

FERNANDO LOPES ALBERTO

**Avaliação do Transcriptoma da Leucemia Mielóide Crônica
por ORESTES (*Open Reading Frame Expression Sequence Tags*)**

Tese de doutorado apresentada à Comissão de Pós-graduação da Faculdade de Ciências Médicas da Universidade Estadual de Campinas para obtenção do título de Doutor em Clínica Médica, área de concentração, Clínica Médica.

CAMPINAS

2002

FERNANDO LOPES ALBERTO

**Avaliação do Transcriptoma da Leucemia Mielóide Crônica
por ORESTES (*Open Reading Frame Expression Sequence Tags*)**

Orientador: Prof. Dr. Fernando Ferreira Costa
Faculdade de Ciências Médicas da Universidade
Estadual de Campinas

Campinas
Faculdade de Ciências Médicas da Universidade Estadual de Campinas

2002

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA FACULDADE DE CIÊNCIAS MÉDICAS DA UNICAMP**

Bibliotecário: Sandra Lúcia Pereira – CRB-8ª / 6044

AL 14a	<p>Alberto, Fernando Lopes</p> <p>Avaliação do transcriptoma da leucemia mielóide crônica por ORESTES (Open Reading Frame Expression Sequence Tags) / Fernando Lopes Alberto. Campinas, SP : [s.n.], 2003.</p> <p style="text-align: center;">Orientador : Fernando Ferreira Costa</p> <p style="text-align: center;">Tese (Doutorado) Universidade Estadual de Campinas. Faculdade de Ciências Médicas.</p> <p style="text-align: center;">1. Reação em Cadeia da Polimerase Via Transcriptase Reversa. 2. Expressão gênica . 3. Leucemia Mielóide Crônica. I. Costa, Fernando Ferreira. II. Universidade Estadual de Campinas. Faculdade de Ciências Médicas. IV. Título.</p>
--------	--

Título em ingles : CML Transcriptome analysis using ORESTES (Open reading frame Expressed Sequence Tags)

Keywords: • Reverse Transcriptase Polymerase Chain Reaction
• Gene Expression
• Leucemia Milóide Crônica

Área de concentração : Clínica Médica

Titulação: Doutorado

Banca examinadora: Prof Dr Fernando Ferreira Costa
Profa. Dra. Maria de Lourdes Ferrari Chauffaille
Prof Dr Luis Eduardo Coelho de Andrade
Prof Dr Gonçalo Amarante Guimarães Pereira
Prof Dr Cármino Antonio de Souza

Data da defesa: 14 /03/2003

Banca Examinadora da Defesa de Tese de Doutorado

Orientador(a): Prof. Dr. Fernando Ferreira Costa

Membros:

1. Profa. Dra. Maria de L^ordes Ferrari Chauffaille _____

2. Prof. Dr. Luís Eduardo Coelho de Andrade _____

3. Prof. Dr. Gonçalo Amarante Guimarães Pereira _____

4. Prof. Dr. Cármino Antonio de Souza _____

Curso de Pós-Graduação em Clínica Médica, área de concentração Clínica Médica, da Faculdade de Ciências Médicas da Universidade Estadual de Campinas.

Data: 14/02/2003

200725895

Dedico especialmente este trabalho

Ao Gustavo e à Isabela, meus filhos.

À Daniella, minha esposa

Juracy e Maria, meus pais.

Agradecimentos

Ao Fernando, meu orientador, que com seu exemplo tem ensinado valores científicos e médicos aos seus alunos, mas que acima de tudo é um homem de caráter irretocável.

Ao Mauro S. Figueiredo pelas oportunidades e postura crítica em vários momentos de minha vida recente, contribuindo sobremaneira para meu crescimento.

Ao Gonçalo Amarante Guimarães Pereira, pelo apoio intelectual em várias das estratégias computacionais empregadas nos momentos embrionários do projeto.

Ao Tarcísio de Souza Peres, “bioinformata”, pelo companheirismo, comprometimento excepcional e importante contribuição para a realização do trabalho.

À Sarita, que também me acolheu no Hemocentro e foi fundamental para que eu me estabelecesse em Campinas.

À Dulcinéia que sempre ofereceu valiosa ajuda nos protocolos experimentais.

Ao André, pela ajuda com pacientes e amostras.

Aos colegas de laboratório, Camila, Anderson, Adriana, Simone, Daniela, Patrícia, Flávia, Ângela, Lena, Natasha, Tereza e Heloísa, pelas amostras, sugestões, apoio incondicional e, principalmente, pela convivência agradável.

A todo o grupo médico do Hemocentro, da Faculdade e do Hospital das Clínicas, particularmente: Margareth, Kátia, Gislaine, Carmen, Malu e Vagner, que conviveu comigo durante esses anos e que contribuiu para o meu amadurecimento.

À Patrícia pela pronta, bem-humorada e sincera ajuda em várias oportunidades.

Aos meus amigos Sérgio, Omar, Teresa, Rendrik, pelo apoio, incentivo e críticas oportunas.

A todos os outros que, de uma forma direta ou indireta, colaboraram e participaram, o meu muito obrigado.

	<i>Pág.</i>
RESUMO	xxvii
ABSTRACT	xxxí
1- INTRODUÇÃO	35
1.1- Câncer e Genômica.....	37
1.2- Leucemia Mielóide Crônica.....	40
2- OBJETIVOS	45
3- MATERIAIS E MÉTODOS	49
3.1- Bioinformática.....	51
3.2- Amostras Biológicas.....	53
3.3- Extração do RNA Total.....	54
3.4- Avaliação de Qualidade do RNA Isolado.....	55
3.5- Transcrição Reversa do mRNA e PCR Quantitativo em Tempo Real (Real-Time PCR).....	56
3.6- Genes Seleccionados para Confirmação de Expressão com Real-Time PCR.....	58
3.7- SAGE – Serial Analysis of Gene Expression (“Análise Seriada da Expressão Gênica”).....	60
4- RESULTADOS	63
4.1- Avaliação por ORESTES.....	65
4.2- Avaliação da Expressão Gênica por SAGE.....	70
4.3- Validação da Expressão Gênica por Real-Time PCR Quantitativo.....	70

5- DISCUSSÃO	87
6- CONCLUSÕES	99
7- REFERÊNCIAS BIBLIOGRÁFICAS	103
8- ANEXOS	115
Anexo A: Termo de Consentimento Informado.....	117
Anexo B: Tabela Comparativa do Número de Seqüências por Cluster ORESTES com o Número de SAGE Tags.....	119
CT	PCR em que a fluorescência é detectada acima ruído de fundo)
cDNA	DNA, cadeia ou fita complementar
DNA	ácido desoxirribonucleico
Dnase	desoxirribonuclease
DO	densidade óptica
EST	expressed sequence tags
FDA	Food and Drug Administration
Gb	(informática) gigabyte, ou um bilhão de bytes
Laser	light amplification by stimulated emission of radiation (amplificação de luz pela emissão estimulada de radiação)
LMC	leucemia mieloide crônica
LPS	lipopolissacáride
M	molar
µg	micrograma
mL	mililitro
mM	milimolar
mRNA	RNA mensageiro

LISTA DAS ABREVIATURAS

ACTH: adrenocorticotrofina

b: (informática) byte

blast: *Basic Local Alignment Search Tool*

cf.: conferir

CT: *threshold cycle* (“ciclo de limiar”, é o ponto da reação de real-time PCR em que a fluorescência é detectada acima ruído de fundo)

cDNA: DNA, cadeia ou fita complementar

DNA: ácido desoxirribonucléico

Dnase: desoxirribonuclease

DO: densidade óptica

EST: *expressed sequence tags*

FDA: Food and Drug Administration

Gb: (informática) gigabyte, ou um bilhão de bytes

Laser: *light amplification by stimulated emission of radiation*
(amplificação de luz pela emissão estimulada de radiação)

LMC: leucemia mielóide crônica

LPS: lipopolissacáride

M: molar

g: micrograma

mL: mililitro

mM: milimolar

mRNA: RNA mensageiro

nM: nanomolar

NTC: non template control (controle sem DNA)

ORESTES: *Open Reading frame ESTs*

pb: par de bases

PCR: *polimerase chain reaction* (reação em cadeia da polimerase)

Relação L:E: relação entre o número absoluto de leucócitos e o de eritroblastos contados no mielograma

RNA: ácido ribonucléico

Rnase: ribonuclease

SAGE: *Serial analysis of gene expression*

SCSI: *Small Computer Standard Interface*

(http://www.stepbystep.nl/technischediensst/tech/pages/tech_info/scsi/scsi.htm).

SNP: *single nucleotide polymorphism* (polimorfismo de única base)

U: unidade(s)

	<i>Pág.</i>
Tabela 1 Relação dos <i>primers</i> utilizados para confirmação da expressão gênica aferida por ORESTES.....	60
Tabela 2 Estatística resumida do número de seqüências por <i>contig</i>	66
Tabela 3 Distribuição percentual dos <i>contigs</i> de acordo com sua categoria funcional.....	67
Tabela 4 Distribuição dos <i>contigs</i> ORESTES de LMC no genoma.....	69
Tabela 5 Resultados de expressão dos 10 genes selecionados por real time PCR em LMC comparativamente ao tecido normal.....	71
Tabela 6 Dados individuais da análise de expressão por real-time PCR para cada um dos genes avaliados.....	72
Figura 2 Classificação funcional dos <i>contigs</i> gerados por ORESTES em LMC.....	66
Figura 3 57% dos <i>contigs</i> analisados foram mapeados no genoma humano.....	68
Figura 4 Expressão de CAMF em células de LMC e controles avaliada por real-time PCR quantitativo.....	75
Figura 5 Expressão de GNAI2 em células de LMC e controles avaliada por real-time PCR quantitativo.....	76
Figura 6 Expressão de ABR em células de LMC e controles avaliada por real-time PCR quantitativo.....	77
Figura 7 Expressão de DEFA3 em células de LMC e controles avaliada por real-time PCR quantitativo.....	78

LISTA DE FIGURAS

	Pág.
Figura 1	Representação espacial das taxas brutas de incidência por 100.000 homens estimadas para o ano de 2002, segundo a Unidade da federação (Leucemias)..... 42
Figura 2	Eletroforese em tampão formamida em gel de agarose e corada por brometo de etídio de amostras de RNA total isoladas para o estudo..... 55
Figura 3	Monitoramento em tempo real das curvas de amplificação e identificação do nível de fluorescência na fase ótima (aumento exponencial da fluorescência) da reação de amplificação..... 57
Figura 4	Representação esquemática da técnica de SAGE empregada nesse trabalho..... 61
Figura 5	Classificação funcional dos <i>contigs</i> gerados por ORESTES em LMC..... 66
Figura 6	87% dos <i>contigs</i> analisados foram mapeados no genoma humano..... 68
Figura 7	Expressão de CAMP em células de LMC e controles avaliada por real-time PCR quantitativo..... 75
Figura 8	Expressão de GNAI2 em células de LMC e controles avaliada por real-time PCR quantitativo..... 76
Figura 9	Expressão de ABR em células de LMC e controles avaliada por real-time PCR quantitativo..... 77
Figura 10	Expressão de DEFA3 em células de LMC e controles avaliada por real-time PCR quantitativo..... 78

Figura 11	Expressão de TPT1 em células de LMC e controles avaliada por real-time PCR quantitativo.....	79
Figura 12	Expressão de CSF3R em células de LMC e controles avaliada por real-time PCR quantitativo.....	80
Figura 13	Expressão de LTF em células de LMC e controles avaliada por real-time PCR quantitativo.....	81
Figura 14	Expressão de MLL2 em células de LMC e controles avaliada por real-time PCR quantitativo.....	82
Figura 15	Expressão de S100A8 em células de LMC e controles avaliada por real-time PCR quantitativo.....	83
Figura 16	Expressão de MS4A3 em células de LMC e controles avaliada por real-time PCR quantitativo.....	84
Figura 17	Curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de b-actina (ACTB).....	85

RESUMO

RESUMO

ALBERTO, F.L. **Avaliação do Transcriptoma da Leucemia Mielóide Crônica por ORESTES (*Open Reading Frame Expression Sequence Tags*)**. Campinas, 2002. Tese (doutorado) - Faculdade de Ciências Médicas, Universidade Estadual de Campinas.

Câncer é o resultado de anomalias no genoma acumuladas ao longo do tempo e corresponde a doença genética mais freqüente na população humana. Os últimos anos têm presenciado enormes avanços tecnológicos que facilitaram o desenvolvimento de técnicas para análise em larga escala do conjunto dos genes expressos nos diferentes tecidos, sendo responsável pela explosão do conhecimento em medicina molecular. No presente trabalho avaliamos o transcriptoma da leucemia mielóide crônica através da técnica de ORESTES (*Open Reading Frame Expressed Sequence Tags*). As seqüências geradas foram agrupadas de acordo com sua homologia através do pacote de bioinformática Phrap. Essa estratégia aumentou o grau de cobertura genômica, já que as seqüências originais apresentavam tamanho máximo de 500 a 700 bases. Após a retirada dos contaminantes, obtivemos 2933 agrupamentos de seqüências (contigs) e 11568 seqüências isoladas. Cada contig representa um gene em potencial e o seu conjunto foi submetido a comparação com o banco de dados público *nr* para possibilitar a anotação e classificação funcional. Os agrupamentos ORESTES foram classificados em comunicação celular (19%); crescimento, desenvolvimento e manutenção celular (9%); metabolismo celular (30%); estrutura, motilidade e transporte celular (15%); função desconhecida (28%). Adicionalmente, estabelecemos o mapeamento cromossômico dos contigs obtidos, observando distribuição relativamente homogênea e proporcional representada em todos os cromossomos humanos. O mapeamento dos contigs gerados por ORESTES apresentou-se com padrão de distribuição semelhante ao observado por Venter e colaboradores, 2001. Comparamos o padrão de expressão obtido com o transcriptoma de medula óssea de paciente portador de anemia hemolítica hereditária (doença hematológica não neoplásica) gerado por SAGE (*Serial Analysis of Gene Expression*). Os genes encontrados foram ordenados de acordo com sua representatividade ou nível de expressão. Não houve correlação entre os padrões de expressão observados por SAGE e por ORESTES. Do conjunto dos 100 genes mais representados nos ORESTES, dez foram selecionados para validação da expressão por PCR quantitativo em tempo real, utilizando o agente fluorescente SybrGreen I. Identificamos e confirmamos o aumento da expressão em nove genes na LMC. Destes, sete tinham aumento superior a 10 vezes o controle (CAMP, ABR, DEFA3, CSF3R, MLL2, S100A8 e MS4A3), com dois deles (DEFA3 e ABR) com aumento de cerca de 50

vezes ou mais. A análise funcional particularizada para cada um destes genes em LMC tem o potencial de identificar possíveis alvos para novas intervenções terapêuticas. O método mostrou-se poderoso para evidenciar e validar as diferenças de expressão gênica observadas pelas técnicas de seqüenciamento utilizadas anteriormente. Os dados deste trabalho constituem a primeira descrição do transcriptoma da leucemia mieloide crônica.

SUMMARY

ALBERTO, F.L. **CML Transcriptome Analysis Using ORESTES (Open Reading Frame Expression Sequence Tags)**. Campinas, 2002. Tese (doutorado) - Faculdade de Ciências Médicas, Universidade Estadual de Campinas.

The complete collection of transcripts generated from the human genome cannot be predicted from the genome sequence, but should be directly determined for each tissue, due to variations of gene expression in different tissues and disease states, and because genes can encode multiple transcripts derived from alternate splicing and polyadenylation sites. As part of larger project that produced up to 1.2 million ORESTES (open reading frames expressed sequence tags) from different cancer tissues, we constructed a set of cDNAs obtained from bone marrow cells of patients with CML, that represented partial expressed gene sequences that are biased toward the central coding regions of the resulting transcripts (Dias-Neto E et al, Proc Nat Acad Sci USA 97:3491, 2000). The 27,361 ESTs were assembled into contigs using the Phrap package producing 2933 clusters containing 2 to 210 ESTs (leaving 11568 isolated sequences). Functional categories were attributed to the contigs on the basis of the annotation of the matched sequences in *nr*, into 5 major functional categories and 22 subcategories (cell communication (19%); cell growth, development and maintenance (9%); metabolism (30%); cell motility, structure and transport (15%); unknown function (28%). The abundance of ESTs that matched the contigs formed by the larger number of EST in bone marrow cells was compared with a control bone marrow sample obtained from a non-neoplastic hematologic condition (hereditary hemolytic anemia) and evaluated with SAGE (serial analysis of gene expression). Ten out of 100 most representative contigs had their expression data evaluated further by real-time quantitative PCR with Sybr Green I reagent, confirming the ORESTES data. Seven out of the 10 genes were more than 10-fold upregulated (CAMP, ABR, DEFA3, CSF3R, MLL2, S100A8 e MS4A3) and two additional genes were 50-fold or higher upregulated in CML (DEFA3 e ABR). The wealthy of information provided by this approach demonstrates its usefulness for the analysis of gene expression in specific hematopoietic tissues and diseases.

1. Introdução

1.1 Câncer e Genômica

Câncer pode ser entendido como o resultado de anomalias no genoma acumuladas ao longo do tempo, capazes de causar comportamento fenotípico anômalo da célula envolvida. Corresponde à doença genética mais freqüente na população humana e dados recentes revelam que mais de cinco milhões de pessoas irão morrer de câncer por ano, um número apenas superado pelas doenças cardiovasculares (causa de 30% de todas as mortes) e por doenças infecto-parasitárias (19% de todas as mortes, excluindo-se as causas externas) (FUTREAL et al, 2001; PARKIN, 2001). No mundo, em 1995 foram gastos com o tratamento de câncer nos EUA cerca de US\$ 41 bilhões, o dobro da quantia investida dez anos antes (NCI, 2001), dos quais mais de US\$ 1 bilhão foram direcionados para o tratamento de leucemias de uma forma geral. Adicionalmente, em países em desenvolvimento, estima-se um aumento da incidência de câncer da ordem de 71% de 1990 até o ano de 2010, em virtude do envelhecimento da população (PISANI et al, 1999). Esses números não só ressaltam a importância do câncer como problema de saúde pública, como apontam para a necessidade de maiores e mais eficazes medidas para acelerar o diagnóstico, definir os grupos prognósticos diferentes e individualizar a terapêutica, diminuindo morbidade e custos.

A última década presenciou grande explosão do conhecimento em medicina molecular, um fenômeno cujos ícones podem ser identificados há mais de duas décadas: em 1982, por exemplo, a insulina humana

recombinante foi aprovada pelo FDA para utilização em seres humanos e desde então mais de uma centena de drogas produzidas com a tecnologia de DNA recombinante está disponível para a aplicação clínica (COLLINS, 1999). A publicação das primeiras versões do genoma humano na transição do século XX para o XXI (LANDER et al, 2001; VENTER et al, 2001) constitui um divisor de águas no conhecimento das vias metabólicas celulares complexas e de seus desvios, culminando com o fenômeno neoplásico (COLLINS, 1999). Abordagens metodológicas para estudo molecular de processos bioquímicos singulares cedem lugar para tecnologias que permitem avaliação de dezenas de milhares de vias metabólicas simultaneamente (GUTTMACHER e COLLINS, 2002), facilitando o entendimento não só no processo de carcinogênese, mas também de variáveis como o prognóstico e até mesmo a classificação funcional dos tumores (ALIZADEH et al, 2000; DHANASEKARAN et al, 2001; FERRANDO et al, 2002; GOLUB et al, 1999; VAN DE VIJVER et al, 2002; YEOH et al, 2002).

Entre as abordagens genômicas de comparação dos perfis de expressão gênica global de tecidos normais e malignos estão o seqüenciamento em larga escala de ESTs (ADAMS et al, 1991; STRAUSBERG et al, 2000) e variantes como ORESTES. Projetos de seqüenciamento de genes que utilizavam de técnicas clássicas de geração de ESTs apresentavam viés na informação gerada, que era predominantemente oriunda das extremidades 5' (2/3) e 3' (1/3) dos genes expressos, com relativa escassez de informação das porções centrais dos

genes (DE SOUZA et al, 2000; DIAS-NETO et al, 2000). ORESTES (*open reading frame EST*) é baseado na geração de cDNA com a utilização de *primers* aleatórios e arbitrários, porém não degenerados, em condições de baixa estringência na reação de transcrição. Nessas condições, uma determinada molécula de mRNA apresentará probabilidade significativamente maior de receber ligação dos oligonucleotídeos nas suas porções mais centrais do que nas extremidades, com conseqüente retrotranscrição preferencial dessa região na molécula de cDNA (DE SOUZA et al, 2000; DIAS-NETO et al, 2000; NETO et al, 1997). No entanto, para análise comparativa de expressão gênica, ORESTES tem a desvantagem teórica de favorecer a normalização experimental dos mRNA transcritos em menor quantidade em relação aos mais abundantes (CAMARGO et al, 2001; DIAS et al, 2000b). Essa característica permite que haja uma relativa facilidade de identificação de mensagens mais raras em determinada amostra, o que torna o método particularmente útil para a descoberta de novos genes ou para investigação de amostras escassas, com conteúdo reduzido de RNA (CAMARGO et al, 2001; DIAS et al, 2000a; NETO et al, 1997). Seja qual for o método empregado para a geração de ESTs, informações valiosas são obtidas para o conhecimento genômico, definindo os genes ativos no organismo ou tecido, evidenciando isoformas de determinados genes que seriam o fruto de maneiras alternativas de processamento do RNA (*alternative splicing*) e imprimindo velocidade tanto na análise experimental, como na análise *in silico*, já que, nos organismos

superiores, os genes expressos representam 3 a 5% do conteúdo de DNA de uma célula.

