknown6	2466	5366	4242	990.5	474	68735.5	6568	7738	64306	5953	830.5	1161	1553	3825	11.5	62195	18	239	11018	2233

Spot dilution series of compounds across wells to allow dose response profiling

# Discovery – setting

## Prioritize compounds

- Focus on gene signature
  - +/- 20 disease genes
- Rank compounds
  - multivariate approach
- Tool compound (TC: reference compound)
  - match or surpass TC

**Select/prioritize** compounds based  
on a **gene set**

# Compound prioritization

## 1. Gene specific analysis

→ Dose response modelling

## 2. Aggregate absolute AC50 estimates

→ Absolute AC50 matrix

# Compound prioritization

## 3. Summarize AC50 matrix

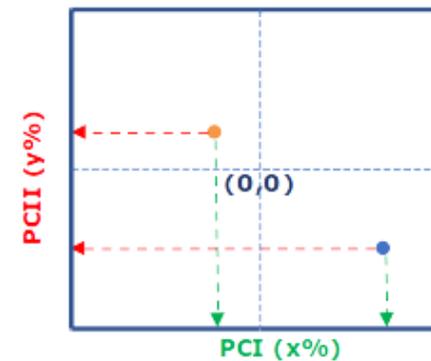
→ Principal Component Analysis

## 4. Ranking

→ Ranking metric vs ranking metric tool compound

- \* Metric based on weighted PCA scores
- \* Signs relative to position of toolcompound

	Gene 1	Gene 2	...
CPD 1	AC50	AC50	
CPD2	AC50	AC50	
...			

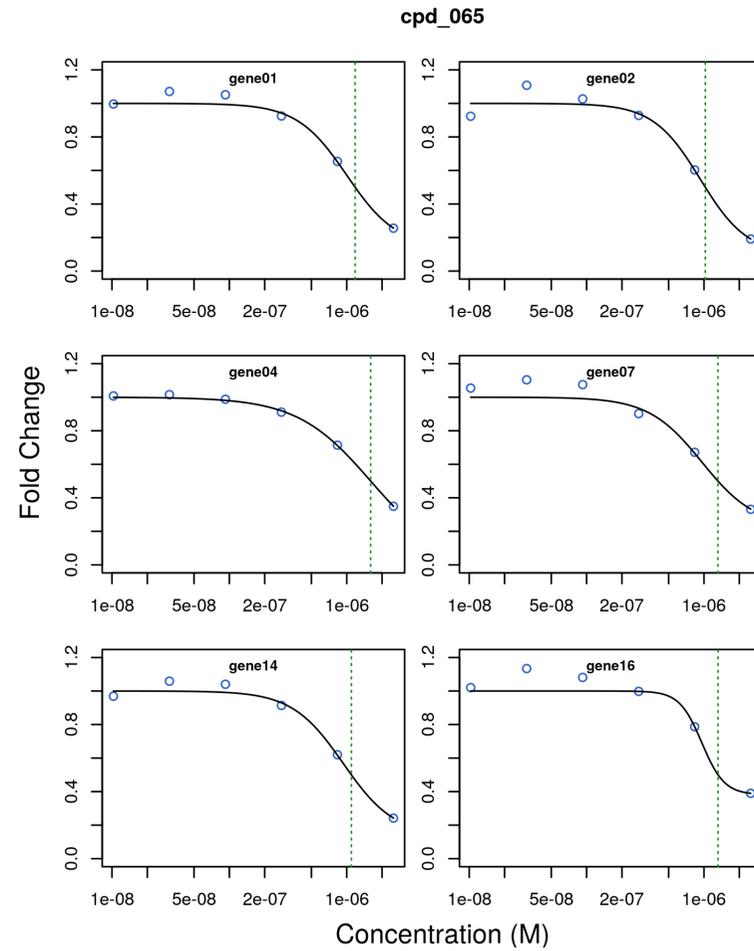
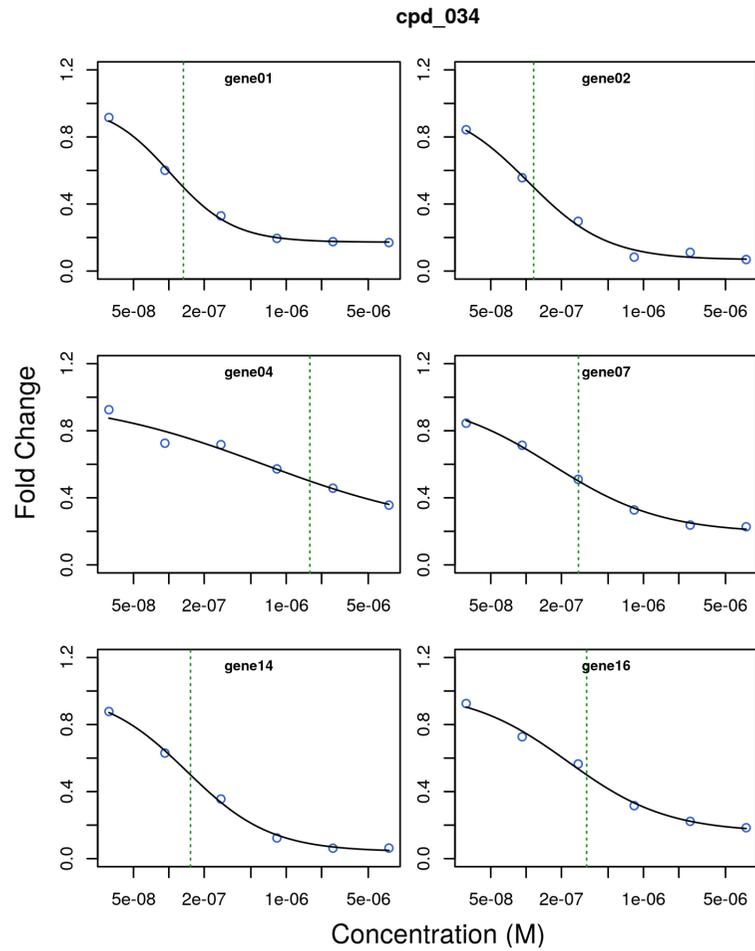


