

Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development

Minoru S. H. Ko^{1,2,*}, John R. Kitchen¹, Xiaohong Wang¹, Tracy A. Threat¹, Xueqian Wang¹, Aki Hasegawa³, Tong Sun¹, Marija J. Grahovac^{1,2}, George J. Kargul^{1,2}, Meng K. Lim^{1,2}, YuShun Cui¹, Yuri Sano², Tetsuya Tanaka², Yuling Liang¹, Scott Mason¹, Paul D. Paonessa¹, Althea D. Sauls¹, Grace E. DePalma¹, Rana Sharara¹, Lucy B. Rowe⁴, Janan Eppig⁴, Chris Morrell⁵ and Hirofumi Doi^{3,6,*}

¹ERATO Doi Bioasymmetry Project, JST, Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48202, USA

²Developmental Genomics and Aging Section, Laboratory of Genetics, National Institute on Aging, NIH, Baltimore, MD 21224-6820, USA

³ERATO Doi Bioasymmetry Project, JST, WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-7112, Japan

⁴The Jackson Laboratory, Bar Harbor, Maine 04609, USA

⁵Statistics and Experimental Design Section, Research Resources Branch, National Institute on Aging, NIH, Baltimore, MD 21224-6820, USA

⁶Fujitsu Labs Ltd, Mihama-ku, Chiba 261-0023, Japan

*Authors for correspondence (e-mail: kom@grc.nia.nih.gov or doi@bioa.jst.go.jp)

Accepted 28 January; published on WWW 21 March 2000

SUMMARY

Little is known about gene action in the preimplantation events that initiate mammalian development. Based on cDNA collections made from each stage from egg to blastocyst, 25438 3'-ESTs were derived, and represent 9718 genes, half of them novel. Thus, a considerable fraction of mammalian genes is dedicated to embryonic expression. This study reveals profound changes in gene expression that include the transient induction of transcripts at each

stage. These results raise the possibility that development is driven by the action of a series of stage-specific expressed genes. The new genes, 798 of them placed on the mouse genetic map, provide entry points for analyses of human and mouse developmental disorders.

Key words: Mouse development, Preimplantation, cDNA analysis, EST, Stage-specific gene, Gene mapping

INTRODUCTION

Preimplantation development of mammalian embryos is marked by many critical and unique events, including the start of zygotic transcription, the first cell differentiation and the initiation of specific cell-cell adhesion (reviewed in Hogan et al., 1994; Pedersen, 1986; Rossant, 1986; Watson et al., 1992). Analysis of these processes is fundamental for the understanding of organ formation, for practical techniques such as veterinary cloning of animals (Wakayama et al., 1998; Wilmut et al., 1997), and for clinical applications such as in vitro fertilization (IVF) and the assessment of fetal well-being. In spite of its importance, very little is known about the molecular events during this early phase of development.

At a macro level, the outline of the process has become clear for several organisms. During preimplantation development, the embryo, confined within the zona pellucida, does not change in overall size. Rather, an increase in cell number is compensated by a decrease in cell size, giving rise to the terminology of 'cleavage stages'. In most species studied, including *Drosophila*, *C. elegans*, *Xenopus*, sea urchins and fish, morphological changes and cell differentiation rely at this

stage of development mainly on maternally stored mRNAs and proteins to drive the regional differentiation of embryonic cells (Wieschaus, 1996). But mammals significantly modify this program. Some oocyte mRNAs are translated, but fertilization triggers massive mRNA degradation (Piko and Clegg, 1982). Two major events then occur. One is the transcriptional activation of the zygotic genome. The timing of this transition is regulated by a 'zygotic clock' (Nothias et al., 1995; Schultz, 1993), and is somewhat species-dependent (at the late one-cell stage in mouse, 4- to 8-cell in human and 8- to 16-cell stage in sheep; Schultz, 1993). The second major event, compaction, occurs at the 8- to 16-cell stage, when cells that were previously loosely associated begin to adhere in the tightly organized cell mass of the morula. This is the starting point for cell differentiation into Inner Cell Mass (ICM), which eventually becomes the embryo, and Trophectoderm, which eventually becomes the placenta. By the 32- to 64-cell stage (blastocyst), the two cell types are clearly distinguishable. Now released from the zona pellucida, the blastocyst is implanted in the uterus.

A primary obstacle that has delayed molecular analysis of this developmental program is the difficulty of collecting and

analyzing large numbers of eggs and embryos. In early studies, expression patterns of a limited number of genes observed by several methods led to the idea that, notwithstanding the dynamic changes, gene expression is monotonous: once gene expression has begun, it is not switched off, and the encoded proteins then accumulate as development proceeds to the blastocyst stage (reviewed in Kidder, 1992; Schultz and Heyner, 1992; Watson et al., 1992). Later analyses with high-resolution two-dimensional protein gels revealed dynamic changes in quantities of many proteins during the 1- to 4-cell stages (Latham et al., 1991) and the 8-cell to blastocyst stages (Shi et al., 1994), but only a limited number of genes have been identified so far (Schultz, 1999). Similarly some attempts to construct cDNA libraries (Adjaye et al., 1997, 1998; Rothstein et al., 1992, 1993; Sasaki et al., 1998; Taylor and Piko, 1987) and examine gene expression patterns by mRNA differential display (Oh et al., 1999; Schultz, 1999) have provided only short bits of transcripts and fragmentary information.

Aiming at a global survey of gene expression and a definition of the number of genes that are preimplantation-specific, we have adapted techniques to generate cDNA libraries from each stage of preimplantation mouse embryos, carried out large-scale sequencing of cDNAs from each stage, and mapped 798 of the novel species on the mouse genome. The results support the inferences that (1) a significant fraction of the genome is dedicated to genes expressed specifically in early development, adding considerably to the nascent catalogue of mammalian genes; (2) genes coexpressed in the same stage tend to cluster in the genome; and (3) the expressed genes include cohorts acting in a stage-specific manner that may suggest a 'hit and run cascade' model for the developmental process.

MATERIALS AND METHODS

Mouse preimplantation embryo collection

Eggs and embryos were collected by standard methods (Hogan et al., 1994). C57BL/6J female mice were superovulated and mated with C57BL/6J male mice. Unfertilized eggs were collected without mating. Embryos from all the other stages were collected by killing the pregnant mice at 0.5, 1.5, 2.5 and 3.5-days post coitum (d.p.c.). Embryos were staged by visual inspection under the stereomicroscope. To avoid undesirable effects of culturing the preimplantation embryos, all the embryos up to the blastocyst stages were collected by flushing the oviduct and uterus.

Construction of stage-specific cDNA libraries

The seven cDNA libraries were constructed from each of seven stages of preimplantation development in essentially the same manner as previously described (Takahashi and Ko, 1994). The normalization and mechanical shortening of cDNA inserts steps were omitted. In brief, total RNAs were extracted from 1528 unfertilized eggs and double-stranded cDNA was synthesized by a kit (Life Technology, Superscriptase II) with an oligo(dT)*NotI* primer (5'-pGACTAGTTC TAGATCGCGAGCGGCCGCC15(T)-3') from 2.7 µg of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal3 (LL-Sal3A: 5'-pGCTATTGACGTCG-ACTATCC-3', LL-Sal3B: 5'-pGGATAGTCGACGTC AAT-3'). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high-fidelity PCR using Ex Taq polymerase (Takara) for 25 cycles

under the following conditions: denature at 94°C for 20 seconds, 25 cycles of 94°C for 10 seconds, 68°C for 10 minutes (plus 20 seconds for each additional cycle), and a final extension at 72°C for 10 minutes, on a Perkin-Elmer GeneAmp PCR system 9600. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were double-digested with *SalI* and *NotI* enzymes. Next, the cDNAs were purified by phenol/chloroform extraction and ethanol-precipitated. Then, the cDNAs were size-selected by Size Fractionation Column (Life Technology, Fraction 8 to 10). The cDNAs were ethanol-precipitated and cloned into the *SalI/NotI* site of pSPORT1 plasmid vector. The DH10B *E. coli* host was transformed with the ligation mixture by chemical methods.

The other libraries were constructed essentially in the same manner. For the fertilized egg library, double-stranded cDNA was synthesized from 5.4 µg of total RNA extracted from 1137 fertilized eggs. For the 2-cell library, double-stranded cDNA was synthesized from 1.2 µg of total RNA extracted from 397 embryos. For the 4-cell library, double-stranded cDNA was synthesized from 2.6 µg of total RNA extracted from 32 embryos. For the 8-cell library, double-stranded cDNA was synthesized from 4.3 µg of total RNA extracted from 230 embryos. For the 16-cell library, double-stranded cDNA was synthesized with an oligo(dT)GC primer 5'-pGACTAGTTC TAGATCGCG-AGCGGCCGCGC15(T)-3' from 2.1 µg of total RNA extracted from 42 embryos. For the blastocyst library, double-stranded cDNA was synthesized with an Oligo(dT)-1 primer 5'-GAGAGAGACTAG-TTCTAGATCGCGAGCGGCCCG18(T)-3' from 1.5 µg of total RNA extracted from 40 embryos.

A single-path sequencing of cDNA clones

A single-path cDNA sequencing was conducted as described (Ko et al., 1998). The 96-well microtiter plates were thawed and cDNA clones were inoculated into a 1 ml deep-well 96-microtiter plate (Beckman). Plasmid preparations from the cDNA clones were performed with Qiagen's 96-well format REAL-prep system. The plasmid DNAs were resuspended in 50 µl TE (8.0) buffer. 5 µl of DNA were used for cycle-sequencing reactions. The first 2000 Blastocyst ESTs were sequenced using standard dye primer chemistry (Perkin-Elmer-ABI). The ESTs from all other libraries, and the remaining 4000 Blastocyst ESTs, were sequenced using ET-dye primer chemistry (Amersham). All sequencing reactions were performed by an ABI Prism 877 Integrated Thermal Cycler (Perkin-Elmer-ABI).

Sequence data analyses

Clustering of 3'-EST sequences was done using the Blast2 program (Altschul et al., 1990). The criteria for identifying the unique gene set will be described elsewhere (A. H. and H. D., in preparation). In brief, all the 3'-ESTs were searched against each other for sequence similarities. For each EST, hits were sorted according to the score and the difference of scores between that EST and each of the hits were examined. Hits with a score greater than 70% of the highest score (generated by an EST's homology to itself) in each list were classified to the same group. All the ESTs below this threshold were classified to other gene sets.

