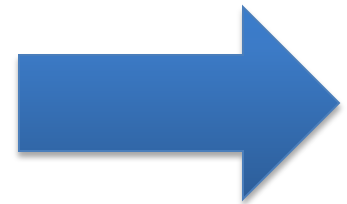
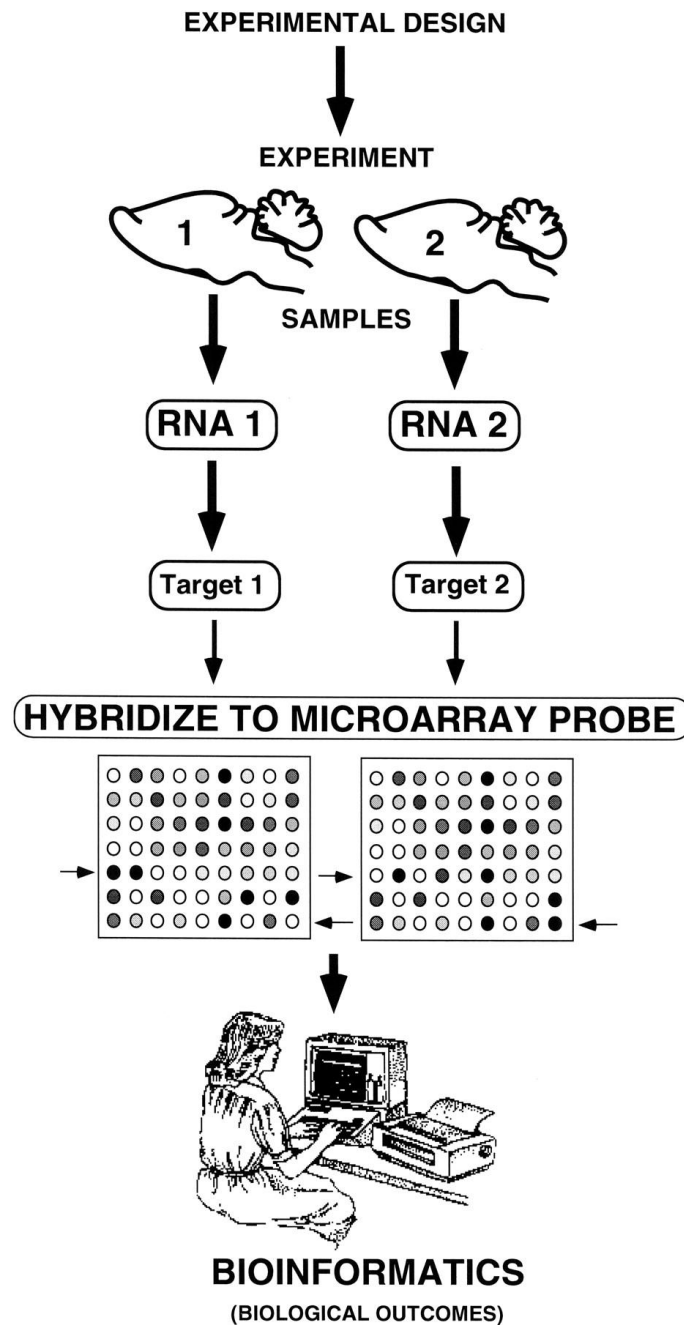
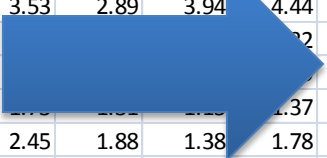
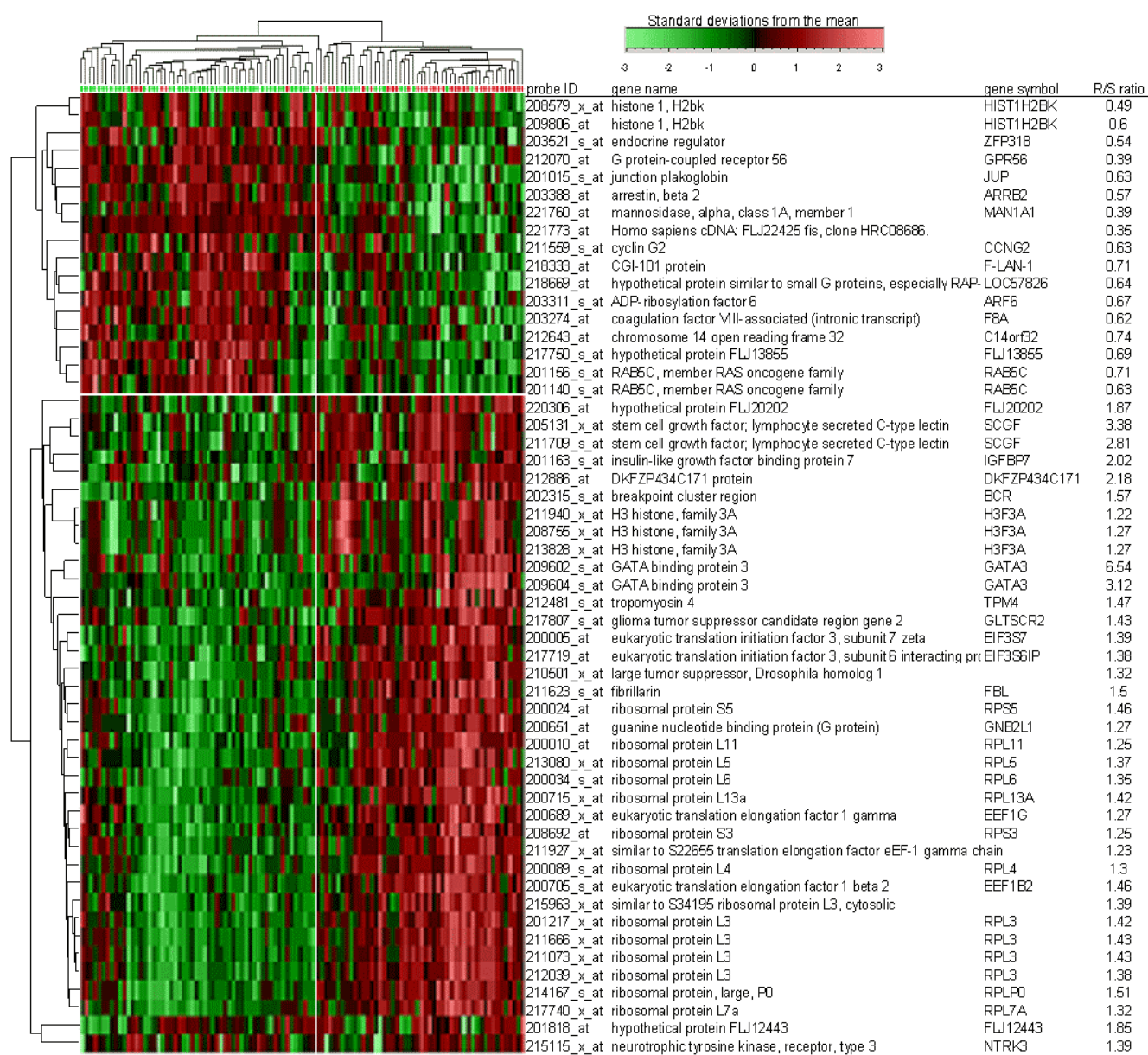