$$\text{PCI dist} * x\% + \text{PCII dist} * y\%$$

$$\text{PCI dist} * x\% + \text{PCII dist} * y\%$$

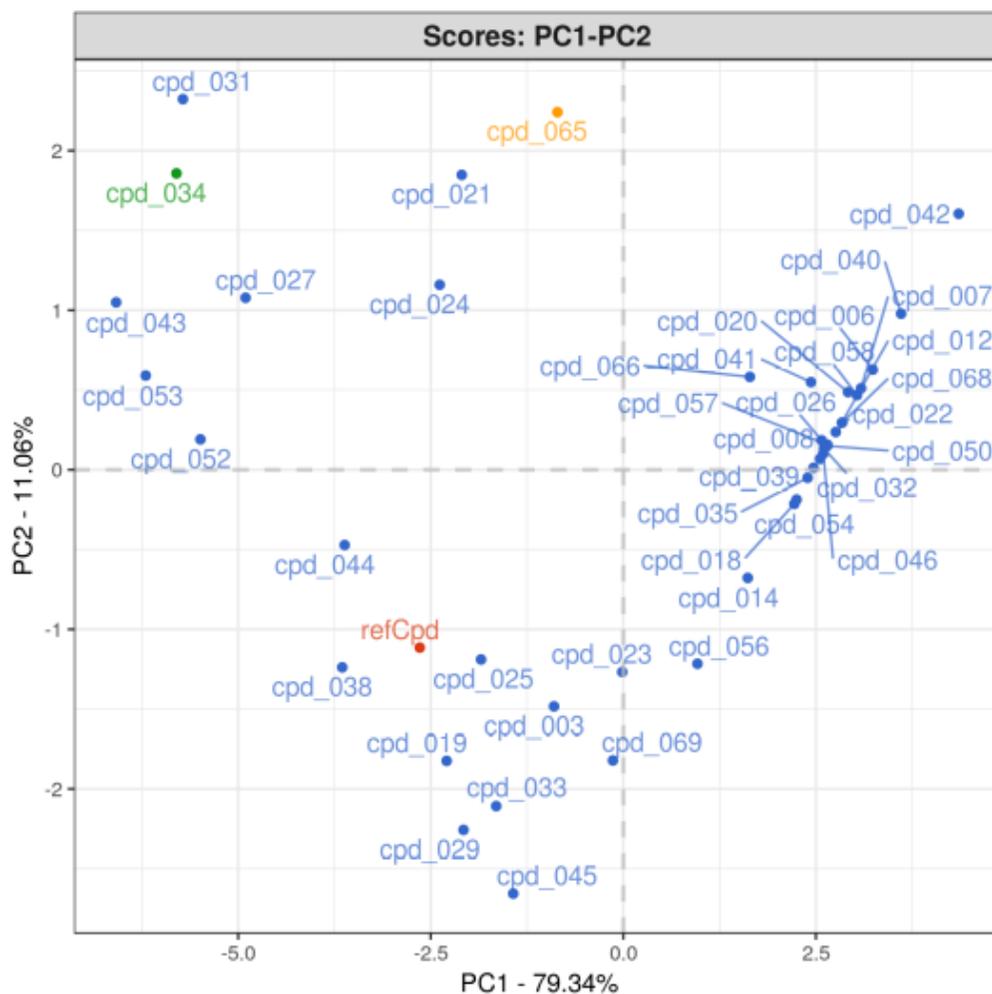
...

# Compound prioritization – example





# Compound prioritization – example



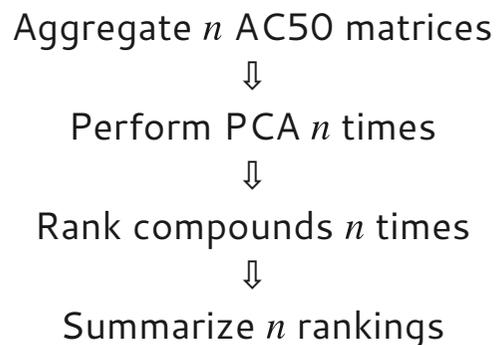
compound	rank
cpd_043	1
cpd_053	2
cpd_034	3
cpd_052	4
cpd_031	5
cpd_027	6
cpd_038	7
cpd_044	8
refCpd	9
cpd_019	10
...	
cpd_065	18
...	

