

ARTICLE

Application of microarrays to the analysis of the inactivation status of human X-linked genes expressed in lymphocytes

Ian W Craig^{*1}, Jonathan Mill¹, Gavin M Craig¹, Caroline Loat¹ and Leonard C Schalkwyk¹

¹SGDP Centre, Institute of Psychiatry, Box PO 82, Denmark Hill, London SE5, UK

Dosage compensation in mammalian females is achieved by the random inactivation of one X chromosome early in development; however, inactivation is not complete. In addition to a majority of pseudoautosomal loci, there are genes that are expressed from both the active and the inactive X chromosomes, and which are interspersed among other genes subject to regular dosage compensation. The patterns of X-linked gene expression in different tissues are of great significance for interpreting their impact on sex differences in development. We have examined the suitability and sensitivity of a microarray approach for determining the inactivation status of X-linked genes. Biotinylated cRNA from six female and six male lymphocyte samples were hybridised to Affymetrix HG-U133A microarrays. A total of 36 X-linked targets detected significantly higher levels of female transcripts, suggesting that these corresponded to sequences from loci that escaped, at least partly, from inactivation. These included genes for which previous experimental evidence, or circumstantial evidence, existed for their escape, and some novel candidates. Six of the targets were represented by more than one probe set, which gave independent support for the conclusions reached.

European Journal of Human Genetics (2004) 12, 639–646. doi:10.1038/sj.ejhg.5201212

Published online 28 April 2004

Keywords: X-inactivation; sex-differences; gene expression

Introduction

Dosage compensation in mammalian females is achieved by the inactivation of one X chromosome early in development leading to the equality of X-linked gene products between male and female cells.¹ Not all X-linked genes are inactivated. Most loci in the pseudoautosomal regions on the tips of the long and short arms are active; there are, in addition, genes interspersed throughout the remainder of the chromosome that are expressed from both the active and, at varying levels, the inactive chromosomes.^{2–4} Some estimates indicate that 10–20% of X-linked genes escape inactivation, with the majority

located on Xp. Some of these have Y homologues, but many do not.^{5,6} Furthermore, there may be other genes on the X chromosome which are differentially imprinted, that is, selectively silenced, depending on their parent of origin.⁷ Loci escaping inactivation appear to be found both in clusters and in isolation.^{8,9} The patterns of X-linked gene expression in different tissues and, in particular, different regions of the brain are of great significance for interpreting their impact on behavioural sex differences. Due to the paucity of current data, determination of the inactivation status of additional genes remains critical for the formulation and resolution of rival theories concerning the nature and spread of the inactivation process.

An efficient means of determining the activity of an X-linked gene is via the analysis of mouse–human somatic cell hybrids retaining only the inactive human X chromosome in a rodent background.⁵ As inactivation is effected at

*Correspondence: Professor I Craig, SGDP Centre, Box PO 82, Institute of Psychiatry, Denmark Hill, London SE5, UK. Tel: +44 207 848 0018; Fax: +44 207 848 0407; E-mail: i.craig@iop.kcl.ac.uk
Received 26 June 2003; revised 25 February 2004; accepted 19 March 2004

the level of transcription, detection of human-specific transcripts by RT-PCR is suggestive of an escape from inactivation. Inactivation can be confirmed by quantitation of the relevant transcripts from XO, XX, XXX or XXXX human cell lines.^{9,10} Other strategies are possible including observations on tissues, or cell lines, from females with skewed inactivation and we have previously presented an independent RT-PCR-based approach examining transcribed polymorphisms using cDNA from cloned human cell lines.¹¹

Although much useful data have been generated by the above techniques, they are labour intensive and/or generally applicable only to a restricted number of loci. The majority of available data reflects inactivation status in established cell lines or somatic cell hybrids, and thus may not reflect the situation *in vivo*. Recourse to microarray 'chip technology' enables the screening of all X-linked genes for which sufficient sequence data are available. We have therefore investigated the potential of this system employing commercially available microarrays to assess quantitatively the comparative expression of X-linked genes in female lymphocytes to that in males. Sudbrak *et al*¹² were the first to demonstrate the value of the microarray approach in such studies, employing X-linked EST clone inserts to interrogate transcripts from lymphoblastoid cell lines.

Materials and methods

Lymphocytes, RNA extraction and cDNA synthesis

Six male and six female volunteers provided 20 ml blood samples and lymphocytes were isolated using Lymphoprep (Axis-Shield-PoC AS, Oslo). Total RNA was isolated subsequently with RNA-Bee (AMS Biotechnology). RNA was treated with DNase I (Qiagen, Crawley, UK) to avoid DNA contamination. The quality and purity of total RNA was assayed in a 2% agarose gel and the recovery was calculated

after measuring absorbance with a spectrophotometer at 260 nm.