Assessing Individual Differences in Genome-Wide Gene Expression in Human Whole Blood: Reliability Over Four Hours and Stability Over 10 Months

Emma L. Meaburn, Cathy Fernandes, Ian W. Craig, Robert Plomin, and Leonard C. Schalkwyk
Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom

Twin Research and Human Genetics
2009



PlatePosition	Title	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
HK1A1	excision repair cross-complementing	1.60	1.58	1.67	1.79	1.99	1.27	1.37	1.58	2.36	1.61	1.88	2.33	2.64	1.66	2.13
HK1A10	ESTs	1.64	2.91	1.18	1.45	1.64	2.36	1.36	1.47	1.54	1.37	1.42	52.35	6.15	3.27	6.15
HK1A11	EphB4	2.51	4.72	4.00	5.66	3.10	6.41	2.89	1.45	1.96	2.69	1.65	1.89	2.38	1.77	1.75
HK1A2	early development regulator 2 (homo	1.75	2.62	1.62	2.71	3.36	2.01	1.99	2.06	1.88	1.52	2.11	2.01	1.77	2.11	1.66
HK1A4	EphA1	2.64	5.28	1.40	1.65	3.58	3.12	2.06	4.11	1.74	2.25	2.95	1.82	1.78	2.41	3.71
HK1A5	sema domain, seven thrombospondin	6.02	2.41	3.07	9.92	4.92	5.21	3.63	2.73	4.76	2.38	3.14	2.13	2.11	2.23	3.51
HK1A6	ESTs	2.77	5.90	5.28	3.53	3.51	2.55	2.53	4.86	5.46	7.16	3.81	2.43	1.82	1.75	2.23
HK1A7	thioredoxin reductase 1	4.56	7.94	19.70	11.96	3.97	3.53	51.63	699.41	17.75	76.64	10.56	4.69	10.63	3.86	7.11
HK1A8	actin related protein 2/3 complex, sub	10.85	2.38	1.52	2.35	1.74	1.58	1.66	1.96	1.88	1.71	2.43	1.87	1.71	2.68	1.83
HK1A9	ribosomal protein L39	5.24	3.68	1.78	2.25	3.51	2.81	2.11	2.25	2.39	7.11	16.00	4.06	4.89	3.63	5.03
HK1B1	heat shock protein, neuronal DNAJ-like	1.46	1.87	2.28	3.39	2.64	1.29	3.34	1.47	1.71	1.32	1.77	1.48	1.20	1.45	1.47
HK1B10	ESTs	5.39	4.59	1.97	3.23	4.59	2.22	2.43	3.84	2.35	2.31	1.85	2.71	1.91	1.99	2.95
HK1B11	interferon gamma receptor 2 (interfer	2.64	3.53	2.68	3.71	2.51	2.03	4.69	6.36	2.04	2.83	5.03	3.27	2.04	2.97	4.53
HK1B12	adenine nucleotide translocator 3 (liv	2.11	3.05	2.14	4.38	4.92	1.65	4.69	1.85	1.56	2.51	2.57	5.54	2.14	2.58	1.89
HK1B2	N-acetylgalactosaminidase, alpha-	1.47	1.75	2.51	2.28	2.38	1.44	2.20	1.72	1.80	1.28	1.71	1.48	1.36	1.64	1.37
HK1B3	KIAA0295 protein	6.32	2.73	1.82	2.62	4.35	2.41	39.95	2.48	2.30	1.78	1.84	4.32	1.93	3.66	3.94
HK1B4	proteasome (prosome, macropain) 26	3.14	5.21	2.83	2.53	7.01	2.33	12.47	5.62	2.87	7.41	3.53	2.89	3.94	4.44	2.99
HK1B5	chaperonin containing TCP1, subunit 2	1.35	1.64	1.30	1.28	1.92	1.69	1.27	1.68	1.65	1.99		2.72	1.78		