Métodos como SAGE (*Serial Analysis of Gene Expression*, Lal et al., 1999; Velculescu et al., 1999) e *microarrays* de cDNA e de oligonucleotídeos (SCHENA et al, 1995) são ainda mais poderosos na característica de fornecer elementos comparativos para a expressão gênica entre duas ou mais situações. A melhoria tecnológica e a redução de custos evidentes, não somente nas máquinas de produção e hidridização dessas matrizes de genes – os *microarrays* - como também nas ferramentas de análise dos dados gerados tem popularizado sua utilização em todo o mundo. Até janeiro de 2003, uma busca no site do Pubmed utilizando-se as palavras-chave “*microarray*” e “*cancer*” fornecia 829 entradas para referências bibliográficas, um número cerca de oito vezes maior do que o encontrado em dezembro de 2001. SAGE, ao contrário dos *microarrays*, tem a vantagem de permitir a caracterização da expressão gênica em determinado tecido dispensando o conhecimento prévio das seqüências dos genes envolvidos. Entretanto, por ser extremamente laboriosa, essa técnica impossibilita que sejam analisadas um grande número de amostras, e seus resultados geralmente fazem referência a uma amostra na situação controle (tecido normal, por exemplo) contra uma amostra característica de determinado processo biológico (neoplasia, por exemplo) (BOON et al, 2002; VELCULESCU et al, 1995; VELCULESCU et al, 2000b).

Particularmente em câncer, metodologias de análise genômica proporcionam a identificação de numerosos processos biológicos que têm

caracterizado, à maneira de uma impressão digital, os diferentes tumores que afetam o ser humano tanto quanto ao seu diagnóstico e prognóstico (DHANASEKARAN et al, 2001; FERRANDO et al, 2002; GOLUB et al, 1999; HERNAN et al, 2003; REIS e VERJOVSKI-ALMEIDA, 2000; RISINGER et al, 2003; VAN DE VIJVER et al, 2002; YEOH et al, 2002; ZHANG et al, 2002), como quanto a possíveis alvos para terapia (OKUTSU et al, 2002; SUZUKI et al, 2003; WALLQVIST et al, 2002). Para a maior parte dos métodos de análise genômica é possível a detecção de cerca de 1% de genes com expressão diferencial comprovada por segundo método “padrão-ouro” (como *northern blot* ou RT-*real-time* PCR) entre o tecido normal e o tumor (ZHANG et al, 1997); muitos desses novos genes apresentam, de fato, particularidades no controle de sua expressão que podem torná-los candidatos para o uso de drogas com o objetivo de modular sua função. Essa linha de investigação abre novas fronteiras, portanto, para o diagnóstico, avaliação prognóstica definição terapêutica e desenho de novas drogas para tratamento das diferentes neoplasias humanas. Recentemente, van de Vijver e colaboradores examinaram o nível de expressão de 70 genes previamente escolhidos em decorrência do seu papel prognóstico em tumores de mama em 295 pacientes portadores de câncer de mama. O perfil de expressão dos 70 genes permitiu agrupar esses pacientes em dois grupos: 115 casos no grupo de bom prognóstico e 180 para a categoria de mau prognóstico (VAN DE VIJVER et al, 2002). Um fato interessante foi que o prognóstico revelou-se independente do envolvimento de linfonodos, ressaltando a precariedade do estadiamento clínico dos tumores,

habitualmente praticado em câncer de mama, com vistas à decisão terapêutica.

1.2 Leucemia Mielóide Crônica

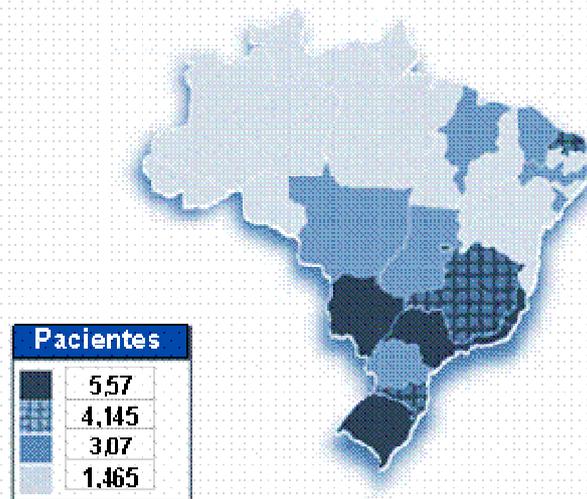
Policitemia vera, metaplasia mielóide agnogênica, trombocitemia essencial e leucemia mielóide crônica (LMC) são todas conhecidas como doenças mieloproliferativas, das quais LMC é a mais comum e apresenta, de forma consistente, o cromossomo Philadelphia (Ph). LMC é caracterizada por granulocitose progressiva, hiperplasia acentuada na medula óssea com elevada relação L:E e esplenomegalia, resultado da infiltração dos tecidos hematopoéticos pelo clone neoplásico que se origina a partir de um precursor hematopoético anormal na medula óssea. O desvio fenotípico está presente nas três linhagens oriundas da célula progenitora hematopoética, quais sejam, série eritróide, granulocítica e linfocitária. O cromossomo Ph representa a translocação recíproca entre o braço longo do cromossomo 9 e o braço longo do cromossomo 22 que ao nível molecular significa a fusão do oncogene ABL (cromossomo 9) com o gene BCR, no cromossomo 22: t(9:22)(q34;q11). Esse gene quimérico produz, de maneira constitutiva, produto com atividade de tirosina quinase, protagonista dos processos metabólicos que culminam com a hematopoese descontrolada da LMC (SAVAGE e ANTMAN, 2002; SAWYERS, 1999). Mais raramente o paciente portador de LMC não apresenta o cromossomo Ph porém, ou existem translocações variantes que envolvem um terceiro cromossomo além do 9 e

22, ou ocorre uma translocação críptica do 9q34 e 22q11 que não pode ser demonstrada pela citogenética convencional; nessas situações a utilização de métodos moleculares como o FISH ou RT-PCR podem detectar o gene quimérico. Pacientes portadores de LMC “negativos” também para o rearrajo BCR/ABL representam fenômeno mais raro e, geralmente com comportamento biológico mais desfavorável (SAWYERS, 1999).

Dados do Instituto Nacional do Câncer prevêem a ocorrência de 2195 casos novos de leucemias no Estado de São Paulo em 2002 (INCA - MINISTÉRIO DA SAÚDE, 2002), com a distribuição dos valores médios para o sexo e região do país resumidos na figura 1.

LEUCEMIAS

Representação espacial das taxas brutas de incidência por 100.000 homens estimadas para o ano 2002, segundo a Unidade da Federação (leucemias)



Fonte: INCA, 2002

Figura 1: Valores acima representam a média para ambos os sexos.

Se considerarmos que cerca de 15-20% desses casos são de LMC (JAFFE ES, 2001), teremos incidência de pelo menos 400 novos diagnósticos de LMC em nosso estado até o final de 2002. Habitualmente, a doença apresenta evolução clínica com duas ou três fases: crônica ou indolente, quando geralmente o diagnóstico é realizado, acelerada, em que há agravamento do quadro clínico do paciente, e crise blástica, com comportamento de doença aguda e refratária. (JAFFE ES, 2001)

LMC ainda não é curável com a terapêutica convencional, mas esse objetivo pode ser atingido com o transplante alogênico de medula óssea

para a maioria dos casos (SAWYERS, 1999). Esse índice de sucesso vai decrescendo à medida que a doença progride. A progressão da doença para os estadios mais avançados e de maior refratariedade à terapêutica está associada à aquisição de novas alterações genéticas em adição ao rearranjo BRC/ABL e o achado de alterações citogenéticas adicionais é clássico, ocorrendo para cerca de 80% dos pacientes (JAFFE ES, 2001). As alterações citogenéticas mais comuns encontradas nessa fase são: cromossomo Filadélfia adicional, trissomia do cromossomo 8, ou inversão do isocromossomo 17q. Há casos em que a fase crônica evolui diretamente para a crise blástica, caracterizando quadro de leucemia aguda típica, com pouca ou nenhuma resposta aos tratamentos quimioterápicos usuais (SAVAGE e ANTMAN, 2002; SAWYERS, 1999). À análise molecular, na fase de transformação têm sido descritas alterações de TP53, RB1, MYC, RAS, AML1, EVI-1, p16^{INK4a}, sem, no entanto, ficar claro o seu papel nesse processo (JAFFE ES, 2001).

LMC é considerada o paradigma da doença neoplásica e, de fato, é um dos primeiros tumores humanos curáveis com quimioterapia mieloablativa e transplante alogênico de medula óssea (FADERL et al, 1999; SAWYERS, 1999). Além disso, representa um modelo de doença molecular, em que uma proteína tirosina-quinase quimérica característica, resultante da justaposição da porção da porção 3' do gene ABL com a porção 5' do gene BCR~, desequilibra mecanismos finamente regulados de transdução de sinal na célula, interferindo com o controle da divisão e morte celular. O fato de esse rearranjo ser característica quase que exclusiva da célula neoplásica

suscitou o desenvolvimento de drogas com atividade específica anti-tirosina-quinase como o mesilato de imatinib (Mesilato de imatinibe). Mesilato de imatinibe é capaz de ocupar o sítio da molécula protéica ligante de ATP, impedindo essa ligação e abolindo o efeito enzimático resultante (SAVAGE e ANTMAN, 2002). Diversos estudos clínicos têm comprovado sua eficácia inicial no controle dessa doença, sendo demonstrado, inclusive, novo controle da doença na fase aguda (SAVAGE e ANTMAN, 2002). Entretanto, o papel do Mesilato de imatinibe como terapêutica inicial da LMC, assim como coadjuvante de outras modalidades de tratamento (como o transplante alogênico de medula óssea), ainda não está estabelecido, mas sugere ser bem promissor (SAVAGE e ANTMAN, 2002).

O Mesilato de imatinibe é, portanto, um exemplo de abordagem terapêutica eficaz, fruto do conhecimento molecular de doença complexa como o câncer. O que se antevê para os próximos anos como consequência do conhecimento dos padrões de expressão gênica global em diferentes situações fisiológicas e patológicas, além do conhecimento da variabilidade inter-pessoal (conhecimento dos SNPs) e certamente os novos avanços no proteoma, é uma era de profundas mudanças na maneira de diagnosticarmos, definirmos grupos de risco e prognóstico e acompanharmos os tumores humanos.

Nas leucemias crônicas e particularmente na leucemia mielóide crônica, pouco ou nenhuma informação a respeito do seu transcriptoma é conhecida. Hashimoto e colaboradores descreveram recentemente o transcriptoma de subpopulações leucocitárias humanas, contribuindo para

entender as mudanças no perfil de expressão dos diferentes tipos celulares frente aos diversos estímulos a que foram submetidas (HASHIMOTO et al, 2003). Nesse trabalho foram utilizados métodos de análise genômica de seqüenciamento e avaliação de expressão para investigação do transcriptoma de células de LMC. Adicionalmente, foram comparados os padrões de expressão de LMC com medula óssea não neoplásica e confirmada a expressão de genes potencialmente importantes na fisiopatologia da célula tumoral. Alguns dos genes analisados podem ainda ser candidatos a alvo de novas drogas capazes de interferir em vias metabólicas importantes na célula tumoral.

2. Objetivos

No presente trabalho tivemos por objetivos:

- Avaliar o conjunto dos genes expressos em LMC pela técnica de ORESTES, realizar o mapeamento cromossômico desses genes e organizá-los de acordo com sua categoria funcional.
- Ordenar os *contigs* gerados por ORESTES de acordo com seu conteúdo numérico de seqüências, em ordem decrescente e identificar os 100 mais expressos representativos. Desse conjunto, validar o nível de expressão, por método padrão-ouro (*real-time* PCR), comparando com tecido normal.
- Discutir possíveis mecanismos na tumorigênese envolvendo os genes estudados.
- Comparar o conjunto dos dados obtidos por ORESTES com o transcriptoma de medula óssea de paciente portador de doença hematológica não neoplásica (anemia hemolítica hereditária, anemia falciforme) avaliado pr SAGE.

3. Materiais e Métodos

3.1. Bioinformática

Do total de 1.188.862 seqüências ORESTES obtidas no projeto Genoma Humano do Câncer FAPESP/Ludwig, 57108 correspondiam àquelas oriundas de neoplasias hematológicas, leucemias mielóides e linfóides crônicas (CAMARGO et al, 2001). Desse conjunto, 27361 seqüências foram o resultado da análise dos genes expressos em leucemia mielóide crônica (LMC) e constituíram o objeto de análise desse trabalho. Os arquivos “phd” (arquivos-texto com cada seqüência e sua respectiva informação de qualidade) dessas 27361 seqüências foram fornecidos por J. P. Kitajima, pesquisador do Laboratório de Bioinformática do Instituto de Computação da UNICAMP. Essas seqüências já estavam livres de contaminantes com informações de vetores de clonagem, *primers* e contaminantes bacterianos eventuais. As seqüências obtidas no projeto Genoma Humano do Câncer FAPESP/Ludwig confirmaram a expectativa de se referirem às porções centrais dos genes expressos, conforme trabalho preliminar realizado com as primeiras 10.000 seqüências (DIAS-NETO et al, 2000).

Em testes preliminares realizamos a comparação dos resultados obtidos com dois programas adequados a essa finalidade: Phrap (BRENT EWING et al, 2002) e Cap3 (HUANG e MADAN, 1999). Os resultados foram semelhantes e optamos por continuar as análises com o pacote de

programas Phrap (Phragment Assembly Program), de distribuição livre para o meio acadêmico. Foi obtido diretamente com seu autor, Phil Green, através da assinatura do termo de acordo de usuário acadêmico disponível em http://www.phrap.org/consed/academic_agreement.txt. Basicamente, Phrap tem o papel de montagem digital de seqüências oferecidas em agrupamentos maiores, os *contigs* (BRENT EWING et al, 2002). Entre suas características principais, destacam-se:

- Melhor aproveitamento das seqüências oferecidas, muitas vezes com aproveitamento de regiões de menor qualidade;
- Permite a utilização da combinação de parâmetros próprios com informações inseridas pelo usuário, o que aumenta a eficiência de montagem na presença de regiões repetitivas;
- O *contig* gerado representa um mosaico de regiões de qualidade elevada de cada seqüência analisada e não apenas uma região de consenso;
- Oferece informações detalhadas a respeito da montagem realizada, incluindo valores de qualidade, o que auxilia na resolução de eventuais problemas de montagem;
- Pode ser utilizado para manuseio de grandes conjuntos de dados.

As seqüências-consenso de cada *contig* gerado utilizando-se o programa Phrap no ambiente abaixo descrito foram utilizadas para comparação local com banco de dados público NR, atualização de setembro de 2002 (ALTSCHUL et al, 1997). A melhor descrição para cada

comparação entre seqüências foi escolhida baseada no melhor “valor esperado” (*e-value*). O “valor esperado” é o parâmetro que descreve o número de resultados que se poderia esperar ao acaso numa procura em um banco de dados de determinado tamanho. Esse valor decresce exponencialmente na proporção do aumento do escore resultante da similaridade entre duas seqüências. Um “valor esperado” de *zero* equivale a dizer que duas seqüências são idênticas. Um “valor esperado” de *um* para um resultado significa afirmar que em banco de dados com o tamanho em questão há a chance de encontro de *uma* seqüência com escore semelhante ao acaso (ALTSCHUL et al, 1997; NCBI, 2002).

Ao resultado do blast foi aplicado segundo filtro, padronizados por nós, para retirada de eventuais contaminantes bacterianos, virais e vetores de clonagem, que continha os seguintes verbetes: “expression vector”, “cloning”, “vector”, “cosmid”, “clone image”, “Cytophaga”, “Ferroplasma”, “virus”, e “plasmid”. Essas palavras-chave foram escolhidas após avaliação minuciosa do conteúdo das listas de genes identificados através das seqüências dos *contigs*.

Para o entendimento da função dos transcritos gerados por ORESTES em LMC foram utilizadas basicamente as ferramentas disponíveis *online* no NCBI: Entrez, PubMed, OMIM, LocusLink (NCBI, 2002). Informações adicionais foram também coletadas *online* no Gene Ontology *online* (ASHBURNER et al, 2000).

As análises foram conduzidas em uma estação de trabalho Sun (SunBlade1000) com as seguintes características:

- Dois processadores Sparc (RISC) 750 MHz
- Três Gb de memória RAM
- 120 Gb de memória em disco rígido (quatro discos SCSI de 30 Gb cada, acessados paralelamente)
- Sistema operacional Solaris (proprietário, derivado de UNIX)

3.2. Amostras Biológicas

a) Caracterização clínica.

As coletas das amostras utilizadas neste estudo foram realizadas após a autorização por escrito dos pacientes e indivíduos controle, mediante leitura e entendimento do Termo de Consentimento Livre e Esclarecido aprovado pela Comissão de Ética do Hospital do Câncer A. C. Camargo.

Para a geração dos ORESTES, utilizamos as amostras de MMC, 38 anos, masculino e JMC, 47 anos, masculino, que tiveram o diagnóstico de LMC em fase crônica confirmado com base nos achados citológicos, anatomopatológicos, citogenéticos e moleculares (JAFFE ES, 2001). Para a análise de expressão com *real-time* PCR utilizamos amostras de sangue periférico de MMR, 28 anos, sexo feminino, também portadora de LMC em fase crônica, e DMA, 28 anos, sexo feminino, hígida. Também foi utilizada para comparação amostra de medula óssea de JSS, 31 anos, portador de anemia falciforme. As amostras foram coletadas no momento da punção e biópsia diagnóstica. Para a realização do SAGE (cf. adiante), amostra de medula óssea foi obtida de JSS, 31 anos, portador de anemia falciforme.

Todos os pacientes são acompanhados no Ambulatório de Hematologia do Hemocentro da UNICAMP e assinaram termo de consentimento informado (ANEXO A) aprovado pela Comissão de Ética do Hospital das Clínicas da Faculdade de Ciências Médicas da UNICAMP e pela Comissão de Ética do Hospital do Câncer A. C. Camargo (Anexo A).

3.3. Extração do RNA Total

RNA foi isolado utilizando TRIzol Reagent (Life Technologies, USA). O TRIzol apresenta uma solução monofásica de fenol e isotiocianato de guanidina. A extração de RNA com este reagente é uma adaptação do método desenvolvido por Chomczinski e Sacchi (CHOMCZYNSKI e SACCHI, 1987). Para a extração de RNA total a camada leucocitária é homogeneizada em TRIzol na proporção de 1,0 ml para cada $1,0 \times 10^7$ células (CHOMCZYNSKI e SACCHI, 1987). A quantificação do RNA obtido em solução aquosa foi realizada através da leitura de uma alíquota da amostra em espectrofotômetro em comprimento de onda de 260 nm no espectrofotômetro de luz UV GeneQuant (Pharmacia).

Nessas condições, uma unidade de DO equivale a 40 µg/ml de RNA. A relação entre as leituras realizadas a 260 e 280 nm é utilizada como parâmetro na estimativa do grau de contaminação do RNA por proteínas, e resultou, invariavelmente, entre 1,6 e 1,8 (JOSEPH SAMBROOK, 2001).

3.4. Avaliação de Qualidade do RNA Isolado

Além da razão adequada entre os resultados das leituras das amostras de RNA em 260 nm e 280 nm (descrita anteriormente), a qualidade das amostras de RNA foi testada por meio da eletroforese em agarose em tampão formamida. Um μg de RNA em tampão formamida 80% (12 M) foi aquecido a 80 °C e aplicado em cada poço no gel de agarose a 1% em 0,5% TBE com 0,1 $\mu\text{g/ml}$ de brometo de etídio (figura 2). Qualidade adequada para o manuseio subsequente das amostras pode ser definida pela presença de maior intensidade de coloração na banda referente ao RNA ribossomal 28S em relação ao 18S.

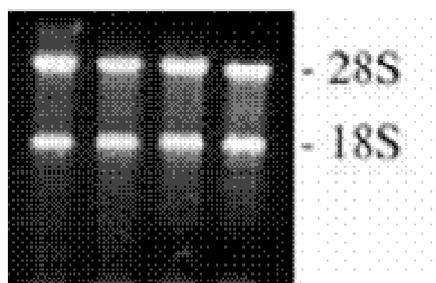


Figura 2: Eletroforese em tampão formamida em gel de agarose e corada por brometo de etídio de amostras de RNA total isoladas para o estudo.

Eventual degradação do RNA das amostras é preferencialmente percebida na fração 28S, invertendo essa relação. Nessa última situação, o que se espera é que também ocorra prejuízo do mRNA, objeto do estudo, devendo a amostra ser descartada da análise subsequente. As amostras utilizadas nesse estudo apresentavam o padrão característico de qualidade do RNA ribossomal (figura 2).