Biotin labelled cRNA was prepared from 5 µg of each total RNA, employing the manufacturer's reagents and protocols with the experiments being conducted in Affymetrix Laboratories, High Wycombe, England. cRNA products were subjected to quality control steps as follows: after spectrophotometric quantification they were hybridised to 'Test 3 arrays' providing information on 3' compared to 5' signal ratios for the human housekeeping genes. Each sample was hybridised to a single HG-U133A chip and processed using the standard protocol EukGE-WS2v4. The 12 chips were all scanned with the same scanner. Quality indicators from these hybridisations are summarised in Table 1. On this basis, it was decided to eliminate female sample 3 and, to balance the numbers of male and female samples, male 7 was also excluded. In the remaining 10 samples, the range of scale factors was 0.85–2.1. The intensities were analysed with the Affymetrix Microarray Suite (MAS) version 5.0 'statistical algorithm', and the resulting signal values were scaled to give a trimmed mean of 100 for each chip. These values were then imported into the Affymetrix Data mining tool and also the R statistical environment. The HG-U133A microarray displays sequences corresponding to approximately 18 400 different transcripts and 14 500 known genes, of which 772 are known X chromosome probe sets. Male and female samples were tested for difference using the Mann–Whitney test.

Results

Employing the Mann–Whitney test provided with the Affymetrix Data Mining Tool, a total of 36 probes with a female to male ratio above 1 ($P < 0.05$) were identified. Although this two-tailed nonparametric test is not corrected for multiple testing, it is otherwise extremely conservative. The overall female/male ratio for all sequences represented, based on five microarrays for each

Table 1 Quality control data for the 12 cRNA samples

Sample	cRNA yield (µg)	Scale factor	Background	Noise	No. present (%)	GAPDH	Actin
1	28.8	0.854	110.28	9.68	37.7	1	1.12
2	43.44	1.288	68.59	3.63	40	0.91	1.2
3	8.96	3.913	57.94	2.19	27.40	1.33	3.06
4	29.08	1.1	60.54	2.82	43.60	0.94	1.13
5	21.76	1.215	64.03	3.47	41.10	0.88	1.17
6	13.2	1.561	65.06	3.78	40.20	1.14	1.2
7	17.16	1.87	71.98	3.9	34	1.1	1.68
8	52.4	1.036	38.37	3.43	46.30	1.1	1.1
9	56.4	1.037	101.16	4.7	42.10	0.99	1.1
10	48.72	1.7	87.37	4.61	40.10	0.94	1.1
11	29.2	1.216	91.13	7.65	36.80	0.95	0.9
12	16.56	2.1	60.41	2.95	33.30	1.1	1.7

Quality control parameters are as follows: GAPDH and actin refer to the ratios of 3'/5' representation of cRNA sequences obtained following amplification.

sex, was 0.91. As an additional check on the validity of the identification process, means and standard deviations were calculated for the standardised signal outputs for the five female and five male samples for loci identified as giving significantly higher female:male ratios. These data together with information on chromosomal location (in Mb from Xpter) and probe derivation (Unigene accession details) are provided in Table 2. All probe sets represented on the microarray are listed and those with significant values are in bold.

Examination of those loci having evidence of escape from inactivation revealed four classes. The first is represented by genes with well-established records as escapees, the second by those for which previous weak or preliminary evidence for their escape is available, the third by loci for which previous evidence suggested normal inactivation and the fourth by candidates with no previous information available concerning their inactivation status.

Loci with substantial pre-existing data supporting their escape from inactivation

X-linked genes that have active Y homologues

These are discussed sequentially in reported cytogenetic (Xpter–Xqter) order below.

Zinc-finger protein, X-linked, *ZFX*, (Xp22.2–p21.3) has a similar gene organisation and transcription profile to that of *ZFY* and escapes X inactivation.^{9,12–15}

DEAD/H Box 3, X-linked, *DDX3*, (Xp11.3–p11.23) is one of a group of five X-linked loci with Y homologues in the nonpseudautosomal region, which escape inactivation.^{16,17}

Ubiquitously transcribed tetratricopeptide repeat gene on X chromosome, *UTX*, (Xp11.2) is homologous to the murine X-linked locus *Utx*. Both are reported to escape inactivation.^{18,19} Carrel *et al*⁵ observed the expression of *UTX* from inactive X chromosomes in all six hybrid cell lines tested. (see: <http://genetics.gene.cwru.edu/willard/data/xin/genesurvey-all.xls> for information on hybrid analysis).

X-linked genes that have inactive Y homologues

Arylsulphatase C, *ARSC2*, comprises two microsomal isoenzyme forms, s and f, with distinct transcripts. The coding sequences for both are localised at Xpter–p22.32, close to PAR1 and escape partially from inactivation.²⁰ The extent to which the locus escapes inactivation may vary in different tissues with an overall ratio of female to male activity of 1.6, and ratios for peripheral blood white cells being particularly low.²¹ In the current examination, we obtained weak evidence for its escape from inactivation with one probe set indicating a female to male ratio of 1.33 with a *P*-value of 0.075.