Estimation of gene expression levels by EST frequency

The EST data sets from each cDNA library were subjected to Blast2 analyses against the set of 9718 unique genes. Then, the frequency of EST appearance for each gene was tabulated. The 95% confidence interval for each EST frequency in a total 3000 EST set is as follows: 0 EST matches, sample proportion 0, (0, 0.0012); 1 EST matches, sample proportion 0.0003 (0, 0.0018); 2 EST matches, sample proportion 0.007, (0.0001, 0.0023); 9 EST matches, sample proportion 0.0030, (0.0013, 0.0055). Therefore, differences among 0 matches, 1 match and 2 matches are not statistically significant. According to Fisher's exact test results, 7 EST matches in one library are required to have a statistically significant difference from 1 EST

match in another library (1-sided; $P=0.035$). Similar results were obtained by the formula that has been developed to test the significance of differential gene expression in EST/SAGE projects (Audic and Claverie, 1997; Claverie, 1999). Application of this formula to a total 3000 EST set in each cDNA library indicates a differential gene expression with a probability greater than 0.96 and less than 0.97 in the following combinations of EST matches: 5 EST matches in one library and 0 EST matches in another library, 7 EST matches in one library and 1 EST match in another library, 9 EST matches in one library and 2 EST matches in another library, 11 EST matches in one library and 3 EST matches in another library, 13 EST matches in one library and 4 EST matches in another library, and so on.

RT-PCR analyses

For each stage, 10 embryos were collected under the microscope and stored in 10 μ l BGJb Medium (Life Technology). Embryos were collected and directly lysed in 0.05% NP40. Samples were sequentially diluted in fivefold steps and subjected to RT-PCR. Reverse transcription and PCR amplification were performed using EzrTth RNA PCR kit (Perkin-Elmer) in 50 μ l reaction mixtures containing 25 mM manganese acetate, 0.2 units rTth DNA polymerase, 10 ng/ μ l primers (for the Alpha03732 gene: 5'-GTTCCAGGAGACTAAGTTTCCGTG-3', 5'-AGGCTGTCCATCAGAAAGTTGCT-3'; for the gamma-actin gene: 5'-TTCCTGCGCAGATCGCAA-3', 5'-GTGACAATGCCGTG-TTCGATAGG-3'), 10 mM dNTPs, 250 mM Bicine (pH 8.2), 575 mM potassium acetate and 10 μ l RNAs. Reactions were incubated at 60°C for 30 minutes for reverse transcription, 94°C for 1 minute for preheating, 40 cycles of PCR at 94°C for 15 seconds and 58°C for 30 seconds, followed by the final extension at 58°C for 7 minutes. The PCR products were electrophoresed on a 3% agarose gel and the gel was stained with SYBR Green. The gel was analyzed with a STORM phosphor Imager (Molecular Dynamics).

Genetic mapping of new ESTs

New ESTs were mapped on the mouse genetic map by using The Jackson Laboratory BSS Interspecific Backcross Panel (Rowe et al., 1994). PCR primer pairs were developed from approximately 350 bp of the most 3'-end of the cDNA sequences to increase the chance of having sequence polymorphisms between C57BL/6J and *M. spretus* (Takahashi and Ko, 1993). Primers were designed as a batch in a semi-automatic manner on a Sun Workstation, UNIX platform. The Unix version of PRIMER program developed by the WI/MIT Mouse Genome Center (<http://www-genome.wi.mit.edu>) was used as a core engine of our primer design program. The front end and the back end of the programs were written by our group. A total of 4500 primer pairs were developed during the course of this work. The primer pairs are available from the Research Genetics (<http://www.resgen.com/>). To test for sequence polymorphisms, genomic DNAs of C57BL/6J, *M. spretus*, and an equimolar mixture of C57BL/6J and *M. spretus*, were amplified by the each primer pair. The PCR products were run

on a customized polyacrylamide gel electrophoresis system using 10% non-denaturing polyacrylamide gels. Electrophoresis was performed for 1 hour at 250 volts. The gels were stained with ethidium bromide and photographs were taken on a UV transilluminator. Only the PCR primer pairs that exhibited heteroduplex bands were used in the gene mapping study. Approximately 1000 primer pairs fell into this category.

Assembly of the PCR reactions was performed with the Biomek1000 robotic workstation (Beckman). Genotyping of The Jackson Laboratory BSS Interspecific Backcross DNAs (94 N2 animals plus C57BL/6Jei and SPRET/Ei parental DNAs) was scored by visual inspection and analyzed by the Map Manager computer program (Manly, 1993).

Data and cDNA clones access

All cDNA clones reported in this paper are available from the American Type Culture Collection (ATCC: <http://www.atcc.org/>) or RIKEN DNA Bank (<http://www.rtc.riken.go.jp/DNA/HTML/engsearch.html> and http://www.rtc.riken.go.jp/DNA/mouse_info.html). cDNA sequence information is available through Entrez and BLAST servers at NCBI,

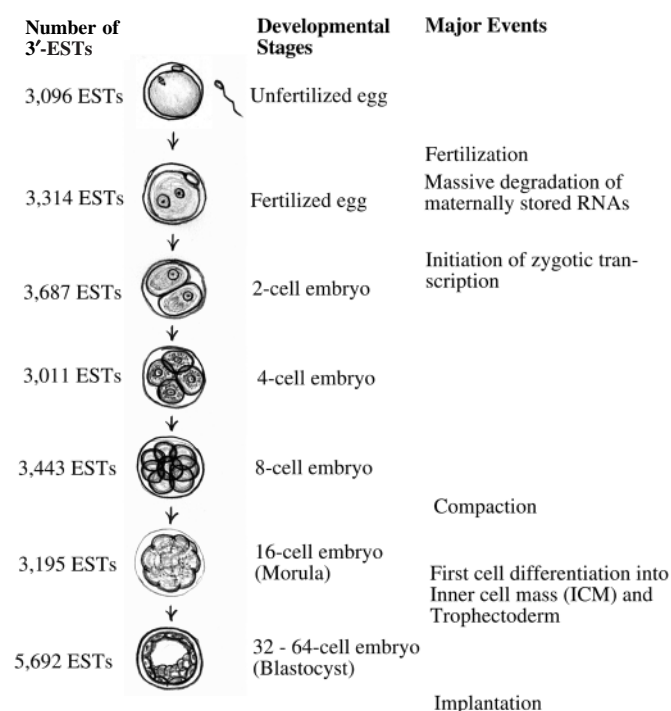


Fig. 1. Stages of preimplantation development and number of collected 3'-ESTs.

Table 1. Summary of ESTs

	Cluster number	Embryonic stage							Total
		Unfertilized egg	Fertilized egg	2-cell	4-cell	8-cell	16-cell	Blastocyst	
Classified (Alpha clusters)	9718	2823	2983	3068	2793	3194	2945	5349	23155
ESTs matched to named gene	955	266	473	432	496	579	536	994	3776
ESTs matched to other libraries	888	253	450	418	485	556	517	978	3657
ESTs not matched to other libraries	67	13	23	14	11	23	19	16	119
ESTs not matched to named gene	8763	2557	2510	2636	2297	2615	2409	4355	19379
ESTs matched to other libraries	4412	1458	1837	1824	1740	2060	1918	3012	13849
ESTs not matched to other libraries	4351	1099	673	812	557	555	491	1343	5530
Unclassified (Beta clusters)	28	273	331	619	217	249	250	340	2279
Others (Gamma clusters)	2	0	0	0	1	0	0	3	4
Total	9748	3096	3314	3687	3011	3443	3195	5692	25438