3.5. Transcrição Reversa para PCR Quantitativo em Tempo Real (*Real-Time* PCR)

Três a 5 µg das amostras de RNA foram incubadas com 1 U de DNaseI (Life technologies) por 15 minutos à temperatura ambiente e EDTA foi adicionado para uma concentração final de 2 mM com o intuito de interromper a reação. A enzima foi inativada por incubação a 65 °C por 10 minutos. As amostras de RNA tratadas com DNaseI foram submetidas à transcrição reversa com 200 U de SuperScriptII (Life Technologies) por 50 minutos a 42 °C. Duas unidades de RNaseH (Life Technologies) foram adicionadas a seguir e as amostras foram incubadas a 37 °C por 20 minutos. As amostras de cDNA foram quantificadas através do espectrofotômetro de luz UV GeneQuant (Pharmacia).

3.5. PCR Quantitativo em Tempo Real (*Real-Time* PCR)

A reação de amplificação em tempo real foi realizada com o objetivo de confirmar o elevado nível de expressão encontrado em determinados genes com a técnica de ORESTES e comparar com a respectiva expressão em tecido mielóide não neoplásico. Para isso RNA total isolado de leucócitos do sangue periférico de paciente portador de LMC foi submetido à análise comparativa com RNA extraído de medula óssea de paciente portador de condição não-neoplásica (anemia falciforme). *Real-time* PCR foi conduzido na plataforma ABI 5700 Sequence Detector System (Applied Biosystems). O

princípio do método está baseado na detecção de fluorescência no tubo de reação à medida que DNA dupla fita é gerado, em virtude da concentração do agente SybrGreen I entre as cadeias de DNA geradas. A presença de DNA dupla fita na solução é capaz de aumentar essa emissão de luz em até 100 vezes para uma mesma concentração de SybrGreen I (APPLIED BIOSYSTEMS INC., 2001). O equipamento utiliza uma lâmpada halógena como fonte de luz capaz de excitar o fluorocromo envolvido na reação. A fluorescência eventualmente produzida pela amostra é detectada por câmara CCD com amplitude de detecção de 530 a 580 nm. O momento da reação de PCR em que a fluorescência de determinada amostra é detectada inequivocamente acima do ruído de fundo (*background*) é denominado CT (*threshold cycle*). Adotaremos essa nomenclatura com a ressalva de que ela pode dar margem a interpretações equivocadas do processo, uma vez que o CT freqüentemente não é um ciclo, mas um momento na reação, que pode ocorrer em qualquer intervalo entre um ciclo e outro do PCR (figura 3).

Erro!

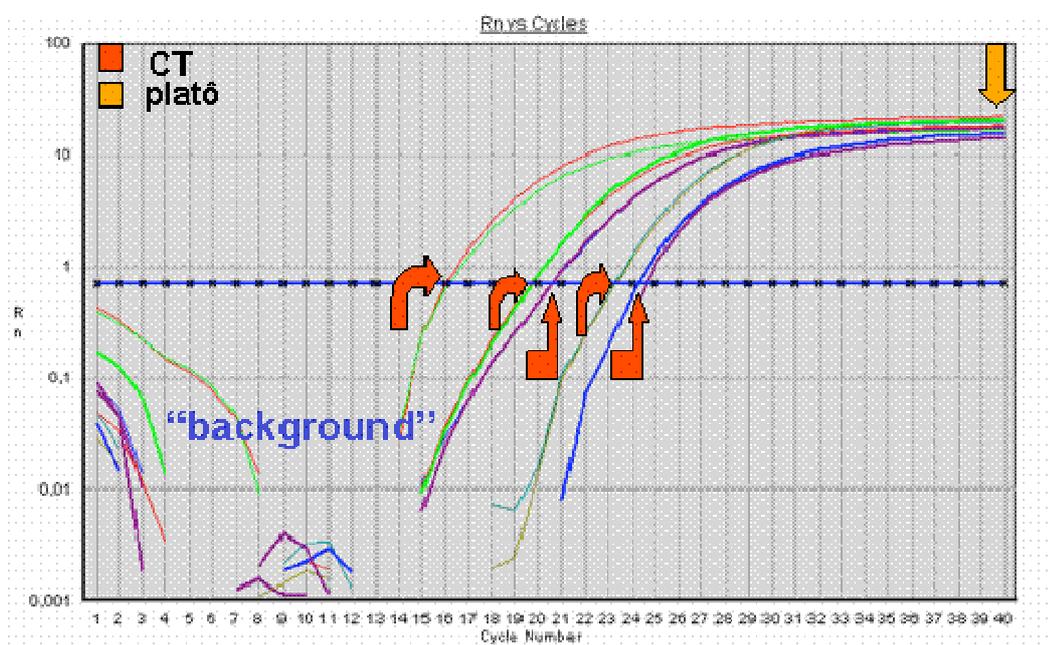


Figura 3: Monitoramento em tempo real das curvas de amplificação proporciona estabelecimento de nível de fluorescência na fase ótima (aumento exponencial da fluorescência) da reação de amplificação. O nível de fluorescência computado para cada amostra é aquele suficiente para atingir um limiar (traço horizontal azul) de detecção igual para cada conjunto de *primers*/amostras testado e por convenção denominado de CT. A figura mostra cinco conjuntos de *primers* e uma única amostra. Para cada conjunto de *primers* os experimentos foram realizados em duplicadas (visível na figura através da diferença de cores entre as curvas de amplificação, embora com coeficiente de variação reduzido). A fluorescência emitida antes do nível do CT para cada curva é considerada ruído de fundo (*background*). Nota-se que, para algumas curvas de amplificação, o valor do CT para o *threshold* que foi definido não é um número inteiro.

Para cada amostra foram empregados 12,5 μ l de solução de SybrGreen I 2x (SybrGreen PCR Master Mix - Applied Biosystems) para o volume final de 25 μ l, numa adaptação do protocolo original, que sugeria volume final de reação de 50 μ l (APPLIED BIOSYSTEMS INC., 2001). Dez a 20 ng de cDNA e 300 a 900 nM dos *primers* senso e antissenso foram

empregados em cada reação de amplificação. Os *primers* foram desenhados por programa de computador dedicado (PrimerExpress – Applied Biosystems) adotando-se as condições padrão recomendadas pelo fabricante (APPLIED BIOSYSTEMS INC., 2001). Alguns cuidados foram tomados no desenho dos oligonucleotídeos com o objetivo de garantir a melhor qualidade dos resultados:

- Sempre que possível foram desenhados *primers* em éxons diferentes, ou na transição entre dois éxons.
- O tamanho dos produtos gerados por cada conjunto de *primers* era de 50 a 150 pb.
- Nas etapas de padronização os tamanhos dos fragmentos amplificados foram confirmados em gel de agarose a 1,5%. Essa precaução ajudou a confirmar a qualidade da reação de PCR, permitindo identificar possíveis produtos inespecíficos, ou produtos gerados através da amplificação de dímeros de *primers*.
- A temperatura teórica de desnaturação dos *primers* (T_m) foi em torno de 60 °C.

3.6. Genes Selecionados para Confirmação de Expressão com Real-Time PCR

A tabela 1 revela a seqüência de cada um dos oligonucleotídeos empregados nesse trabalho. Os *primers* foram construídos tendo por base os *contigs* gerados pela montagem das seqüências de ORESTES. A escolha

dos genes para avaliação subsequente por *real-time* PCR foi feita analisando-se minuciosamente os 100 genes mais representados pelos *contigs* ORESTES, isto é aqueles que apresentavam o maior número de seqüências por *contig* (ANEXO B). Foram escolhidos genes de diferentes categorias funcionais, como ciclo celular e sinalização, metabolismo celular e resposta a estímulos externos. O nível de expressão de cada transcrito foi avaliado sempre em relação ao resultado da análise de expressão da β -actina realizada na mesma amostra, nas mesmas condições e ao mesmo tempo. A β -actina foi escolhida pelo seu padrão de expressão constitutivo (APPLIED BIOSYSTEMS INC., 2001).

Os ensaios foram realizados sempre em duplicata e com as amostras de sangue periférico de paciente com LMC, medula óssea de paciente com anemia falciforme e sangue periférico de indivíduo hígido.

Reações controles, sem adição de cDNA ("*no template controls* – *NTC*") foram incluídas para cada conjunto de *primers*. Ao final da reação de amplificação foi estabelecido ciclo de variação crescente de temperatura (de 60 °C a 95 °C) e monitoramento da fluorescência de forma a avaliar a temperatura de dissociação do produto de PCR (*melting curve analysis*). Esse resultado, expresso em graus Celsius, é característica de cada produto de PCR, dependente de seu tamanho e composição química, e essa abordagem pode revelar a presença de mais de um produto em determinada amostra (amplificação inespecífica) ou mesmo de interferentes da reação de amplificação, como dímeros de *primers* que eventualmente concatenam-se naquelas condições, produzindo quimeras (APPLIED BIOSYSTEMS INC.,

2001). Para padronização da concentração ótima de *primers*, foram realizadas reações em duplicata contendo diferentes diluições de *primers*, adotando-se aquela que proporcionava o CT mais baixo (maior sensibilidade) sem a geração de produtos inespecíficos, avaliados pela curva de dissociação e por eletroforese em gel de agarose. Após a padronização, cada teste foi realizado em duplicata, no mínimo.

Tabela 1: Relação dos *primers* utilizados para confirmação da expressão gênica aferida por ORESTES. A coluna *contigs* revela o número do *contig* cujos detalhes encontram-se no ANEXO B e cujo gene symbol encontra-se adiante, coluna *Sigla*.

<i>Contig/gene</i>	<i>primers senso</i>	<i>primers antissenso</i>	Conc. (nM)	Sigla
1301	GGAGGCGGTAGAGGTTAGCAT	GGAAGCTGTGCTTCGTGCTAT	800	CAMP
1519	TGCTCCACCAGTGATGGAGTT	GTAGTCTGAGAGGCCAGGTATACC	800	BPI
2267	TGGCCGCTCAAGGGAAT	CGCTCCAGGTCGTTCCAGGTA	800	GNAI2
2757	ACACAGACAGCCCGCTTTG	GCTGGACTCAGGCGGAAA	800	ABR
283	AGCCCCGGAGCAGATTG	CAAGCTTTCGTCATGCA	800	DEFA3
3014	TGACTCGCTCATTGGTGGAA	GTGATTACTGTGCTTTCGGTACCTT	400	TPT1
3038	ACAGCAGGAGGCCGAACAG	GTCTCACCCTGATGACCTTGA	800	CSF3R
347	GCCCCAGTGTAGCCGTAGTATC	GCGACGAGCAGGGTGAGA	800	LTF
2638	GGCTTGGCACACTTTGCATT	TGCCTGGACACTGCTCTGACT	800	MLL2
1262	CAAGTCCGTGGGCATCATG	AGTACTTGTGGTAGACGTCGATGATAG	600	S100A8
400	AGGAGAACATAAGCCTGCTCGTA	TCAGTCAAACGTCCAAGGTTCTT	800	MS4A3
β -actina	TCACCGAGCGCGGCT	TAATGTCACGCACGATTTCCC	400	ACTB

3.7. SAGE – *Serial Analysis of Gene Expression* (“Análise Seriada da Expressão Gênica”)

A técnica de SAGE baseia-se na variabilidade da extremidade 3' não-codificante dos mRNAs e na possibilidade de identificação de genes por pequenas seqüências específicas (*tags*) de 9- 15 pb desta região em cada transcrito (LAL et al, 1999; VELCULESCU et al, 2000a; VELCULESCU et al, 1995; VELCULESCU, 1999). O princípio do método é baseado no isolamento, seqüenciamento e contagem dessas *tags* únicas (LAL et al, 1999; VELCULESCU et al, 2000a; VELCULESCU et al, 1995; VELCULESCU, 1999). A figura 4 representa o método em detalhes, de acordo com o protocolo utilizado nesse trabalho (VELCULESCU et al, 2000a).

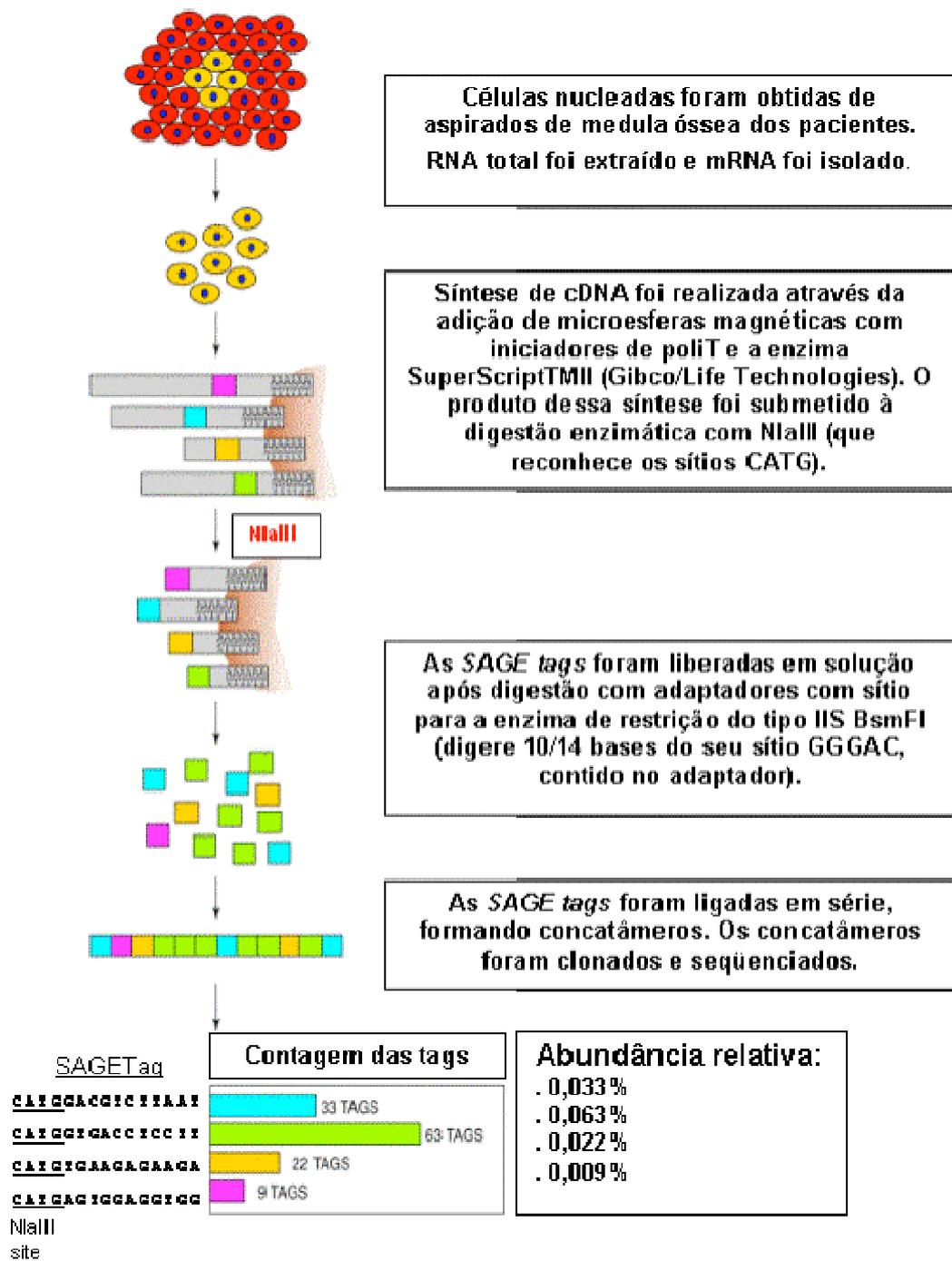


Figura 4: Representação esquemática da técnica de SAGE empregada nesse trabalho. Ao final da reação de seqüenciamento, as SAGE tags são contadas com a utilização de programa dedicado (SAGE2000, Invitrogen) que é capaz de identificar e quantificar cada uma das tags. A identificação é feita pela confrontação das seqüências obtidas com banco de dados próprio (Hs) (NCBI/NIH, 2002) utilizando o programa Microsoft Access.

4. Resultados

4.1 Avaliação por ORESTES

Das 27361 seqüências oriundas de células leucêmicas estudadas, 3573 *contigs* e 11568 *singlets* foram gerados pelos algoritmos de bioinformática empregados (BRENT EWING et al, 2002; HUANG e MADAN, 1999). Após a retirada de contaminantes utilizando os filtros descritos na seção anterior foram separados para análise 2933 *contigs*. A princípio, cada *contig* representa um transcrito – ou gene expresso – único, o que se comprovou na maioria dos casos. O número de seqüências únicas em cada *contig* variou de dois a 210 (média: 4,3 seqüências/*contig*; mediana: 2 seqüências/*contig*), conforme ilustra a tabela 2 e o ANEXO B.

Após a análise comparativa de cada uma dessas seqüências com o banco de dados *nr* através do algoritmo blast (*Basic Local Alignment Search Tool*) (, 2002; ALTSCHUL et al, 1997), o melhor “valor esperado” para cada gene foi selecionado. Essa foi a base para o processo de anotação do transcriptoma da LMC avaliado por ORESTES e permitiu, pela primeira vez, a classificação funcional dos genes expressos identificados pelos *contigs* que eram exclusivo de leucemia mielóide crônica (figura 5). Os percentuais distribuídos em cada categoria funcional podem ser conferidos na tabela 3. O critério de classificação foi o mesmo adotado por Dias-Neto et al que teve como objetivo coletar o maior número de funções ou processos celulares envolvidos. (DIAS-NETO et al, 2003) A localização cromossômica dos *contigs* foi também determinada e a distribuição dos *contigs* por cromossomo encontra-se detalhada na tabela 4. De maneira semelhante à

Tabela 3: Distribuição percentual dos *contigs* de acordo com sua categoria funcional.

Comunicação Celular	19
Adesão	3
Reconhecimento	1
Sinalização Intercelular	1
Transdução de Sinal	9
Resposta a Estimulos Externos	5
Crescimento, Desenvolvimento & Manutenção	9
Ciclo Celular	6
Diferenciação Celular e Tissular	1
Homeostase	1
Apoptose	1
Metabolismo Celular	30
Metabolismo de DNA	2
Metabolismo de RNA	5
Metabolismo de Proteínas	8
Metabolismo de Lipídeos	2
Metabolismo de Carboidratos	2
Metabolismo Geral	4
Regulação de Transcrição	7
Estrutura , Mobilidade & Transporte	15
Organização Citoplasmática & Biogênese	1
Mobilidade	1
Estrutura & Organização Intracelular	4
Matriz Extracelular, Membrana Celular, Citoesqueleto	7
Transporte	2
Função Desconhecida	28

Mapeamento cromossômico dos *clusters* ORESTES



Figura 6: 87% dos *contigs* analisados foram mapeados no genoma humano.

Tabela 4: Distribuição dos *contigs* ORESTES de LMC no genoma. Resumidamente, a seqüência consenso originada de cada *contig* foi comparada com algoritmo de alinhamento local Blast com banco de dados *nr* (atualização de outubro de 2002). O mapeamento cromossômico ocorreu para 87% das seqüências e foi representado em todos os cromossomos. Em destaque, a maior densidade encontrada no cromossomo 19 e a menor, no cromossomo 13 (coincidente com Venter e colaboradores).

Cromossomo	Número de contigs	Tamanho (Mpb)*	Densidade de Genes#	Venter et al**
1	222	220	1,01	16
2	165	240	0,69	12
3	129	200	0,65	12
4	83	166	0,50	10
5	102	182	0,56	11
6	138	172	0,80	13
7	107	146	0,73	12
8	95	146	0,65	11
9	104	113	0,92	13
10	110	130	0,85	12
11	166	132	1,26	16
12	134	134	1,00	14
13	44	99	0,44	10
14	87	87	1,00	14
15	99	80	1,24	15
16	120	75	1,60	19
17	180	78	2,31	19
18	88	79	1,11	10
19	197	58	3,40	29
20	100	61	1,64	16
21	44	33	1,33	13
22	82	36	2,28	23
X	100	128	0,78	11
Y	14	19	0,74	11
Média	113	117	1	14
Mapeados	2710			
Não Mapeados	402			

* Tamanho dos cromossomos humanos em milhões de pares de bases; # número de genes (ou contigs) por milhão de pares de bases; ** Densidade de genes por milhão de pares de bases obtida por Venter e colaboradores utilizando algoritmos computacionais de predição de genes em estruturas gênicas (VENTER et al, 2001).

4.2. Avaliação da Expressão Gênica por SAGE

Cerca de 50 mil SAGE *tags* foram geradas a partir de RNA de medula óssea de paciente com anemia falciforme. A distribuição das *tags*, de acordo com o seu nível de expressão, está apresentada detalhadamente no ANEXO B. Nesse documento ainda é possível realizar a comparação entre o nível de expressão obtido por ORESTES (LMC) com o método de referência na avaliação da expressão em larga escala (SAGE) (LAL et al, 1999; VELCULESCU et al, 2000a; VELCULESCU, 1999). Não houve correlação entre o padrão de expressão gênica demonstrado por ORESTES e o perfil revelado por SAGE.

4.3. Validação da Expressão Gênica por *Real-Time* PCR Quantitativo

Para estabelecer o grau de expressão de cada um dos transcritos escolhidos para a etapa de validação (tabela 1) utilizamos como referência o nível de expressão da β -actina, gene de expressão constitutiva (APPLIED BIOSYSTEMS INC., 2001) e que se presta como referência de expressão entre as diferentes amostras. Dessa forma, o nível de expressão do gene de interesse é relatado em relação ao do gene de referência para cada amostra (APPLIED BIOSYSTEMS INC., 2001).

A tabela 5 e as figuras 7 a 16 nos mostram os resultados de expressão dos 10 genes selecionados, por *real-time* PCR em LMC e no tecido normal.