HK1B6	serine hydroxymethyltransferase 1 (se	4.89	4.76	2.39	3.41	4.82	3.86	2.87	3.18	4.17	6.11		3.68			
HK1B7	deoxycytidine kinase	2.00	1.83	1.49	1.51	3.29	1.85	2.22	1.60	1.80	1.85		1.32			
HK1B8	D123 gene product	3.25	3.97	4.41	2.58	3.43	3.39	5.94	2.68	2.06	3.25	2.45	1.88	1.38	1.78	2.23
HK1B9	protein tyrosine phosphatase, recepto	2.93	8.46	3.03	2.62	3.36	2.00	2.69	2.69	3.81	1.93	1.96	2.71	1.52	1.68	2.13
HK1C10	keratin 7	4.08	3.34	3.07	2.75	2.83	2.87	2.95	2.71	2.75	1.72	2.99	11.00	6.15	2.48	3.05
HK1C2	cyclin G2	3.86	4.03	5.86	3.84	5.46	3.48	5.86	7.41	5.70	3.16	4.76	9.51	2.30	13.27	10.06
HK1C3	methyltransferase-like 1	1.19	1.83	1.35	1.17	1.33	1.08	1.25	1.16	1.20	1.64	1.20	1.71	1.15	1.29	1.34
HK1C4	integrin, beta 4	1.84	1.78	1.36	2.11	4.32	2.45	1.66	2.73	2.08	1.87	2.87	2.97	1.37	1.77	2.27
HK1C5	small nuclear RNA activating complex,	2.97	2.64	1.67	2.35	2.25	2.08	3.68	2.23	2.01	2.08	41.07	3.10	2.00	2.33	2.07
HK1C6	matrix metalloproteinase 9 (gelatinas	1.85	2.22	1.66	2.81	1.97	2.20	2.03	2.64	2.13	2.10	2.39	2.01	2.39	1.95	2.25
HK1C7	ubiquitin-activating enzyme E1 (A159T	2.46	3.03	2.53	2.64	3.36	2.03	1.96	2.97	1.73	2.00	2.75	3.73	1.74	1.47	2.50
HK1C8	clathrin, heavy polypeptide-like 2	2.97	2.75	2.01	3.84	3.01	2.97	1.92	2.19	2.55	2.14	1.93	1.75	1.59	1.61	2.17
HK1C9	Homo sapiens nuclear antigen H731-li	2.64	3.76	1.92	2.03	3.81	3.05	1.37	2.27	2.07	1.82	2.30	4.50	2.03	3.41	4.92
HK1D10	ribosomal protein L7a	3.03	4.63	2.79	1.92	2.81	2.04	2.36	2.20	2.69	2.99	3.36	3.23	1.96	4.32	3.61
HK1D11	nucleosome assembly protein 1-like 1	1.99	1.57	3.71	3.03	1.71	1.57	2.14	1.44	1.85	1.54	1.75	1.38	1.30	1.91	1.74
HK1D2	Human mRNA for hepatoma-derived g	2.04	2.83	2.06	1.91	4.44	2.13	3.27	1.78	1.96	1.84	2.69	1.55	1.31	1.78	2.68
HK1D3	KIAA0669 gene product	2.23	1.89	3.48	2.06	1.71	2.64	2.27	2.79	2.51	1.71	2.22	2.07	2.71	2.36	2.33
HK1D5	MADS box transcription enhancer fact	2.36	3.43	2.11	2.30	6.63	4.79	2.30	3.20	2.64	2.62	2.62	5.03	2.19	2.08	6.45
HK1D6	O-linked N-acetylglucosamine (GlcNA	2.64	6.06	2.19	3.07	4.11	5.21	4.26	1.96	2.36	2.45	3.39	2.08	1.77	2.03	2.41
HK1E1	forkhead (Drosophila)-like 1	9.25	4.35	3.94	2.79	3.25	6.82	97.01	4.41	1.65	8.00	4.56	2.81	2.30	1.59	1.85
HK1E10	steroidogenic acute regulatory protein	5.35	99.04	3.27	6.96	5.31	2.17	41.07	2.23	2.64	2.17	3.51	2.08	1.28	1.99	2.25
HK1E11	ESTs, Weakly similar to hsp 27 [H.sapie	2.16	1.83	1.72	2.75	2.57	2.06	2.08	2.03	1.93	1.79	2.46	1.46	1.56	1.85	1.89