# Risk-based compound prioritization

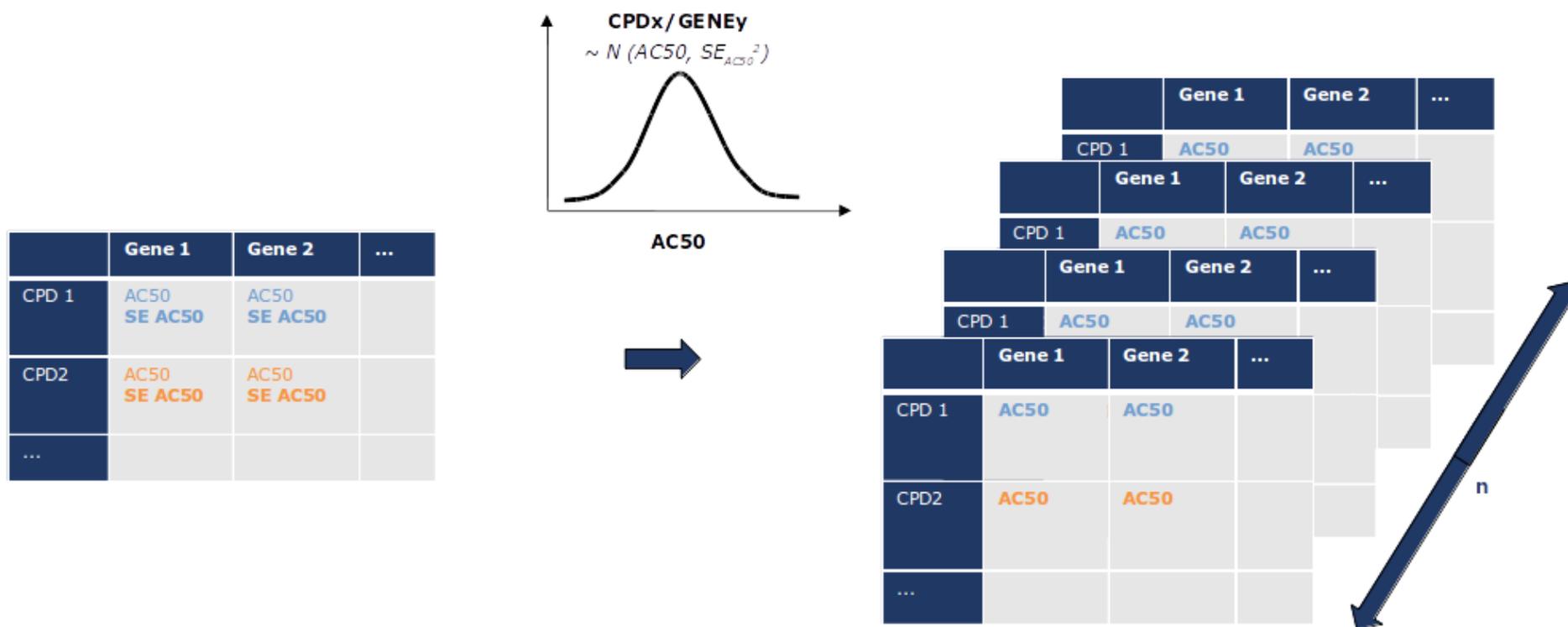
## Uncertainty of dose response fitting

- Model **estimates** have associated **uncertainty**: standard errors (SE)
- So far, PCA and ranking based on absolute AC50 estimates only
- Use estimates *and* SEs to define a normal distribution for each compound/gene combination
- Randomly **sample**  $n$  AC50 estimates from each distribution

	Gene 1	Gene 2	...
CPD 1	AC50 SE AC50	AC50 SE AC50	
CPD2	AC50 SE AC50	AC50 SE AC50	
...			

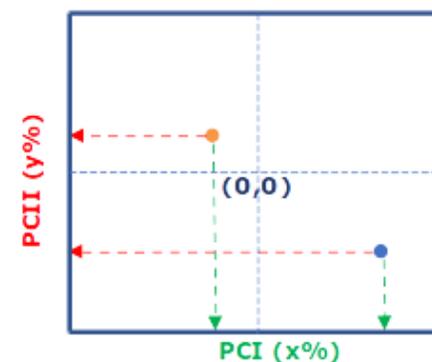


# Risk-based compound prioritization



# Risk-based compound prioritization

	Gene 1	Gene 2	...
CPD 1	AC50	AC50	
CPD 1	AC50	AC50	
CPD 1	AC50	AC50	
CPD 2	AC50	AC50	
...			