Estabelecidos os CT de cada amostra, o cálculo do nível de expressão é realizado através da fórmula (APPLIED BIOSYSTEMS INC., 2001):

$$\text{Expressão do Gene de Interesse (GOI)} = 2^{-\Delta\Delta\text{CT}}$$

Onde $\Delta\text{CT} = (\text{CT}_{\text{gene interesse}}) - (\text{CT}_{\beta\text{-actina}})$ para cada amostra em duplicata

$$\Delta\Delta\text{CT} = (\Delta\text{CT}_{\text{LMC}}) - (\Delta\text{CT}_{\text{controle}})$$

Tabela 5: Resultados de expressão dos 10 genes selecionados, por *real-time* PCR em LMC comparativamente ao tecido normal. Destacado em itálico encontra-se LTF (lactoferrina) cuja expressão foi semelhante nos dois tipos celulares.

genes	LMC	Controle	Figura correspondente
CAMP	27,76	1,00	7
GNAI2	5,66	1,00	8
ABR	46,85	1,00	9
DEFA3	74,29	1,00	10
TPT1	2,61	1,00	11
CSF3R	24,93	1,00	12
<i>LTF</i>	<i>0,96</i>	<i>1,00</i>	<i>13</i>
MLL2	26,63	1,00	14
S100A8	10,97	1,00	15
MS4A3	16,97	1,00	16

Os dados brutos para cada gene estão mostrados na tabela 6 a seguir. Foi a partir deles que foram construídos os gráficos das figuras 7 a 16.

Tabela 6: dados individuais da análise de expressão por real-time PCR para cada um dos genes avaliados. O CT é apresentado como valores individuais para cada tubo e a média das duplicadas é empregada para o cálculo. A coluna $2^{(-\Delta\Delta CT)}$ revela o aumento de expressão para cada um dos genes da coluna seguinte (*Gene*).

Amostra	CT	Média (CT)	Des Pad	ΔCT	$\Delta\Delta CT$	$2^{(-\Delta\Delta CT)}$	Gene
NTC	40.00						
NTC	40.00	40.00	0.00				
LMC	35.95						
LMC	35.68	35.82	0.19				ACTB
Normal	29.98						
Normal	30.42	30.20	0.31				ACTB
NTC	40.00						
NTC	40.00	40.00	0.00				
LMC	35.75						
LMC	35.78	35.77	0.02				ACTB
Normal	30.22						
Normal	30.28	30.25	0.04				ACTB
NTC	40.00						
NTC	40.00	40.00	0.00				
LMC	38.47						
LMC	39.20	38.84	0.52	3.02	-4.08	16.97	MS4A3
Normal	37.25						
Normal	37.36	37.31	0.08	7.11	0.00	1.00	MS4A3
NTC	40.00						
NTC	40.00	40.00	0.00				
LMC	26.08						
LMC	26.11	26.10	0.02	-9.72	-6.22	74.29	DEFA3
Normal	27.58						
Normal	25.81	26.70	1.25	-3.51	0.00	1.00	DEFA3
NTC	40.00						
NTC	40.00	40.00	0.00				
LMC	33.62						
LMC	34.98	34.30	0.96	-1.52	-5.55	46.85	ABR
Normal	34.24						
Normal	34.23	34.24	0.01	4.04	0.00	1.00	ABR
NTC	40.00						
NTC	40.00	40.00	0.00				
LMC	37.14						
LMC	36.53	36.84	0.43	1.07	-0.05	1.04	LTF
Normal	31.40						
Normal	31.35	31.38	0.04	1.13	0.00	1.00	LTF
NTC	40.00						
NTC	40.00	40.00	0.00				
LMC	37.56						
LMC	38.68	38.12	0.79	2.31	-4.64	24.93	CSF3R
Normal	37.02						
Normal	37.27	37.15	0.18	6.95	0.00	1.00	CSF3R
NTC	40.00						

NTC	40.00	40.00	0.00					
LMC	31.98							
LMC	32.61	32.30	0.45	-3.52	-1.38	2.61	TPT1	
Normal	28.02							
Normal	28.11	28.07	0.06	-2.14	0.00	1.00	TPT1	
NTC	40.00							
NTC	40.00	40.00	0.00					
LMC	38.29							
LMC	39.06	38.68	0.54	2.86	-4.74	26.63	MLL2	
Normal	37.33							
Normal	38.26	37.80	0.66	7.60	0.00	1.00	MLL2	
NTC	40.00							
NTC	40.00	40.00	0.00					
LMC	39.76							
LMC	40.00	39.88	0.17	4.07	-2.50	5.66	GNAI2	
Normal	36.51							
Normal	37.02	36.77	0.36	6.57	0.00	1.00	GNAI2	
NTC	40.00							
NTC	40.00	40.00	0.00					
LMC	35.75							
LMC	35.51	35.63	0.17	-0.19	-4.80	27.76	CAMP	
Normal	34.27							
Normal	35.35	34.81	0.76	4.61	0.00	1.00	CAMP	
NTC	40.00							
NTC	40.00	40.00	0.00					
LMC	31.48							
LMC	31.57	31.53	0.06	-4.29	-3.46	10.97	S100A8	
Normal	29.20							
Normal	29.53	29.37	0.23	-0.84	0.00	1.00	S100A8	

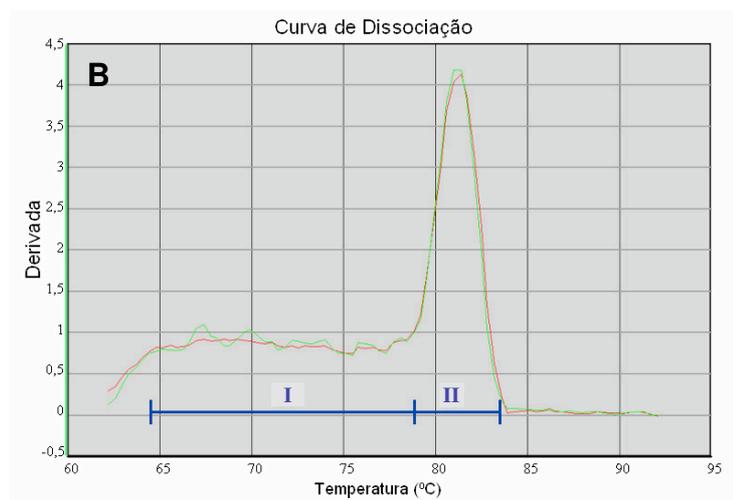
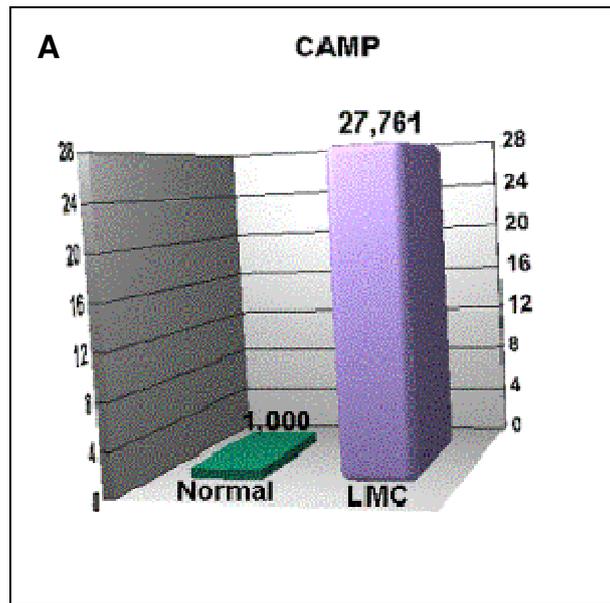


Figura 7: Expressão de CAMP em células de LMC e controles avaliada por real-time PCR quantitativo. **A**: CAMP foi 27,8 vezes mais expresso em LMC. **B**: curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de CAMP; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

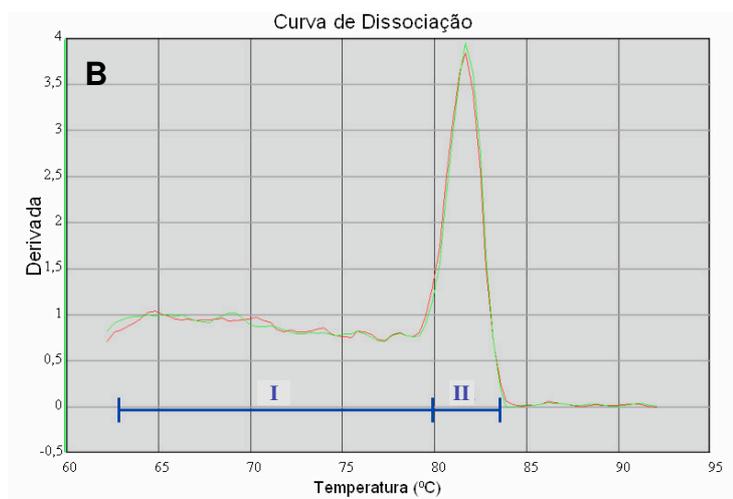
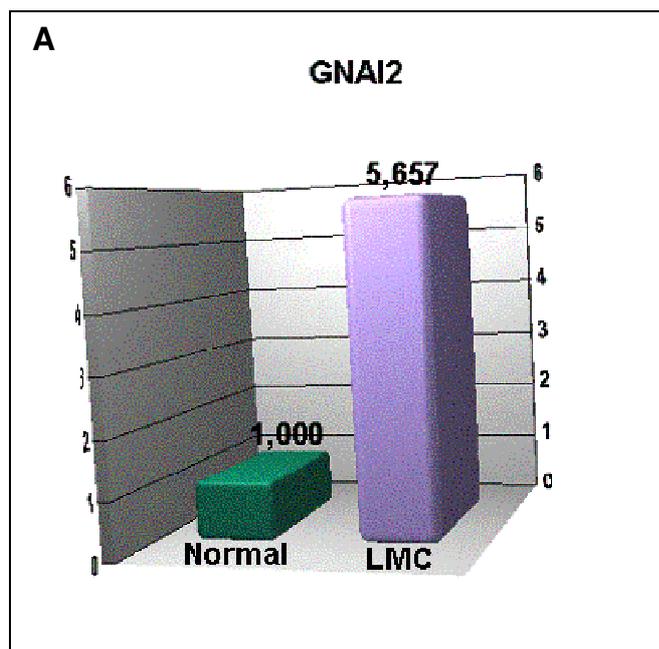


Figura 8: Expressão de GNAI2 em células de LMC e controles avaliada por real-time PCR quantitativo. **A**: GNAI2 foi 27,8 vezes mais expresso em LMC. **B**: curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de GNAI2; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

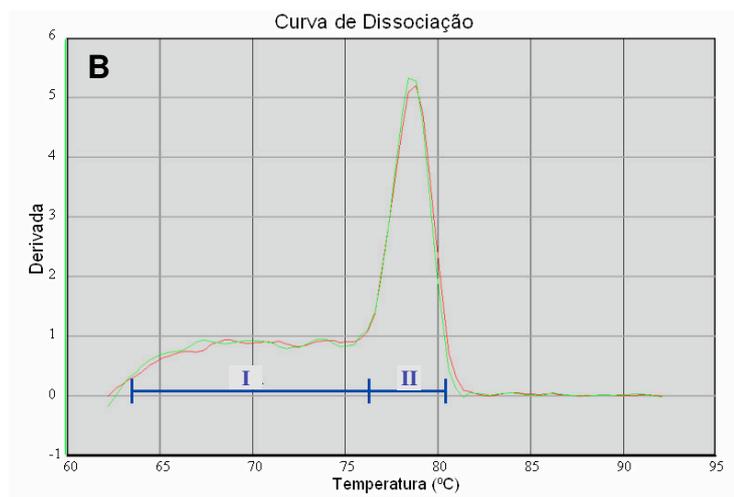
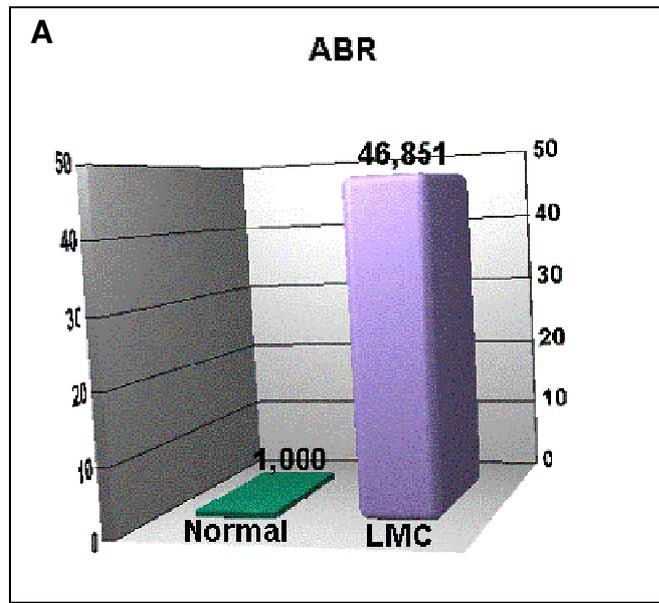


Figura 9: Expressão de ABR em células de LMC e controles avaliada por real-time PCR quantitativo. **A:** ABR foi 27,8 vezes mais expresso em LMC. **B:** curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de ABR; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

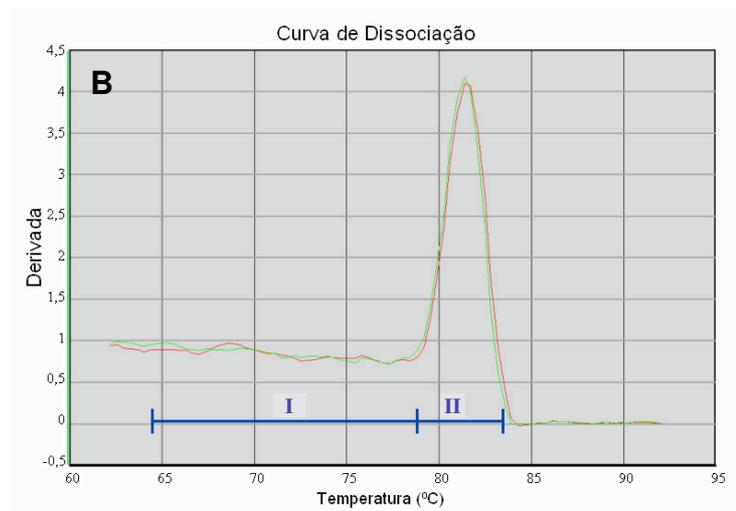
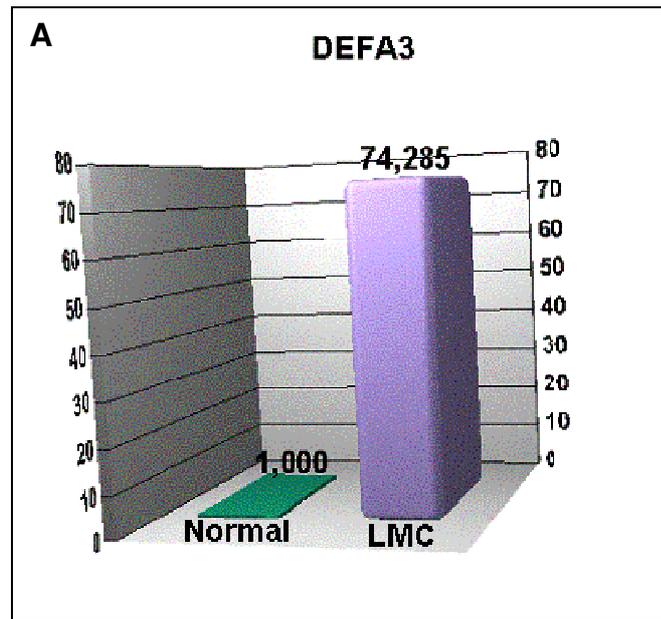


Figura 10: Expressão de DEFA3 em células de LMC e controles avaliada por real-time PCR quantitativo. **A**: DEFA3 foi 27,8 vezes mais expresso em LMC. **B**: curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de DEFA3; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

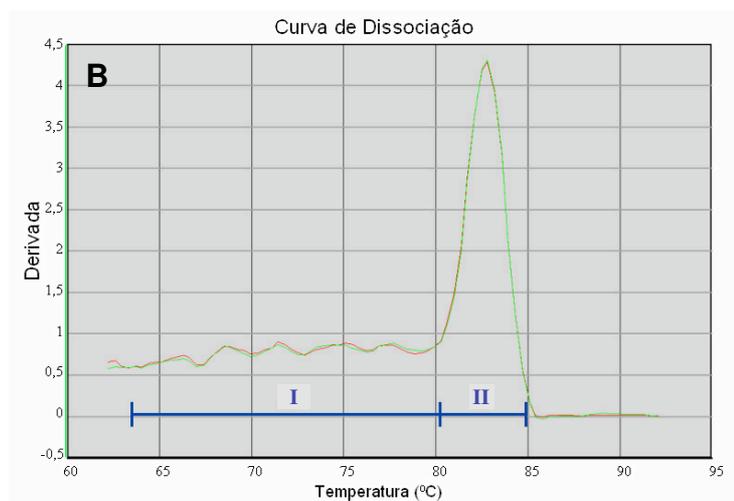
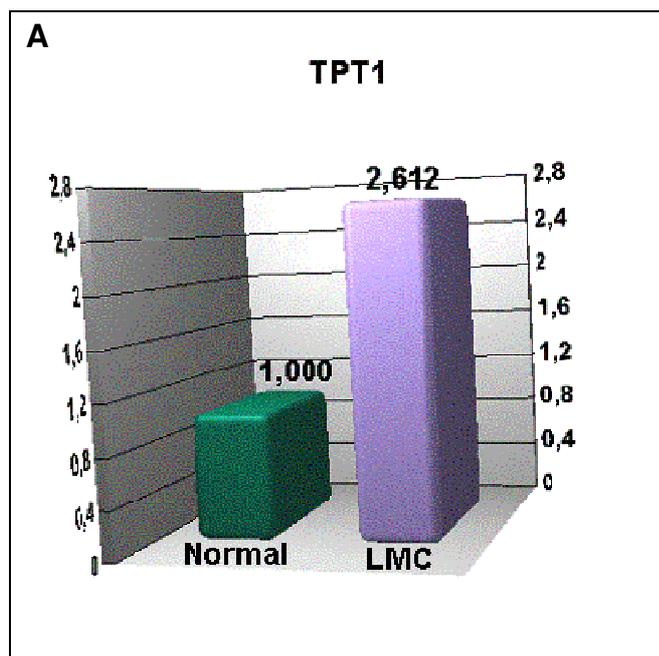


Figura 11: Expressão de TPT1 em células de LMC e controles avaliada por real-time PCR quantitativo. **A**: TPT1 foi 27,8 vezes mais expresso em LMC. **B**: curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de TPT1; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

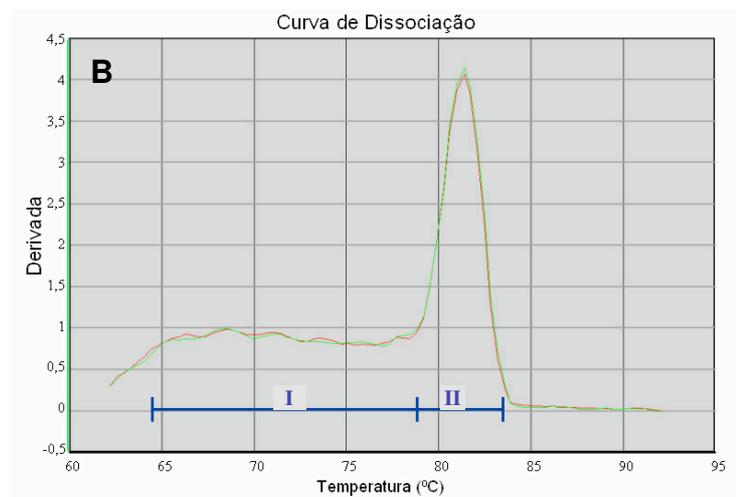
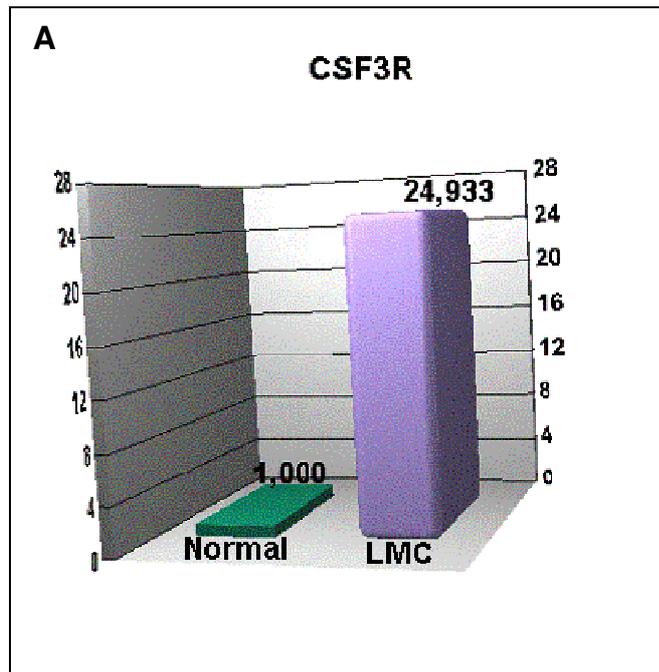