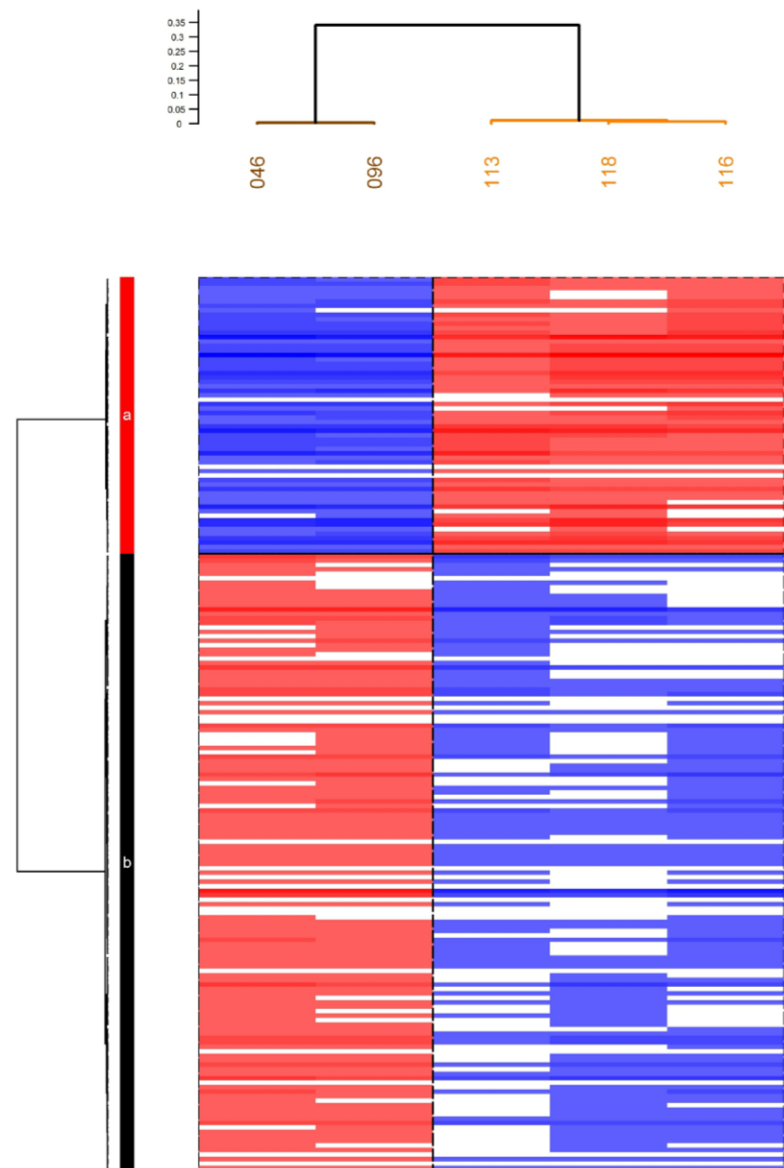
R/S ratio: ratio of the median expression in drug resistant versus sensitive ALL.

