OriGene Technologies

June 2005

Volume 1, Issue 1

GeneDex: A Journal Dedicated to Innovative Technologies for Gene Function Discovery

Inside this issue:

The scientists at OriGene Technologies, Inc. are looking forward to sharing their skills and knowledge in the field of human gene discovery through this journal. Each issue will focus on a new topic that is relevant to gene-centered biology. You will find reports on the hottest new genes, interesting genome facts, information on new product development and helpful laboratory protocols.

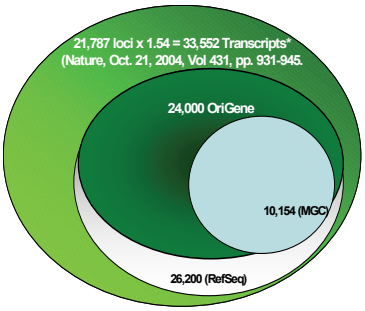


Figure 1: Venn diagram of cDNA clone coverage of the NCBI RefSeq transcript database

OriGene has been diligently adding new human full-length cDNA clones into its TrueClone Collection for over 6 years, and it all started with a random sequencing approach. When the redundancy of new gene hits became evident, a new and proprietary, targeted cloning approach was employed (OriGene's own RapidScreen). Between the two approaches, the TrueClone Collection now stands above the rest in terms of overall human gene content. The Venn diagram on the right depicts the overlapping coverage of the entire human cDNA transcriptome. While the RefSeq curated reference database (NCBI) is the gold standard for non-redundant human transcript sequences, OriGene's TrueClone collection represents the largest actual cDNA clone set.

Special points of interest:

- Most Comprehensive collection: over 24,000 human cDNA clones to date
- Authentic: no PCR artifacts
- Expression Ready: All clones in CMV vector
- Quick Delivery: Shipped to you within 48 hours

FEATURED GENE: How to clone LRRK2, a large (9kb), rare gene implicated in Parkinson's Disease

Two independent groups have recently reported that mutations in the LRRK2 (Leucine rich kinase 2) gene cause Parkinson's disease (Neuron 44, 601-607 & 575-577 (2004)), which is the second most common neurodegenerative disorder affecting more than one million people in the United States. The LRRK2 gene encodes a large protein with multiple functional domains, including a small GTPase domain and a Ser/Thr/Tyr kinase domain. The full-length sequence is predicted from overlapping ESTs or three RT-PCR fragments across the open reading frame, but the full-length gene has not been cloned. Here we describe how we used our proprietary technology, RapidScreen, and our comprehensive cDNA clone collection to isolate a 9.2 kb full-length cDNA clone and two novel splice variants of LRRK2. Full-length cDNA clones are the most important tools for functional analysis of human genes. However, it remains difficult to isolate a physical