X-linked genes with existing evidence for escape from inactivation and that have no Y homologues

X-linked loci escaping inactivation for which there are no apparent Y-linked homologues are represented in Table 2 by the eight loci detailed below. The eukaryotic translation initiation factor 2, gamma, *EIF2S3*, (Xp22.2–p22.1) unlike the mouse locus, apparently does not have a Y homologue; nevertheless, it appears to escape inactivation.^{22,23,5} Sedlin, *SEDL*, (Xp22.2–p22.1), which is mutated in spondyloepiphyseal dysplasia tarda, also escapes inactivation along with its closest flanking partners.²⁴ Carrel *et al*⁵ reported its expression in all nine inactive-X-only hybrids tested. A similar result was reported for the gene encoding the cofactor required for SP1 transcriptional activation, subunit 2, *CRSP2*, (Xp11.4–p11.2), also referred to as *TRAP170*.^{25,26} Ubiquilin 2, *UBQLN2*, was localised to Xp11.23–p11.1 by Kaye *et al*²⁷ and reported by Carrel *et al*⁵ to escape from inactivation in two of nine hybrid cell lines. Another gene, structural maintenance of chromosomes 1-like, *SMC1L1* is localised to a similar region (Xp11.22–p11.21) and reported to escape inactivation.⁵ On the long arm, the locus for armadillo repeat protein, *ALEX2*, (Xq21.33–q22.2) was found to escape inactivation in two of nine somatic cell hybrids tested.⁵ Also located on the long arm (Xq22) is the locus for collagen, type IV, alpha-6, *COL4A6*, which is mutated in some patients with Alport's syndrome.^{28,29} Higher female expression for *COL4A6* was detected in two probe sets in the current analysis and it was found to escape inactivation in three of nine somatic hybrids examined by Carrel *et al*⁵ The same approach suggests that the *Homo sapiens* hypothetical protein, *FLJ21174*, localised to Xq22.1, escapes inactivation in some (two of nine) hybrid backgrounds.

Loci with existing weak, conflicting or circumstantial evidence for escapee status

Montini *et al*³⁰ identified a human cDNA, subsequently attributed to the locus, Sex comb on midleg, *Drosophila*, homologue-like 2, *SCML2*. Although Carrel *et al*⁵ failed to detect significant expression in the hybrid system, it is a possible escapee candidate by virtue of its chromosomal localisation at Xp22 in a region containing the *MIC2* gene family and the sulphatase gene cluster. A similar argument applies to the gene encoding transducin-beta-like 1, X-linked, *TBL1X*, located in Xp22.3.³¹ The Chloride channel 4, *CLCN4*, locus is also located at Xp22.3.³² Interestingly, in *Mus spretus* the locus is X-linked, but autosomal (chromosome 7) in C57BL/J6, suggesting a recent evolutionary X rearrangement close to the pseudoautosomal boundary consistent with the gene's possible escapee status. The Rho guanine nucleotide exchange factor 6, *ARHGEF6*, has been identified at the X breakpoint in a mother and son having an apparently balanced X:21 translocation (tXq26:21p11). Notably, wild-type mRNA for the locus was detected in the mother, consistent with