Figura 12: Expressão de CSF3R em células de LMC e controles avaliada por real-time PCR quantitativo. **A:** CSF3R foi 27,8 vezes mais expresso em LMC. **B:** curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de CSF3R; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

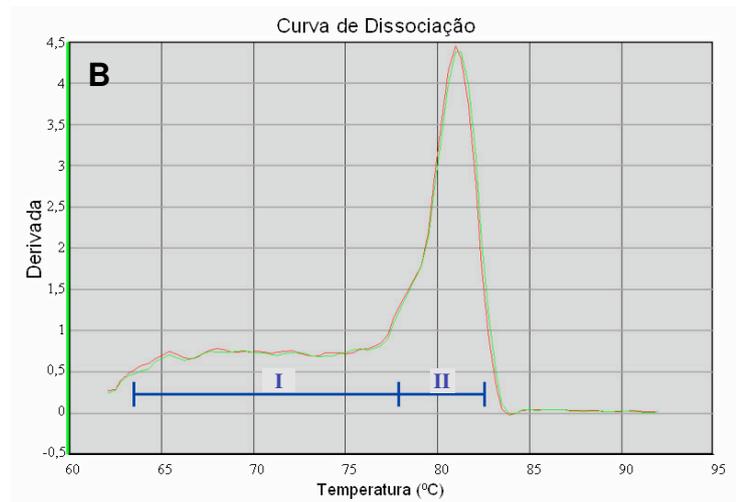
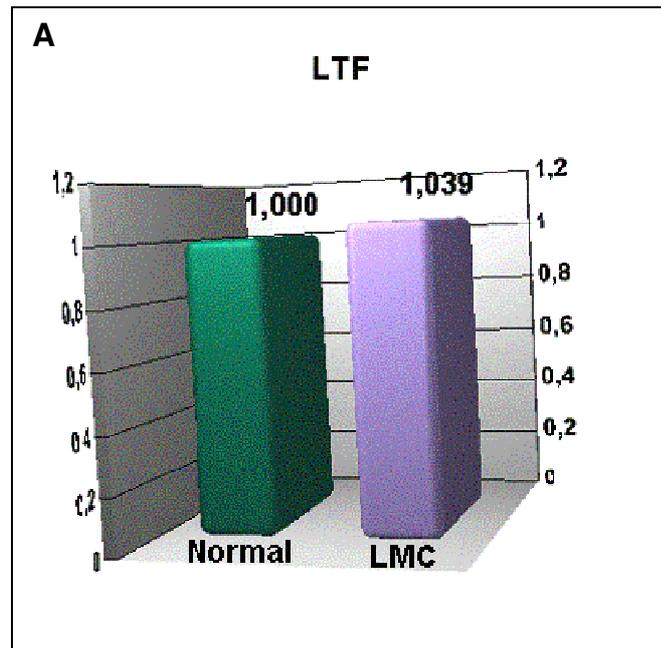


Figura 13: Expressão de LTF em células de LMC e controles avaliada por real-time PCR quantitativo. **A**: LTF foi 27,8 vezes mais expresso em LMC. **B**: curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de LTF; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

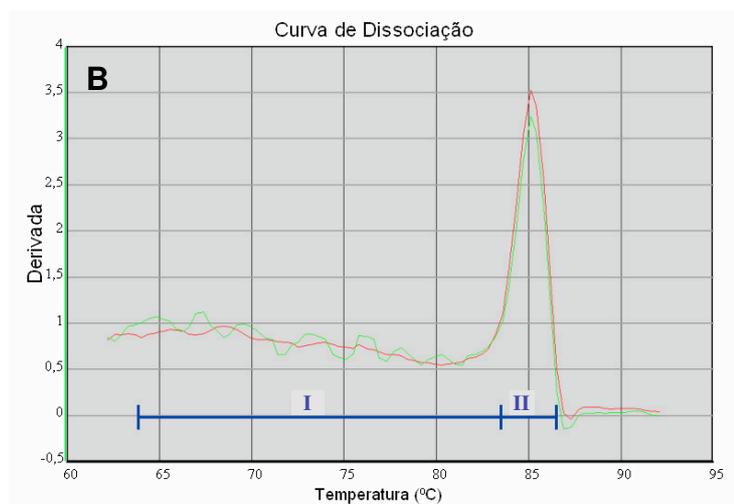
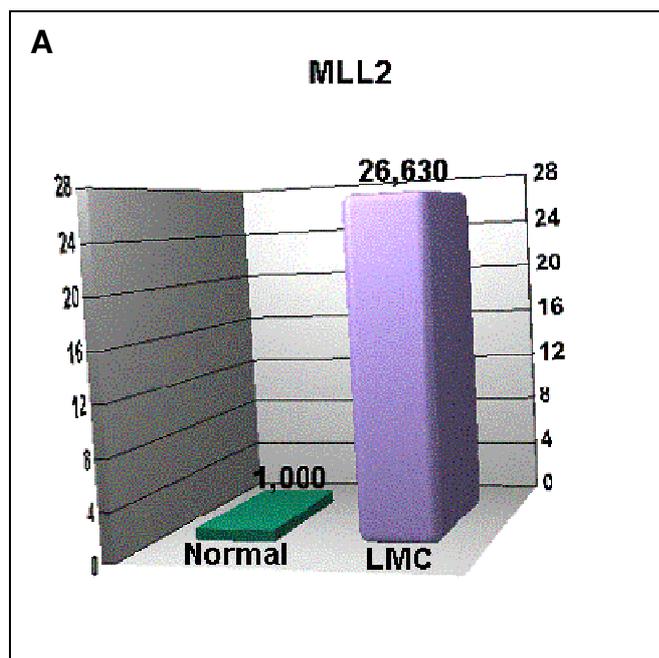
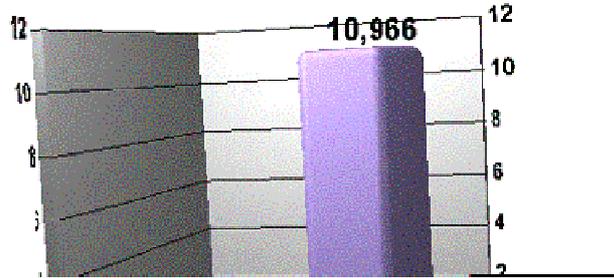


Figura 14: Expressão de MLL2 em células de LMC e controles avaliada por real-time PCR quantitativo. **A**: MLL2 foi 27,8 vezes mais expresso em LMC. **B**: curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de MLL2; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

S100A8



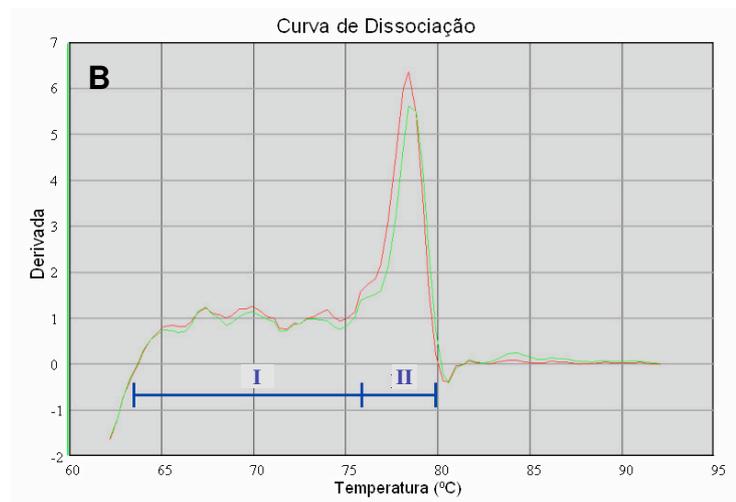
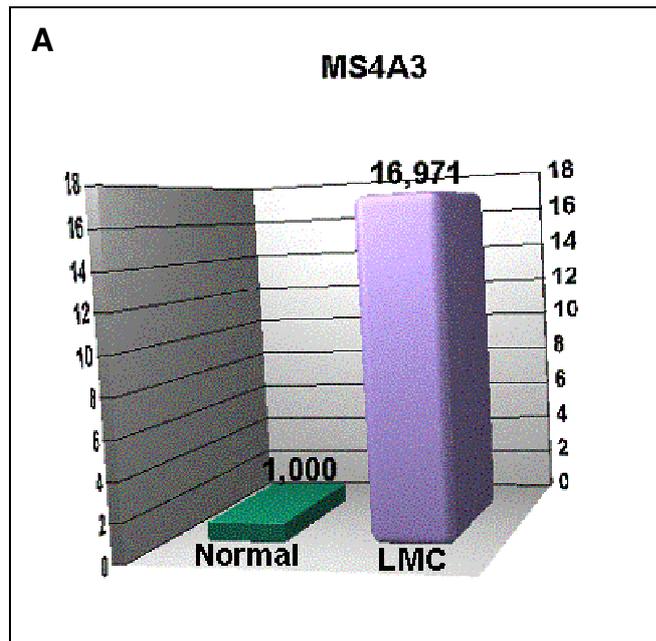


Figura 16: Expressão de MS4A3 em células de LMC e controles avaliada por real-time PCR quantitativo. **A**: MS4A3 foi 27,8 vezes mais expresso em LMC. **B**: curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de MS4A3; observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

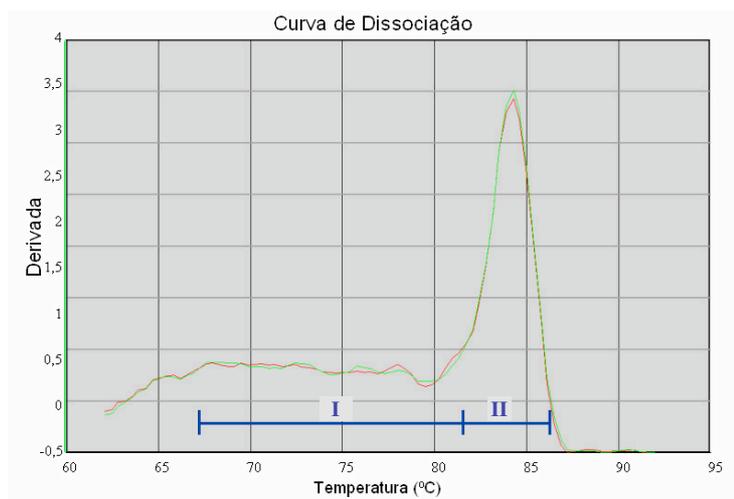


Figura 17: Curva de dissociação demonstrada em gradiente de temperatura realizado ao final da reação de amplificação de β -actina (ACTB); observa-se o baixo ruído de fundo (segmento I) que precede a emissão de luz específica do produto gênico amplificado (segmento II).

5. Discussão

O padrão básico de expressão gênica em células eucarióticas já é bem estabelecido (HASTIE e BISHOP, 1976; HEREFORD e ROSBASH, 1977). Em uma típica célula de mamífero a massa total de RNA consiste de aproximadamente 300.000 moléculas, que correspondem a cerca de 12.000 transcritos que se dividem em três classes de abundância. Um número muito pequeno de mRNAs (cerca de 10) são expressos em níveis muito elevados (3.000 a 15.000 cópias por célula); um número maior de mRNAs (cerca de 500) atinge um nível de expressão da ordem de 100 a 500 cópias por célula. Finalmente, a maioria dos mRNAs (mais que 10.000) têm um nível de expressão muito reduzido (10 a 100 cópias por célula). Esse é o padrão esperado em análises globais de expressão gênica, por métodos como SAGE, *microarrays* e mesmo ORESTES. Por meio do seqüenciamento em larga escala das porções codificantes centrais dos genes expressos em LMC foram produzidas 27361 seqüências. Fomos capazes de agrupá-las em *contigs* (2933, após a retirada dos contaminantes), utilizando ferramentas de bioinformática para o trabalho de análise de expressão. O agrupamento (“clusterização”) das seqüências geradas tem como principal objetivo estender a cobertura da informação, já que os métodos de seqüenciamento atuais apresentam limitação de cerca de 500 a 700 bases para uma única leitura (GRUBER, 2002). Após esse tratamento preliminar, realizamos a anotação dos genes identificados, o que permitiu classificá-los os genes de acordo com sua categoria funcional (figura 5). Essa é a primeira análise detalhada do conjunto dos genes expressos globalmente, ou transcriptoma,

em LMC. Gnatenko e colaboradores avaliaram o transcriptoma de plaquetas do sangue periférico utilizando microarrays e SAGE. No entanto, devido a particularidades na expressão gênica dos trombócitos, como a ausência de núcleo, a presença de mitocôndrias e o grande número de seqüências de globina encontradas, a análise de transcritos mais raros ficou comprometida (GNATENKO et al, 2002). Mesmo assim esses autores identificaram 11% dos genes encontrados como relacionados ao metabolismo, 11% relacionados a receptores e sinalização 25% “miscelânea” e 32% com função desconhecida, este último número considerado excessivo. No presente trabalho identificamos 30% e 19% dos genes como envolvidos com metabolismo e receptores/sinalização, respectivamente, e 28% com função desconhecida. Recentemente Hashimoto e colaboradores também avaliaram o perfil de expressão (SAGE) em subtipos de leucócitos (HASHIMOTO et al, 2003).

A utilização de SAGE como referência traz como vantagens o fato desta técnica permitir a definição qualitativa e quantitativa dos genes expressos em um tecido sem o prévio conhecimento da seqüência completa de seu transcriptoma, assumindo que cada *tag* corresponde a um único transcrito e que a freqüência de cada *tag* reflete diretamente a abundância do mRNA correspondente. No entanto, limitações do método têm sido descritas valendo salienter o tamanho reduzido da SAGE *tag* (YAMAMOTO et al, 2001), o que pode conferir certa inespecificidade ao procedimento. Além disso, há a restrição do número de amostras analisadas em cada

projeto, limitando, de certa forma, uma análise estatística mais robusta dos achados.

O mapeamento cromossômico realizado nesse trabalho (figura 6 e tabela 4) evidenciou distribuição dos *contigs* ao longo de todos os cromossomos. Venter e colaboradores, em seu trabalho original com o seqüenciamento do genoma humano também observaram algumas peculiaridades que obtivemos aqui, como a maior densidade de genes por milhão de pares de bases no cromossomo 19 e a menor no cromossomo 13 (VENTER et al, 2001). Considerando que as abordagens metodológicas foram distintas, os resultados semelhantes reforçam a precisão desse achado.

Os experimentos de *real-time* PCR superam sobremaneira a acurácia de métodos tradicionais de avaliação da expressão gênica e hoje são considerados o padrão-ouro na validação dos achados de expressão gênica diferencial obtidos por outros métodos (APPLIED BIOSYSTEMS INC., 2001; EL RIFAI et al, 2002). Essas características, aliadas à relativa facilidade de análise, permitem que o método seja utilizado na confirmação dos achados de experimentos envolvendo *microarrays* ou *SAGE*, mesmo quando estão envolvidos um maior número de genes. A utilização de SyberGreen I como agente fluorescente e revelador da presença de produto amplificado no tubo de reação é solução de mais baixo custo em comparação com métodos específicos como emprego de sondas TaqMan, em ensaios de 5' exonuclease na reação de real-time PCR. SyberGreen I mostrou-se versátil, específico e permite maior rapidez na análise, uma vez que um mesmo

agente é utilizável em todas as reações. Em nossas análises subseqüentes dos genes selecionados para confirmação, detectamos expressão diferencial significativa em nove dos 10 genes. Apenas lactoferrina, proteína ligadora do ferro e envolvida em processos de defesa a microorganismos em mucosas, não variou significativamente nas células de LMC em relação às normais. A seguir, descreveremos mais detalhadamente os achados revelados para cada gene investigado e suas possíveis implicações.

CAMP (*cathelicidin antimicrobial peptide*, peptídeo antimicrobiano catelicidina)

Os peptídeos com atividade microbicida encontrados em animais podem ser classificados em dois grupos funcionais: aqueles que se acumulam nos grânulos dos fagócitos e matam os microorganismos fagocitados, como a α -defensina (DAHER et al, 1988) e aqueles que ocorrem em circulação nos fluídos corporais ou mucosas, como a defensina 5 (JONES e BEVINS, 1992). CAMP é uma proteína com atividade antimicrobiana encontrada em grânulos específicos dos leucócitos polimorfonucleares, caracterizada em 1996 (LARRICK et al, 1996), quando sua expressão foi demonstrada especificamente em granulócitos. Demonstramos incremento na produção de mRNA para CAMP nas amostras de paciente com LMC da ordem de 28 vezes (tabelas 5 e 6) fenômeno que poderia ser entendido como resultado do predomínio desse tipo celular (granulócitos) nas amostras neoplásicas (LMC) em relação ao paciente sem neoplasia.

GNAI2 (guanine nucleotide-binding protein, alpha-inhibiting activity polypeptide 2)

Em 1991 Hermouet e colaboradores demonstraram que mutações capazes de ativar ou inativar a função do GNAI2 apresentavam efeitos opostos na proliferação de fibroblastos 3T3 (HERMOUET et al, 1991), o que apontava para possível função desse gene durante o ciclo celular. Na mesma época, Lyons e colaboradores demonstraram mutações no GNAI2 que substituíam a arginina na posição 179 por cisteína ou histidina em tumores da córtex adrenal e tumores endócrinos de ovário (LYONS et al, 1990). Esses autores denominaram esse GNAI2 mutante de oncogene GIP2 e postularam ainda que mutações semelhantes poderiam estar associadas à ativação de outras proteínas G. Williamson e colaboradores também encontraram a mutação Arg179 em um dos 32 pacientes portadores de adenomas hipofisários secretores de ACTH (WILLIAMSON et al, 1995).

Nesse trabalho encontramos aumento da expressão de GNAI2 nas células leucêmicas da ordem de seis vezes, sugerindo a possibilidade de possível papel desse gene no descontrole proliferativo característico dessa neoplasia. A investigação de alterações estruturais como a mutação Arg179 poderia ainda contribuir para essa suspeita.

ABR (active BCR-related gene; gene functional relacionado ao BCR)

O gene BCR, localizado no cromossomo 22 está envolvido em translocações recíprocas com o gene ABL, no cromossomo 9 em LMC com

cromossomo Ph (JAFÉ ES, 2001). Heisterkamp e colaboradores descreveram pela primeira vez um gene funcional relacionado ao BCR, mapeado no cromossomo 17 (17p13.3) (HEISTERKAMP et al, 1989; HEISTERKAMP et al, 1993). McDonald e colaboradores encontraram deleção do locus do ABR em 7 de 8 casos de meduloblastoma, o que forneceu evidências para aqueles autores concluírem pelo papel de gene supressor de tumor do ABR, na porção distal do cromossomo 17p, região em que também se localiza o gene supressor de tumor p53 (MCDONALD et al, 1994). A proteína ABR é semelhante ao BCR, porém não apresenta um domínio estrutural que poderia influenciar seu papel biológico regulatório. Tanto ABR como BCR apresentam um domínio de proteína ativadora de GTPase semelhante àqueles encontrados em outras proteínas que estimulam a hidrólise do GTP por membros da família Rho de proteínas ligantes de GTP, assim como uma região de homologia com o domínio de estimulação de dissociação de guanina do produto do oncogene DBL (CHUANG et al, 1995). Chuang e colaboradores ainda sugerem que as proteínas BCR e ABL poderiam interagir simultaneamente com membros da família Rho para regular e coordenar a sinalização celular.

Foi notável o aumento de 47 vezes da expressão do ABR nas células leucêmicas observado nesse trabalho, sugerindo um papel importante dessa classe de proteínas no controle da proliferação celular dessas células leucêmicas. Embora não tenha sido testado, não é possível afastar a possibilidade de que exista interação mútua do ABR com o gene quimérico BCR/ABL na fisiopatologia da LMC, tal como o alto nível de expressão

detectado poderia sugerir. Essa possível associação permite conjecturas sobre a possibilidade de desenvolvimento de drogas que possam modular seus efeitos, com o objetivo de modificar o comportamento da doença. Finalmente, essa abordagem pode ser particularmente útil nos casos de doença refratária ao tratamento com mesilato de imatinibe (SAVAGE e ANTMAN, 2002).

DEFA3 (defensina alfa 3)

As defensinas representam um grupo de peptídeos citotóxicos e microbicidas produzidos pelos neutrófilos. Três peptídeos dessa família, conhecidos como defensinas 1, 2 e 3 constituem cerca de 30% do conteúdo total de proteína dos grânulos dos grânulos neutrofílicos. As defensinas têm cerca de 29 (DEFA2) a 30 (DEFA 1 e 3) aminoácidos e suas seqüências são idênticas, exceto quanto ao aminoácido N-terminal (alanina na DEFA1 e ácido aspártico na DEFA3). Zhang e colaboradores também demonstraram que linfócitos CD8 também produzem e secretam as defensinas e recentemente as defensinas têm sido implicadas também na resposta do hospedeiro à infecção pelo HIV, particularmente naqueles indivíduos que apresentam evolução mais indolente para AIDS (CHUANG et al, 1995; ZHANG et al, 2002). Halder e colaboradores ainda sugerem um papel de proteção contra destruição dos linfócitos produtores de defensinas no organismo (HALDER et al, 2000).

A expressão aumentada de defensina em células de leucemias mielóides agudas e crônicas já havia sido descrita anteriormente (HALDER

et al, 2000; HERWIG et al, 1996). No presente trabalho observamos aumento da ordem de 74 vezes na expressão de defensina 3 em LMC.