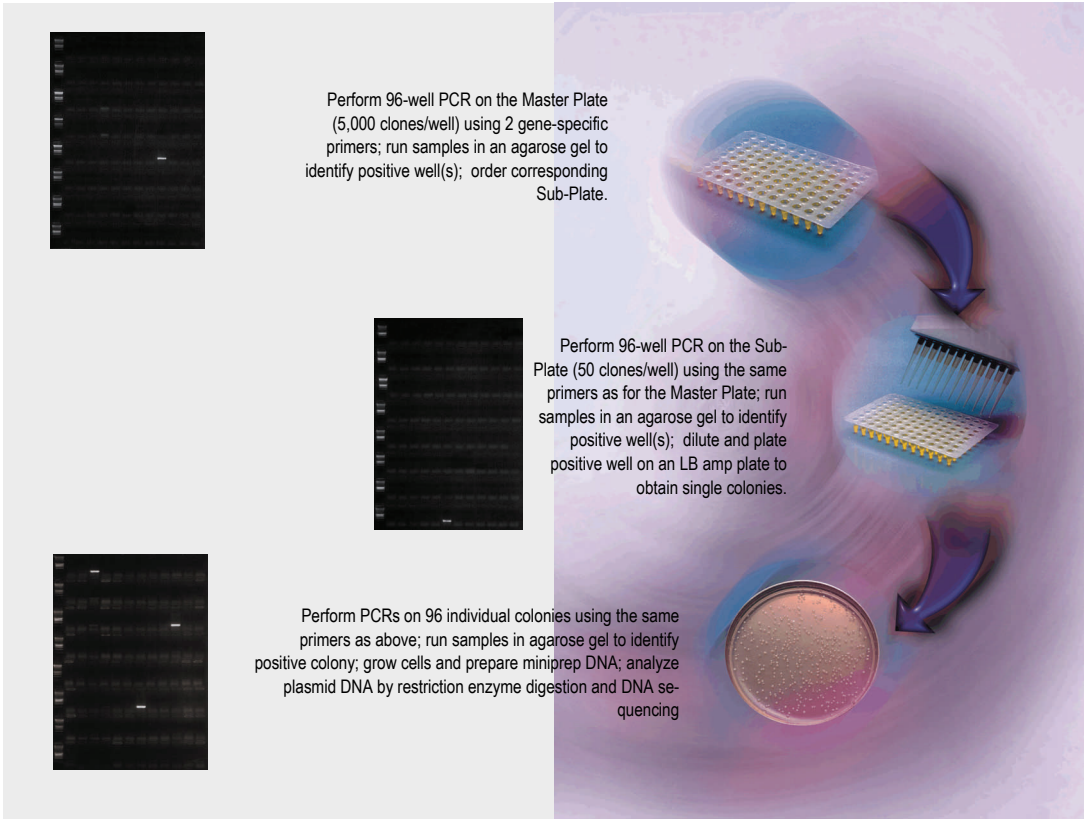
FEATURED GENE: The cloning of LRRK2

"LRRK2 may play a central role in the pathogenesis of several major neurodegenerative disorders associated with parkinsonism."

cDNA clone from complex cDNA libraries, especially for a long or rare transcript. Large-scale cDNA and genomic sequencing has created thousands of reference sequences in the NCBI database. While the DNA sequence provides important structural information, accession to a full-length clone remains the key if one is to proceed to study the function of that gene or its interaction with other proteins. Many of the compiled gene sequences in the public databases do not have a physical full-length clone available to support further investigation.

There are several approaches to clone a full-length gene. The conventional method which includes trapping and colony-lift, is labor-intensive, time-consuming and expensive. The PCR-based cloning method has its own limitations including frequent copying errors (inactivating mutations) and failure of amplification of long and rare transcripts from a certain RNA source. In vitro gene synthesis is an alternative approach to obtain a full-length cDNA clone assuming the predicted sequence is correct, but this approach is also expensive and time-consuming, especially for a long transcript.

OriGene Technologies, Inc. has developed 20 high-quality human Rapid-Screen™ arrayed cDNA library panels (US patent 6,316,193) to favor the cloning of long transcripts. A great deal of foresight, ingenuity and innovation has also been incorporated into the development and screening of these cDNA library panels. Besides the Rapid-Screen™ arrayed cDNA



Continued on Page 4

GENE FACTS: How many human genes and transcripts are there, now that the human genome is finished?

Abstract

As the final number of human genes begins to come into focus following the sequencing of the human genome, the natural variations that exist for each bring the total number of transcripts to a level that is consistent with the complexity of humans. The transcriptome is best illustrated by the worldwide efforts to create non-redundant transcript sequence databases, but this is by no means complete. In order to effectively begin systems biology over-expression screening, the content of a cDNA clone collection is key.

Introduction

As the Human Genome Sequencing effort approached completion, a new scientific focus, Functional Genomics, was born which was to exploit the genome sequence in a manner that would allow new gene discovery to lead the way in fighting human disease. Unfortunately, a vital step was overlooked in this transition, namely the creation of a physical set of cDNA clones representing the entire human transcriptome. Laboratory experiments could not be done on a genome-wide scale without first having authentic copies, or clones, for each of those transcripts. In fact, a final sequence reference set or description of all human genes and their variants is still a work in progress at the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI), and the DNA Data Bank of Japan (DDBJ). A quality reference set is primary and pivotal for the creation of a quality clone collection. The use of genome-wide collections for high throughput functional analyses is emerging as a powerful systems biology tool (1).

The Number of Human Genes

The completion of the human genome sequence has not led to the creation of a complete human transcript reference set, the transcriptome. The genome sequence provides the data necessary to determine the number of genes, but not the number of transcripts that differ due to post-transcriptional alternative splicing. We are still learning more about the regulation of these splicing mechanisms but the inherent complexities create a real challenge for prediction programs. Even single base differences within an exon or intron can cause the inclusion or exclusion of the exon (2). By mapping cDNA sequence data to the genome, it is possible to evaluate and adjust certain algorithms to make educated, but not perfect, predictions in all cases.

The most recent human genome build at NCBI, #35, boasts a large scale improvement over the previous builds and is likely to become the working genome for quite some time, providing a more stable gene reference set. The International Human Genome Sequencing Consortium (IHGSC) has recently published their analysis. While the total number of gene loci (individual genomic locations) is currently believed to be 21,787, there are (only) 2,188 of these that are predicted. The expected total number of transcripts including the splice variants is expected to be 1.54 times that, or 33,552 (3). For purposes of this discussion, genes are defined as those genomic regions or loci that are capable of being transcribed into mRNA. Transcripts, therefore, are the mature mRNA molecules that have undergone alternative splicing. Messenger RNA molecules can be copied by reverse transcriptase to create more stable DNA copies called cDNA. When linked to a functional vector, the cDNA is called a clone. OriGene is dedicated to the task of collecting a cDNA clone for every gene variant in the human transcriptome, linked to a consistent mammalian expression vector for use in a systems biology tool.