Table 2. Examples of classification of ESTs based on sequence similarity

Unique gene ID	Sequence ID	Description	Embryonic stage						
			U	F	2	4	8	16	B
Alpha03578	gb AF000982 HSAF000982	<i>Homo sapiens</i> dead box, X isoform (DBX) mRNA, alternative transcript 2	5	1	0	1	0	0	0
Alpha01399	gb U05341 IRRU05341	<i>Rattus norvegicus</i> p55CDC mRNA, complete cds	3	1	0	0	1	0	0
Alpha01806	gb AF016099 AF016099	Mouse L1 repetitive element; <i>Mus musculus</i> glycine receptor beta-subunit gene, partial cds	2	0	0	0	0	0	1
Alpha06817	gb M18252 MUSIAPL31	Mouse retrovirus-like intracistronic type A particle element DNA, clone L31	2	0	0	0	0	0	0
Alpha06441	gb U02599 MMU02599	Mouse A12 mRNA; <i>Mus musculus</i> Balb/c xlr3a mRNA, complete cds	2	1	0	0	0	0	0
Alpha02151	gb U17088 MMU17088	<i>Mus musculus</i> MT transposon-like element clone MTI6	2	1	1	0	0	0	0
Alpha03119	gb U94402 MMU94402	<i>Mus musculus</i> ubiquitin conjugating enzyme UBC9 mRNA, complete cds	3	0	0	0	0	0	1
Alpha06127	gb U72059 MMU72059	<i>Mus musculus</i> chloride channel regulator 1cIn (1cIn) pseudogene	2	0	1	0	1	0	0
Alpha01784	gb S63728 S63728	JAK1 protein=protein tyrosine kinase [mice, eye, mRNA, 4191 nt]	2	1	0	0	1	0	0
Alpha01880	emb Z19581 MMSIAH2A	<i>M. musculus</i> siah-2 protein mRNA	2	0	1	0	0	0	0
Alpha04955	emb X57413 MMTGFB2	Mouse mRNA for transforming growth factor-beta2	2	0	1	0	0	0	0
Alpha06934	gb L43326 MUSGC1R	<i>Mus musculus</i> coiled-coil protein (CG-1) mRNA, complete cds	2	0	1	0	0	0	0
Alpha01097	gb U58512 MMU58512	<i>Mus musculus</i> Rho-associated, coiled-coil forming protein kinase p160 ROCK-1 mRNA, complete cds	4	1	0	0	0	0	0
Alpha03973	gb S78271 S78271	SB1.8/DXS423E=mitosis-specific chromosome segregation protein SMC1 homolog	5	0	0	0	0	0	0
Alpha00486	gb M22995 HUKMKREV1A	Human ras-related protein (Krev-1) mRNA, complete cds	4	0	0	0	0	0	0
Alpha00110	dbj D50264 D50264	Mouse mRNA for phosphatidylinositol glycan class F, complete cds	3	0	0	0	0	0	0
Alpha02232	gb U20159 MMU20159	<i>Mus musculus</i> 76 kDa tyrosine phosphoprotein SLP-76 mRNA, complete cds	3	0	0	0	0	0	0
Alpha02292	emb Y15740 MMSV15740	Moloney murine sarcoma virus mRNA for mos gene	2	0	0	0	0	0	0
Alpha03168	gb J04022 RATATPSRA	Rat brain Ca ²⁺ -ATPase mRNA, complete cds	2	0	0	0	0	0	0
Alpha04477	gb U95116 MMU95116	<i>Mus musculus</i> lissencephaly-1 protein (LIS-1) mRNA, complete cds	2	0	0	0	0	0	0
Alpha05510	emb X60672 MMRAD	<i>M. musculus</i> mRNA for radixin	2	0	0	0	0	0	0
Alpha06107	gb M80631 MUSGNA14A	Mouse G protein alpha subunit (GNA-14) mRNA, complete cds	2	0	0	0	0	0	0
Alpha06532	emb Y08460 MMMDDES	<i>Mus musculus</i> mRNA for Mdes transmembrane protein	2	0	0	0	0	0	0
Alpha01370	gb U53456 MMU53456	<i>Mus musculus</i> protein phosphatase 1gamma (PP1gamma) mRNA; Mouse mRNA for PP1gamma	2	0	1	0	1	1	1
Alpha00724	emb X13986 MMPOINTIN	Mouse mRNA for minopontin; Murine gene for osteopontin	3	9	0	0	1	0	1
Alpha03605	gb U27323 MMU27323	<i>Mus musculus</i> Cdc25a (cdc25a) mRNA, complete cds	2	2	0	0	0	0	1
Alpha02058	gb U67187 MMU67187	<i>Mus musculus</i> G protein signaling regulator RGS2 (rgs2) mRNA, complete cds	3	3	1	0	0	0	0
Alpha00280	gb AE000664 MMAE000664	<i>Mus musculus</i> TCR beta locus; Mouse virus-like (VL30) retro-element	3	6	0	0	0	0	0
Alpha01214	gb M58567 MUSHSD3B	<i>Mus musculus</i> delta-5-3-beta-hydroxysteroid dehydrogenase/delta-5-> delta-4 isomerase (HsdCb)	2	5	0	0	0	0	0
Alpha03336	gb U58883 MMU58883	<i>Mus musculus</i> c-Cbl associated protein CAP mRNA, complete cds	2	3	0	0	0	0	0
Alpha01217	emb X62940 MMTSC22	<i>M. musculus</i> TSC-22 mRNA	3	5	4	0	0	0	0
Alpha03842	gb U57343 MMU57343	<i>Mus musculus</i> homeobox protein Meis2 mRNA, complete cds	2	4	2	0	0	0	0
Alpha04360	dbj D86728 D86728	Mouse mRNA for topoisomerase-inhibitor suppressed, complete cds	0	2	0	0	1	0	1
Alpha00707	gb M14044 MUSCALP	Mouse mRNA for protein-tyrosine kinase substrate p36 (calpactin I heavy chain), complete cds	0	5	0	1	0	0	1
Alpha03821	gb M96823 MUSNUCLEOB	Mouse nucleobindin mRNA, complete cds	0	2	0	1	0	0	1
Alpha03514	gb J03750 MUSBP9P	Mouse single stranded DNA binding protein p9 mRNA, complete cds	0	3	0	0	0	1	0
Alpha02997	gb U49350 MMU49350	<i>Mus musculus</i> CTP synthetase mRNA, complete cds	0	2	0	0	0	1	0
Alpha06386	emb X04017 MMSPARCR	Mouse mRNA for cysteine-rich glycoprotein SPARC; Mouse p2-4 mRNA for SPARC/osteonectin	0	2	0	0	0	1	0
Alpha06401	gb U96726 MMU96726	<i>Mus musculus</i> vibrator critical region, phosphatidylinositol transfer protein alpha (Pitpn)	0	2	0	1	1	0	0
Alpha03511	gb U37720 MMU37720	<i>Mus musculus</i> CDC42 mRNA, complete cds	0	4	1	0	1	0	0
Alpha03762	gb U75361 RNU75361	<i>Rattus norvegicus</i> Munc13-3 mRNA, complete cds	1	2	1	0	1	0	0
Alpha00937	emb X51438 MMVIM	Mouse mRNA for vimentin	0	4	0	0	1	0	0
Alpha01106	gb M21065 MUSIRF1B	Mouse interferon regulatory factor 1 mRNA, complete cds	0	3	0	0	1	0	0
Alpha01861	gb U39302 MMU39302	<i>Mus musculus</i> 26S proteasome subunit 4 ATPase mRNA, complete cds	0	3	0	0	1	0	0
Alpha01916	emb X67644 MMGLY96	<i>M. musculus</i> gly96 mRNA	0	6	1	1	0	0	0
Alpha04172	gb AF004107 AF004107	<i>Mus musculus</i> unknown protein mRNA, complete cds	0	2	1	1	0	0	0
Alpha01332	emb X51829 MMMDPRMR	Mouse myeloid differentiation primary response mRNA encoding MyD116 protein	0	4	0	1	0	0	0
Alpha01954	gb S72537 S72537	zebrin II [mice, C57BL/6J inbred, P20 cerebella, mRNA, 1587 nt]	0	2	0	1	0	0	0
Alpha01522	gb U16818 MMU16818	<i>Mus musculus</i> UDP glucuronosyltransferase (UGT1-06) mRNA, complete cds	0	3	1	0	0	0	0
Alpha00165	dbj D45860 MUSMDPPB4B	Mouse mRNA for magnesium dependent protein phosphatase (protein phosphatase 2C) beta-4	1	2	1	0	0	0	0
Alpha01375	dbj D12622 MMU121A	Mouse mRNA encoding unknown protein, complete cds	0	2	1	0	0	0	0
Alpha03477	gb U51907 MMU51907	<i>Mus musculus</i> TRAF family member associated NF-kappa B activator (TANK) mRNA, complete cds	0	2	1	0	0	0	0
Alpha00933	gb U32745 U32745	<i>Haemophilus influenzae</i> Rd section 60 of 163; <i>E. coli</i> tmh gene for rRNAs and tRNAs	0	3	0	0	0	0	0
Alpha01224	gb J04596 MUSSPKC	Mouse platelet-derived growth factor-inducible KC protein mRNA, complete cds	0	3	0	0	0	0	0
Alpha01260	gb M12253 CRUTUBAB	Chinese hamster alpha-tubulin II mRNA; Mouse alpha-tubulin isotype M-alpha-6 mRNA	0	3	0	0	0	0	0
Alpha02846	gb M64085 MUSSP2A	Mouse spi2 proteinase inhibitor (spi2/eb1) mRNA, 3' end	0	3	0	0	0	0	0
Alpha03366	gb S43105 S43105	Cyeb1=cyclin B1 [mice, mRNA, 2387 nt]	0	3	0	0	0	0	0
Alpha04823	gb M84361 RATCSF1A	Mouse macrophage colony-stimulating factor (4 kb) mRNA; Rat CSF-1 protein mRNA, complete cds	0	3	0	0	0	0	0
Alpha01484	gb U67187 MMU67187	<i>Mus musculus</i> G protein signaling regulator RGS2 (rgs2) mRNA, complete cds	1	2	0	0	0	0	0
Alpha03369	emb Z71173 MMIP3R2	<i>M. musculus</i> mRNA for inositol 1,4,5-trisphosphate receptor (type 2)	1	2	0	0	0	0	0
Alpha05298	gb U67874 MMU67874	<i>Mus musculus</i> fat facets homolog (Fam) mRNA, complete cds	1	2	0	0	0	0	0
Alpha00608	gb L76155 MUSMHBAT4R	<i>Mus musculus</i> Bat-4 gene, complete cds	0	2	0	0	0	0	0
Alpha01107	emb X55957 MMINAS	<i>M. musculus</i> mRNA for inhibin alpha subunit	0	2	0	0	0	0	0
Alpha01532	gb U87557 MMU87557	<i>Mus musculus</i> phosphatidylcholine-specific phospholipase D2 (mPLD2) mRNA, complete cds	0	2	0	0	0	0	0
Alpha01874	emb X71327 MMMTF1	<i>M. musculus</i> mRNA for MRE-binding transcription factor	0	2	0	0	0	0	0
Alpha02056	emb X78682 MMBAP32	<i>M. musculus</i> mRNA for B-cell receptor associated protein (BAP) 32	0	2	0	0	0	0	0
Alpha03281	gb U01063 MMU01063	<i>Mus musculus</i> pLk serine/threonine kinase mRNA; <i>Mus musculus</i> protein kinase (Plk) mRNA	0	2	0	0	0	0	0
Alpha03347	gb U13838 MMU13838	<i>Mus musculus</i> vacuolar adenosine triphosphatase subunit B gene, complete cds	0	2	0	0	0	0	0
Alpha03577	gb U38690 MMU38690	<i>Mus musculus</i> DAZ-like putative RNA binding protein mRNA; <i>Homo sapiens</i> dead box, X isoform (DBX)	0	2	0	0	0	0	0
Alpha04706	dbj D16333 MUSCPP	Mouse mRNA for coproporphyrinogen oxidase, complete cds	0	2	0	0	0	0	0
Alpha04707	gb U36757 MMTHRECO2	<i>Mus musculus</i> thrombin receptor (CF2r) gene, exon 2 and complete cds	0	2	0	0	0	0	0
Alpha04749	gb AF001688 AF001688	<i>Mus musculus</i> U4/U6 snRNP 90 kDa protein gene, complete cds	0	2	0	0	0	0	0
Alpha05109	gb AC000399 AC000399	Genomic sequence from Mouse 9, complete sequence (<i>Mus musculus</i>)	0	2	0	0	0	0	0
Alpha05121	dbj D38517 MUSDHMP1P	Mouse mRNA for Dhm1 protein, complete cds	0	2	0	0	0	0	0
Alpha06366	gb M29324 MUSL1A1	Mouse L1Md-A13 repetitive sequence	0	2	0	0	0	0	0
Alpha06891	gb U48972 MMU48972	<i>Mus musculus</i> spindlin (Spin) mRNA, complete cds	0	2	0	0	0	0	0
Alpha06987	gb U04672 MMU04672	<i>Mus musculus</i> type I receptor BRK-1 mRNA, complete cds	0	2	0	0	0	0	0
Alpha00574	emb X64070 MMCDMPR	<i>M. musculus</i> gene for cation-dependent mannose-6-phosphate receptor	0	2	3	1	1	0	1
Alpha02751	emb X14607 MM24P3	Mouse SV-40 induced 24p3 mRNA	0	2	4	1	0	0	0
Alpha02003	gb AC000398 AC000398	Genomic sequence from Mouse 11, complete sequence (<i>Mus musculus</i>)	1	2	2	1	0	0	0
Alpha02539	emb X57413 MMTGFB2	Mouse mRNA for transforming growth factor-beta2	0	3	2	0	0	0	0
Alpha04709	emb Y00711 MITOMM	Mouse mitochondrial genome	0	3	3	2	1	1	1
Alpha01875	dbj D14077 MUSGSP2	Mouse mRNA for sulfated glycoprotein-2; <i>Mus musculus</i> alpha-clustrin and beta-clustrin mRNA	1	2	3	6	1	0	0
Alpha02801	emb X64837 MMOATMR	<i>M. musculus</i> Oat mRNA for ornithine aminotransferase	0	2	2	3	0	0	0
Alpha00830	gb M20495 MUSPROL	Mouse cathepsin L gene; Mouse mRNA for major excreted protein (MEP); Mouse cysteine proteinase	0	3	7	5	6	2	1