TPT1 (*tumor protein, translationally-controlled 1*)

MacDonald e colaboradores identificaram e purificaram a TPT1, um fator de liberação de histamina dependente de imunoglobulina E, produzido por linfócitos de crianças atópicas e presente em líquidos biológicos de pacientes alérgicos (MACDONALD et al, 1995). Mais recentemente o mesmo grupo têm sugerido o papel desse gene na asma alérgica (MACDONALD, 1997; MACDONALD et al, 1999). Nesse trabalho observamos incremento apenas moderado da expressão de TPT1 em células leucêmicas.

CSF3R (*colony-stimulating factor 3 receptor, granulocyte; receptor do fator estimulador de colônias 3 de granulócitos*)

O fator estimulador de colônias de granulócitos (G-CSF) é glicoproteína de massa 20000 a 25000, produzida pelos macrófagos estimulados por LPS e que exerce um papel importante na granulopoese durante o processo inflamatório. G-CSF é um transdutor de sinal que regula a proliferação, diferenciação e sobrevivência das células mielóides, exercendo suas ações através de receptores específicos na superfície celular presentes não apenas nos granulócitos e seus progenitores, mas também em células da placenta, endoteliais e várias linhagens de carcinoma (FUKUNAGA et al, 1990). O receptor do G-CSF (CSF3R) pode ser dividido em 5 domínios: um

domínio semelhante a imunoglobulina, um homólogo a um receptor de citocina, três domínios de fibronectina tipo III, um domínio transmembrana e uma região citoplasmática (SETO et al, 1992).

Nesse trabalho encontramos mRNA do CSF3R aumentado em cerca de 25 vezes nas células leucêmicas, teoricamente tornando essas células mais suscetíveis ao fator circulante. É possível que o CSF3R possa ter um papel na evolução clonal da LMC.

LTF (lactoferrina; lactotransferrina)

Assim como a transferrina e o antígeno tumoral p97 do melanoma, a LTF pertence à família das proteínas ligantes de ferro, que modulam o metabolismo desse metal, além de participarem da hematopoese e de reações imunológicas. De fato, LTF é secretado em mucosas e no leite materno e diversos autores têm demonstrado seu papel na proteção do hospedeiro contra microorganismos (QIU et al, 1998; SINGH et al, 2002). Nesse trabalho não encontramos diferenças significativas no conteúdo de mRNA para LTF em células de LMC, quando comparadas a células não neoplásica).

MLL2 (*mixed lineage leukemia 2*)

O gene MLL foi originalmente identificado como oncogene freqüentemente envolvido em várias neoplasias hematológicas através do rearranjo cromossômico envolvendo a região 11q23 (JAFFE ES, 2001; YOKOYAMA et al, 2002). Alterações no gene MLL são encontradas em 5 a

10% dos rearranjos cromossômicos presentes em crianças e adultos com leucemia linfóide aguda, leucemia mielóide aguda, leucemias bifenotípicas ou com pouca diferenciação e em síndromes mielodisplásicas (JAFFE ES, 2001; YOKOYAMA et al, 2002). MLL2 é um gene relacionado ao MLL, mapeado em 12q12-q14, sem função conhecida, mas aparentemente também envolvido em duplicações e translocações associados ao câncer (PRASAD et al, 1997). A molécula protéica é constituída por uma longa cadeia polipeptídica, com 5262 aminoácidos e apresenta homologia com o gene TRX de *Drosophila*. MLL2 expressa molécula de mRNA de cerca de 18 kb em diversos tecidos, além das células da linhagem hematopoética. Nesse trabalho encontramos a expressão de MLL2 aumentada em cerca de 27 vezes em células leucêmicas, o que pode constituir um forte indício de seu papel na LMC.

S100A8 (proteína ligante de cálcio S100 A8)

A família das proteínas ligadoras de cálcio S100 contém 20 membros identificados e que participam de processos celulares como crescimento e motilidade celular, regulação do ciclo celular, transcrição e diferenciação, e há ainda descrição de membros da família que participam do fenômeno metastático em meduloblastoma (EL RIFAI et al, 2002; HERNAN et al, 2003). Alguns dos membros dessa família têm sido apontados como potenciais genes supressores de tumor, com supressão de sua expressão quando do fenótipo neoplásico. Não é clara, entretanto, a participação de S100A8 na fisiopatologia do câncer. El-Rifai e colaboradores descreveram

recentemente através da análise de expressão diferencial em câncer gástrico utilizando SAGE, que diversos membros dessa família de proteínas, inclusive a S100A8 apresentavam elevação de sua expressão (EL RIFAI et al, 2002). Em nossa análise detectamos aumento de expressão de S100A8 da ordem de 11 vezes nas células leucêmicas, apontando que esse gene poderia desempenhar um papel significativo em diferentes tipos de leucemias.

MS4A3 (membrane-spanning 4-domains, subfamily A)

MS4A3 representa família de proteínas com função pouco conhecida e que contém 4 domínios transmembrana altamente conservados, flanqueados pelas regiões N-terminal e C-terminal citoplasmáticas. Através de northern blot sua expressão foi confirmada como exclusiva do tecido hematopoético e se propõe, baseado em sua semelhança com CD20 e FCER1B que apresente papel em transdução de sinal (ADRA et al, 1994).

Nesse trabalho encontramos aumento da expressão de MS4A3 da ordem de 17 vezes no tecido da LMC.

6. Conclusões

- Os dados desse trabalho constituem a primeira descrição do transcriptoma da leucemia mielóide crônica (LMC); as seqüências foram geradas pela técnica de ORESTES (*Open Reading Frame Expressed Sequence Tags*). Em relação à classificação funcional, não há diferenças significativas, em termos qualitativos, com os genes anotados por SAGE em medula óssea não neoplásica. Também não houve correlação entre os resultados de expressão obtidos por SAGE e por ORESTES.
- O mapeamento cromossômico dos *contigs* gerados por ORESTES apresentou-se com padrão de distribuição semelhante ao observado por Venter e colaboradores.
- Cerca de 28% dos genes encontrados apresentaram função desconhecida.
- Identificamos e confirmamos o aumento da expressão gênica em nove genes em LMC. Desses, oito tinham aumento superior a 10 vezes o normal (CAMP, ABR, DEFA3, CSF3R, MLL2, S100A8 e MS4A3), com dois deles (DEFA3 e ABR) com aumento de cerca de 50 vezes o normal ou mais. A análise funcional particularizada para cada um desses genes em LMC tem o potencial de identificar possíveis alvos para novas intervenções terapêuticas.

7. Referências Bibliográficas

Referências Bibliográficas

National Center for Biotechnology Information - NCBI. **NCBI/NIH**. 2002.

ADAMS, M. D., KELLEY, J. M., GOCCAYNE, J. D. et al. Complementary DNA sequencing: expressed sequence tags and human genome project. **Science**. 252(5013): 1651-1656, 1991.

ADRA, C. N., LELIAS, J. M., KOBAYASHI, H. et al. Cloning of the cDNA for a hematopoietic cell-specific protein related to CD20 and the beta subunit of the high-affinity IgE receptor: evidence for a family of proteins with four membrane-spanning regions. **Proc. Natl. Acad. Sci. U.S.A.** 91(21): 10178-52, 1994.

ALIZADEH, A. A., EISEN, M. B., DAVIS, R. E. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. **Nature**. 403(6769): 503-511, 2000.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.** 25(17): 3389-3402, 1997.

APPLIED BIOSYSTEMS INC. User Bulletin #2.

<http://www.appliedbiosystems.com/support/>. 2001.

ASHBURNER, M., BALL, C. A., BLAKE, J. A. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nat.Genet.** 25(1): 25-29, 2000.

BOON, K., OSORIO, E. C., GREENHUT, S. F. et al. An anatomy of normal and malignant gene expression. **Proc.Natl.Acad.Sci.U.S.A.** 99(17): 11287-11292, 2002.

BRENT EWING, PHIL GREEN, e DAVID GORDON. Phred, Phrap, and Consed-Autofinish. 2002.

CAMARGO, A. A., SAMAIA, H. P., DIAS-NETO, E. et al. The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. **Proc.Natl.Acad.Sci.U.S.A.** 98(21): 12103-12108, 2001.

CHOMCZYNSKI, P. e SACCHI, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Anal.Biochem.** 162(1): 156-159, 1987.

CHUANG, T. H., XU, X., KAARTINEN, V. et al. Abr and Bcr are multifunctional regulators of the Rho GTP-binding protein family. **Proc.Natl.Acad.Sci.U.S.A.** 92(22): 10282-10286, 1995.

COLLINS, F. S. Shattuck lecture--medical and societal consequences of the Human Genome Project. **N.Engl.J.Med.** 341(1): 28-37, 1999.

DAHER, K. A., LEHRER, R. I., GANZ, T. et al. Isolation and characterization of human defensin cDNA clones. **Proc.Natl.Acad.Sci.U.S.A.** 85(19): 7327-7331, 1988.

DE SOUZA, S. J., CAMARGO, A. A., BRIONES, M. R. et al. Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. **Proc.Natl.Acad.Sci.U.S.A.** 97(23): 12690-12693, 2000.

DHANASEKARAN, S. M., BARRETTE, T. R., GHOSH, D. et al. Delineation of prognostic biomarkers in prostate cancer. **Nature.** 412(6849): 822-826, 2001.

DIAS, Neto E., GARCIA, Correa R., VERJOVSKI-ALMEIDA, S. et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. **Proc.Natl.Acad.Sci.U.S.A.** 97(7): 3491-3496, 2000b.

DIAS, Neto E., GARCIA, Correa R., VERJOVSKI-ALMEIDA, S. et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. **Proc.Natl.Acad.Sci.U.S.A.** 97(7): 3491-3496, 2000a.

DIAS-NETO, E., GARCIA, Correa R., VERJOVSKI-ALMEIDA, S. et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. **Proc.Natl.Acad.Sci.U.S.A.** 97(7): 3491-3496, 2000.

DIAS-NETO, E., TAJARA, E. H., COSTA, F. F. et al. Transcriptome in head and neck cancers. 2003.

EL RIFAI, W., MOSKALUK, C. A., ABDRAHBO, M. K. et al. Gastric cancers overexpress S100A calcium-binding proteins. **Cancer Res.** 62(23): 6823-6826, 2002.

FADERL, S., TALPAZ, M., ESTROV, Z. et al. The biology of chronic myeloid leukemia. **N.Engl.J.Med.** 341(3): 164-172, 1999.

FERRANDO, A. A., NEUBERG, D. S., STAUNTON, J. et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. **Cancer Cell.** 1(1): 75-87, 2002.

FUKUNAGA, R., SETO, Y., MIZUSHIMA, S. et al. Three different mRNAs encoding human granulocyte colony-stimulating factor receptor. **Proc.Natl.Acad.Sci.U.S.A.** 87(22): 8702-8706, 1990.

FUTREAL, P. A., KASPRZYK, A., BIRNEY, E. et al. Cancer and genomics. **Nature.** 409(6822): 850-852, 2001.

GNATENKO, D. V., DUNN, J. J., MCCORKLE, S. R. et al. Transcript profiling of human platelets using microarray and serial analysis of gene expression. **Blood.** 2002.

GOLUB, T. R., SLONIM, D. K., TAMAYO, P. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. **Science.** 286(5439): 531-537, 1999.

GRUBER, A. Latin American Course on Bioinformatics for Tropical Diseases Research. <http://icb.ime.usp.br/tdr/material/arthur/assembly.pdf>. 2002.

GUTTMACHER, A. E. e COLLINS, F. S. Genomic medicine--a primer. **N.Engl.J.Med.** 347(19): 1512-1520, 2002.

HALDER, T. M., BLUGGEL, M., HEINZEL, S. et al. Defensins are dominant HLA-DR-associated self-peptides from CD34(-) peripheral blood mononuclear cells of different tumor patients (plasmacytoma, chronic myeloid leukemia). **Blood.** 95(9): 2890-2896, 2000.

HASHIMOTO, S. I., NAGAI, S., SESE, J. et al. Gene expression profile in human leukocytes. **Blood.** 2003.

HASTIE, N. D. e BISHOP, J. O. The expression of three abundance classes of messenger RNA in mouse tissues. **Cell.** 9(4 PT 2): 761-774, 1976.

HEISTERKAMP, N., KAARTINEN, V., VAN SOEST, S. et al. Human ABR encodes a protein with GAPrac activity and homology to the DBL nucleotide exchange factor domain. **J.Biol.Chem.** 268(23): 16903-16906, 1993.

HEISTERKAMP, N., MORRIS, C., e GROFFEN, J. ABR, an active BCR-related gene. **Nucleic Acids Res.** 17(21): 8821-8831, 1989.

HEREFORD, L. M. e ROSBASH, M. Number and distribution of polyadenylated RNA sequences in yeast. **Cell**. 10(3): 453-462, 1977.

HERMOUET, S., MERENDINO, J. J., Jr., GUTKIND, J. S. et al. Activating and inactivating mutations of the alpha subunit of Gi2 protein have opposite effects on proliferation of NIH 3T3 cells. **Proc.Natl.Acad.Sci.U.S.A.** 88(23): 10455-10459, 1991.

HERNAN, R., FASHEH, R., CALABRESE, C. et al. ERBB2 up-regulates S100A4 and several other prometastatic genes in medulloblastoma. **Cancer Res.** 63(1): 140-148, 2003.

HERWIG, S., SU, Q., ZHANG, W. et al. Distinct temporal patterns of defensin mRNA regulation during drug-induced differentiation of human myeloid leukemia cells. **Blood**. 87(1): 350-364, 1996.

HUANG, X. e MADAN, A. CAP3: A DNA sequence assembly program. **Genome Res.** 9(9): 868-877, 1999.

INCA - MINISTÉRIO DA SAÚDE. Estimativa da Incidência e Mortalidade por Câncer no Brasil. **INCA - Instituto Nacional do Câncer**. 2002.

JAFFE ES, Harris NL Stein H Vardiman JW Eds. World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. 20-26, 2001.

JONES, D. E. e BEVINS, C. L. Paneth cells of the human small intestine express an antimicrobial peptide gene. **J.Biol.Chem.** 267(32): 23216-23225, 1992.

JOSEPH SAMBROOK, David W. Russell Joe Sambrook. Molecular Cloning: A Laboratory Manual. 3rd edition 2001.

LAL, A., LASH, A. E., ALTSCHUL, S. F. et al. A public database for gene expression in human cancers. **Cancer Res.** 59(21): 5403-5407, 1999.

LANDER, E. S., LINTON, L. M., BIRREN, B. et al. Initial sequencing and analysis of the human genome. **Nature.** 409(6822): 860-921, 2001.

LARRICK, J. W., LEE, J., MA, S. et al. Structural, functional analysis and localization of the human CAP18 gene. **FEBS Lett.** 398(1): 74-80, 1996.

LYONS, J., LANDIS, C. A., HARSH, G. et al. Two G protein oncogenes in human endocrine tumors. **Science.** 249(4969): 655-659, 1990.

MACDONALD, S. M. Human recombinant histamine-releasing factor. **Int.Arch.Allergy Immunol.** 113(1-3): 187-189, 1997.

MACDONALD, S. M., PAZNEKAS, W. A., e JABS, E. W. Chromosomal localization of tumor protein, translationally-controlled 1 (TPT1) encoding the

human histamine releasing factor (HRF) to 13q12-->q14. **Cytogenet.Cell Genet.** 84(1-2): 128-129, 1999.

MACDONALD, S. M., RAFNAR, T., LANGDON, J. et al. Molecular identification of an IgE-dependent histamine-releasing factor. **Science.** 269(5224): 688-690, 1995.

MCDONALD, J. D., DANESHVAR, L., WILLERT, J. R. et al. Physical mapping of chromosome 17p13.3 in the region of a putative tumor suppressor gene important in medulloblastoma. **Genomics.** 23(1): 229-232, 1994.

NCBI, NIH. National Center for Biotechnology Information - NCBI. **NCBI/NIH.** 2002.

NCBI/NIH. Serial Analysis of Gene expression Tag to Sequence Mapping (SAGEMap). <http://www.ncbi.nlm.nih.gov/sage/>. 2002.

NCI, NIH. 2001 Cancer Progress Report. **NCI, NIH.** 2001.

NETO, E. D., HARROP, R., CORREA-OLIVEIRA, R. et al. Minilibraries constructed from cDNA generated by arbitrarily primed RT-PCR: an alternative to normalized libraries for the generation of ESTs from nanogram quantities of mRNA. **Gene.** 186(1): 135-142, 1997.

OKUTSU, J., TSUNODA, T., KANETA, Y. et al. Prediction of chemosensitivity for patients with acute myeloid leukemia, according to expression levels of 28 genes selected by genome-wide complementary DNA microarray analysis. **Mol.Cancer Ther.** 1(12): 1035-1042, 2002.

PARKIN, D. M. Global cancer statistics in the year 2000. **Lancet Oncol.** 2(9): 533-543, 2001.

PISANI, P., PARKIN, D. M., BRAY, F. et al. Estimates of the worldwide mortality from 25 cancers in 1990. **Int.J.Cancer.** 83(1): 18-29, 1999.

PRASAD, R., ZHADANOV, A. B., SEDKOV, Y. et al. Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to *Drosophila trithorax*. **Oncogene.** 15(5): 549-560, 1997.

QIU, J., HENDRIXSON, D. R., BAKER, E. N. et al. Human milk lactoferrin inactivates two putative colonization factors expressed by *Haemophilus influenzae*. **Proc.Natl.Acad.Sci.U.S.A.** 95(21): 12641-12646, 1998.

REIS, E. M. e VERJOVSKI-ALMEIDA, S. Global analysis of differential gene expression between prostate cancer and normal prostate tissues using cDNA microarray of open-reading frame expressed sequence tags (ORESTES). **Prostate Cancer Prostatic.Dis.** 3(S1): S362000.

RISINGER, J. I., MAXWELL, G. L., CHANDRAMOULI, G. V. et al. Microarray Analysis Reveals Distinct Gene Expression Profiles among Different Histologic Types of Endometrial Cancer. **Cancer Res.** 63(1): 6-11, 2003.

SAVAGE, D. G. e ANTMAN, K. H. Imatinib mesylate--a new oral targeted therapy. **N.Engl.J.Med.** 346(9): 683-693, 2002.

SAWYERS, C. L. Chronic myeloid leukemia. **N.Engl.J.Med.** 340(17): 1330-1340, 1999.

SCHENA, M., SHALON, D., DAVIS, R. W. et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science.** 270(5235): 467-470, 1995.

SETO, Y., FUKUNAGA, R., e NAGATA, S. Chromosomal gene organization of the human granulocyte colony-stimulating factor receptor. **J.Immunol.** 148(1): 259-266, 1992.

SINGH, P. K., PARSEK, M. R., GREENBERG, E. P. et al. A component of innate immunity prevents bacterial biofilm development. **Nature.** 417(6888): 552-555, 2002.

STRAUSBERG, R. L., BUETOW, K. H., EMMERT-BUCK, M. R. et al. The cancer genome anatomy project: building an annotated gene index. **Trends Genet.** 16(3): 103-106, 2000.

SUZUKI, M., KOBAYASHI, H., TANAKA, Y. et al. Bikunin target genes in ovarian cancer cells identified by microarray analysis. **J.Biol.Chem.** 2003.

VAN DE VIJVER, M. J., HE, Y. D., VAN'T VEER, L. J. et al. A gene-expression signature as a predictor of survival in breast cancer. **N.Engl.J.Med.** 347(25): 1999-2009, 2002.

VELCULESCU, V., ZHANG, L., ZHOU, W. et al. Serial Analysis of Gene Expression - Detailed Protocol. 2000a.

VELCULESCU, V. E. Essay: Amersham Pharmacia Biotech & Science prize. Tantalizing transcriptomes--SAGE and its use in global gene expression analysis. **Science.** 286(5444): 1491-1492, 1999.

VELCULESCU, V. E., VOGELSTEIN, B., e KINZLER, K. W. Analysing uncharted transcriptomes with SAGE. **Trends Genet.** 16(10): 423-425, 2000b.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. et al. Serial analysis of gene expression. **Science.** 270(5235): 484-487, 1995.

VENTER, J. C., ADAMS, M. D., MYERS, E. W. et al. The sequence of the human genome. **Science.** 291(5507): 1304-1351, 2001.

WALLQVIST, A., RABOW, A. A., SHOEMAKER, R. H. et al. Establishing connections between microarray expression data and chemotherapeutic cancer pharmacology. **Mol.Cancer Ther.** 1(5): 311-320, 2002.

WILLIAMSON, E. A., INCE, P. G., HARRISON, D. et al. G-protein mutations in human pituitary adrenocorticotrophic hormone-secreting adenomas. **Eur.J.Clin.Invest.** 25(2): 128-131, 1995.

YAMAMOTO, M., WAKATSUKI, T., HADA, A. et al. Use of serial analysis of gene expression (SAGE) technology. **J.Immunol.Methods.** 250(1-2): 45-66, 2001.

YEOH, E. J., ROSS, M. E., SHURTLEFF, S. A. et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. **Cancer Cell.** 1(2): 133-143, 2002.

YOKOYAMA, A., KITABAYASHI, I., AYTON, P. M. et al. Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. **Blood.** 100(10): 3710-3718, 2002.

ZHANG, L., YU, W., HE, T. et al. Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. **Science.** 298(5595): 995-1000, 2002.