Table 2 Comparison of male/female ratios of expression

Probeset	Unigene	Gene symbol	Location	Female		Male		P-value	F/M ratio
				Mean	SD	Mean	SD		
203769_s_at	Hs.79876	STS(ARAS)	6.60	41	4.06	31	11.68	0.075	1.33
203768_s_at		STS(ARAS)	6.60	38	10.44	34	5.66	0.619	1.11
203767_s_at		STS(ARAS)	6.60	39	6.84	40	7.80	0.921	0.97
201867_s_at	Hs.76536	TBL1	9.03	105	12.31	87	6.95	0.044	1.21
213400_s_at		TBL1	9.03	185	17.29	194	45.95	0.484	0.95
201868_s_at		TBL1	9.03	47	5.86	34	25.58	0.619	1.38
201869_s_at		TBL1	9.03	51	21.42	57	38.38	0.766	0.89
213401_s_at		TBL1	9.03	19	4.83	18	14.57	1.000	1.04
205148_s_at	Hs.199250	CLCN4	9.54	22	18.48	6	1.79	0.024	3.74
219351_at	HSSEDL6	SEDL	13.09	189	12.97	134	35.48	0.024	1.41
208174_x_at	Hs.171909	U2AF1RS2	15.17	181	35.11	131	39.66	0.044	1.38
213876_x_at		U2AF1RS2	15.17	165	30.37	136	28.22	0.123	1.22
220873_at	Hs.283049	PRO0386	16.51	36	5.03	28	7.32	0.044	1.28
206147_x_at	Hs.171558	SCML2	17.62	75	11.14	43	7.14	0.004	1.75
205321_at	Hs.211539	EIF2S3	23.43	298	80.92	182	53.00	0.024	1.63
214678_x_at	Hs.2074	ZFX	23.55	57	15.39	21	10.72	0.004	2.74
207920_x_at		ZFX	23.55	49	7.87	32	14.00	0.044	1.54
217176_s_at		ZFX	23.55	18	5.68	11	4.48	0.044	1.74
203923_s_at	Hs.88974	CGD	36.67	589	89.23	475	49.09	0.044	1.24
217431_x_at		CGD	36.67	17	7.03	22	4.09	0.075	0.78
203922_s_at		CGD	36.67	315	46.00	309	68.43	0.619	1.02
218251_at	Hs.296433	STRAIT11499	37.59	162	9.75	104	11.43	0.004	1.56
202611_s_at	Hs.21586	CRSP2	39.54	47	7.94	27	14.73	0.012	1.77
215167_at		CRSP2	39.54	55	17.79	27	14.42	0.044	2.04
202610_s_at		CRSP2	39.54	65	22.23	55	14.88	0.361	1.19
202612_s_at		CRSP2	39.54	17	10.62	17	6.29	0.619	1.02
217120_s_at		CRSP2	39.54	51	17.37	55	26.01	0.921	0.93
201210_at	Hs.147916	DDX3	40.22	928	94.00	647	118.19	0.004	1.43
212514_x_		DDX3	40.22	209	22.25	124	41.63	0.004	1.68
212515_s_at		DDX3	40.22	364	48.92	205	102.26	0.024	1.78
201211_s_at		DDX3	40.22	105	35.22	77	18.94	0.123	1.37
207617_at		DDX3	40.22	12	6.69	8	6.20	0.365	1.42
203990_s_at	Hs.13980	UTX	43.78	65	6.78	44	5.86	0.004	1.48
203992_s_at		UTX	43.78	116	11.35	65	14.77	0.004	1.78
203991_s_at		UTX	43.78	51	9.41	29	14.19	0.024	1.74
213200_at	Hs.75667	SYP	48.07	32	14.13	10	3.17	0.012	3.13
217349_s_at		SYP	48.07	18	12.05	16	5.37	0.921	1.14
207739_s_at	Hs.278444	GAGE5	48.24	20	5.13	14	4.23	0.044	1.43
201589_at	Hs.211602	SMC1L1	53.37	291	27.50	194	37.19	0.004	1.50
215884_s_at	Hs.4552	UBQLN2	55.56	324	38.48	269	44.09	0.044	1.21
212727_at	Hs.11101	KIAA1232	68.43	65	17.00	42	16.40	0.044	1.53
212729_at		KIAA1232	68.43	63	16.53	56	10.58	0.365	1.14
212728_at		KIAA1232	68.43	10	1.82	10	1.77	0.615	0.97
219650_at	Hs.89306	FLJ20105	70.29	8	3.41	4	1.77	0.024	2.26

Table 2 continued

Probeset	Unigene	Gene symbol	Location	Female		Male		P-value	F/M ratio
				Mean	SD	Mean	SD		
208859_s_at	Hs.96264	ATRX	75.52	135	49.24	85	16.63	0.044	1.59
208860_s_at		ATRX	75.52	131	35.33	101	22.84	0.187	1.30
208861_s_at		ATRX	75.52	239	17.15	201	64.62	0.187	1.19
211022_s_at		ATRX	75.52	43	13.74	37	13.91	0.266	1.18
214615_at	Hs.296433	P2Y10	76.96	44	5.87	34	6.42	0.044	1.29
216809_at	Hs.307358	CYLC1	81.89	52	10.37	38	7.28	0.024	1.38
216779_at		CYLC1	81.89	34	9.02	46	11.31	0.044	0.73
216778_s_at		CYLC1	81.89	11	2.48	13	12.54	0.841	0.91
201132_at	Hs.278857	HNRPH2	99.43	236	32.35	168	23.13	0.004	1.40
203404_at	Hs.48924	ALEX2	99.58	43	9.05	28	14.29	0.075	1.55
211276_at	Hs.17481	my048	100.05	13	2.81	6	2.16	0.004	1.99
219335_at	Hs.119699	FLJ12969	100.53	58	10.36	39	5.41	0.004	1.49
202371_at	Hs.194329	FLJ21174	101.50	66	9.70	46	13.14	0.024	1.45
210945_at	Hs.408	COL4A6	106.17	20	2.21	12	5.32	0.024	1.67
213115_at		COL4A6	106.17	87	9.90	109	16.98	0.024	0.80
213992_at		COL4A6	106.17	46	10.04	37	17.53	0.619	1.22
217343_at		COL4A6	106.17	13	4.55	13	4.68	0.921	1.02
212414_s_at	Hs.90998	SEPT6	117.40	480	61.15	321	83.46	0.004	1.50
214298_x_at		SEPT6	117.40	223	42.46	154	27.87	0.024	1.45
212413_at		SEPT6	117.40	143	21.84	113	12.75	0.044	1.26
213666_at		SEPT6	117.40	131	42.97	98	22.70	0.123	1.34
212415_at		SEPT6	117.40	216	54.71	164	53.08	0.187	1.32
218499_at	Hs.23643	MST4	129.76	279	52.41	210	52.35	0.044	1.33
218668_s_at	Hs.225979	LOC57826	129.95	130	15.25	90	25.00	0.024	1.45
218669_at		LOC57826	129.95	149	9.54	152	41.97	0.619	0.98
219626_at	Hs.24078	FLJ12649	134.00	89	19.54	49	19.39	0.012	1.79
209539_at	Hs.79307	ARHGEF6	134.46	321	23.43	289	10.34	0.004	1.11
214633_at	Hs.157429	SOX3	138.29	23	7.95	7	4.12	0.012	3.38
216592_at	Hs.247705	MAGEC1	139.69	13	4.04	5	1.73	0.012	2.73
206609_at		MAGEC1	139.69	26	7.72	22	9.59	0.766	1.18
206216_at	Hs.104865	STK23	151.51	64	10.67	34	17.49	0.024	1.86