"The use of genome-wide collections for high throughput functional analyses is emerging as a powerful systems biology tool"

To date, the estimate of 21,787 gene loci underrepresents the total number of genes, and includes some sequence prediction errors. The complexity of this estimation is two-fold. First, the recognition of all transcribed regions within the genome (a.k.a. human genes) is hampered by the imprecise capability of predicting intron/exon boundaries. The 5'-most exons (200bp on average) can be difficult to predict because intervening sequences (introns) can be decidedly larger (up to 30 kb), therefore, that first exon can be difficult to predict, especially without cDNA supporting evidence. Comparative Genomics uses the genome sequence from closely related organisms to help validate gene predictions, and the CAGE (Cap Analysis Genome Expression) project by Yoshihide

Continued on Page 6

FEATURED GENE: The cloning of LRRK2

library panels, we have also collected and sequenced over 800,000 individual clones from our non-amplified cDNA libraries at the 5' end.

Most of these libraries are enriched with insert sizes above 5 kb for the isolation of large size inserts.

Rapid-Screen panels were designed as a fast, inexpensive, and easy way to clone full-length cDNAs. The cloning process has been made faster and more economical by having the libraries arrayed in 96-well plates and developing a screening procedure that is PCR based. Since the screening is PCR-based, vector and gene-specific primers can be used in concert to determine the size of the various cDNA inserts after the very first PCR screen, thus eliminating the need to isolate many clones and then find out which, if any, are full-length.

"Since the functions of parental LRRK2 and its two splice variants are not known, we also constructed these two variants containing either the insertion or deletion."

A clone of interest can be easily identified using Rapid-Screen™ with a minimum of three rounds of PCR. An overview of the screening procedure is shown in Figure 1. Each panel is a series of arrayed cDNA library clones. The Master Plate for each library panel contains purified plasmid DNA from 500,000 cDNA clones divided into 96 pools and is initially screened in one simple 96-well PCR. This first round of PCR identifies the Sub-plate(s), which contains the clone of interest. The Sub-plates are *Escherichia coli* glycerol stocks contained in standard 96-well microtiter dishes. Each well contains 50 cDNA clones with approximately 5,000 clones per plate. The Sub-plates are screened in a second round of 96-well PCR to identify the positive well(s). Cells from a positive well(s) are then plated on LB/ampicillin agar and the resulting bacterial colonies are screened by PCR (third and final round) in order to obtain the desired clone. An overview of the screening procedure is as shown in Figure 1.

There are two key factors that determine the successful isolation of a long or rare transcript. One is the quality of the cDNA library; and the other is an efficient approach to fish out a full-length cDNA clone from a complex cDNA library of millions of members. cDNA library construction is a multiple step process and every step of a cDNA library construction, through its preparation of intact mRNAs, first-strand and second-strand cDNA synthesis, size fractionation of cDNA, and ligation and transformation, is biased against the formation of very long full-length clones. If a full-length cDNA clone were not present in the original cDNA library it would be impossible to isolate a full-length clone from that library no matter what approach you choose. From sequencing of 800,000 individual clones, OriGene has identified 24,000 unique

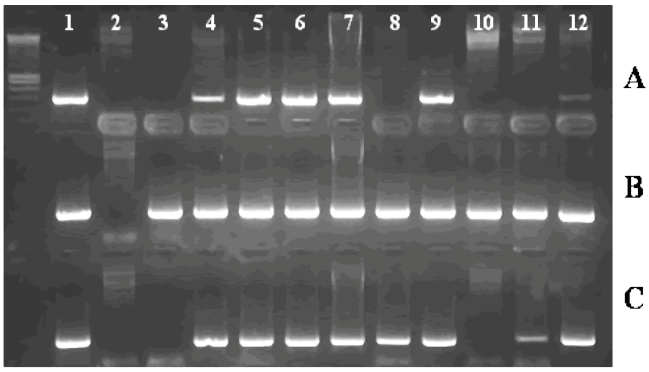


Figure 2. Lanes 1-12 represent twelve different RapidScreen cDNA libraries (1. Fetal Brain; 2. Adult Brain; 3. Heart; 4. Kidney; 5. Liver; 6. Lung; 7. Prostate; 8. Peripheral Blood; 9. Placenta; 10. Small Intestine; 11. Spleen; 12. Testis). Rows A, B and C are PCR reactions using 5', 3' and middle primers.

full-length cDNA clones to match the RefSeq sequence. This has proven the superior quality of our Rapid-Screen arrayed cDNA library panels.