Hollemann et al.

Gene-expression patterns in drug resistant acute lymphoblastic leukemia cells and response to treatment.

N Engl J Med. 2004 Aug 5;351(6):533-42.

Figure S1: Hierarchical clustering using 200 probe sets that distinguish between molecular apocrine tumors with and without germline PTEN mutations. Tumors 46 and 96: molecular apocrine carcinoma without PTEN mutation; tumors 113, 116 and 118, breast carcinoma occurring on Cowden disease. Table S6 lists the corresponding genes. Blue: under expressed; red: over expressed.



Question:

- How many genes show reliable and stable expression pattern over time (4 hrs; 10 months)
- This experiment studies stability of individual, person-specific differences, not technical or technological reliability.
- Technology used for measuring gene expression levels: **Affymetrix HG-U133 plus 2.0** microarray assessing **54,675 probesets**

Experiment:

Family	Individual	Test time: T		Test time: T + 10 months	
		10am	2pm	10am	2pm
A	A1	X ₁	X ₂	Y ₁	Y ₂
	A2	X ₁	X ₂	Y ₁	Y ₂
B	B1	X ₁	X ₂	Y ₁	Y ₂
	B2	X ₁	X ₂	Y ₁	Y ₂
C	C1	X ₁	X ₂	Y ₁	Y ₂
	C2	X ₁	X ₂	Y ₁	Y ₂
D	D1	X ₁	X ₂	Y ₁	Y ₂
	D2	X ₁	X ₂	Y ₁	Y ₂
E	E1	X ₁	X ₂	Y ₁	Y ₂
	E2	X ₁	X ₂	Y ₁	Y ₂

5 healthy MZ twin pairs, all 12 years old

4 female pairs, 1 male pair

Whole blood sample was drawn at 10am and 2 pm

Experiment was repeated after 10 months

Quality control:

INITIAL DATA: 40 arrays and 54,675 probesets (tested transcripts)

- Microarray images were inspected for the presence of image artefacts
- Each array was examined and compared for signal quality differences, average background intensity, scaling factor, percent present call rate and 3'/5' hybridization intensity ratios.
- Probe-level model fitting was performed to assess relative log expression (RLE) values and unscaled standard errors (NUSE)
- Probes with signal present in < 50% arrays were removed

USED DATA: 36 arrays and 25,864 *present* probesets

Reliability experiment:

Family	Individual	Test time: T		Test time: T + 10 months	
		10am	2pm	10am	2pm
A	A1	x_1	x_2	y_1	y_2
	A2	x_1	x_2	y_1	y_2
B	B1	x_1	x_2	y_1	y_2
	B2	x_1	x_2	y_1	y_2
C	C1	x_1	x_2	y_1	y_2
	C2	x_1	x_2	y_1	y_2
D	D1	x_1	x_2	y_1	y_2
	D2	x_1	x_2	y_1	y_2
E	E1	x_1	x_2	y_1	y_2
	E2	x_1	x_2	y_1	y_2



4 hr reliability test 1
CORREL()

4 hr reliability test 2
CORREL()

Reliability tests:

- Mean correlation 0.338 and 0.474
- Statistically significant correlations ($r > 0.55$; $p < 0.05$, one-tailed) were detected for 9,238 and 13,156 probesets (**36%** and **51%** of all *present* probesets)
- Highly significant correlations ($r > 0.7$; power 80%) + highly variable probesets ($CV > \text{medianCV}$) were detected for 3,414 and 5,174 probesets
- Overlap between these two reliability tests: **1,752 probesets**

Reliability tests:

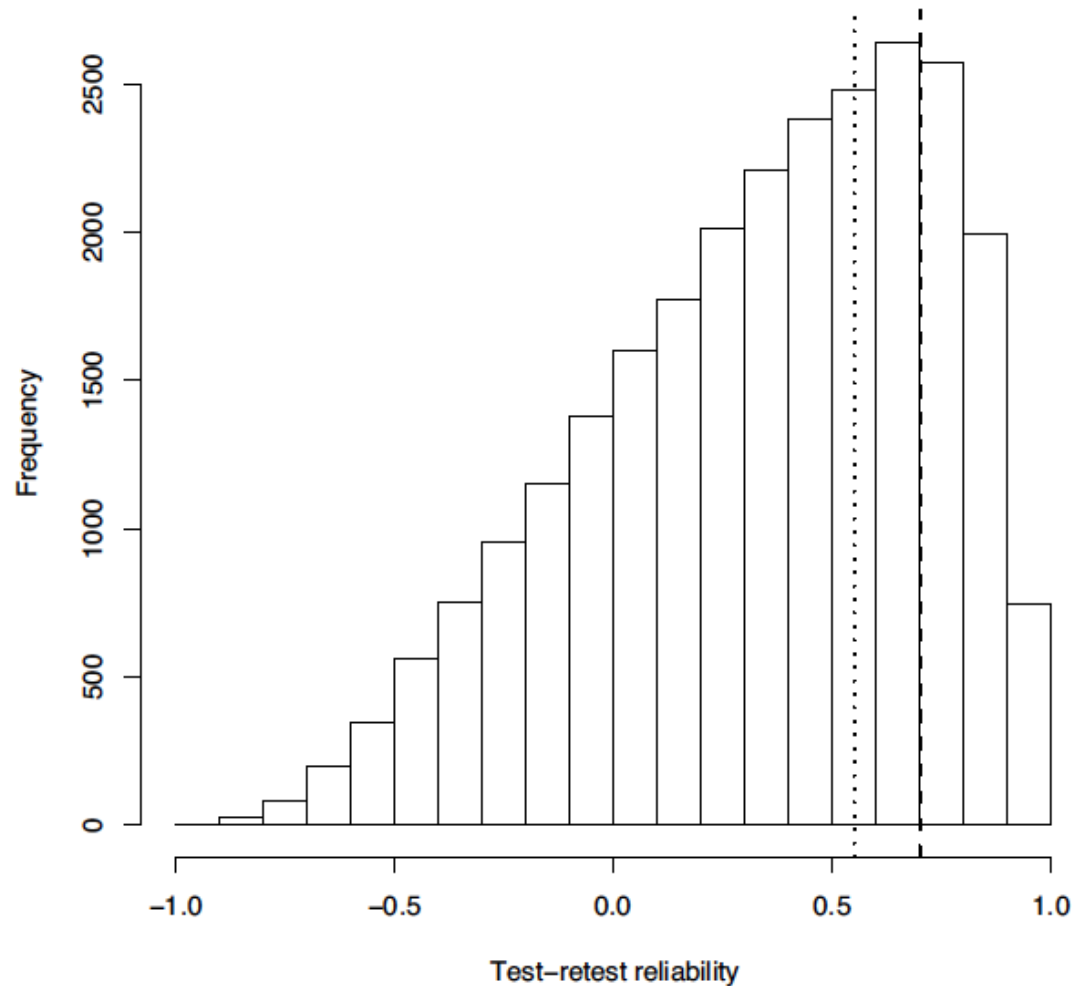


Figure 1

Distribution of test-retest reliabilities for 25,864 'present' probesets. The dotted line indicates a test-retest correlation of .55 ($p < .05$, 50% power), the dashed line designates a test-retest correlation of .70 ($p < .05$, 80% power).

Stability experiment:

Family	Individual	Test time: T		Test time: T + 10 months	
		10am	2pm	10am	2pm
A	A1	x_1	x_2	y_1	y_2
	A2	x_1	x_2	y_1	y_2
B	B1	x_1	x_2	y_1	y_2
	B2	x_1	x_2	y_1	y_2
C	C1	x_1	x_2	y_1	y_2
	C2	x_1	x_2	y_1	y_2
D	D1	x_1	x_2	y_1	y_2
	D2	x_1	x_2	y_1	y_2
E	E1	x_1	x_2	y_1	y_2
	E2	x_1	x_2	y_1	y_2



10 month stability test

Stability tests:

1,291 probesets out of 1,752 probesets (74%)
are stable at $r > 0.55$ after 10 months

1,019 probesets out of 1,752 probesets (58%)
are stable at $r > 0.7$ after 10 months

Stability tests:

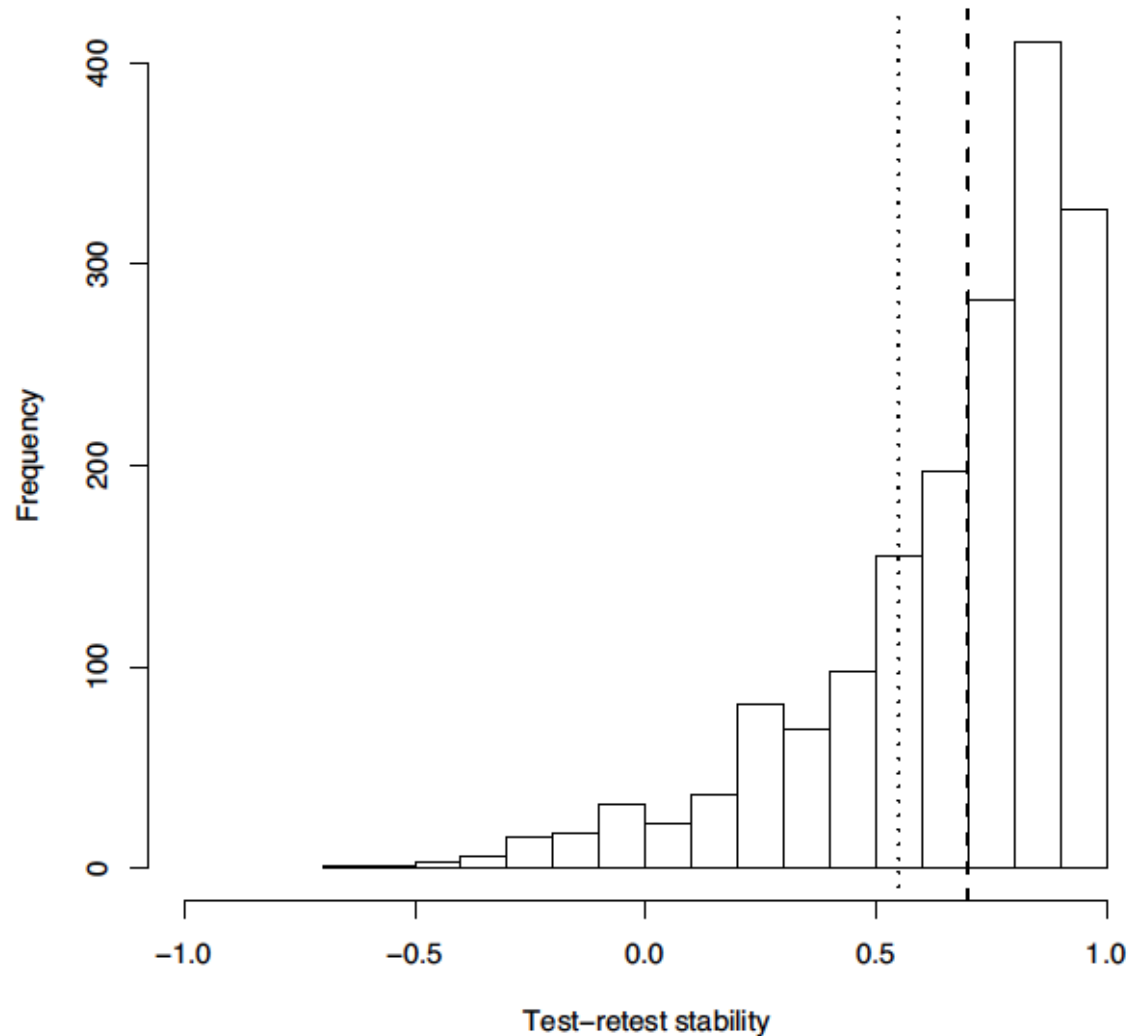


Figure 2

Distribution of 10-month stability correlations for individual differences in gene expression for 1,752 probesets that reliably detect individual differences in gene expression across a 4-hour period from blood collection on two occasions 10 months apart. The dotted line indicates test-retest correlation of .55, the dashed line indicates a test-retest correlation of .70.