Expression level data for recognised X-linked genes are detected in the screen for X loci escaping inactivation. Where multiple probe sets correspond to a given locus, values indicating those with significantly higher female expression are in bold. Male and female values represent the mean of five observations. SD = standard deviation and F/M ratio = mean female expression/mean male expression. The values presented are expressed relative to an arbitrary target intensity (TGT) of 100, averaging data across the whole microarray. The P-values represent the probability of observing the observed deviation from equality of male–female expression by chance. Locations are given in Mb from Xpter. (Note the pter–qter order may differ from those derived from reported cytogenetic positions).

her normal phenotype,^{33,34} which is contrary to the normal expectation that genes on the intact X in translocation carrier females are inactive. Somatic cell hybrid studies, nevertheless, suggest that *ARHGEF6* is normally inactivated. In our analysis, the observed female/male ratio (1.11) suggested only very partial escape,

but with a highly significant P-value (0.004). SRY-related HMG-box gene 3, *SOX3*, (Xq26–q27) is homologous to the *SRY* locus on the Y chromosome and its conservation in marsupials suggests that *SOX3* represents the ancestral gene from which *SRY* was derived; it may therefore have an unusual inactivation profile.^{35,36}

Loci for which previous evidence suggested normal inactivation

Earlier reports concerning the following loci detected in our current screen failed to provide evidence of significant escape from inactivation: chronic granulomatous disease, *CGD*, (Xp21.1),^{37,38} alpha thalassemia/mental retardation syndrome, X-linked, *ATRX*, (Xq13), which results from mutations in the helicase 2/RAD54 gene^{5,39,40} and septin 6, *SEPT6*.⁵ Ste20-like kinase, *MST4*, (Xq26) also appeared to be regularly inactivated in studies on X-only-somatic cell hybrids.⁴¹

Established loci lacking evidence either for, or against, inactivation

These include U2 small nuclear ribonucleoprotein auxiliary factor, small subunit 2, *U2AF1RS2*, (Xp22.1) a human homologue of the imprinted mouse gene *U2af1rs1*,⁴² synaptophysin, *SYP*, (Xp11.23–p11.22) heterogeneous ribonuclear protein H2, *HNRPH2*, (Xq22), cylicin 1, *CYCL1*, (Xq21.1), the melanoma antigen – family C, *MAGEC1*, (Xp26) and the serine/threonine kinase 23, *STK23*, (Xq28).

Finally, there are several poorly characterised X-linked loci, which have been identified in this screen as potential escapees. These are: the loci encoding the *Homo sapiens* hypothetical proteins, PRO0386, (Xp22.31); STRAIT 11499, (Xp11.4); G antigen 5, *GAGE5/7*, (Xp11.4–p11.2); *FLJ20105*, (Xq12); the putative purinergic receptor, P2Y10, (Xq21.1); *FLJ2969*, (Xq22.1–q22.3); the brain my048 protein, locus also at Xq22.1–q22.3; the *Homo sapiens* hypothetical proteins, *LOC5726*, at (Xq25) and *FLJ12649*, (Xq26.3) and the *Homo sapiens* protein, *KIAA1232*, for which no subregional localisation has been reported.

(Unigene accession numbers for these and other loci without literature references are provided in Table 2).

Discussion

Two of the most common previously employed strategies to assess inactivation status are to analyse methylation profiles at, or adjacent to, the locus of interest and to examine the expression of transcripts from the inactive X chromosome isolated in somatic cell hybrids employing RT-PCR. The former is based on the assumption that methylation is associated with inactive status, but does not assay expression directly and cannot provide information on the extent to which loci might escape the inactivation process. The latter is highly adaptable; however, it can only relate to the inactivation status in a cultured somatic cell hybrid background.^{2,5,6} Various attempts to overcome both of these limitations have been made. The recent availability of reliable microarray technology provides a powerful approach to examine transcription profiles and advances in both the preparation of arrays and in the software available for the analysis of signal intensity, have

brought the realistic possibility of detecting fractional rather than fold differences in expression. This in turn provides an opportunity to examine subtle sex differences in gene expression and, in this investigation, we have analysed the relative expression of X-linked loci to evaluate the potential of the approach to investigate inactivation status.