Once a good library is in hand the next step is to consider how to pinpoint a cDNA and isolate

FEATURED GENE: The cloning of LRRK2

that full-length cDNA clone from a complex cDNA library in both an economical and timely fashion. Our RapidScreen approach enables us to identify and isolate a single full-length cDNA clone from multiple cDNA libraries. 20 of our cDNA libraries (10 million cDNA members) can be screened simultaneously and the full-length clones can be isolated within 2-3 days.

LRRK2 may play a central role to the pathogenesis of several major neurodegenerative disorders associated with parkinsonism. In order to develop new therapeutics, future research will have to focus on elucidating the function of LRRK2, for example the ultimate substrate of this kinase. To study the function and the consequences of the mutation on disease pathway, a full-length cDNA clone is absolutely essential.

To isolate a full-length clone of LRRK2, we designed three pairs of PCR primers covering the initiation site, the middle of the gene and the termination site respectively based on the RefSeq reference, XM_058513. Examination of 12 cDNA libraries using these three pairs of primers indicated that the full-length clones were present in a number of cDNA libraries (Figure 2). Surprisingly, the LRRK2 is not expressed in adult brain but it was in all other libraries examined. Next we chose human placenta cDNA library for the isolation of a full-length cDNA clone.

Through the master plate, subplate and colony screening, a single colony was isolated using 5' end primers. Sequencing of both ends and the insert size digestion (7.3 kb) indicated that this clone contains the complete 5' end, but the 3' is truncated (missing 380 bp) due to mispriming of oligo d(T) during first-strand cDNA synthesis. Instead of isolating another clone from other subplates of placenta library or other cDNA libraries, we decided to ligate this 7.3 kb clone with an overlapping 3' partial cDNA clone from our 800,000 individual clone collection. The full-length sequencing of the newly ligated clone showed that the insert is 9193 bp long and contains the complete ORF of AY792511. The 3' UTR also matched the predicted XM_058513. Blast analysis against our in-house 800,000 ESTs showed that a number of ESTs, mainly from breast cancer and bone marrow cDNA libraries, also matched the reference AY792511, but some of these ESTs contain either an insertion or deletion compared with published reference. Alignment with human genome indicated that the insertion and deletion are addition or deletion of independent exons. The in-frame insertion of 126 bp exon resulted in early termination due to two termination codons in the insertion. A deletion of 188 bp in the other clone causes frame shift and truncated proteins. Since the functions of parental LRRK2 and its two splice variants are not known we also constructed these two variants using the 9193 bp insert and these two ESTs containing either the insertion or deletion.

The technical limitation of any cDNA library is the representation of very long clones (for example >10 kb) in the library. Depending upon the composition and length of the cDNA, a full-length cDNA may not exist in a certain cDNA library. Our sequenced 800,000 high quality brute-force clones provided us unique resource to construct some of the very long and difficult cDNAs in the human genome. Like LRRK2, multiple splice variants can also be constructed using these ESTs.

"The technical limitation of any cDNA library is the representation of very long clones (for example >10 kb)."

Zairen Sun, PhD
V.P. Molecular Biology
OriGene Technologies, Inc

GENE FACTS: How many human genes and transcripts are there?

Hayashizaki et al. at the Genome Science Center at RIKEN is addressing this problem through the identification of transcription start points (4). However, the ultimate validation will always be found in the sequenced cDNA molecules. Secondly, despite all efforts, current technology is not able to sequence through 252 gaps in the genome representing approximately 37.2 million bp of euchromatic sequence (gene-rich regions). With an average gene size of 10-15kb and intergenic distance of 25-30kb, these gaps could easily produce another 1500 genes and gene variants. While the genome sequencing project has been a phenomenal success and has provided a superb template, the best representation of all human genes must come from the actual transcript copies themselves, cDNA.

"The current data still do not allow a total representative gene or transcript number to be reported, and reflect a current uncertainty in spite of major genomic advances in the last 5 years."

The Number of Human Transcripts

The NCBI, EBI and DDBJ institutions accept and freely exchange cDNA sequence information from many laboratories. Each program evaluates this information to curate their own non-redundant transcriptome database. At NCBI (http://www.ncbi.nlm.nih.gov/) this database is called RefSeq which currently contains 24,911 mRNA references annotated from overlapping cDNA sequences. ENSEMBL (www.ensembl.org), a joint project between EMBL-EBI and the Sanger Institute, has done a parallel analysis and reports that there are 24,194 gene loci producing 35,845 gene transcripts. Both of these data sets are similar in their analysis of the recent genome build, #35. Therefore, the correct answer must still be posted as a range. The number of gene loci is 22-24k and the number of transcripts resulting from the alternative splicing of these genes is 33-35k. The current data still do not allow a total representative gene or transcript number to be reported, and reflect a current uncertainty in spite of major genomic advances in the last 5 years.