●  $PCI\ dist * x\% + PCII\ dist * y\%$

●  $PCI\ dist * x\% + PCII\ dist * y\%$

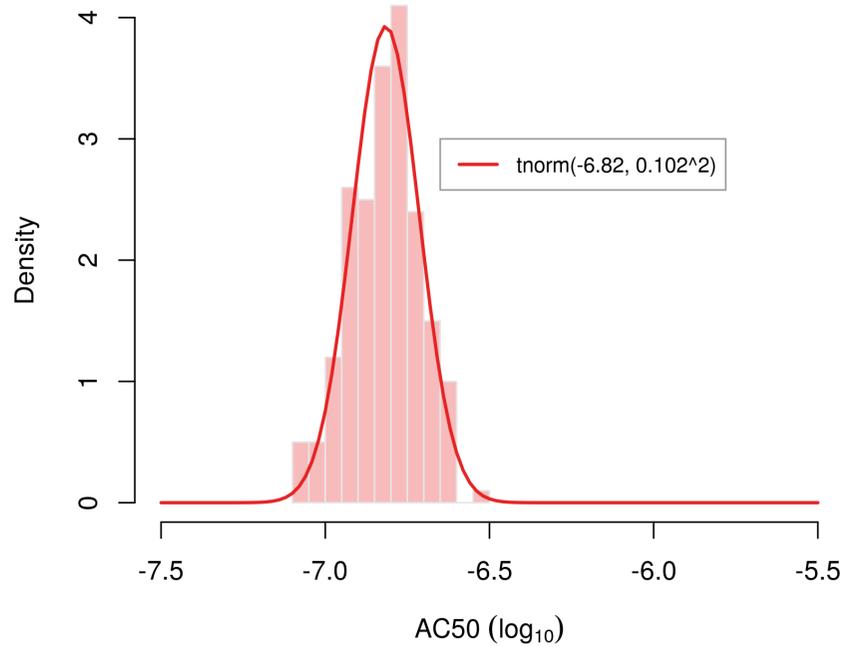
...