The selection of lymphocytes as the tissue of choice was dictated by ease of access and in lymphoblastoid cell lines derived from them, about 50% of the 5184 genes represented on the microarrays used in one study were expressed at sufficient levels for analysis.⁴³ Although such expression studies indicated some natural variation in human gene expression, with some familial aggregation – suggesting the existence of high and low expressing alleles – our investigations reported only results for which there were consistent patterns of expression differentials between the five male and female replicates. Although possible cross-hybridisation to Y-related sequences may be difficult formally to eliminate in some cases (and would in any case tend to reduce female/male ratios), the detection of loci for which there was previous evidence of their escapee status and for which no Y homologues have been reported supports the general potential of this approach. Indeed, given the reproducibility of signals observed and that in several examples the same loci have been identified by more than one probe set, we have not attempted replication by other techniques, which are not directly comparable and unlikely to detect subtle differences in expression levels; but rather have provided an overall evaluation of the method's potential by comparisons with the extensive data already available in the literature.

Sudbrak *et al*,¹² employed cDNAs bound to microarrays to interrogate transcripts from lymphoblastoid lines. Their overall pattern of observation was similar to those reported here, with about 4% escapees – equally distributed between short and long arms, and with many of the same loci being identified. Clearly, the microarray approach will enable an efficient strategy to obtain a better understanding of the tissue and individual differences in inactivation profiles as more studies are completed.

The results obtained in this investigation have been extremely encouraging in that, of the 20 well-studied loci which we identified as having significantly higher female expression levels, 11 had previous supporting evidence for their escape from inactivation and five had provisional, or circumstantial, evidence. Only four characterised loci had previous contrary evidence to the conclusions reached in our study. The remaining loci, the majority of which encoded hypothetical proteins, had no associated information concerning their inactivation status.

We also examined the data for some genes that were not detected in the screen for X loci escaping inactivation, but for which previous data supported their inactivation status as either escapees, or as normally inactivated.

Table 3 Expression level data for recognised X-linked genes believed to escape inactivation, but which were not detected by microarray analysis

	Probe ID	GENE	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Mean	SD	Female/male ratio
Female	208823_s_at	PCTK1	134	128	47	112	68	97.8	38.3	
Male	208823_s_at		90	80	53	81	134	87.3	29.3	1.12
Female	200964_at	UBE1	423	476	369	304	549	424.3	94.3	
Male	200964_at		282	374	480	339	235	342.0	93.5	1.24
Female	200797_at	PDHA1	1814	1911	1980	1579	1809	1818.6	151.7	
Male	200797_at		1228	2337	1912	1474	1221	1634.4	482.8	1.11
Female	200980_s_at	PDHA1	186	208	125	167	178	172.8	30.7	
Male	200980_s_at		181	144	188	206	158	175.4	24.6	0.99
Female	207660_at	DMD	43	65	42	46	39	47.0	10.4	
Male	207660_at		71	33	50	56	57	53.4	13.8	0.88

Expression level data for recognised X-linked genes, which did not reach the criteria in the screen expected for X loci escaping inactivation, but for which previous data support their inactivation status as either escapees or as normally inactivated. *PDHA1* and *DMD* are included for comparison, as representing loci known to be regularly inactivated (eg see text and Carrel *et al*⁵).

Website references: <http://genetics.gene.cwru.edu/willard/data/xin/genesurvey-all.xls>

Unfortunately, two well-established escapee loci, *MIC2* (cell surface antigen MIC2) and *RPS4X* (ribosomal protein S4) were not represented on the microarray. PCTAIRE-1, cdc2-related protein kinase (*PCTK1*) and ubiquitin-activating enzyme (*UBE1*) were represented and have previously been reported as escaping inactivation^{5,6} although for the latter there have been conflicting reports concerning its inactivation status.^{4,44} Our data indicate that *UBE1* is expressed at significant levels in lymphocytes and that there appears to be an excess of female transcripts (see Table 3), which may reflect partial escape from inactivation. *PCTK1* also shows a slight excess of female transcripts; however, as for *UBE1*, the standard deviation intervals overlap and the interpretation concerning their inactivation status remains tentative. In contrast, pyruvate dehydrogenase subunit A1, *PDHA1*, and the Duchenne muscular dystrophy locus, *DMD*, are known to be regularly inactivated and the ratios we observed for both were consistent with a normal inactivation pattern.

Several loci detected in the screen had female: male ratios of >2; however, these all had low absolute detection levels viz *ZFX*, *SYP*, *SOX3* and *CLCN4* (with values c. <10) and the data for these should be regarded as preliminary, even though the reproducibility of their expression levels in the two sexes was robust. Clearly, the number of loci for which we have provisional information for their escape from inactivation based on the microarray approach is considerably below the 10–20% figure predicted from some other studies.^{5,6} Among the factors contributing to this discrepancy are that the reliability of any method depends both on absolute expression levels and the extent to which escape from inactivation occurs. Many loci represented on the microarrays are expressed only at low levels in lymphocytes and their inactivation profiles may not be representative of other tissues. In addition, there may be differences in patterns of inactivation between lymphocytes and cell lines and somatic cell hybrids; the latter, in particular, may not reflect the situation *in vivo*.