A closer look at the quality of these reference sequences suggests that the numbers may be correct but the quality or authenticity of many transcripts remains suspect. Each of these transcripts has an associated coding sequence (CDS) and a recent combined effort has been launched to include quality assessments for each. The new Consensus CDS (CCDS) project reported in Entrez gene (http://www.ncbi.nlm.nih.gov/CCDS/) has released its first set of quality coding sequences for which there are currently 14,975 members from 13,142 genes. This means that a maximum of only 60% of RefSeq references have met the specified quality criteria, while the others remain suspect. This number of validated coding sequences is also significantly less than the 21,787 (x 1.54 variants per loci) predicted in the IHGSC Nature paper (3).

Only 30% of the RefSeq sequences have been qualified as Reviewed or Validated while the remaining are listed as Provisional, Predicted, or Inferred. This information, along with the CCDS data indicates that while the genome has allowed a fairly consistent assessment of the number of genes and transcripts, an accurate picture of the transcriptome can best be described as a work in progress. Another related effort is underway at the Weizmann Institute of Science (Rehovot, Israel). The GeneTide effort combines their GeneCards® with the clustering results of EST sequences and annotated full-length sequences from NCBI and ENSEMBL. While both of these institutions have developed refined clustering algorithms, neither has placed much effort in generating further annotation based on these clusters because of over-clustering and under-

GENE FACTS: How many human genes and transcripts are there?

clustering problems due to sequence repeat regions and partial EST sequences, respectively. However, if the clusters from both algorithms are in agreement then the quality of the assessment increases. The GeneTide effort hopes to wrap all this curated information into a single database that is made available at www.xennexinc.com.

Variants

It would be remiss not to include a discussion of gene variants, a term used broadly but with specific inferences. To be variant means to be different than the original. However, in the case of human genes, there is no original, only perhaps an initial cDNA form that garnered the first accession number. While some forms are more prevalent than others, the variant-specific expression levels are rarely incorporated within public annotation.

In the world of genome sequencing, a variant is inferred to mean a single or simple nucleotide polymorphism (SNP). These are natural differences between the genomes of individuals that may or may not fall within a coding exon or within a gene region altogether. Such variations are not usually accorded a separate transcript record, but are recorded in SNP databases. None-the-less, such variations do cause amino acid substitutions or frame shifts. The linking of these protein-altering changes to human disease is clearly of major interest for purposes of both diagnostics and treatment. A complete understanding of the biological significance of these millions of SNP-variants will take decades of research. In the meantime, it is a truth that all molecular biologists must accept and embrace. Considering the prevalence of human SNPs at 0.1% within 33,000 transcripts having average lengths of 2.5kb, the human transcriptome would have more than 82,500 differences to investigate (2.5 per transcript on average). In fact, there are many more than this already reported from the results of quality sequence clustering.

In the world of cDNA cloning and sequencing, a variant is more often inferred to represent a splice variant arising from a single gene. Each gene is typically comprised of 8 exons. The inclusion of all exons is regulated by a multi-component spliceosome that recognizes splice boundaries, removes the exonic DNA and splices them together. Sometimes the spliceosome is directed to leave out 1 or more exons, thereby creating a different transcript. Certain polymorphisms can influence this process causing a prevalence of one form, which might lead to certain disease phenotypes (2). Because the alternative splice forms usually impose significant changes to the coded protein, each spliced transcript is accorded its own accession number within NCBI's RefSeq.

cDNA Collections: Send in the Clones

While several groups have been active in the creation of clone collections, there is a distinct difference between them. Synthesized clones run the risk of creating an artificial clone that is not biologically relevant, especially since the world's best transcript reference databases are constantly updating reference sequences based on the prevailing sequence data. This is a risky and potentially costly adventure. Prime examples are the collections of PCR products (ORF clones). While the mis-incorporation rates of the best polymerases have improved significantly, the potential for error will always exist. Even more startling is the error rates encountered within the PCR primers themselves. In fact, a frame-shifting insertion or deletion is much more frequently encountered in these regions than in the amplified region between them. This was shown in the prepara-