Table 2. Continued

Unique gene ID	Sequence ID	Description	Embryonic stage						
			U	F	2	4	8	16	B
Alpha00363	emb X06406 MMML40KD	Mouse mRNA for translational controlled 40 kDa polypeptide p40; Mouse laminin receptor mRNA	0	2	5	4	7	2	19
Alpha00210	gb M88335 MUSTUMSEQA	<i>M. musculus</i> mRNA sequence; <i>R. norvegicus</i> putative v-fos transformation effector protein (Fte-1)	0	6	3	5	5	5	14
Alpha00483	gb U52822 MMU52822	<i>Mus musculus</i> ornithine decarboxylase antizyme mRNA, complete cds	1	5	2	3	3	4	3
Alpha03072	dbj D88315 D88315	Mouse mRNA for tetracycline transporter-like protein, complete cds	0	0	2	0	1	0	1
Alpha01538	emb X13605 MMH33REP	Murine mRNA for replacement variant histone H3.3	1	1	3	1	0	0	1
Alpha04396	gb U94593 MMU94593	<i>Mus musculus</i> uncoupling protein homolog (UCPH) mRNA; <i>Mus musculus</i> UCP2 mRNA	0	0	3	1	0	0	1
Alpha05193	gb U45977 MMU45977	<i>Mus musculus</i> calcium-binding protein Cab45a mRNA, complete cds	0	0	2	1	0	0	1
Alpha05193	dbj D50461 D50461	Mouse SDF4 mRNA, complete cds	0	0	2	1	0	0	1
Alpha01066	emb X90875 MMFXR1PRT	<i>M. musculus</i> mRNA for FXR1 protein	0	0	3	0	0	0	1
Alpha01085	gb S71186 S71186	XPBC/ERCC-3=DNA repair gene [mouse, mRNA, 2673 nt]	0	0	3	0	0	1	0
Alpha05436	gb L25255 MUSRANBP1	<i>Mus musculus</i> Ran/TC4 binding protein (RanBP1) mRNA	0	0	2	0	0	1	0
Alpha02606	emb AJ223794 MMU223794	<i>Mus musculus</i> CDC10 gene, exon 13	0	0	2	0	1	0	0
Alpha00697	gb AF003346 AF003346	<i>Mus musculus</i> ubiquitin-conjugating enzyme UbcM2 mRNA; <i>Homo sapiens</i> Xp22 BAC GSHB-257G1	0	0	3	1	0	0	0
Alpha01730	dbj D17571 MUSCYPOR	Mouse mRNA for NADPH-cytochrome P450 oxidoreductase, complete cds	0	0	2	1	0	0	0
Alpha03529	dbj D89063 D89063	<i>Mus musculus</i> mRNA for oligosaccharyltransferase, complete cds	0	0	2	1	0	0	0
Alpha03719	gb U62483 MMU62483	<i>Mus musculus</i> ubiquitin conjugating enzyme (ubc4) mRNA, complete cds	1	1	5	0	0	0	0
Alpha02659	dbj D01034 MUSTFIID	<i>Mus musculus</i> mRNA for TFIID; <i>Mus musculus domesticus</i> transcription factor IID (Tbp) mRNA	0	0	3	0	0	0	0
Alpha03475	gb S43105 S43105	Cybc1=cyclin B1 [mouse, mRNA, 2387 nt]	0	0	3	0	0	0	0
Alpha03879	gb U08440 MMU08440	<i>Mus musculus</i> Balb/c cytochrome c oxidase subunit VIaL mRNA, complete cds	0	0	3	0	0	0	0
Alpha05210	gb U28016 MMU28016	<i>Mus musculus</i> parathion hydrolase (phosphotriesterase)-related protein mRNA, complete cds	0	0	3	0	0	0	0
Alpha01388	gb S59342 S59342	nuclear pore complex glycoprotein p62 [mouse, mRNA, 2411 nt]	0	1	2	0	0	0	0
Alpha01653	gb U89506 MMU89506	<i>Mus musculus</i> Mlark mRNA, complete cds	0	1	2	0	0	0	0
Alpha03727	gb U47737 MMU47737	<i>Mus musculus</i> thymic shared antigen-1 (TSA-1) gene; <i>Mus musculus</i> C57BL/6 Sca-2 precursor	0	1	2	0	0	0	0
Alpha04019	gb S82156 S82156	GST-5=glutathione S-transferase-sperm antigen MSAG-5 fusion protein [3' region] [mouse, testis]	0	1	2	0	0	0	0
Alpha04841	gb M35797 MUSTCPIX	Mouse t-complex protein (Tc-p1) mRNA, 3' end	0	1	2	0	0	0	0
Alpha00121	gb L16846 MUSBTG1X	Mouse BTG1 mRNA, complete cds	0	0	2	0	0	0	0
Alpha00488	gb U31758 MMU31758	<i>Mus musculus</i> transcriptional regulator RPD3 homolog mRNA, complete cds	0	0	2	0	0	0	0
Alpha01229	gb U13262 MMU13262	<i>Mus musculus</i> myelin gene expression factor (MEF-2) mRNA, partial cds	0	0	2	0	0	0	0
Alpha01281	gb U58105 MMU58105	<i>Mus musculus</i> Btk locus, alpha-D-galactosidase A (Ags) and Bruton's tyrosine kinase (Btk) genes	0	0	2	0	0	0	0
Alpha02538	gb U10871 MMU10871	<i>Mus musculus</i> MAP kinase mRNA; Mouse mRNA for p38b	0	0	2	0	0	0	0
Alpha03157	gb L36244 MUSMAT	<i>Mus musculus</i> metalloproteinase matrilysin mRNA, complete cds	0	0	2	0	0	0	0
Alpha03263	gb M60474 MUSMARCKS	Mouse myristoylated alanine-rich C-kinase substrate (MARCKS) mRNA, complete cds	0	0	2	0	0	0	0
Alpha04346	gb AC003996 AC003996	Mouse Cosmid m66a097 from 14D1-D2 (T-Cell Receptor Alpha Locus), complete sequence	0	0	2	0	0	0	0
Alpha04915	emb V00711 MITOMM	Mouse mitochondrial genome	0	0	2	0	0	0	0
Alpha06067	emb X91144 MMRNAPSGL	<i>M. musculus</i> mRNA for P-selectin glycoprotein ligand 1	0	0	2	0	0	0	0
Alpha06638	gb U57692 MMU57692	<i>Mus musculus</i> N-terminal asparagine amidohydrolase (Ntan1) mRNA, complete cds	0	0	2	0	0	0	0
Alpha07088	gb U13371 MMU13371	<i>Mus musculus</i> clone 1.5 novel mRNA from renin-expressing kidney tumor cell line	0	0	2	0	0	0	0
Alpha07324	gb L11651 RATEIF5	<i>Rattus norvegicus</i> eukaryotic initiation factor 5 (eIF-5) mRNA, complete cds	0	0	2	0	0	0	0
Alpha00014	emb X79233 MEEWS	<i>M. musculus</i> EWS mRNA	0	0	2	2	1	0	0
Alpha01337	gb AF011644 AF011644	<i>Mus musculus</i> oral tumor suppressor homolog (Doc-1) mRNA; <i>M. musculus</i> mRNA poly(A) site sequence	0	1	3	2	0	0	0
Alpha02878	gb J03298 MUSULT	Mouse uterine lactotransferrin mRNA	0	0	3	2	0	0	0
Alpha01980	emb X81987 MMTAX107	<i>M. musculus</i> mRNA for TAX responsive element binding protein 107	0	1	2	7	10	4	17
Alpha00237	gb M73436 MUSRSP4	Mouse ribosomal protein S4 (Rps4) mRNA, complete cds	0	1	4	6	5	13	10
Alpha00237	emb X14210 RNRSP4	Rat mRNA for ribosomal protein S4	0	1	4	6	5	13	10
Alpha00503	gb L04280 MUSRPL12A	<i>Mus musculus</i> ribosomal protein (Rpl12) mRNA; Rat mRNA for ribosomal protein L12	0	0	2	4	2	6	7
Alpha00973	emb X52803 MMCYCM	Mouse mRNA for cyclophilin (EC 5.2.1.8); Rat housekeeping protein P31 mRNA	1	1	2	5	4	6	6
Alpha01389	emb X82636 RNUQL40	<i>R. norvegicus</i> mRNA for a fusion protein of ubiquitin and ribosomal protein L40	0	0	4	2	5	4	5
Alpha00130	emb X68282 RNRPL13A	<i>R. norvegicus</i> mRNA for ribosomal protein L13a; <i>Rattus norvegicus</i> hexokinase type III mRNA	0	1	3	2	2	8	3
Alpha00345	dbj D55720 MUSNPTCC	Mouse mRNA for nuclear pore-targeting complex; <i>Mus musculus</i> pendulin (pendulin) mRNA	0	1	3	2	2	6	3
Alpha01052	dbj D17653 MUSHBL2B	Mouse mRNA for HBP15/L22, complete cds	0	1	1	2	0	1	1
Alpha00594	emb X61433 MMSODPOT	<i>M. musculus</i> mRNA for sodium/potassium ATPase beta subunit	0	1	0	5	0	0	1
Alpha02766	gb J03941 MUSFERH	Mouse ferritin heavy chain (MFH) mRNA, complete cds	1	0	1	4	0	0	1
Alpha00839	emb Z31553 MMCTBE	<i>M. musculus</i> (129/Sv) Cctb mRNA for CCT (chaperonin containing TCP-1) beta subunit	0	0	1	4	0	0	1
Alpha05164	emb X99395 MMENOGD	<i>M. musculus</i> gene encoding endonuclease G	0	0	0	2	0	0	1
Alpha06789	gb L12383 RATADPRF4A	<i>Rattus norvegicus</i> ADP-ribosylation factor 4 mRNA, complete cds	0	0	0	2	0	0	1
Alpha02706	emb X02487 MMMURS	Mouse retrovirus-related DNA sequence (MuRRS); Mouse middle repetitive LTR-like DNA sequence	0	0	1	6	0	1	0
Alpha05168	emb Z14249 MMERK1	<i>M. musculus</i> mRNA for mitogen activated protein kinase (erk-1)	0	1	0	2	0	1	0
Alpha04165	emb X74856 MMRNAL28	<i>M. musculus</i> L28 mRNA for ribosomal protein L28	0	0	0	2	0	1	0
Alpha04038	gb U43206 MMU43206	<i>Mus musculus</i> phosphatidylethanolamine binding protein mRNA, complete cds	1	1	0	4	1	0	0
Alpha01109	gb U49351 MMU49351	<i>Mus musculus</i> lysosomal alpha-glucosidase mRNA, complete cds	0	0	0	3	1	0	0
Alpha05063	gb L10652 RATEIF2B	<i>Rattus rattus</i> eukaryotic initiation factor (Eif-2) 67 kDa associated protein mRNA, complete cds	0	0	0	3	1	0	0
Alpha05758	emb X60831 MMUBF	Mouse ubf gene for transcription factor UBF	0	0	0	2	1	0	0
Alpha01215	dbj D17614 D17614	<i>Rattus norvegicus</i> mRNA for 14-3-3 protein theta-subtype, complete cds	0	0	0	7	0	0	0
Alpha01929	gb AF030343 AF030343	<i>Mus musculus</i> peroxisomal/mitochondrial dienoil-CoA isomerase ECH1p (Ech1) mRNA	0	0	0	5	0	0	0
Alpha03573	gb M64301 RATERK3	Rat extracellular signal-related kinase (ERK3) mRNA, complete cds	0	0	0	5	0	0	0
Alpha02521	gb M29462 MUSMDHA	Mouse malate dehydrogenase mRNA, complete cds	0	1	0	4	0	0	0
Alpha03020	gb U70210 MMU70210	<i>Mus musculus</i> TR2L mRNA, partial cds	0	0	0	4	0	0	0
Alpha02095	emb X52046 MMCOL3A1	<i>M. musculus</i> COL3A1 gene for collagen alpha-1; <i>R. norvegicus</i> mRNA for pro alpha 1 collagen type III	0	0	1	3	0	0	0
Alpha00819	gb S80082 S80082	Murine leukemia virus gag protein; gag...env [provirus]	0	0	0	3	0	0	0
Alpha01922	gb M63245 MUSALASH	<i>Mus musculus</i> amino levulinatase synthase (ALAS-H) mRNA, 3' end	0	0	0	3	0	0	0
Alpha02006	gb S45012 S44957S7	Tapa-1=integral membrane protein TAPA-1; <i>M. musculus</i> MD3 mRNA	0	1	1	2	0	0	0
Alpha03253	gb U05809 MMU05809	<i>Mus musculus</i> LAF1 transketolase mRNA, complete cds	0	1	1	2	0	0	0
Alpha00220	gb M12660 MUSH	Mouse Cfh locus, complement protein H gene; Mouse factor H mRNA	1	1	0	2	0	0	0
Alpha00259	gb J02622 MUSASPATM	Mouse mitochondrial aspartate aminotransferase isoenzyme mRNA, complete cds	0	1	0	2	0	0	0
Alpha02018	gb M27938 MUSMEA	Mouse male-enhanced antigen mRNA (Mea), complete cds	0	1	0	2	0	0	0
Alpha07894	gb U83896 RNU83896	<i>Rattus norvegicus</i> sec7B mRNA, complete cds	0	1	0	2	0	0	0
Alpha03403	gb U47328 MMU47328	<i>Mus musculus</i> MHC class I heavy chain precursor (H-2K(b)) mRNA; <i>Mus musculus</i> PYS-2 mRNA	1	0	0	2	0	0	0
Alpha00592	gb U27457 MMU27457	<i>Mus musculus</i> origin recognition complex protein 2 homolog mORC2L mRNA, complete cds	0	0	0	2	0	0	0