# Risk-based compound prioritization – example

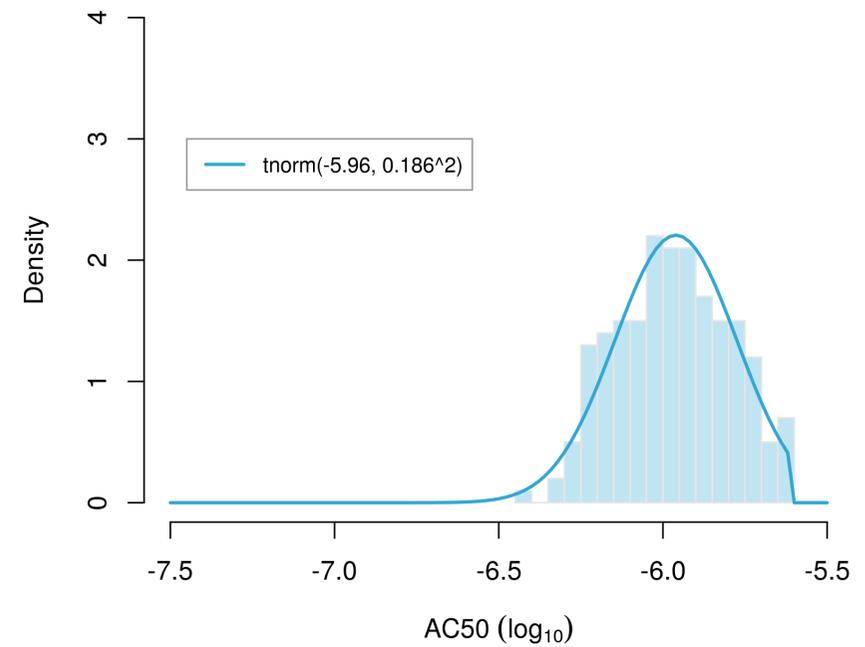
## Random sampling of AC50 estimates

n = 200

cpd\_034/gene14



cpd\_065/gene14

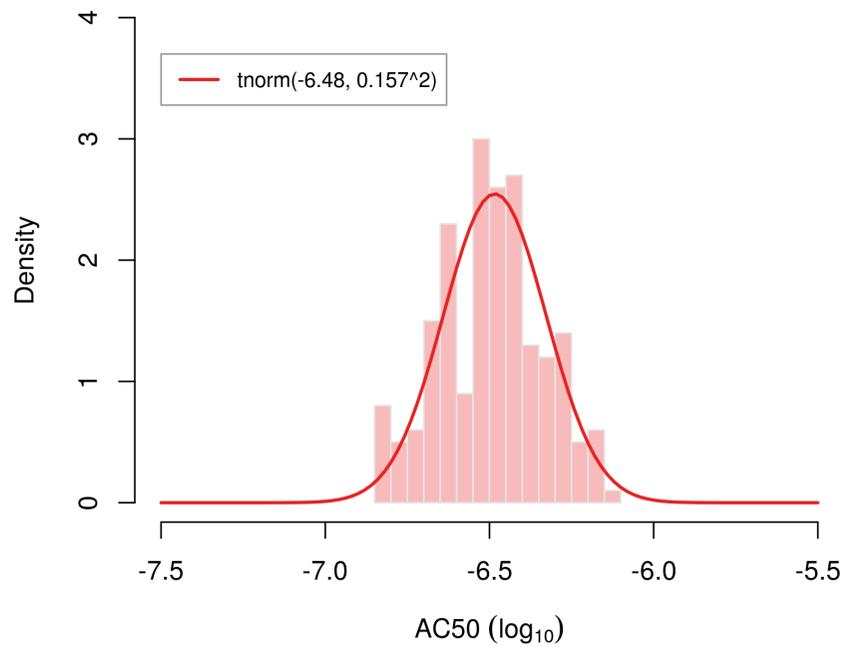


# Risk-based compound prioritization – example

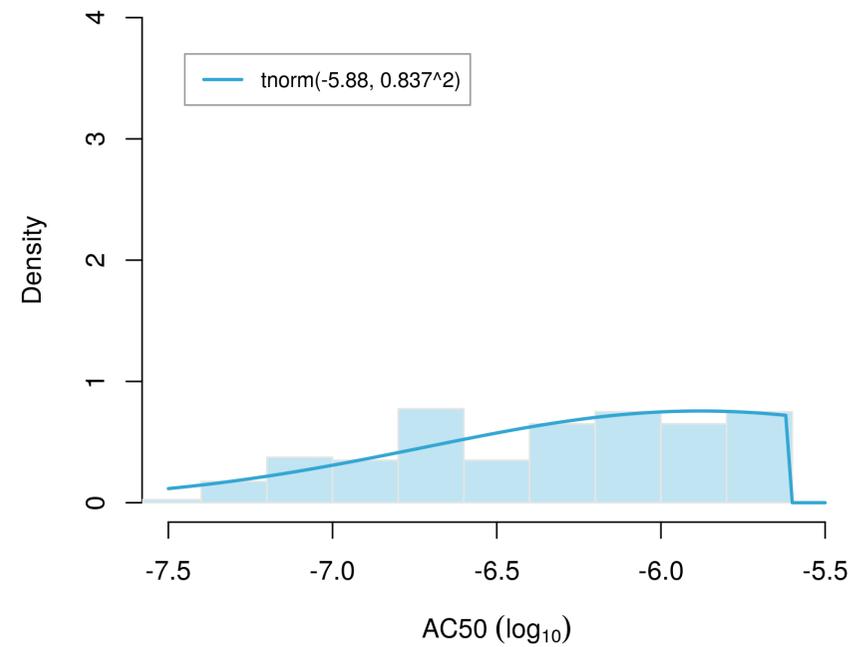
## Random sampling of AC50 estimates

n = 200

cpd\_034/gene16

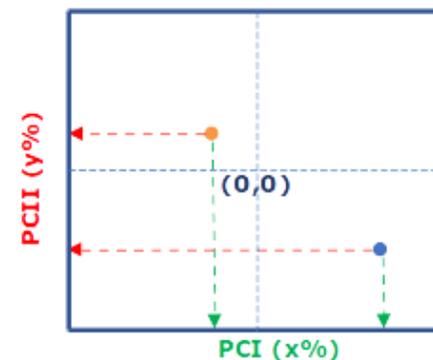


cpd\_065/gene16



# Risk-based compound prioritization – example

	Gene 1	Gene 2	...
CPD 1	AC50	AC50	
CPD 2	AC50	AC50	
...			



●  $PCI\ dist * x\% + PCII\ dist * y\%$

●  $PCI\ dist * x\% + PCII\ dist * y\%$

...