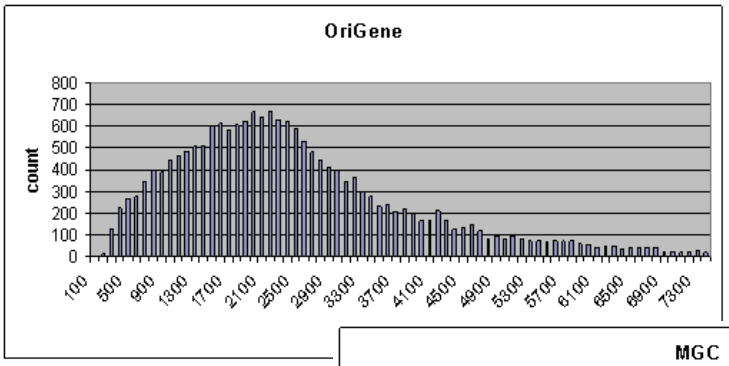
"While the genome sequencing project has been a phenomenal success and has provided a superb template, the best representation of all human genes must come from the actual transcript copies themselves, cDNA."

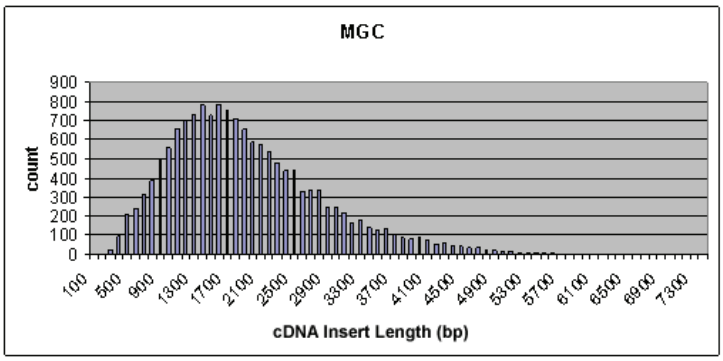
GENE FACTS: How many human genes and transcripts are there?

"OriGene's TrueClone Collection covers 80% of all cDNA clones available worldwide from commercial sources and contains 65% unique clones unavailable from any other single source"

tion of the ORFeome v1.1 collection created at the Dana Farber Cancer Institute, where more than 8000 cDNA transcripts were PCR sub-cloned. One can only imagine the statistical number of errors resulting from the PCR amplification from crude cDNA libraries where the PCR cycle numbers must be increased to compensate for the rarity of the target molecules (5).

To avoid PCR-related mutations, it is best to obtain a transcript clone directly from a cDNA library. There are only two large human clone collections that offer primarily cDNA clones, the Mammalian Gene Collection (MGC), a NIH-funded effort and the TrueClone Collection available from OriGene Technologies, Inc. The current MGC collection has full sequence validation of its members, but the current number of its non-redundant clones is just over 10,000 (5). To make up for a lack in content, the MGC decided in 2004 to move forward by cloning and adding PCR clones, thus leaving OriGene's TrueClones as the single source for genome wide cDNA clones. This collection offers 24,000 transcripts from over 15,000 genes. This covers 80% of all the clones that are commercially available worldwide. Methods like Rapid-Screen highlighted in the LRRK2 article allow OriGene to move forward by cloning authentic cDNA clones for gene-focused research or genome wide screening.





Some might say that the clones in a collection are only as good as the references that they represent. This might be true for PCR-generated or synthesized clones, but TrueClones represent their own independent reference set. As cDNA copies obtained directly from cDNA libraries, they are as close to representing actual mRNA transcripts as current technology allows. It will take many years before the RefSeq database contains all the individual references for all the splice variants and significant polymorphic transcripts. It is for this reason that all cDNA clones today should be considered biologically significant until proven otherwise.

Historical Perspectives

In the late 1990's, Human Genome Sciences, Inc. and Incyte emerged as the leaders in genome-wide transcript discovery through the partial end-sequencing of clones, a.k.a. ex-

GENE FACTS: How many human genes and transcripts are there, now that the human genome is finished?

pressed sequence tags (EST). Original human gene estimates from these programs was on the order of 100,000 to 120,000 non-redundant genes, which spurred the genomics investment bubble. In September 2000, Celera added to the complexity by releasing its single (simple) nucleotide polymorphism (SNP) database that featured 2.8 million possible variant possibilities from roughly five donors.

A few months later this transcriptome complexity came into focus with the publication of the annotated genome in February 2001. The IHGSC quoted 24,500 genes and 30-40,000 total variant transcripts, while Celera quoted 26,588 genes plus 12,000 predicted genes totaling 38,588 genes. An in-depth comparison of the two annotations by Michael Cooke and John Hogenesch of the cutting edge Genomics Institute of Novartis Foundation revealed that only 16,000 genes were in common bringing the gene estimate to about 42,000. Those in common between the sets were believed to be correct while the others remained suspect.