U, unfertilized egg; F, fertilized egg; 2, 4, 8 and 16, 2-, 4-, 8- and 16-cell embryos; B, blastocyst; cds, coding sequence.

NIH (<http://www.ncbi.nlm.nih.gov/>). Detailed map locations of genes are accessible through The Jackson Laboratory Backcross DNA Mapping Resource [<http://www.jax.org/resources/documents/cmdata/bkmap/BSS.html>]. Information about gene clustering and expression profile is available at ERATO Doi Project Home page (<http://www.bioa.jst.go.jp/pge/>). PCR primer pairs are available from the Research Genetics (<http://www.resgen.com/>). Finally the detailed information about cDNA clones, sequences, PCR primer pairs, and the library-specific BLAST search is available at the Laboratory of Genetics home page (<http://lgsun.grc.nia.nih.gov>).

RESULTS

Construction of cDNA libraries and characterization of ESTs

cDNA libraries were constructed from each of seven mouse preimplantation stages (Fig. 1). The cDNAs were directionally cloned, with an average size of insert of about 1.5 kb. cDNA clones from each library were arrayed in 96-well microtiter plates and about 400 bp sequenced from 3' termini. All 25,438 Expressed Sequence Tags (ESTs) obtained were deposited in the public sequence database and have been made available to the scientific community since the summer of 1997 [GenBank accession numbers: C75935-C81630; C85044-C88357, AU014577-AU024803, AU040095-AU046300].

By comparing the EST sequences to the repetitive sequence database, 2279 ESTs containing repeat sequences were identified (Beta clusters in Table 1). ESTs with low complexity sequence information were also identified and excluded from the further analyses (Gamma clusters in Table 1). The rest of the ESTs (23,155, Alpha clusters in Table 1) were condensed to a set of 9718 unique genes based on sequence similarity searches against one another. Similar genes were sought by BlastN search of the non-redundant (nr) public sequence database. Only 10% of the genes (955) showed close matches and were identified as known [named] genes (e.g. Table 2; for more complete information, refer to <http://www.bioa.jst.go.jp/pge/>). Furthermore, when similar ESTs were sought by the BlastN program against NCBI's public EST database (dbEST; Boguski et al., 1993), only 55% (5300) showed close matches to at least one EST from human, mouse, or rat. Considering the large number of ESTs (Adams et al., 1991; Hillier et al., 1996; Hwang et al., 1997; Marra et al., 1999, 1998; Okubo et al., 1992) in the dbEST ($>1 \times 10^6$ for human, $>4 \times 10^5$ for mouse, and $>2 \times 10^5$ for rat), the rate of discovery of new ESTs in these cDNA libraries is very high. It supports the notion that many genes expressed in mammalian preimplantation stages have not been otherwise isolated.

Global changes of gene expression patterns during preimplantation development

For each unique gene, the number of reads from each cDNA library was summed. Table 2 shows an example of this summation, using 'named genes'; complete results are available through the World Wide Web <http://www.bioa.jst.go.jp/pge/>. Since the frequency of ESTs in a particular cDNA library corresponds roughly to the expression level of the gene (Okubo et al., 1992), the data compiled here provide a first approximation of gene expression levels at each stage.

To assess changes in gene expression, the 9718 unique gene set was grouped into four main patterns based on the EST frequency at each stage (Fig. 2). Since one EST in

Group A: Low expression

	U	F	2	4	8	M	B	No. Known	No. Unknown	No. Total	% Total
	█	█	█	█	█	█	█	499	6766	7265	74.76%

Group B: Constitutive expression

	U	F	2	4	8	M	B	No. Known	No. Unknown	No. Total	% Total
	█	█	█	█	█	█	█	1	10	11	0.11%

Group C: Complex expression

	U	F	2	4	8	M	B	No. Known	No. Unknown	No. Total	% Total
	█	█	█	█	█	█	█	143	173	316	3.25%

Group D: Single-peak expression

	U	F	2	4	8	M	B	No. Known	No. Unknown	No. Total	% Total
1	█	█	█	█	█	█	█	23	268	291	2.99%
2	█	█	█	█	█	█	█	7	24	31	0.32%
3	█	█	█	█	█	█	█	2	9	11	0.11%
4	█	█	█	█	█	█	█	0	2	2	0.02%
5	█	█	█	█	█	█	█	45	198	243	2.50%
6	█	█	█	█	█	█	█	4	8	12	0.12%
7	█	█	█	█	█	█	█	3	1	4	0.04%
8	█	█	█	█	█	█	█	1	1	2	0.02%
9	█	█	█	█	█	█	█	3	4	7	0.07%
10	█	█	█	█	█	█	█	35	247	282	2.90%
11	█	█	█	█	█	█	█	3	8	11	0.11%
12	█	█	█	█	█	█	█	0	1	1	0.01%
13	█	█	█	█	█	█	█	0	1	1	0.01%
14	█	█	█	█	█	█	█	8	1	9	0.09%
15	█	█	█	█	█	█	█	42	206	248	2.55%
16	█	█	█	█	█	█	█	7	6	13	0.13%
17	█	█	█	█	█	█	█	0	3	3	0.03%
18	█	█	█	█	█	█	█	1	2	3	0.03%
19	█	█	█	█	█	█	█	6	0	6	0.06%
20	█	█	█	█	█	█	█	59	288	347	3.57%
21	█	█	█	█	█	█	█	7	18	25	0.26%
22	█	█	█	█	█	█	█	7	7	14	0.14%
23	█	█	█	█	█	█	█	49	338	387	3.98%
24	█	█	█	█	█	█	█	12	0	12	0.12%
25	█	█	█	█	█	█	█	110	51	161	1.66%

Fig. 2. Grouping of genes based on the global expression patterns. U, unfertilized; F, fertilized; 2, 4, 8, 2-cell, 4-cell and 8-cell embryos; M, morula; B, blastocyst.

approximately 3000 obtained from each developmental stage may be present by chance particularly for genes with very low level of expression, the initial analyses have focused mainly on genes with relatively abundant expression, i.e. genes represented by more than two independent clones in a cDNA library. Though this criterion is still statistically weak for individual genes (see Materials and Methods section), the groupings provide an indication of global changes.

The majority of genes (75%) were in Group A 'Low expression' throughout preimplantation development. A very small fraction (0.11%) of genes showed constitutive expression throughout preimplantation development (Group B). Some genes (3.25%) showed complex expression patterns (Group C) that may reflect up-and-down regulation but are probably also affected by sampling statistics. The rest of the genes (22%) were classified in 'Single-peak expression' (Group D). This group consists of genes undergoing: (1) gradual degradation from maternally stored mRNAs (3.44%, Groups D1-D4) and (2) constitutive expression once the gene is activated at a certain stage (2.14%, Groups D9, D14, D19, D22, D24 and D25). But it also includes an unexpectedly large number of genes (17.2%) that show apparently stage-specific expression.

Fig. 3. RT-PCR analyses of selected genes. For each stage, total RNAs extracted from ten embryos were sequentially diluted in fivefold steps ($\times 25$, $\times 125$ and $\times 625$) and subjected to RT-PCR. (A) Alpha03732 gene selected as an example of a 2-cell stage-specific expressed gene from Table 2. (B) Cytoplasmic gamma-actin gene selected as an example of a constitutively expressed gene from Table 2.