ZHANG, L., ZHOU, W., VELCULESCU, V. E. et al. Gene expression profiles in normal and cancer cells. **Science**. 276(5316): 1268-1272, 1997.

Anexo A:
Termo de Consentimento Informado



CENTRO DE HEMATOLOGIA E HEMOTERAPIA DA UNICAMP TERMO DE INFORMAÇÃO E CONSENTIMENTO

Para obter um maior conhecimento clínico e científico do câncer, o Corpo Clínico deste Centro (médicos e pesquisadores) desenvolve pesquisa clínica científica. Através dessa pesquisa é possível conhecer melhor os mecanismos da doença e, portanto, oferecer novas possibilidades de diagnóstico e tratamento. Ainda mais, este trabalho envolve a busca, em tumores, de novos genes ou de lesões em genes já existentes. Estes estudos são realizados em fragmentos de tumores removidos em cirurgias, ou em material biológico colhido.

Você está sendo admitido(a) neste Centro para estabelecimento de diagnóstico e/ou tratamento de alguma forma de tumor. Para fins de diagnóstico, fator prognóstico e/ou como parte de seu tratamento, há a necessidade da remoção do tumor e/ou material biológico retirado, para exames clínicos laboratoriais, necessários para um diagnóstico definitivo. O restante do tumor que é retirado, ou do material biológico, não é utilizado, sendo então congelado e armazenado para novos exames se necessários. Caso contrário, são descartados, conforme Legislação Sanitária regulamentar sobre o assunto.

A obtenção e o estudo dos referidos fragmentos de tumor e material biológico não implicarão em riscos adicionais no seu tratamento ou na sua cirurgia, nem tampouco, em aumento no tempo da operação ou na extensão da mesma.

O fragmento de tumor e/ou material biológico será identificado no laboratório por um código formado por números e letras e, portanto, sua privacidade e identidade serão sempre preservadas. A eventual inclusão dos resultados em publicação científica será feita de modo a manter o anonimato do paciente.

Concordando com o uso do material para os fins acima descritos, é necessário esclarecê-lo(a) que não existem quaisquer benefícios ou direitos financeiros a receber sobre os eventuais resultados decorrentes da pesquisa. Se você não concordar em doar o material para pesquisa, sua decisão não influenciará, de nenhum modo, no seu tratamento.

Caso você ainda tenha questões a fazer sobre este Termo de Consentimento ou alguma dúvida que não tenha sido esclarecida pelo seu médico, por gentileza, entre em contato com a Coordenadoria do Hemocentro da UNICAMP pelo telefone: (0XX19) 788-8734.

Você receberá uma cópia deste documento e o original será arquivado em seu prontuário. Somente assine este Termo se consentir.

DECLARAÇÃO

Declaro estar ciente das informações ora prestadas, tendo lido atentamente e concordando com todo o teor.

Campinas (SP),de de

.....
Responsável ou Paciente

Nome:

RG :

RGH :

Anexo B

**Tabela Comparativa do Número de Seqüências por *Cluster*
ORESTES com o Número de *SAGE Tags***

CLUSTER	EVENTOS	ANOTAÇÃO
Contig347	256	gi 467236 gb U07643.1 HSU07643 Human lactoferrin (HLF2) mRNA, complete cds
GCAAACAAC	15	lactotransferrin
AAACCAAAA	2	endoglin (Osler-Rendu-Weber syndrome 1)
GCAAACAAC	15	eukaryotic translation initiation factor 5A
SOMA	32	Metabolismo Geral
Contig752	210	gi 188657 gb M19507.1 HUMMPO Human myeloperoxidase mRNA, complete cds
TGCCATCTGT	3	cyclin B1
CTCAATAAAT	3	synaptogyrin 1
CTTCTTGCC	2	ribosomal protein L14
TCAATAAATG	6	syntaxin 5A
SOMA	14	Resposta a Estimulos Externos
Contig3087	173	gi 13788565 ref NM_000518.3 Homo sapiens hemoglobin, beta (HBB), mRNA
TGGATCCTGA	5	glycyl-tRNA synthetase
GCCCACAAGT	3	hemoglobin, beta
GCAAGAAAGT	1973	hemoglobin, beta
GCAAGAAAGT	1973	glucosidase I
TGGATCCTGA	5	hemoglobin, gamma A
CTTCTTGCC	2	ribosomal protein L14
CAGCTTGTC	2	glycyl-tRNA synthetase
GCCCACAAGT	3	LAG1 longevity assurance homolog 2 (S. cerevisiae)
CAGCTTGTC	2	hemoglobin, gamma A
SOMA	3968	Metabolismo Geral
Contig283	170	gi 20379719 gb BC027917.1 Homo sapiens, defensin, alpha 3, neutrophil-specific, clone MGC:34189 IMAGE:5226993, mRNA, complete cds
CAAGGAAAC	11	defensin, alpha 3, neutrophil-specific
GGACGAAAGC	4	defensin, alpha 3, neutrophil-specific
GACTGCTATT	211	defensin, alpha 3, neutrophil-specific
TTCTTTTCT	3	heme-regulated initiation factor 2-alpha kinase
TTCTTTTCT	3	emopamil binding protein (sterol isomerase)
SOMA	232	Resposta a Estimulos Externos
Contig3014	129	gi 4507668 ref NM_003295.1 Homo sapiens tumor protein, translationally-controlled 1 (TPT1), mRNA
CATTTCTAAA	5	phosphofructokinase, platelet
CATTTCTAAA	5	EGF-like-domain, multiple 5
SOMA	10	Função Desconhecida
Contig1273	125	gi 22507513 gb BC019669.1 Homo sapiens, eukaryotic translation elongation factor 1 alpha 1, clone MGC:25051 IMAGE:4478650, mRNA, complete cds
ACTGTTTGTT	2	major histocompatibility complex, class II, DP beta 1
SOMA	2	Regulação de Transcrição
Contig1578	101	gi 22051561 ref XM_083842.5 Homo sapiens phosphoglycerate mutase 1 (brain) (PGAM1), mRNA
CTGGGGTTTC	2	immunoglobulin superfamily, member 4
TTCTTGTTGC	14	ribosomal protein S11

GAAACCCAG		Homo sapiens isolate donor N clone N168K immunoglobulin kappa light 3 chain
SOMA	19	Metabolismo de Carboidratos
Contig2703	95	gi 14043068 ref NM_000517.3 Homo sapiens hemoglobin, alpha 2 (HBA2), mRNA
GCAAGAAGGT		47 glycyl-tRNA synthetase
CCAACGCGCT		17 hemoglobin, alpha 1
GTGCTGTCTC		59 hemoglobin, alpha 1
TCGTCCACGT		23 hemoglobin, alpha 2
GCAAGAAGGT		47 hemoglobin, gamma A
CCCAACGCGC		705 hemoglobin, alpha 2
SOMA	898	Metabolismo Geral
Contig1545	78	gi 11493492 gb AF130094.1 AF130094 Homo sapiens clone FLC0165 mRNA sequence
CCCTTAGCTT		11 myosin regulatory light chain
CCCTTAGCTT		11 myosin, light polypeptide, regulatory, non-sarcomeric (20kD)
ACAACTTAG		2 calmodulin 1 (phosphorylase kinase, delta)
GAAACAAGAT		3 phosphoglycerate kinase 1
SOMA	27	Função Desconhecida
Contig1519	62	gi 21594898 gb BC032230.1 Homo sapiens, bactericidal/permeability-increasing protein, clone IMAGE:5211259, mRNA
TGTGTATTTT		2 Meis1, myeloid ecotropic viral integration site 1 homolog (mouse)
SOMA	2	Resposta a Estimulos Externos
Contig923	58	gi 21594687 gb BC031906.1 Homo sapiens, similar to ribosomal protein L9; 60S ribosomal protein L9, cloneMGC:29903 IMAGE:4996986, mRNA, complete cds"
ATCTTTCTGG		tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta
SOMA	4	Metabolismo de Proteínas
Contig1262	55	gi 21614543 ref NM_002964.3 Homo sapiens S100 calcium binding protein A8 (calgranulin A) (S100A8), mRNA
CCGTCTACAG		3 S100 calcium binding protein A8 (calgranulin A)
ATGCCACGG		6 S100 calcium binding protein A8 (calgranulin A)
SOMA	9	Ciclo Celular
Contig3038	46	gi 4503080 ref NM_000760.1 Homo sapiens colony stimulating factor 3 receptor (granulocyte) (CSF3R), mRNA
TTCCCTTCTT		7 major histocompatibility complex, class II, DP beta 1
GGGCTGGGGT		70 ribosomal protein L29
TGGAGAAGAG		4 thioredoxin interacting protein
SOMA	81	Sinalização Intercelular
Contig378	42	gi 13278743 gb BC004147.1 BC004147 Homo sapiens, lysozyme (renal amyloidosis), clone MGC:2337 IMAGE:2959387, mRNA, complete cds
AATTAATTA		4 eukaryotic translation initiation factor 4 gamma, 1
TATTATGATT		leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM 3 domains),
SOMA	7	Resposta a Estimulos Externos
Contig113	40	gi 22054529 ref XM_167414.2 Homo sapiens vimentin (VIM), mRNA
AGCTGTTTCT		2 ESTs, Highly similar to S29539 ribosomal protein L13a, cytosolic
CTCCTGAAAA		2 G protein-coupled receptor 39
CTCCTGAAAA		2 ubiquitin-conjugating enzyme E2E 3 (UBC4/5 homolog, yeast)
AGCTGTTTCT		2 ribosomal protein L13a
SOMA	8	Matriz Extracelular, Membrana Celular, Citoesqueleto

Contig2890	38	gi 2529714 gb AF001542.1 AF001542 Homo sapiens clone alpha_est218/52C1 mRNA sequence
TGGTAGTTAC	2	dynein, cytoplasmic, intermediate polypeptide 2
GTAACAGAAA	7	general transcription factor II, i, pseudogene 1
SOMA	9	Função Desconhecida
Contig2267	37	gi 14737457 ref XM_041507.1 Homo sapiens guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2 (GNAI2), mRNA
GTTCTGGTTT	20	ATPase inhibitor precursor
CCAGAACAGA	55	ribosomal protein L30
TGGGCAAAGC	28	eukaryotic translation elongation factor 1 gamma
CTGAGTCTCC	4	guanine nucleotide binding protein (G protein), alpha inhibiting activity
SOMA	107	Transdução de Sinal
Contig746	34	gi 14602961 gb BC009970.1 BC009970 Homo sapiens, transketolase (Wernicke-Korsakoff syndrome), clone MGC:15349IMAGE:4310396, mRNA, complete cds
GGTGACAGTG	3	2,4-dienoyl CoA reductase 2, peroxisomal
SOMA	3	Metabolismo Geral
Contig2997	34	gi 22538474 ref NM_018955.2 Homo sapiens ubiquitin B (UBB), mRNA syntrophin, beta 1 (dystrophin-associated protein A1, 59kD, basic 2 component 1)
TGCAGATCTT	2	septin 6
TGGCTGTAA	2	septin 6
SOMA	4	Metabolismo Geral
Contig2757	33	gi 11641400 ref NM_021962.1 Homo sapiens active BCR-related gene (ABR), transcript variant 1, mRNA
CTGTGCCAG	4	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange)
CTGTGCCAG	4	ribosomal protein L30
SOMA	8	Transdução de Sinal
Contig1525	32	gi 22265992 emb X13015.1 HSEDNEUROH.sapiens mRNA for eosinophil-derived neurotoxin (EDN)
CTGGTGTCTG	2	dyskeratosis congenita 1, dyskerin
TGACAACAGA	4	ribonuclease, RNase A family, 3 (eosinophil cationic protein)
SOMA	6	Resposta a Estimulos Externos
Contig124	32	gi 20988189 gb BC030569.1 Homo sapiens, ribosomal protein S15a, clone MGC:44895 IMAGE:5580542, mRNA, complete cds
ATGCCAGCTG	3	ribosomal protein S15a
ATGCCAGCTG	3	polynucleotide phosphorylase-like
ATGCCAGCTG	3	phospholipase A2, group IVB (cytosolic)
GGGATGGAAG	5	Rho GTPase activating protein 8
SOMA	14	Metabolismo de Proteínas
Contig787	31	gi 18089066 gb BC020653.1 BC020653 Homo sapiens, arginase, liver, clone MGC:22316 IMAGE:4712189, mRNA, complete cds
GAAATTTAAA	4	high-mobility group (nonhistone chromosomal) protein 1
AATGCCAGCA	2	thioredoxin-like, 32kD
TTTTCTGAAA	7	Homo sapiens thioredoxin delta 3 (TXN delta 3) mRNA, partial cds
TAAAGACTTG	2	adenylate kinase 2
AATGCCAGCA	2	multiple endocrine neoplasia I
CTAAAGACTT	2	chondroitin sulfate proteoglycan 3 (neurocan)
SOMA	19	Metabolismo de Proteínas
Contig3125	31	gi 509046 emb X70941.1 SSRPT45 S.scrofaDNA for centromere specific motif

CCACTGCATT	9	SCAN domain-containing 2
ACTACCTTCA	2	px19-like protein
CCACTGCATT	9	inactivation escape 1
SOMA	20	Função Desconhecida
Contig1049	31	gi 32432 emb X17042.1 HSHPCP Human mRNA for hematopoietic proteoglycan core protein
TAAAATATGT		cyclin-dependent kinase 7 (MO15 homolog, Xenopus laevis, cdk-2 activating kinase)
AGTAGCTTGA	5	CD53 antigen
CATATCATTA	8	insulin-like growth factor binding protein 7
CATATCATTA	8	polymerase (RNA) II (DNA directed) polypeptide B (140kD)
AGTAGCTTGA	5	3 amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate
TAAAATATGT	2	CD8 antigen, alpha polypeptide (p32)
SOMA	30	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig1109	30	gi 8051583 ref NM_002003.2 Homo sapiens ficolin (collagen/fibrinogen domain containing) 1 (FCN1), mRNA
TGTGGCCTCC	7	heat shock transcription factor 1
ATCAACTGGA	3	nucleobindin 2
CAGAGATGAA	2	heat shock 70kD protein 1A
GGTACAAATA	3	cAMP responsive element binding protein 3 (luman)
CAGAGATGAA	2	TAR DNA binding protein
GGTACAAATA	3	bile acid beta-glucosidase
SOMA	20	Função Desconhecida
Contig515	29	gi 5453749 ref NM_006317.1 Homo sapiens brain abundant, membrane attached signal protein 1 (BASP1), mRNA
CTTTTCTTCT	6	insulin-like growth factor binding protein 7
TCAGATCTTT	99	ribosomal protein S4, X-linked
CTTTTCTTCT	6	polymerase (RNA) II (DNA directed) polypeptide B (140kD)
CTTTTCTTCT	6	BRCA2 and CDKN1A interacting protein
TGGACAAGCT	3	brain abundant, membrane attached signal protein 1
CAGAAGAAAA	2	protein phosphatase 3 (formerly 2B), regulatory subunit B (19kD), alpha
CTTTTCTTCT	6	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 32
SOMA	128	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig1301	28	gi 902627 emb X89658.1 HSCAP18PN H.sapiens mRNA for CAP-18 protein
TTGTCCAGAG	3	polymerase (RNA) II (DNA directed) polypeptide G
TTCACTGTGA	12	lectin, galactoside-binding, soluble, 3 (galectin 3)
SOMA	15	Homeostase
Contig2036	27	gi 22058885 ref XM_055013.5 Homo sapiens ubiquitin C (UBC), mRNA syntrophin, beta 1 (dystrophin-associated protein A1, 59kD, basic component 1)
TGCAGATCTT	2	component 1)
AGAGGTTGAT	3	major histocompatibility complex, class I, A
SOMA	5	Metabolismo Geral
Contig891	26	gi 21757539 dbi_AK097690.1 Homo sapiens cDNA FLJ40371 fis, clone TEST12034931, highly similar to COFILIN.NON-MUSCLE ISOFORM
CCCCAGTTGC	10	calpain, small subunit 1
TCCATCCCTT	5	collapsin response mediator protein 1
GGAAGGGGGA	6	cofilin 1 (non-muscle)
CTGCCAACTT	3	cofilin 1 (non-muscle)
SOMA	24	Matriz Extracelular, Membrana Celular, Citoesqueleto

	gi 4557885 ref NM_000211.1 Homo sapiens integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit) (ITGB2), mRNA"
Contig3565	26
GACTGTGCCA	5 DKFZP434B103 protein
GGTGAAGACA	2 cathepsin H
GACTGTGCCA	5 dynein, cytoplasmic, light polypeptide
SOMA	12 Adesão
	gi 219536 dbj D90278.1 HUMCGM1B Human CGM1b mRNA for CD66d
Contig3265	26
GGCTGGGGGC	40 chromosome 1 amplified sequence 3
CTGGTGTCTG	2 dyskeratosis congenita 1, dyskerin
SOMA	42 Adesão
	gi 23238210 ref NM_152862.1 Homo sapiens actin related protein 2/3 complex, subunit 2, 34kDa (ARPC2), transcript variant 1, mRNA
Contig301	26
CTTTTCTTTA	4 general transcription factor IIIA
CTTTTCTTTA	4 insulin-like growth factor binding protein 3
CCAAACGTGT	39 ferritin, light polypeptide
CCAAACGTGT	39 H3 histone, family 3A
AGAGGGTGGG	2 DnaJ (Hsp40) homolog, subfamily B, member 1
SOMA	88 Mobilidade
	gi 22597103 gb AF521670.1 Homo sapiens SWI/SNF chromatin remodeling complex subunit OSA1 (OSA1) mRNA, partial cds
Contig952	25
AGAGGGTGGG	2 DnaJ (Hsp40) homolog, subfamily B, member 1
SOMA	2 Regulação de Transcrição
	gi 561638 gb L35848.1 HUMIERB Homo sapiens IgE receptor beta chain (HTm4) mRNA, complete cds
Contig400	25
TGAAATAAAA	11 nucleophosmin (nucleolar phosphoprotein B23, numatrin)
CTGTATTTGA	4 transformer-2 alpha (htra-2 alpha)
SOMA	15 Transdução de Sinal
	gi 5031852 ref NM_005564.1 Homo sapiens lipocalin 2 (oncogene 24p3) (LCN2), mRNA
Contig2834	25
ACACTTCTTT	3 CDw92 antigen
TCAAGAAAGT	SWI/SNF related, matrix associated, actin dependent regulator of chromatin,
SOMA	6 Organização Citoplasmática & Biogênese
	gi 22058218 ref XM_006748.6 Homo sapiens seven transmembrane protein TM7SF3 (TM7SF3), mRNA
Contig1630	25
GAGGGAGTTT	77 ribosomal protein L27a
GAGGGAGTTT	77 ESTs, Highly similar to 2113200C ribosomal protein L27a [H.sapiens]
TTTTTGTATT	20 thioredoxin interacting protein
ATTTTAGAAT	3 RAN binding protein 2
SOMA	177 Transdução de Sinal
	gi 19923220 ref NM_002483.2 Homo sapiens carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen) (CEACAM6), mRNA
Contig1354	25
TAATACCAAG	2 splicing factor, arginine/serine-rich 3
SOMA	2 Adesão
	gi 22060548 ref XM_113343.2 Homo sapiens myosin, light polypeptide, regulatory, non-sarcomeric (20kD) (MLCB), mRNA
Contig1332	25
TCATCTTTGT	2 mitochondrial ribosomal protein L3
CCCTTAGCTT	11 myosin regulatory light chain
TCATCTTTGT	2 thrombospondin 1

CCCTTAGCTT	11	myosin, light polypeptide, regulatory, non-sarcomeric (20kD)
SOMA	26	Mobilidade
Contig1250	25	gii4103561 gb AF025395.1 AF025395Cercopithecus aethiops B42b mRNA, complete cds
GAAGATCCCT	2	myosin IF
SOMA	2	Função Desconhecida
Contig2360	23	gii22063181 ref XM_042765.4 Homo sapiens chromosome 20 open reading frame 3 (C20orf3), mRNA
TGGAAGCTGTG	2	CGI-30 protein
TGGAAGCTGTG	2	sialic acid binding Ig-like lectin 8
SOMA	4	Função Desconhecida
Contig890	22	gii20379495 gb BC027725.1 Homo sapiens, enolase 1, (alpha), clone MGC:23185 IMAGE:4843306, mRNA, complete cds
TGGAAAGTGA	6	v-fos FBJ murine osteosarcoma viral oncogene homolog
SOMA	6	Regulação de Transcrição
Contig656	22	gii6912575 ref NM_012387.1 Homo sapiens peptidyl arginine deiminase, type V (PADI5), mRNA
GAGCTCTTGT	2	Fanconi anemia, complementation group A
CAGCCCTCCC	2	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma
SOMA	4	Metabolismo de Proteínas
Contig352	22	gii4502264 ref NM_001675.1 Homo sapiens activating transcription factor 4 (tax-responsive enhancer elementB67) (ATF4), mRNA
CTGGCCAAGG	6	guanylate kinase 1
GTGGCTGCTG	3	mitochondrial ribosomal protein S36
SOMA	9	Regulação de Transcrição
Contig3198	22	gii337452 gb M29064.1 HUMRNPB1A Human hnRNP B1 protein mRNA
CCTGGCTAAT	3	p53-responsive gene 3
TTAGCCAGGA	3	folate transporter/carrier
CCTGGCTAAT	3	Opa-interacting protein 2
TCACCTGAAA	2	DKFZP564K247 protein
SOMA	11	Metabolismo de RNA
Contig2844	22	gii20149554 ref NM_030666.2 Homo sapiens serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member1 (SERPINB1), mRNA
CTTTTCTTTA	4	general transcription factor IIIA
CTTTTCTTTA	4	insulin-like growth factor binding protein 3
SOMA	8	Função Desconhecida
Contig2205	22	gii21359983 ref NM_003651.3 Homo sapiens cold shock domain protein A (CSDA), mRNA
CATTTTGGAT	2	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta,
CATTTTGGAT	2	glutamate-cysteine ligase, catalytic subunit
SOMA	4	Regulação de Transcrição
Contig606	21	gii7657016 ref NM_015383.1 Homo sapiens hypothetical protein DJ328E19.C1.1 (DJ328E19.C1.1), mRNA
GCTTAACCTG	4	glutamate dehydrogenase 1
SOMA	4	Função Desconhecida
Contig2346	21	gii4502218 ref NM_001665.1 Homo sapiens ras homolog gene family, member G (rho G) (ARHG), mRNA
GATGAACTGG	6	"doublecortex; lissencephaly, X-linked (doublecortin)"
SOMA	6	Ciclo Celular
Contig212	21	gii31321 emb X16863.1 HSEFCGR31 Human Fc-gamma RIII-1 cDNA for Fc-gamma receptor III-1 (CD 16)