After the dust settled, it was time to announce the Cold Spring Harbor contest winner who came closest to guessing the total number of genes at a time when it was generally assumed to be in the 6-digit range. In 2003, Ewan Birney of the EBI awarded the honor to three individuals: Lee Rowen (Inst for Systems Biology; 25,947), Olivier Jaillon (Genoscope; 26,500), and Paul Dear (UK Medical Research Council; 27,462).

As work on the human genome continued, October 2004 marked the publication of build 35 annotation which should stand as the working template for some time. The key statistics boast that 99.9 percent of the euchromatic DNA is covered by quality sequence (one error per 100,000; PHRED >50). Francis Collins initiated a HapMap program in the same month that looks at all human genetic variation, both inside and outside ancestral populations. In much the same vein, the Venter Institute embarked on a human genome re-sequencing project aimed to expand the Celera coverage by sequencing the genomes from other donors in order to generate more SNPs.

Conclusion

The complexity of the human transcriptome is highlighted in three aspects. The first is the primary number of genes which is now focused near 23,000 but may grow by another few thousand as individual chromosomes are completely annotated. The X-chromosome was recently reported (6) to have an additional 399 confirmed novel genes that were not previously reported, and there are another 100 evolutionarily conserved regions (ECRs) that were found as unconfirmed novel genes. Secondly, on average each of these 23,000+ genes produces 1-2 splice variant transcripts. Each splice variant may code for a different protein, confer stability to the mRNA, or have no biological effect. Lastly, the existence of millions of SNPs genome wide or 100,000 SNPs within expressed transcripts themselves may serve to bring the number of true independent transcripts in the transcriptome back to the original estimates (23,000 genes x 1.54 splice variants x 2.5 SNPs = **88,550**).

References

(1) Identification of the Wnt signaling activator leucine-rich repeat in Flightless interaction protein 2 by a genome-wide functional analysis, J. Liu, et al. P.N.A.S. (2005)

(2) Tissue-Specific Reduction in Splicing Efficiency of IKBKAP Due to the Major Mutation Associated with Familial Dysautonomia. M.P. Caujungco, et al. Am. J. Hum. Genet. (2003)

(3) Finishing the Euchromatic Sequence of the Human Genome. F.S. Collins, et al. Nature (2004)

(4) Cap Analysis Gene Expression for High-Throughput Analysis of Transcriptional Starting Point and Identification of Promoter Usage. T Shuraki et al., Genetics (2003)

(5) Human ORFeome Version 1.1: A Platform for Reverse Proteomics. JF Rual et al., Genome Research (2004)

(6) The DNA sequence of the Human X Chromosome. M Ross et al., Nature (2005)

Karl Kovacs, PhD
V.P. Alliance Management
OriGene Technologies, Inc
cdna@origene.com

HuSH: Pre-made shRNA Panels for long-term inhibition of protein expression

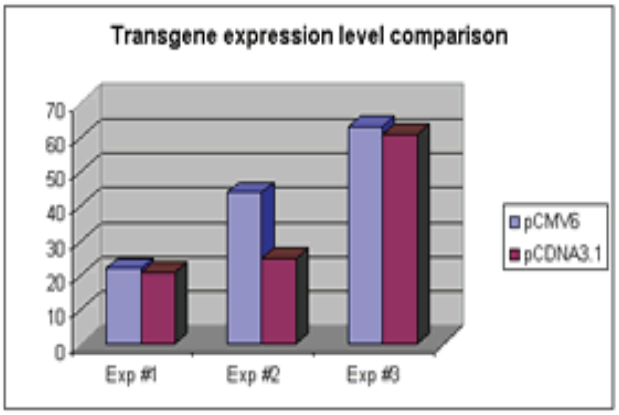
In the course of your research to validate new targets, you will have need of tools to stimulate as well as inhibit protein expression. For the former, OriGene has created the TrueClone cDNA genes in pCMV vectors for uniform mammalian expression. As a complement to this tool, OriGene has also created a set of tools for the inhibition of protein expression. These ready-made shRNA Expression Plasmids save you Time and Labor on Vector Design and Construction and are available with corresponding negative and positive controls.

shRNAs Targeting the Following Families of genes are currently available in our catalog

- Over three hundred (300) Protein Kinases
- Over forty (40) nuclear Hormone Receptors
    - Drug-resistance Marker for Stable Integration Selection
    - Retroviral Backbone for Efficient Generation of Stable Cell Line
    - Amplifiable Supply--Once Ordered, Never Order Again
- Including Key Regulator Genes Like P53, PTEN, TLR2, TLR3, TLR4 and TERT
- Corresponding Full-Length cDNA Clone Available from OriGene
- Corresponding Validated Antibodies Available from OriGene

pCMV Vector provides highest transgene expression

Extensive work has been done to engineer a vector to achieve the highest level of transgene expression level. When compared with another popular expression plasmid, pCDNA3.1 (Invitrogen), pCMV-based plasmids provide comparable if not higher levels of transgene expression. In three independent experiments, a same quantity of plasmid DNA was transfected into COS1 cells and the CAT activity was scored.