Heritability of GE:

Heritability for:

25,864 *present* probesets is **0.34**

1,752 *present and reliable* probesets is **0.68**

1,291 *present, reliable and stable* probesets is **0.76**

Heritability of GE:

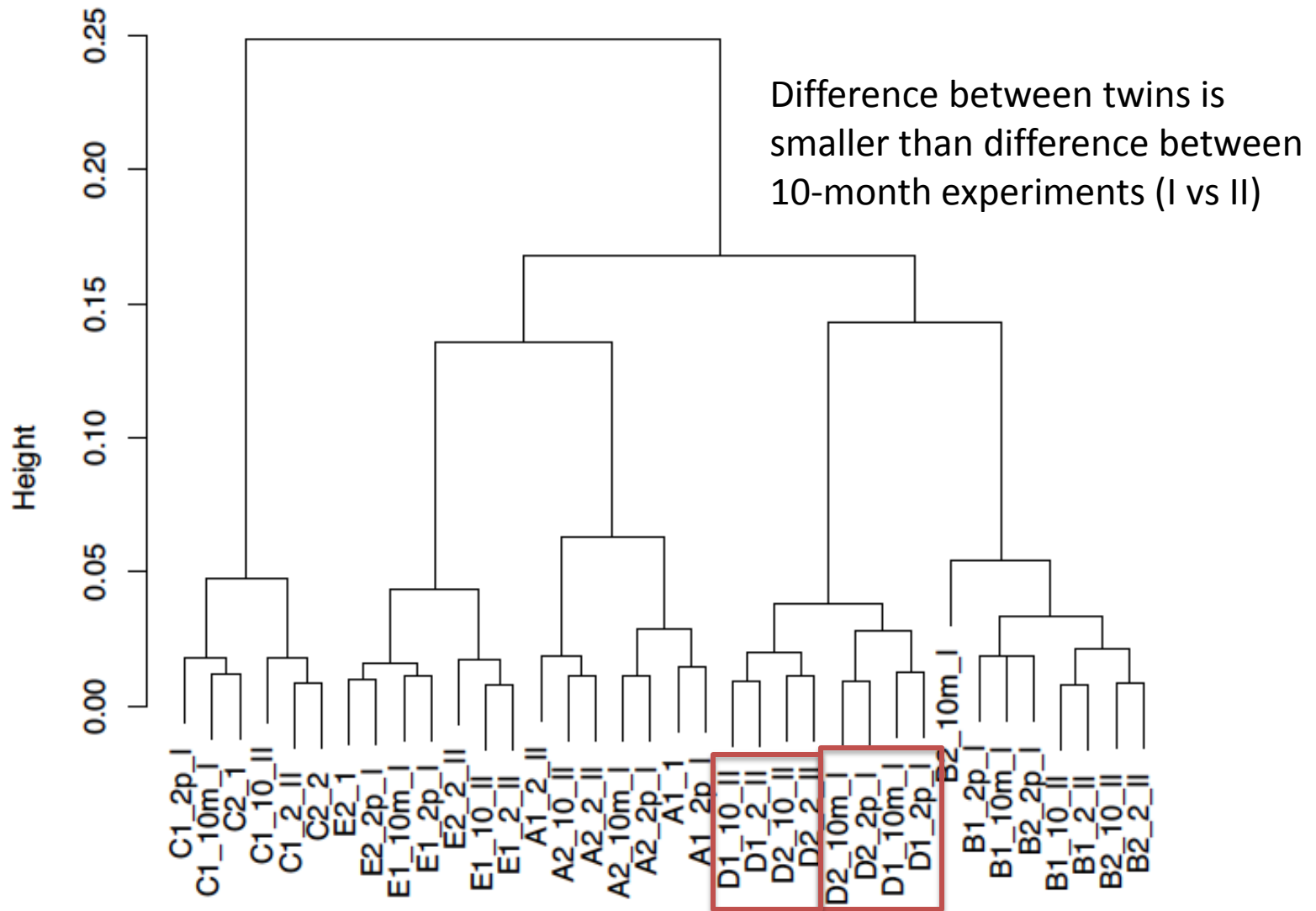


Figure 3

Hierarchical clustering across all 36 arrays for the 1,291 probesets that reliably and stably detect individual differences in gene expression. Five distinct branches can be seen, each representing a twin pair (pairs A to E).

Function of stable genes:

From 1,291 *present, reliable and stable* probesets 775 different genes were detected, most of them are widely expressed in different tissues.

No particular functional themes were identified, characteristic to *stable* probesets.

The top five molecular functions were:
cellular growth and proliferation
cell-to-cell interaction
cell death
cell signaling
molecular transport

This is similar to *present* probesets (25,864 probesets)

Conclusion:

50% of all 54,675 probesets were ***present*** in > 50% microarrays

16-26% of all probesets were ***present and reliable*** over 4 hours

2% of all probesets were ***present, reliable and stable*** over 10 months AND over 4 hours in two experiments.

Could there be larger number of *stable* genes which were not *reliable* in 4hr experiments or not *present* in all arrays?