Naturally, there are also sex differences in expression for non-X-linked genes. Four probesets gave female to male signal ratios of <0.1: 206700_s_at, 205000_at, 201909_at, and 204409_s_at. These have uncorrected *t*-test *P*-values ranging from 0.0002 to 0.02 and they are all unique Y chromosome transcripts. At the other extreme, one probe set presents with a ratio of >10 and corresponds to the HLA class II DRB4 gene with a common null allele, for which all but two of the female subjects are homozygous. These data indicate the general robustness of the approach. Overall, employing the rudimentary criterion of an uncorrected *t*-test *P*-value of 0.1, X chromosome transcripts show a 3.2-fold excess for female to male high expression, compared with a 2.4 excess for the genome as a whole.

Overall, therefore, we believe that the approach described here will have an important role in the investigation of inactivation patterns in a variety of tissues including the brain and may lead to insights into possible sex differences in behaviour that may result from such dosage differentials.

Acknowledgements

We acknowledge the support of Professors Peter McGuffin and Robert Plomin in facilitating this investigation, which could not have taken place without the active cooperation of Deborah Keenan, Claire Holdaway and Geoff Scopes, Affymetrix UK Ltd, High Wycombe, Berks.

References

- 1 Lyon MF: Some milestones in the history of X-chromosome inactivation. *Annu Rev Genet* 1992; 26: 16–28.
- 2 Distche CM: Escape from X inactivation in human and mouse. *Trends Genet* 1995; 11: 17–22.
- 3 Miller AP, Gustashaw K, Wolff DJ *et al*: Three genes that escape X chromosome inactivation are clustered within a 6 Mb YAC contig and STS map in Xp11.21–p11.22. *Hum Mol Genet* 1995; 4: 731–739.