# Risk-based compound prioritization – example

## Summarize $n$ rankings

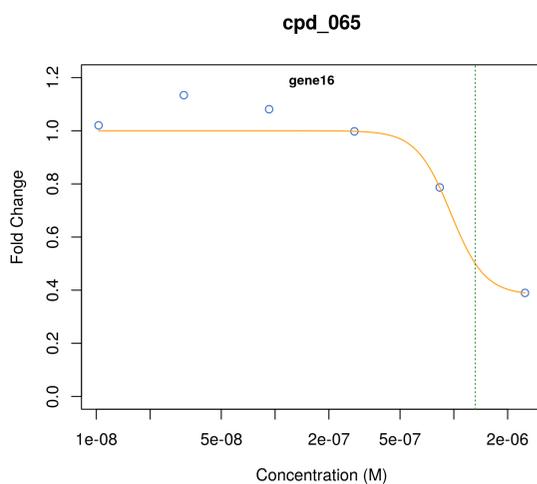
- Median rank across  $n$  rankings
- Times ranked above tool compound (%)
- Consensus ranking

compound ↕	rank ▲	medRank ↕	above refCpd ↕
cpd_043	1	1	100
cpd_053	2	2	100
cpd_034	3	4	100
cpd_052	4	4	100
cpd_031	5	5	100
cpd_027	6	6	100
cpd_038	7	6	100
cpd_044	8	8	99
refCpd	9	11	
cpd_019	10	9	77
...			
cpd_065	18	16	14
...			

# Risk-based compound prioritization – example

## Summarize $n$ rankings

- Median rank across  $n$  rankings
- Times ranked above tool compound (%)
- Consensus ranking



compound ↕	rank ▲	medRank ↕	above refCpd ↕
cpd_043	1	1	100
cpd_053	2	2	100
cpd_034	3	4	100
cpd_052	4	4	100
cpd_031	5	5	100
cpd_027	6	6	100
cpd_038	7	6	100
cpd_044	8	8	99
refCpd	9	11	
cpd_019	10	9	77
...			
cpd_065	18	16	14
...			

# Summary

- **QuantiGene plex** assay allows for **multiplexed** gene expression quantitation
- From **gene-specific** dose-response analysis to **multivariate** summary:  
*PCA and ranking*
- Take into account **associated uncertainty**:  
*sampling procedure*
- Repeat PCA and ranking  $n$  times
- **Risk-based** compound selection:  
*median/average rank*

Thank you!

Any Questions?

# Back up

## Abstract

One of the first steps in the drug discovery process is to check the activity of newly synthesized compounds on a particular target. Only compounds with the most desirable activity profile will be further processed to understand more of their characteristics. Often, pharmacodynamic (PD) markers are used to monitor biological activity on the targets. The Quantigene™ Plex Assay allows for the simultaneous measurement of up to 80 genes (markers) of interest in a single well. Dilution series of compounds spotted across the wells, allow to investigate the dose response profiles of each of the markers in one single experiment. Hence, instead of selecting compounds based on one activity estimate, this technology allows to look at multiple estimates at the same time.

To be able to select most promising compounds using all information available, multivariate techniques like principal component analysis (PCA) can be used on the matrix of AC50 values. Nevertheless, such approach does not consider the uncertainty associated with the estimation of these AC50 values from the single experiment. However, this information can be very valuable to build confidence on the actual selection of compounds you make. To include the uncertainty in the decision making, a random sampling approach is proposed based on the AC50 values and corresponding standard errors. These figures can act as parameters, defining a distribution from which a random sample is obtained. The resulting bootstrapped AC50 matrices can be supplied to PCA to obtain a set of compound rankings. Finally, a summarized rank across this set of rankings is obtained for each compound incorporating the uncertainty associated with the available estimates from your one experiment.

This proposal allows to assess the risks associated with the decision to move some compounds forward. A concept that is often ignored during the drug discovery process and could be more broadly applied.