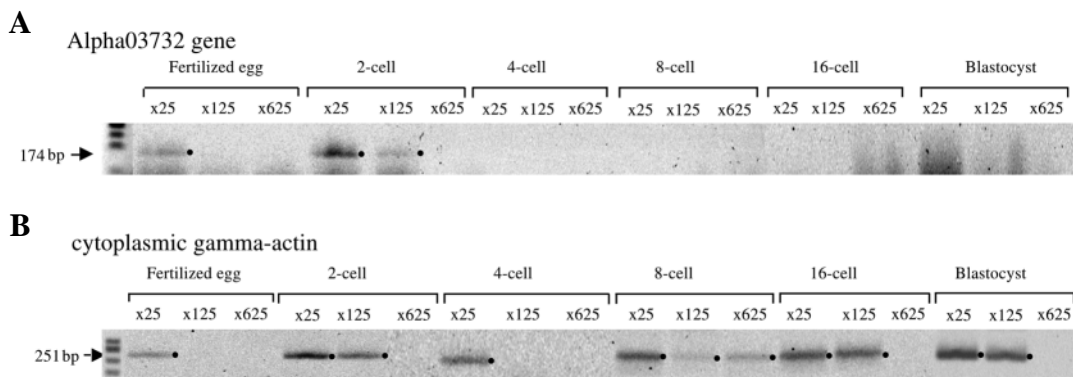


Table 3. Examples of stage-specific expressed genes

Gene ID	Gene Name	Unfertilized Egg (3096)	Fertilized Egg (3314)	2-cell Embryo (3684)	4-cell Embryo (3011)	8-cell Embryo (3444)	16-cell Embryo (3195)	Blastocyst (5692)
Alpha00603	Unknown, but similar to Human E2 ubiquitin conjugating enzyme UbcH5C mRNA	13	3	5	0	0	0	0
Alpha02663	Unknown, but similar to autoantigen [human, thyroid associated ophthalmopathy]	7	1	0	0	0	0	0
Alpha01916	Mouse gly96 mRNA	0	6	1	1	0	0	0
Alpha00797	Unknown	0	4	0	0	0	0	0
Alpha00096	Unknown	0	5	0	0	0	1	0
Alpha01107	Mouse inhibin alpha-subunit exon 2	0	2	0	0	0	0	0
Alpha03732	Unknown	0	0	7	0	0	0	0
Alpha01766	Unknown	0	0	0	8	0	0	0
Alpha01215	Rat mRNA for 14-3-3- protein theta-subtype protein kinase regulator	0	0	0	7	0	0	0
Alpha04688	Unknown, but similar to Human insulin-like growth factor binding protein-4 (IGFBP4)	0	0	0	6	0	0	0
Alpha01866	Mouse alpha-1 protease inhibitor 5 (alpha-1 PI-5)	0	0	0	6	2	0	0
Alpha01929	Mouse peroxisomal/mitochondrial dienyol-CoA	0	0	0	5	0	0	0
Alpha03573	Rat extracellular signal-related kinase (ERK3) mRNA	0	0	0	5	0	0	0
Alpha01262	Unknown	0	0	0	5	0	0	0
Alpha04643	Mouse gene for sphingomyelin Phosphodiesterase	0	0	0	0	8	0	0
Alpha01192	Unknown	0	0	0	0	6	0	0
Alpha06657	Unknown	0	0	0	0	6	0	0
Alpha01500	Unknown	0	0	0	0	5	0	0
Alpha00739	Unknown	0	0	0	0	2	9	2
Alpha01031	Mouse mRNA for 2,3-bisphosphoglycerate mutase	0	0	0	0	0	7	0
Alpha01563	Unknown	0	0	0	0	0	7	0
Alpha00199	Unknown	0	0	0	0	0	6	0
Alpha00425	Unknown	0	0	0	0	0	6	0
Alpha00754	Unknown	0	0	0	0	0	6	0
Alpha01891	Mouse S-adenosyl homocysteine hydrolase	0	0	0	0	0	6	0
Alpha03190	Unknown, but similar to bone morphogenetic protein type 1A receptor	0	0	0	0	0	6	0
Alpha02045	Mouse zinc finger protein Requiem (req) mRNA	0	0	0	0	0	5	0
Alpha02117	Unknown	0	0	0	0	0	5	0
Alpha02259	Unknown	0	0	0	0	0	5	0
Alpha03351	Unknown	0	0	0	0	0	5	0
Alpha01216	<i>M. musculus</i> mRNA for radixin	0	2	0	0	0	3	0
Alpha01188	Human ring zinc-finger protein (ZNF127-Xp) gene	0	0	0	0	1	0	15
Alpha01005	Mouse mRNA for heparin binding protein-44	0	0	0	2	0	0	14
Alpha00537	Mouse DNA for t-haplotype-specific elements	0	0	0	0	0	0	11
Alpha00913	Unknown, but similar to Mouse gene for topoisomerase I	0	0	0	0	0	0	6
Alpha01176	Rat mRNA for CIC-K1 protein	0	0	0	0	0	0	6
Alpha00893	<i>M. musculus</i> mRNA for connexin31	0	0	0	0	0	0	3
Alpha00075	Mouse cytoplasmic gamma-actin	3	5	1	4	4	4	14
Alpha04585	Mouse mRNA for elongation factor 1-alpha (EF 1-alpha)	4	9	8	25	30	28	71
Alpha01597	Mouse mRNA for G protein beta subunit homologue	3	2	3	5	5	6	15
Alpha00195	Mouse nucleolar protein N038 mRNA	3	1	3	3	6	3	3
Alpha00285	<i>Mus musculus</i> ATP synthase beta-subunit (beta-F1 ATPase)	2	1	5	3	2	13	10

The total number of ESTs at each embryonic stage are given in parentheses.

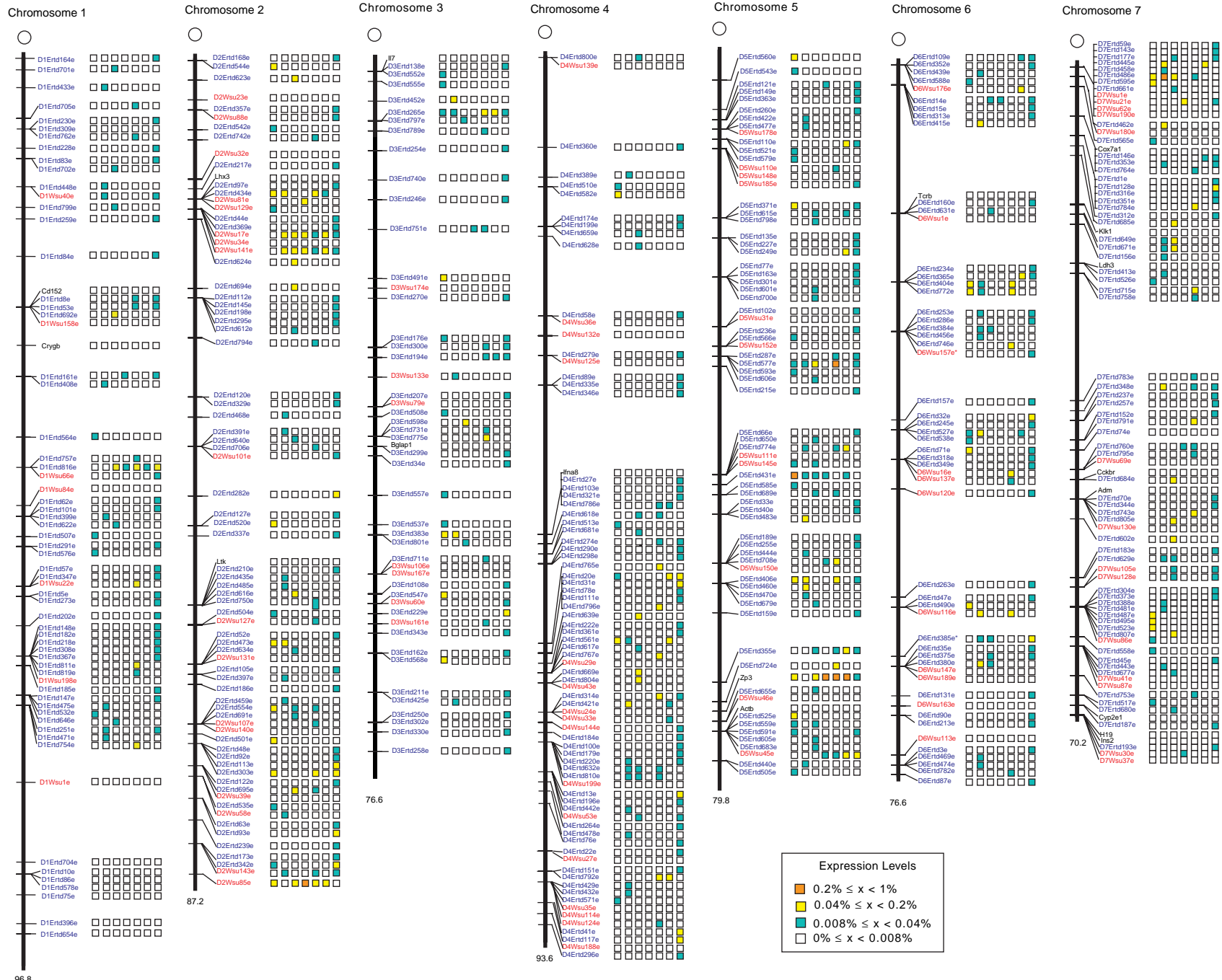


Fig. 4. For legend see p. 1746.

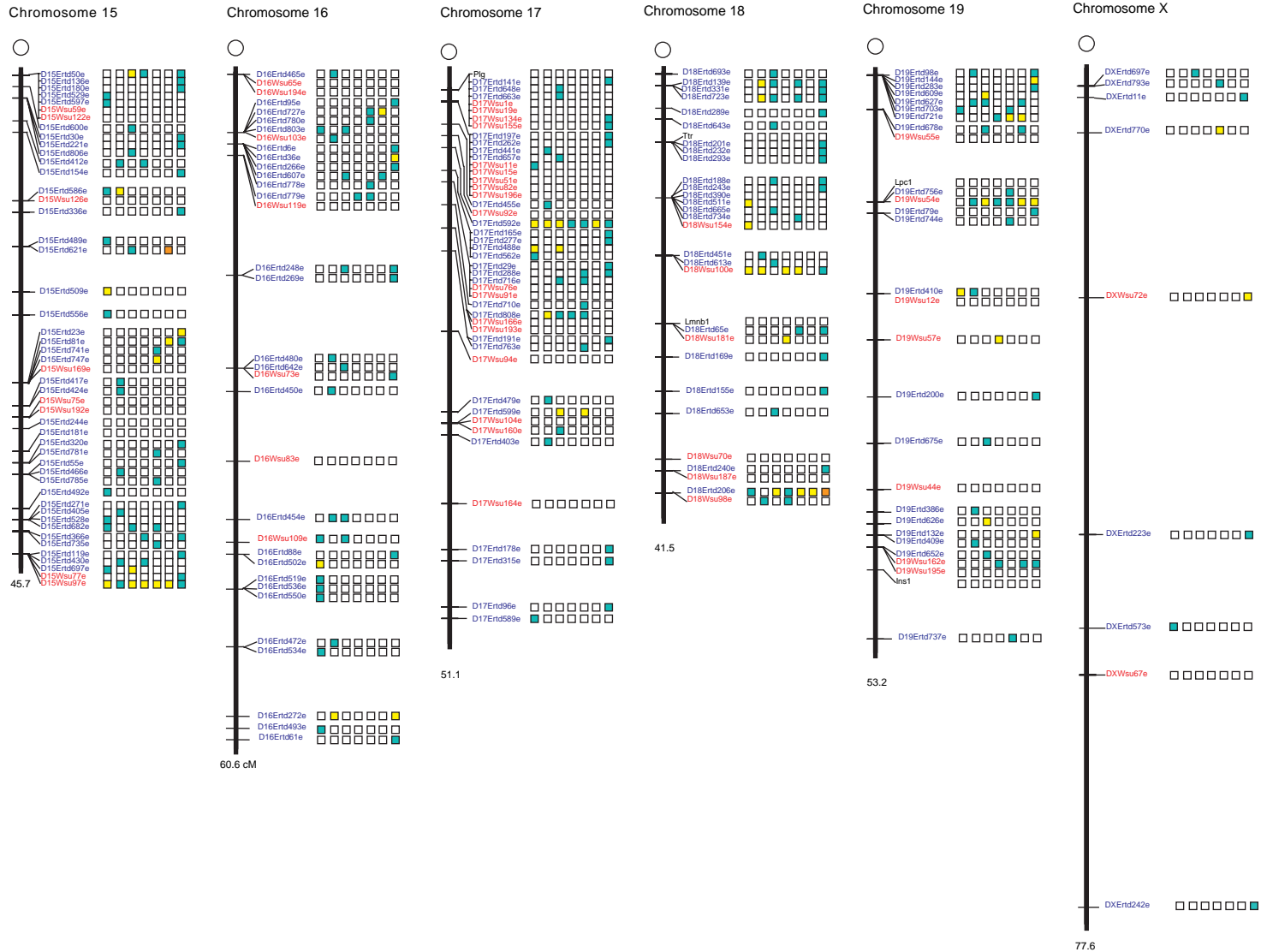


Fig. 4. Gene map with expression profiles. Expression levels of individual genes are shown in seven small boxes next to locus names and are color coded based on the frequency of ESTs in each cDNA library. The cDNA libraries represented by each box are from right to left: unfertilized egg, fertilized egg, 2-cell embryo, 4-cell embryo, 8-cell embryo, 16-cell embryo and blastocyst. Ertid markers in blue are newly mapped in this

paper. Wsu markers in red are derived from extraembryonic tissues of 7.5-d.p.c. embryos (Ko et al., 1998). Markers in black are some known genes we mapped previously (Ko et al., 1994). The total chromosome length is shown in cM at the bottom of each chromosome.