- 4 Carrel L, Clemson CM, Dunn JM *et al*: X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse. *Hum Mol Genet* 1996; 5: 391–401.
- 5 Carrel L, Cottle AA, Goglin KC, Willard HF: A first-generation X-inactivation profile of the human X chromosome. *Proc Natl Acad Sci USA* 1999; 96: 14440–14444, plus updated data: <http://genetics.gene.cwru.edu/willard/data/xin/genesurvey-all.xls>.
- 6 Disteche CM: Escapees on the X chromosome. *Proc Natl Acad Sci USA* 1999; 96: 14180–14182.
- 7 Skuse DH, James RS, Bishop DV *et al*: Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature* 1997; 387: 705–708.
- 8 Brown CJ, Carrel L, Willard HF: Expression of genes from the human active and inactive X chromosomes. *Am J Hum Genet* 1997; 60: 1333–1343.
- 9 Schneider-Gadicke A, Beer-Romero P, Brown LG, Nussbaum R, Page DC: ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation. *Cell* 1989; 57: 1247–1258.
- 10 Fisher EM, Beer-Romero P, Brown LG *et al*: Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* 1990; 63: 1205–1218.
- 11 Benjamin D, Van Bakel I, Craig IW: A novel expression based approach for assessing the inactivation status of human X-linked genes. *Eur J Hum Genet* 2000; 8: 103–108.
- 12 Sudbrak R *et al*: X chromosome-specific cDNA arrays: identification of genes that escape inactivation and other applications. *Hum Mol Genet* 2001; 10: 77–83.
- 13 Page DC, Mosher R, Simpson EM *et al*: The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* 1987; 51: 1091–1104.
- 14 Page DC, Disteche CM, Simpson EM *et al*: Chromosomal localization of ZFX – a human gene that escapes X inactivation – and its murine homologs. *Genomics* 1990; 7: 37–46.
- 15 Palmer MS, Berta P, Sinclair AH, Pym B, Goodfellow PN: Comparison of human ZFY and ZFX transcripts. *Proc Natl Acad Sci USA* 1990; 87: 1681–1685.
- 16 Chung J, Lee S-G, Song K: Identification of a human homolog of a putative RNA helicase gene (mDEAD3) expressed in mouse erythroid cells. *Kor J Biochem* 1995; 27: 193–197.
- 17 Park SH, Lee SG, Kim Y, Song K: Assignment of a human putative RNA helicase gene, DDX3, to human X chromosome bands p11.3→p11.23. *Cytogenet Cell Genet* 1998; 81: 178–179.
- 18 Greenfield A, Scott D, Pennisi D *et al*: An H-YDb epitope is encoded by a novel mouse Y chromosome gene. *Nat Genet* 1996; 14: 474–478.
- 19 Greenfield A, Carrel L, Pennisi D *et al*: The UTX gene escapes X inactivation in mice and humans. *Hum Mol Genet* 1998; 7: 737–742.
- 20 Chang PL, Mueller OT, Lafrenie RM *et al*: The human arylsulfatase-C isoenzymes: two distinct genes that escape from X inactivation. *Am J Hum Genet* 1990; 46: 729–737.
- 21 Craig IW, Tolley E: Steroid sulphatase and the conservation of mammalian X chromosomes. *Trends Genet* 1986; 2: 201–204.
- 22 Gaspar NJ, Kinzy TG, Scherer BJ, Humbelin M, Hershey JW, Merrick WC: Translation initiation factor eIF-2. Cloning and expression of the human cDNA encoding the gamma-subunit. *J Biol Chem* 1994; 269: 3415–3422.
- 23 Ehrmann IE, Ellis PS, Mazyrat S *et al*: Characterization of genes encoding translation initiation factor eIF-2gamma in mouse and human: sex chromosome localization, escape from X-inactivation and evolution. *Hum Mol Genet* 1998; 7: 1725–1737.
- 24 Tiller GE, Hannig VL, Dozier D *et al*: A recurrent RNA-splicing mutation in the SEDL gene causes X-linked spondyloepiphyseal dysplasia tarda. *Am J Hum Genet* 2001; 68: 1398–1407.
- 25 Gu W, Malik S, Ito M *et al*: A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol Cell* 1999; 3: 97–108.
- 26 Yoshikawa H, Fujiyama A, Nakai K, Inazawa J, Matsubara K: Detection and isolation of a novel human gene located on Xp11.2–p11.4 that escapes X-inactivation using a two-dimensional DNA mapping method. *Genomics* 1998; 49: 237–246.
- 27 Kaye FJ, Shows TB: Assignment of ubiquilin2 (UBQLN2) to human chromosome Xp11.23–Xp11.1 by GeneBridge radiation hybrids. *Cytogenet Cell Genet* 2000; 89: 116–117.
- 28 Zhou J, Mochizuki T, Smeets H *et al*: Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in inherited smooth muscle tumors. *Science* 1993; 261: 1167–1169.
- 29 Oohashi T, Sugimoto M, Mattei MG, Ninomiya Y: Identification of a new collagen IV chain, alpha 6(IV), by cDNA isolation and assignment of the gene to chromosome Xq22, which is the same locus for COL4A5. *J Biol Chem* 1994; 269: 7520–7526.
- 30 Montini E, Buchner G, Spalluto C *et al*: Identification of SCML2, a second human gene homologous to the Drosophila sex comb on midleg (Scm): a new gene cluster on Xp22. *Genomics* 1999; 58: 65–72.
- 31 Bassi MT, Ramesar RS, Caciotti B *et al*: X-linked late-onset sensorineural deafness caused by a deletion involving OA1 and a novel gene containing WD-40 repeats. *Am J Hum Genet* 1999; 64: 1604–1616.
- 32 van Slegtenhorst MA, Bassi MT, Borsani G *et al*: A gene from the Xp22.3 region shares homology with voltage-gated chloride channels. *Hum Mol Genet* 1994; 3: 547–552.
- 33 Manser E, Loo TH, Koh CG *et al*: PAK kinases are directly coupled to the PIX family of nucleotide exchange factors. *Mol Cell* 1998; 1: 183–192.
- 34 Kutsche K, Yntema H, Brandt A *et al*: Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat Genet* 2000; 26: 247–250.
- 35 Stevanovic M, Lovell-Badge R, Collignon J, Goodfellow PN: SOX3 is an X-linked gene related to SRY. *Hum Mol Genet* 1993; 2: 2013–2018.
- 36 Foster JW, Graves JA: An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proc Natl Acad Sci USA* 1994; 91: 1927–1931.
- 37 Royer-Pokora B, Kunkel LM, Monaco AP *et al*: Cloning the gene for an inherited human disorder – chronic granulomatous disease – on the basis of its chromosomal location. *Nature* 1986; 322: 32–38.
- 38 Windhorst DB, Holmes B, Good RA: A newly defined X-linked trait in man with demonstration of the Lyon effect in carrier females. *Lancet* 1967; 1: 737–739.
- 39 Gibbons RJ, Suthers GK, Wilkie AO, Buckle VJ, Higgs DR: X-linked alpha-thalassemia/mental retardation (ATR-X) syndrome: localization to Xq12–q21.31 by X inactivation and linkage analysis. *Am J Hum Genet* 1992; 51: 1136–1149.
- 40 Gibbons RJ, Picketts DJ, Villard L, Higgs DR: Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* 1995; 80: 837–845.
- 41 Qian Z, Lin C, Espinosa R, LeBeau M, Rosner MR: Cloning and characterization of MST4, a novel Ste20-like kinase. *J Biol Chem* 2001; 276: 22439–22445.
- 42 Kitagawa K, Wang X, Hatada I *et al*: Isolation and mapping of human homologues of an imprinted mouse gene U2af1-rs1. *Genomics* 1995; 30: 257–263.
- 43 Cheung VG, Conlin LK, Weber TM *et al*: Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* 2003; 33: 420–425.
- 44 Zacksenhaus E, Sheinin R: Molecular cloning, primary structure and expression of the human X linked A1S9 gene cDNA which complements the ts A1S9 mouse L cell defect in DNA replication. *EMBO J* 1990; 9: 2923–2929.