Comparison of transgene expression levels in pCMV6- and pCDNA3.1-based plasmids. CAT gene was cloned downstream of the promoters in the pCMV6 and pCDNA3.1 vectors.

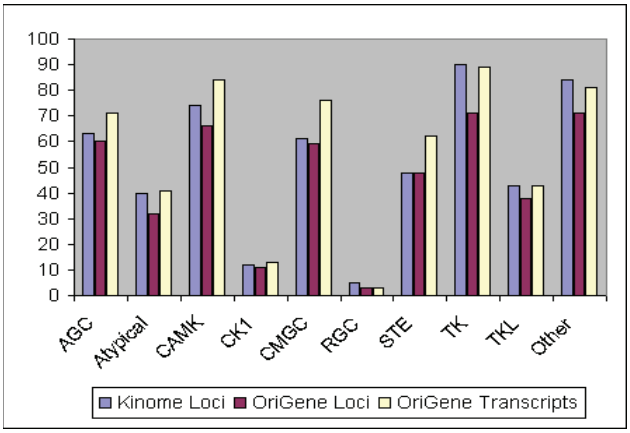
Key Functional Features of pCMV vectors:

- **Promoters for *in vivo* expression in mammalian cells and *in vitro* cell free systems**
- **Cell lines suitable for transfection:** COS, 293, Hela, CHO, NIH3T3, Mouse L cell, etc.
- **Transcription termination and polyadenylation signals**: from human growth hormone (hGH) gene

The Kinome: Protein Kinase CloneSet™

With the completion of the human genome projects, scientists have been able to catalogue the "kinome" - the entire family of kinase genes in the human genome. According to a recent survey (Manning, G. et al., Science 298:1912, 2002), the human "kinome" (complete kinase collection) contains 518 gene loci, with multiple alternative splicing variants for some of the loci, and represents 1.7% of the human gene repertoire.

The OriGene Kinase CloneSet consists of 564 full-length human protein kinase cDNA clones representing 460 gene loci. All genes in the Kinase CloneSet are cloned in expression vectors directly downstream of a CMV promoter and are ready for protein expression in mammalian cells.

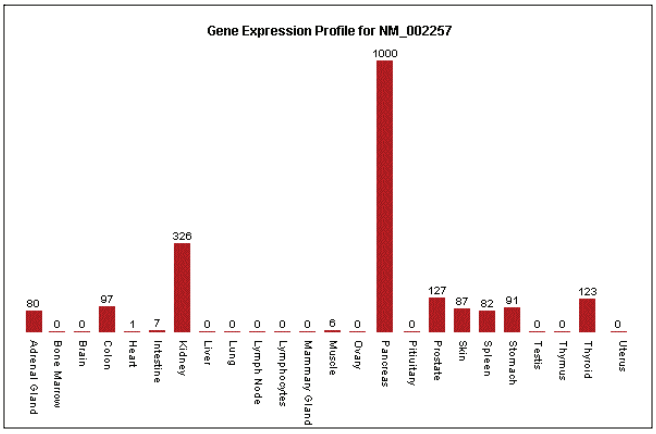


OriGene Coverage of the Human Kinome

Rapid-Scan™ Expression Panels

High-Quality cDNA from normal and diseased tissue, arrayed and normalized to actin mRNA message levels for gene expression profiling. These easy to use panels can be screened for your gene of interest in less than three (3) hours. They are convenient and well cited:

- MYND-less splice variants of AML1-MTG8 (RUNX1-CBFA2T1) are expressed in leukemia with t(8;21) **Genes, Chromosomes & Cancer, 2005**
- Expression of Constitutive Androstane Receptor Splice Variants in Human Tissues and Their Functional Consequences **J. Pharmacol. Exp. Ther., Nov 2004**
- Human Tribbles, a Protein Family Controlling Mitogen-activated Protein Kinase Cascades, **J. Biol. Chem., Oct 2004**



OriGene has published thousands of Rapid-Scan expression profiles on its website (e.g. KLK1; Kallikrein-1)

OriGene Technologies

OriGene Technologies, Inc.
Six Taft Court, Suite 100
Rockville, MD 20850

Phone: 301-340-3188
Fax: 301-340-8606
Web: www.origene.com

The most complete collection of human full-length cDNA clones in the world

ORIGENE
Your Gene Company

Six Taft Court, Suite 100
Rockville, MD 20850 USA

PRSRT STD
U.S. POSTAGE
PAID
HMS PRINTING