Stage-specific expressed genes

Expression patterns with a sharp peak in only one stage were observed for each stage. For successive stages, fertilized egg-specific genes (Group D5) comprised 2.50% of the total RNAs; 2-cell-specific (Group D10), 2.90%; 4-cell-specific (Group D15), 2.55%; 8-cell-specific (Group D20), 3.57%; Morula-specific (Group D23), 3.98%; and Blastocyst-specific (Group D25), 1.66%. Table 3 shows examples of such stage-specific expressed genes selected for a statistically significant level of expression that is comparable to the level for classical constitutively expressed genes like actin.

Of course, the EST frequencies in each cDNA library only roughly correlate with the expression level of genes, and there are two possible artifacts that could make these results deviate from the actual expression level of genes: (1) distortion of levels during PCR amplification of cDNA mixtures, and (2) statistical sampling variations. Two lines of evidence, however, argue against these artifacts influencing our results. First, the expression patterns of known genes that have been previously independently studied are consistent with those reported here. For example, the EST frequency of S-adenosylhomocysteine hydrolase is significantly increased at the 16-cell embryo stage and decreased again at the blastocyst stage (Table 3). This gene has been identified as the causative gene for the mouse lethal *nonagouti* (a^x) mutation and our results are consistent with the reported expression of this gene in preimplantation stages (Miller et al., 1994). Another example is connexin31, a gap-junction protein. Blastocyst-specific expression based on EST analysis (Table 3) is consistent with the previous report (Reuss et al., 1997). High expression of radixin at the Morula stage is also consistent with its function as a cell adhesion molecule and its reported expression pattern (Funayama et al., 1991), and blastocyst-stage specific appearance of placental lactogen II (Jackson et al., 1986) is also consistent with placenta-specific expression of the gene.

Second, reverse transcriptase (RT)-PCR analyses on staged embryos confirmed the expression patterns for selected genes, e.g. the gene Alpha03732 (Fig. 3A). Low levels of transcripts were also detected in unfertilized eggs and fertilized eggs, but transcripts were most abundant at the 2-cell stage, confirming the observed EST frequency that showed 2-cell stage expression of this gene. Gamma-actin as a control showed a comparable level of transcripts in all stages (Table 3, Fig. 3B).

Genetic mapping of novel ESTs

Primer pairs were designed and synthesized for 4500 ESTs, selected from the large fraction of new ESTs that were not 'named genes' in the public sequence database. The PCR products of all primer pairs were tested for sequence polymorphisms between C57BL/6J and *M. spretus* by a heteroduplex assay (Ko et al., 1998). About 800 primer pairs were found to produce sequence polymorphisms, and these were genotyped on The Jackson Laboratory BSS Interspecific Backcross Panel. The map location of these ESTs, along with known genes (Ko et al., 1994) and ESTs (Ko et al., 1998) that we previously mapped on the same panel, are shown in Fig. 4. Fig. 4 also indicates the expression patterns of individual genes, assessed by counting the representation of these ESTs at each stage of development. More detailed map information, including the raw data and the relative positions of these markers to other markers in this mapping cross, is accessible

through the World Wide Web [<http://www.jax.org/resources/documents/cmdata/bkmap/BSS.html>]. It is interesting to note that some genes with similar developmental expression patterns have been shown to cluster on the mouse genetic map. For example, there are clusters of blastocyst-specific genes in chromosomes 1 and 8, and unfertilized egg-specific genes in chromosomes 7 and 16 (Fig. 4).

DISCUSSION

Large-scale cDNA sequencing projects have successfully produced more than 1 million ESTs (Adams et al., 1991; Hillier et al., 1996; Hwang et al., 1997; Marra et al., 1999, 1998; Okubo et al., 1992). However, the majority of cDNA libraries have been derived from adult organs and tissues. Because many genes are only expressed at limited times and places, and often at low levels, the gene catalogue thus remains incomplete. In particular, limited information has been obtained about transcripts in the stages of preimplantation mammalian development (Adjaye et al., 1997, 1998; Rothstein et al., 1992, 1993; Sasaki et al., 1998). The optimization of a PCR-based cDNA library construction method (Takahashi and Ko, 1994) has provided seven cDNA libraries used here, and although the libraries were not normalized, a high rate of new gene discovery was seen (9718 unique genes from a total of 25,438 ESTs). This very likely reflects the high complexity of mRNA species in preimplantation embryos. Furthermore, about 50% of the 9718 unique genes were seen for the first time in this study, presumably because these preimplantation mammalian embryonic stages have not been extensively used in other EST projects.

Discussion of the genes showing very interesting expression patterns is largely outside the scope of this paper, but one can easily illustrate some of the usefulness of the data. For example, ERCC3, which is part of TFIIH that has helicase activity and is involved in base excision repair of transcribed DNA (Weeda et al., 1990), is particularly abundant at the 2-cell stage (Table 2). This is concurrent with the initiation of zygotic transcription when many nuclear genes are starting to be expressed. Another example is the BMAL1 gene, which was recently identified as a partner for heterodimer formation with CLOCK genes and plays a critical role in mammalian circadian rhythm (Darlington et al., 1998). Transcripts were present in unfertilized eggs, 2-cell and 16-cell stages (Table 2, <http://www.bioa.jst.go.jp/pge/>). Although further analyses are required, this pattern suggests that the gene is expressed intermittently, skipping some developmental stages. This finding implies that the timing of cell division at the cleavage-stage mouse embryos may be controlled by the same pathway as the circadian rhythm in the adult mouse.

The newly mapped genes will provide a valuable resource for positional cloning of mouse genes. Given their substantial homology in gene organization, the mouse data should also help the positional cloning of human genes. In addition, many of these genes are apparently unique to early mammalian embryos. Consequently, the gene-mapping efforts presented here will provide a complement to the ongoing large-scale EST mapping projects in human and mouse. Finally, from the EST PCR primer pairs described here, approximately 2500 are now being mapped on the T31 radiation hybrid panel at the MRC UK

Mouse Genome Center (Paul Denny and Steve Brown, personal communication, http://www.mgc.har.mrc.ac.uk/est_maps.htm). The 798 genes mapped here in The Jackson Backcross Panel will help to anchor the genetic and radiation hybrid maps.

The map also provides a genome-wide view of the distribution of genes with information on expression levels at each stage (Fig. 4). One significant feature of the map is that genes with similar expression patterns appear to cluster on the mouse genome. This supports the previous suggestion that mammalian genomes are evolved so that coexpressed genes often tend to cluster (Ko et al., 1998). Physical proximity may provide the embryos with an efficient means of coordinately regulating the expression of many genes.

Two general methods for global expression analyses – a high resolution two-dimensional protein gel analysis and mRNA differential display – have been applied to the preimplantation mouse development. Comprehensive two-dimensional gel analyses during the 1-, 2- and 4-cell stages have identified about 1500 protein spots, some 38 of which show a transient increase in a 2-cell stage-specific manner (Latham et al., 1991). Another study has examined 1674 protein spots between compacted 8-cell and blastocyst-stage mouse embryos and identified 43 protein spots that are present only at 8-cell stage and 75 protein spots that are present only at blastocyst-stage (Shi et al., 1994). Because the two-dimensional protein gels represent the protein biosynthesis and/or the modification of pre-existing proteins (Oh et al., 1999), the overall pattern changes could reflect differential regulations at the translational or post-translational level. Another inherent drawback is a difficulty in isolating genes that correspond to each protein spot. In fact, only a few protein spots have been identified at a gene level (Schultz, 1999). In contrast, mRNA differential display has been successfully used to identify some 2-cell, 8-cell and blastocyst-specific cDNA fragments (Zimmermann and Schultz, 1994). However, only one new gene (eIF-4C) has been identified as a 2-cell stage specific gene (Davis et al., 1996; Schultz, 1999). The mRNA differential display has inherent difficulties for gene identification, because isolated cDNA fragments are usually too short to provide clear identity and too fragmentary to recover full-length cDNA clones. In this sense, the subtraction cDNA library method has been most successfully used to isolate stage-specific expressed genes (Oh et al., 1999), although it does not provide much information about global changes of gene expression patterns. Four new genes that are present in the 1-cell and 2-cell embryos, but are absent in the 8-cell embryos, have been identified in this manner (Oh et al., 1999). Considering these facts, the EST approach may have advantages over the other methods, because (1) cDNA clones are readily available for the detailed analysis of genes; (2) the expression levels of genes are monitored as the relative abundance of mRNAs; and (3) it provides the information about global changes of gene expression patterns.

The EST data sets from each stage-specific cDNA library provide a first gene-based index of overall gene expression patterns during preimplantation development. For early cleavage-stage embryos, the data presented here confirm earlier analyses based on smaller numbers of genes (reviewed in Kidder, 1992; Schultz and Heyner, 1992; Watson et al., 1992). For example, the results support the previous inference that most maternally stored mRNAs in unfertilized eggs are degraded by the two-cell stage (Piko and Clegg, 1982). The

individually isolated genes identify many known and unknown genes that may contribute to a more detailed understanding of processes like RNA degradation. Approximately 50% of all the genes in our collection initiated expression at the 1-cell and 2-cell stages. This is also consistent with previous studies (Piko and Clegg, 1982). Because the majority of studies of zygotic gene activation have been done by analyzing the transcriptional activity of exogenously injected genes (reviewed in Kaneko and DePamphilis, 1998), endogenous genes found here will provide the additional means with which to analyze the zygotic gene activation. An obvious extension of the studies will be to see whether specific cohorts of genes are activated earlier than others.

In contrast, the data presented here would suggest that the paradigm for the late cleavage-stage embryos needs to be revised. The existing paradigm is that although the onset of expression of individual genes is varied, almost all of them continue to be expressed once expression starts (Kidder and McLachlin, 1985; Levy et al., 1986; Kidder, 1992; Schultz and Heyner, 1992; Watson et al., 1992). Such findings have prompted researchers to believe that 'the transcription of most genes in preimplantation development is not temporally linked with the morphogenetic transitions they participate in' (Kidder, 1992). However, the new finding here of many apparently stage-specific expressed genes may challenge this conventional view of the regulation of gene expression during early mouse development. Preimplantation development usually takes 4 to 6 days, and each cell division is therefore rather slow, and each cell cycle/division has time to generate new and unusual gene expression patterns from the selective destruction and activation of many genes.

The existence of many stage-specific expressed genes has two primary implications. First, developmental processes may lead to significant differences in gene expression in the transition from 2 cells to 8 cells, even though the process appears to be simply a division of cells. Second, stage-specific expressed genes may actively promote the advancement of embryos from one stage to the next. Considering the rapid and selective turnover of these particular mRNAs, there is likely to be selection for their rapid degradation. The requirement of function of that particular gene is apparently transient, which suggests a 'hit-and-run' type of mechanism for expression of cascades of genes. It will be of interest to see, for example, if specific inactivation of phase-specific genes causes arrest of development at stages of oocyte formation or cleavage.

We would like to thank Drs David Schlessinger, Dan Longo, and Ramaiah Nagaraja for critical reading of the manuscript, Ms Shoshana Stern for assisting with the art work, and Ms Mary Barter for assisting with the genetic map figure. This work was supported by Science and Technology Agency in Japan through the ERATO/JST program, and received additional support from NICHD/NIH grant (RO1HD32243) and the NIA/NIH intramural research program.

REFERENCES

- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F. et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651-1656.
- Adjaye, J., Daniels, R., Bolton, V. and Monk, M. (1997). cDNA libraries from single human preimplantation embryos. *Genomics* **46**, 337-344.

- Adjaye, J., Daniels, R. and Monk, M. (1998). The construction of cDNA libraries from human single preimplantation embryos and their use in the study of gene expression during development. *J. Assist. Reprod. Genet.* **15**, 344-348.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Audic, S. and Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Res.* **7**, 986-995.
- Boguski, M. S., Lowe, T. M. J. and Tolstoshev, C. M. (1993). dbEST-database for 'expressed sequence tags'. *Nature Genetics* **4**, 332-333.
- Claverie, J. M. (1999). Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Genet.* **8**, 1821-1832.
- Darlington, T. K., Wager-Smith, K., Ceriani, M. F., Staknis, D., Gekakis, N., Steeves, T. D. L., Weitz, C. J., Takahashi, J. S. and Kay, S. A. (1998). Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science* **280**, 1599-1603.
- Davis, W., Jr., De Sousa, P. A. and Schultz, R. M. (1996). Transient expression of translation initiation factor eIF-4C during the 2-cell stage of the preimplantation mouse embryo: identification by mRNA differential display and the role of DNA replication in zygotic gene activation. *Dev. Biol.* **174**, 190-201.
- Funayama, N., Nagafuchi, A., Sato, N. and Tsukita, S. (1991). Radixin is a novel member of the band 4.1 family. *J. Cell Biol.* **115**, 1039-1048.
- Hillier, L. D., Lennon, G., Becker, M., Bonaldo, M. F., Chiapelli, B., Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W. et al. (1996). Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6**, 807-828.
- Hogan, B., Beddington, R., Costantini, F. and Lacy, E. (1994). *Manipulating the Mouse Embryo: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.
- Hwang, D. M., Dempsey, A. A., Wang, R. X., Rezvani, M., Barrans, J. D., Dai, K. S., Wang, H. Y., Ma, H., Cukerman, E., Liu, Y. Q. et al. (1997). A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. *Circulation* **96**, 4146-4203.
- Jackson, L. L., Colosi, P., Talamantes, F. and Linzer, D. I. (1986). Molecular cloning of mouse placental lactogen cDNA. *Proc. Natl. Acad. Sci. USA* **83**, 8496-8500.
- Kaneko, K. J. and DePamphilis, M. L. (1998). Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. *Dev. Genet.* **22**, 43-55.
- Kidder, G. M. (1992). The genetic program for preimplantation development. *Dev. Genet.* **13**, 319-325.
- Kidder, G. M. and McLachlin, J. R. (1985). Timing of transcription and protein synthesis underlying morphogenesis in preimplantation mouse embryos. *Dev. Biol.* **112**, 265-275.
- Ko, M. S. H., Threat, T. A., Wang, X., Horton, J. H., Cui, Y., Pryor, E., Paris, J., Wells-Smith, J., Kitchen, J. R., Rowe, L. B. et al. (1998). Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and underrepresentation on the X chromosome. *Hum. Mol. Genet.* **7**, 1967-1978.
- Ko, M. S. H., Wang, X., Horton, J. H., Hagen, M. D., Takahashi, N., Maezaki, Y. and Nadeau, J. H. (1994). Genetic mapping of 40 cDNA clones on the mouse genome by PCR. *Mamm. Genome* **5**, 349-355.
- Latham, K. E., Garrels, J. I., Chang, C. and Solter, D. (1991). Quantitative analysis of protein synthesis in mouse embryos. I. Extensive reprogramming at the one- and two-cell stages. *Development* **112**, 921-932.
- Levy, J. B., Johnson, M. H., Goodall, H. and Maro, B. (1986). The timing of compaction: control of a major developmental transition in mouse early embryogenesis. *J. Embryol. Exp. Morphol.* **95**, 213-237.
- Manly, K. F. (1993). A Macintosh program for storage and analysis of experimental genetic mapping data. *Mamm. Genome* **4**, 303-313.
- Marra, M., Hillier, L., Kucaba, T., Allen, M., Barstead, R., Beck, C., Blistain, A., Bonaldo, M., Bowers, Y., Bowles, L. et al. (1999). An encyclopedia of mouse genes. *Nature Genet.* **21**, 191-194.
- Marra, M. A., Hillier, L. and Waterston, R. H. (1998). Expressed sequence tags—ESTablishing bridges between genomes. *Trends Genet.* **14**, 4-7.
- Miller, M. W., Duhl, D. M., Winkes, B. M., Arredondo-Vega, F., Saxon, P. J., Wolff, G. L., Epstein, C. J., Hershfield, M. S. and Barsh, G. S. (1994). The mouse lethal nonagouti (a(x)) mutation deletes the S-adenosylhomocysteine hydrolase (Ahc) gene. *EMBO J.* **13**, 1806-1816.
- Nothias, J. Y., Majumder, S., Kaneko, K. J. and DePamphilis, M. L. (1995). Regulation of gene expression at the beginning of mammalian development. *J. Biol. Chem.* **270**, 22077-22080.
- Oh, B., Hwang, S. Y., De Vries, W. N., Solter, D. and Knowles, B. B. (1999). Identification of genes and processes guiding the transition between the mammalian gamete and embryo. In *A Comparative Methods Approach to the Study of Oocytes and Embryos* (ed. J. D. Richter), pp. 101-127. New York: Oxford University Press.
- Okubo, K., Hori, N., Matoba, R., Niijima, T., Fukushima, A., Kojima, Y. and Matsubara, K. (1992). Large scale cDNA sequencing for gene expression analysis: quantitative and qualitative aspects of gene expression in a liver cell line, Hep G2. *Nature Genet.* **2**, 173-179.
- Pedersen, R. A. (1986). Potency, lineage, and allocation in preimplantation mouse embryos. In *Experimental Approaches to Mammalian Embryonic Development* (ed. J. Rossant and R. A. Pedersen), pp. 3-33. Cambridge: Cambridge University Press.
- Piko, L. and Clegg, K. B. (1982). Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. *Dev. Biol.* **89**, 362-378.
- Reuss, B., Hellmann, P., Traub, O., Butterweck, A. and Winterhager, E. (1997). Expression of connexin31 and connexin43 genes in early rat embryos. *Dev. Genet.* **21**, 82-90.
- Rossant, J. (1986). Development of extraembryonic cell lineages in the mouse embryo. In *Experimental Approaches to Mammalian Embryonic Development* (ed. J. Rossant and R. A. Pedersen), pp. 97-120. Cambridge: Cambridge University Press.
- Rothstein, J. L., Johnson, D., DeLoia, J. A., Skowronski, J., Solter, D. and Knowles, B. (1992). Gene expression during preimplantation mouse development. *Genes Dev.* **6**, 1190-1201.
- Rothstein, J. L., Johnson, D., Jessee, J., Skowronski, J., DeLoia, J. A., Solter, D. and Knowles, B. B. (1993). Construction of primary and subtracted cDNA libraries from early embryos. *Methods Enzymol.* **225**, 587-610.
- Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J. and Birkenmeier, E. H. (1994). Maps from two interspecific backcross DNA panels are available as a community genetic mapping resource. *Mamm. Genome* **5**, 253-274.
- Sasaki, N., Nagaoka, S., Itoh, M., Izawa, M., Konno, H., Carninci, P., Yoshiki, A., Kusakabe, M., Moriuchi, T., Muramatsu, M. et al. (1998). Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones. *Genomics* **49**, 167-179.
- Schultz, G. A. and Heyner, S. (1992). Gene expression in pre-implantation mammalian embryos. *Mutat. Res.* **296**, 17-31.
- Schultz, R. M. (1993). Regulation of zygotic gene activation in the mouse. *BioEssays* **15**, 531-538.
- Schultz, R. M. (1999). Gene expression in mouse embryos: use of mRNA differential display. In *A Comparative Methods Approach to the Study of Oocytes and Embryos* (ed. J. D. Richter), pp. 148-156. New York: Oxford University Press.
- Shi, C. Z., Collins, H. W., Garside, W. T., Buettger, C. W., Matschinsky, F. M. and Heyner, S. (1994). Protein databases for compacted eight-cell and blastocyst-stage mouse embryos. *Mol. Reprod. Dev.* **37**, 34-47.
- Takahashi, N. and Ko, M. S. H. (1993). The short 3'-end region of complementary DNAs as PCR-based polymorphic markers for an expression map of the mouse genome. *Genomics* **16**, 161-168.
- Takahashi, N. and Ko, M. S. H. (1994). Toward a whole cDNA catalog: Construction of an equalized cDNA library from mouse embryos. *Genomics* **23**, 202-210.
- Taylor, K. D. and Piko, L. (1987). Patterns of mRNA prevalence and expression of B1 and B2 transcripts in early mouse embryos. *Development* **101**, 877-892.
- Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369-374.
- Watson, A. J., Kidder, G. M. and Schultz, G. A. (1992). How to make a blastocyst. *Biochem. Cell. Biol.* **70**, 849-855.
- Weeda, G., van Ham, R. C., Masurel, R., Westerveld, A., Odijk, H., de Wit, J., Bootsma, D., van der Eb, A. J. and Hoeijmakers, J. H. (1990). Molecular cloning and biological characterization of the human excision repair gene ERCC-3. *Mol. Cell. Biol.* **10**, 2570-2581.
- Wieschaus, E. (1996). Embryonic transcription and the control of developmental pathways. *Genetics* **142**, 5-10.
- Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. and Campbell, K. H. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810-813.
- Zimmermann, J. W. and Schultz, R. M. (1994). Analysis of gene expression in the preimplantation mouse embryo: use of mRNA differential display. *Proc. Natl. Acad. Sci. USA* **91**, 5456-5460.