CAGAAAGAGA	2 actin, alpha 1, skeletal muscle
TTTGGAATGT	4 matrin 3
AGTCCTGTGT	2 Homo sapiens clone 23785 mRNA sequence
TTTGGAATGT	4 inositol polyphosphate-4-phosphatase, type I, 107kD
SOMA	12 Reconhecimento
Contig339	20 gi 337998 gb M63573.1 HUMSCYLP Human secreted cyclophilin-like protein (SCYLP) mRNA, complete cds
AGAAAGATGT	9 annexin A1
TGGTGTGG	6 peptidylprolyl isomerase A (cyclophilin A)
SOMA	15 Estrutura & Organização Intracelular
Contig3171	20 gi 4503598 ref NM_001983.1 Homo sapiens excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)(ERCC1), mRNA
CCCGAGGAAG	2 poly(A) binding protein, cytoplasmic 4 (inducible form)
SOMA	2 Metabolismo de DNA
Contig1249	20 gi 20543776 ref XM_045140.5 Homo sapiens prosaposin (variant Gaucher disease and variant metachromaticleukodystrophy) (PSAP), mRNA
TCAACAGCCA	2 legumain
TCAACAGCCA	2 eukaryotic translation initiation factor 3, subunit 5 (epsilon, 47kD)
TCCCTGGCTG	8 leukodystrophy)
TCCCTGGCTG	8 A thiolase)
SOMA	20 Metabolismo de Lipídeos
Contig1221	20 gi 11275977 gb AF308601.1 AF308601 Homo sapiens NOTCH 2 (N2) mRNA, complete cds
CTGTTGGCAT	ESTs, Highly similar to S55913 ribosomal protein L21, cytosolic 5 [H.sapiens]
SOMA	5 Diferenciação Celular e Tissular
Contig1010	20 gi 21756970 dbj AK097275.1 Homo sapiens cDNA FLJ39956 fis, clone SPLEN2024990, highly similar to L-PLASTIN
TGGCTCCTCC	12 calreticulin
SOMA	12 Função Desconhecida
Contig897	19 gi 4505028 ref NM_000895.1 Homo sapiens leukotriene A4 hydrolase (LTA4H), mRNA
CATCCCGTGA	7 leukotriene A4 hydrolase
TTGCTTTGTG	2 HT014
CACCAGCATT	3 CREBBP/EP300 inhibitory protein 1
SOMA	12 Metabolismo Geral
Contig2350	19 gi 4503532 ref NM_001417.1 Homo sapiens eukaryotic translation initiation factor 4B (EIF4B), mRNA
GAGCTCTGAG	5 dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)
TCAGATAGGA	5 Homo sapiens clone 124-1V1, mRNA sequence
TCAGATAGGA	5 mitogen-activated protein kinase-activated protein kinase 3
SOMA	15 Regulação de Transcrição
Contig223	19 gi 12025669 ref NM_001102.2 Homo sapiens actinin, alpha 1 (ACTN1), mRNA
TAAAATATGT	2 CD8 antigen, alpha polypeptide (p32)
SOMA	2 Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig959	18 gi 6996020 ref NM_000397.2 Homo sapiens cytochrome b-245, beta polypeptide (chronic granulomatous disease)(CYBB), mRNA
TCTTCTGCTT	3 hepatocellular carcinoma-associated antigen 59

CCAGCTGCCA	3 mutL homolog 3 (E. coli)
SOMA	6 Resposta a Estimulos Externos
Contig814	18 gij6005813 ref NM_007271.1 Homo sapiens serine/threonine kinase 38 (STK38), mRNA
TCTGTCTTAG	3 grancalcin, EF-hand calcium binding protein
SOMA	3 Metabolismo de Proteínas
Contig2938	18 gij15990556 gb BC015690.1 BC015690 Homo sapiens, ribosomal protein, large, P0, clone MGC:23431 IMAGE:4660892, mRNA, complete cds
GGCCAGGACT	3 formyl peptide receptor 1
GTGGCTGCTG	3 mitochondrial ribosomal protein S36
SOMA	6 Metabolismo de Proteínas
Contig2681	18 gij16307048 gb BC009601.1 BC009601 Homo sapiens, peroxiredoxin 3, clone MGC:14693 IMAGE:4134666, mRNA, complete cds
ATGAATAAAA	2 microtubule-associated protein, RP/EB family, member 2
SOMA	2 Função Desconhecida
Contig2464	18 gij17981701 ref NM_079421.1 Homo sapiens cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4) (CDKN2D), transcript variant 2, mRNA
TGTGGCCTGC	2 ubiquitin-like 4
TGTGGCCTGC	2 ECSIT
TGTGGCCTGC	2 ribosomal protein S4, X-linked
SOMA	6 Ciclo Celular
Contig2177	18 gij4757825 ref NM_004048.1 Homo sapiens beta-2-microglobulin (B2M), mRNA
GTGCTGTCTC	59 hemoglobin, alpha 1
TTTGATGTAT	3 beta-2-microglobulin
SOMA	62 Função Desconhecida
Contig18	18 gij20380703 gb BC027732.1 Homo sapiens, LGN protein, clone MGC:29478 IMAGE:5020191, mRNA, complete cds
TTTTTAGAAT	2 programmed cell death 4 (neoplastic transformation inhibitor)
AATACTTCTC	2 T cell receptor beta locus
SOMA	4 Transdução de Sinal
Contig909	17 gij32153 emb X02457.1 HSHLAA2 Human mRNA fragment for HLA-A2 class I antigen
GGAGGTGGGG	10 granulin
CTGACCTGTG	99 Cw1 antigen
CTGACCTGTG	99 nascent-polypeptide-associated complex alpha polypeptide
CTGACCTGTG	99 major histocompatibility complex, class I, C
GTGCACTGAG	30 major histocompatibility complex, class I, A
SOMA	337 Função Desconhecida
Contig3407	17 gij602451 gb L27943.1 HUMCYDE Homo sapiens cytidine deaminase (CDA) mRNA, complete cds
ACGGTCCAGG	3 cytidine deaminase
GGCTGGGGCC	21 elastase 2, neutrophil
CCCCCTGCCC	3 E3 ubiquitin ligase SMURF1
SOMA	27 Metabolismo de DNA
Contig33	17 gij8745103 emb AJ251593.1 HSA251593 Homo sapiens mRNA for human leucocyte antigen B (HLA-B gene, HLA-B44Nullallele)
GTGATCTCCG	3 DnaJ (Hsp40) homolog, subfamily C, member 5
TGGAGAAGAG	4 thioredoxin interacting protein
GGGCCTGTGC	5 solute carrier family 16 (monocarboxylic acid transporters), member 3

TGTGGTGGTG	2 MLN51 protein
AGAACCTTCC	20 major histocompatibility complex, class I, A
SOMA	34 Reconhecimento
Contig1985	17 gi 16552240 dbj_AK056750.1 Homo sapiens cDNA FLJ32188 fis. clone PLACE6002056, highly similar to GUANINENUCLEOTIDE-BINDING PROTEIN G(I)/G(S)/G(T) BETA SUBUNIT 2
GGGGTGGGGC	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and 2 dihydroorotase
SOMA	2 Transdução de Sinal
Contig1586	17 gi 1262818 emb_X95240.1 HSCRISP3GH.sapiens mRNA for cysteine-rich secretory protein-3
TTTTTACAGT	2 ATPase, H+ transporting, lysosomal 70kD, V1 subunit A, isoform 1
AATGAAAAA	3 RAD51 homolog C (S. cerevisiae)
CCAGGATCTG	2 protein phosphatase 2, regulatory subunit B (B56), gamma isoform
CAAAACTGTA	6 glycoporphin E
TTTTTACAGT	2 putative secreted protein
CTTAATCTTG	5 stromal cell-derived factor 1
SOMA	20 Função Desconhecida
Contig1279	17 gi 31101 emb_X63526.1 HSEF1G H.sapiens mRNA for protein homologous to elongation factor 1-gamma from A.salina
ACTCCAGAAA	3 eukaryotic translation initiation factor 3, subunit 6 (48kD)
ACTCCAGAAA	3 ribosomal protein S15
ACTCCAGAAA	ESTs, Highly similar to R3HU15 ribosomal protein S15, cytosolic 3 [H.sapiens]
TGGACAAGCT	3 brain abundant, membrane attached signal protein 1
CTGGGTCTCC	"ESTs, Highly similar to ribosomal protein L13; 60S ribosomal protein 2 L13;"
CTGGGTCTCC	2 ribosomal protein L13
SOMA	16 Metabolismo de RNA
Contig739	16 gi 6006013 ref NM_000632.2 Homo sapiens integrin, alpha M (complement component receptor 3, alpha; also known as CD11b (p170), macrophage antigen alpha polypeptide) (ITGAM), mRNA"
GTGTGAGTGT	2 makorin, ring finger protein, 1
SOMA	2 Adesão
Contig3427	16 gi 4505226 ref NM_002432.1 Homo sapiens myeloid cell nuclear differentiation antigen (MND4), mRNA
TTTTGTAAAT	2 GL004 protein
SOMA	2 Diferenciação Celular e Tissular
Contig3357	16 gi 21739936 emb_AL834322.1 HSM805370 Homo sapiens mRNA; cDNA DKFZp761L24121 (from clone DKFZp761L24121)"
CGCTGGTTCC	34 eukaryotic translation elongation factor 1 alpha 1
SOMA	34 Função Desconhecida
Contig2838	16 gi 21758446 dbj_AK098434.1 Homo sapiens cDNA FLJ25568 fis. clone JTH03907, highly similar to HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN E E*0101/E*0102 PRECURSOR
GTGAAACCCC	92 "kangai 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2"
CTGACCTGTG	99 Cw1 antigen
CTGACCTGTG	99 nascent-polypeptide-associated complex alpha polypeptide
CTGACCTGTG	99 major histocompatibility complex, class I, C
SOMA	389 Função Desconhecida
Contig203	16 gi 4885234 ref NM_005248.1 Homo sapiens Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog (FGR), mRNA
CATCCCCACC	2 pre-B-cell leukemia transcription factor 2

SOMA	2	Metabolismo de Proteínas
Contig1360	16	gi 16198356 gb BC015822.1 BC015822 Homo sapiens lactotransferrin clone MGC:13618 IMAGE:4251222, mRNA, complete cds
GCAAAACAAC	15	lactotransferrin
AAACCAAAA	2	endoglin (Osler-Rendu-Weber syndrome 1)
GCAAAACAAC	15	eukaryotic translation initiation factor 5A
SOMA	32	Metabolismo Geral
Contig885	15	gi 14141191 ref NM_005617.2 Homo sapiens ribosomal protein S14 (RPS14), mRNA
AGGGCCGACT	2	antigen identified by monoclonal antibody Ki-67
AAGGAGCTGG	2	YME1-like 1 (S. cerevisiae)
AAGGAGCTGG	2	zinc finger protein-like 1
AGGGCCGACT	2	filamin C, gamma (actin binding protein 280)
SOMA	8	Metabolismo de Proteínas
Contig3492	15	gi 22507396 ref NM_002473.2 Homo sapiens myosin, heavy polypeptide 9, non-muscle (MYH9), mRNA
ATTCAGAGCT	45	hemoglobin, gamma A
ATTCAGAGCT	45	glycyl-tRNA synthetase
SOMA	90	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig3421	15	gi 22054442 ref XM_096333.5 Homo sapiens core promoter element binding protein (COPEB), mRNA
GCCGTGAGCA	3	surfactant, pulmonary-associated protein C
TCAAAGACC	2	v-fos FBJ murine osteosarcoma viral oncogene homolog
SOMA	5	Transdução de Sinal
Contig3378	15	gi 337929 gb M22146.1 HUMSCAR Human scarprotein mRNA, complete cds
AGCACCTCCA	30	eukaryotic translation elongation factor 2
TAACAAAAT	2	proteasome (prosome, macropain) 26S subunit, non-ATPase, 10
CTGTTGGTGA	54	ribosomal protein S23
GTTTGTGCC	3	ribosomal protein S4, X-linked
SOMA	89	Metabolismo de Proteínas
Contig31	15	gi 20561345 ref XM_042224.2 Homo sapiens hypothetical protein FLJ11151 (FLJ11151), mRNA
GTGAACTCC	5	heat shock 70kD protein 5 (glucose-regulated protein, 78kD)
CGCCTGTAAT	5	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)
GGAGGCTGAG	5	DKFZP434F1735 protein
SOMA	15	Função Desconhecida
Contig3183	15	gi 19923846 ref NM_030813.2 Homo sapiens suppressor of potassium transport defect 3 (SKD3), mRNA
TCAAAGACC	2	v-fos FBJ murine osteosarcoma viral oncogene homolog
SOMA	2	Transporte
Contig2841	15	gi 311373 emb X57351.1 HS18D Human 1-8D gene from interferon-inducible gene family
ACCATTCTGC	16	interferon induced transmembrane protein 2 (1-8D)
CTGGGCCTGG	4	arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)
CAAATCAGG	9	cyclin I
CTGGGCCTGG	4	similar to vaccinia virus HindIII K4L ORF
GATTTGGGC	2	choline/ethanolaminephosphotransferase
CCCTGATTTT	10	CD1D antigen, d polypeptide
ACCTGTATCC	12	interferon induced transmembrane protein 3 (1-8U)
SOMA	57	Resposta a Estimulos Externos

		gi 22477183 gb BC036723.1 Homo sapiens, Fc fragment of IgG, low affinity IIIa, receptor for (CD16), clone MGC:34195 IMAGE:5206097, mRNA, complete cds
Contig209	15	
CATCCCCACC		2 pre-B-cell leukemia transcription factor 2
TCAGTTCCAA		11 Fc fragment of IgG, low affinity IIIa, receptor for (CD16)
SOMA	13	Resposta a Estimulos Externos
		gi 14714544 gb BC010407.1 BC010407 Homo sapiens, mitogen-activated protein kinase-activated protein kinase 3, clone MGC:15263 IMAGE:3833667, mRNA, complete cds
Contig1903	15	
TACTCTGGGT		29 ribonuclease, RNase A family, 3 (eosinophil cationic protein)
CCTAAGGCTA		8 E2F transcription factor 4, p107/p130-binding carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and 2 dihydroorotase
GGGGTGGGGC		
SOMA	39	Transdução de Sinal
		gi 22760737 dbj AK074956.1 Homo sapiens cDNA FLJ90475 fis, clone NT2RP3002721, highly similar to Porcine citrate synthase mRNA
Contig1334	15	
TGAGGCAGGG		2 syntaxin 5A
TTCTTTTCT		3 heme-regulated initiation factor 2-alpha kinase
TATCAGGTAC		2 phytanoyl-CoA hydroxylase (Refsum disease)
GGGGCTTCTG		3 legumain
SOMA	10	Função Desconhecida
		gi 404044 gb L20688.1 HUMLYGDI Human GDP-dissociation inhibitor protein (Ly-GDI) mRNA, complete cds
Contig2839	14	
GGCAAATTAT		membrane-spanning 4-domains, subfamily A, member 3 (hematopoietic 4 cell-specific)
CCTGCTCCCT		2 Homo sapiens clone 23718 mRNA sequence
SOMA	6	Metabolismo de Proteínas
		gi 23397667 ref NM_005719.2 Homo sapiens actin related protein 2/3 complex, subunit 3, 21kDa (ARPC3), mRNA
Contig2472	14	
AAGAGACAGT		2 polymerase (RNA) III (DNA directed) (62kD)
TTGCTTTGTG		2 HT014
TCTTTACTTG		6 actin related protein 2/3 complex, subunit 3B (21 kD)
CCCTTCACTG		4 nucleolar RNA-associated protein
SOMA	14	Mobilidade
		gi 5713320 ref NM_000655.2 Homo sapiens selectin L (lymphocyte adhesion molecule 1) (SELL), mRNA
Contig170	14	
GAAATTTAAA		4 high-mobility group (nonhistone chromosomal) protein 1
SOMA	4	Adesão
		gi 22450814 gb BC018662.1 Homo sapiens, protein tyrosine phosphatase type IVA, member 2, clone IMAGE:4108553, mRNA
Contig1634	14	
GCTTTATTTG		29 dynactin 4 (p62)
GCTTTATTTG		29 actin, gamma 1
SOMA	58	Matriz Extracelular, Membrana Celular, Citoesqueleto
		gi 4503778 ref NM_002029.1 Homo sapiens formyl peptide receptor 1 (FPR1), mRNA
Contig1602	14	
AAGTGAGGAG		9 F-box and WD-40 domain protein 7 (archipelago homolog, Drosophila)
TACCTTCTGC		331 defensin, alpha 3, neutrophil-specific
GGCCAGGACT		3 formyl peptide receptor 1
AAGTGAGGAG		9 WW domain binding protein 2
SOMA	352	Transdução de Sinal
		gi 4758393 ref NM_004475.1 Homo sapiens flotillin 2 (FLOT2), mRNA
Contig1295	14	
GGAGACAGAG		5 calmodulin 3 (phosphorylase kinase, delta)

SOMA	5	Adesão
Contig1173	14	gi 1136139 emb X89109.1 HSCORONINH.sapiens mRNA for coronin
CCACTCCTCA	5	defender against cell death 1
GACCCTGCCC	4	fibroblast growth factor receptor 4
GACCCTGCCC	4	FK506 binding protein 8 (38kD)
SOMA	13	Apoptose
Contig327	13	gi 23110949 ref NM_001909.3 Homo sapiens cathepsin D (lysosomal aspartyl protease) (CTSD), mRNA
GGAACTGAA	2	ribosomal protein L3
GGAGGTGGGG	10	granulin ESTs, Highly similar to S34195 ribosomal protein L3, cytosolic
GGAACTGAA	2	[H.sapiens]
GCCCAGCTGG	2	associated molecule with the SH3 domain of STAM
GTATTGGCCT	3	transmembrane 9 superfamily member 2 eukaryotic translation elongation factor 1 delta (guanine nucleotide
GCCCAGCTGG	2	exchange
SOMA	21	Metabolismo de Proteínas
Contig2816	13	gi 188264 gb K01171.1 HUMMHDRA Human HLA-DR alpha-chain mRNA
CAATATTACA	2	guanylate binding protein 5
SOMA	2	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig2760	13	gi 10437796 dbj AK025307.1 Homo sapiens cDNA: FLJ21654 fis, clone COL08589, highly similar to HUMCAPA Homosapiens mitochondrial carnitine palmitoyltransferase I mRNA
AATTAATTA	4	eukaryotic translation initiation factor 4 gamma, 1
SOMA	4	Função Desconhecida
Contig2586	13	gi 15209314 emb AL161784.13 AL161784 Human DNA sequence from clone RP11-342H3 on chromosome 9, complete sequence[Homo sapiens]
GTGGAGTTCA	2	adenylate cyclase 7
GTGGAGTTCA	2	bromodomain-containing 7
SOMA	4	Função Desconhecida
Contig2458	13	gi 20336744 ref NM_004893.2 Homo sapiens H2A histone family, member Y (H2AFY), transcript variant 2, mRNA
CTTTTGTTT	2	transmembrane 9 superfamily member 2
AATTTGCAAC	4	H2A histone family, member Y
ATGAAACCCT	5	SOCS box-containing WD protein SWiP-1
CACCACAACA	4	eukaryotic translation elongation factor 1 gamma
SOMA	15	Regulação de Transcrição
Contig2404	13	gi 21359838 ref NM_003461.2 Homo sapiens zyxin (ZYX), mRNA
GCCCTCGGCC	2	protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform
SOMA	2	Adesão
Contig2216	13	gi 16877500 gb BC017004.1 BC017004 Homo sapiens, tubulin alpha 1, clone MGC:22108 IMAGE:4430457, mRNA, complete cds
ACTGGTACGT	9	ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit f,
TGTACCTGTA	69	tubulin, alpha, ubiquitous
GCTGCCCTTG	8	tubulin, alpha, ubiquitous
GCTGCCCTTG	8	tubulin alpha 6
TATGAGGAGG	14	DNA segment on chromosome X (unique) 9928 expressed sequence
SOMA	108	Estrutura & Organização Intracelular

Contig1992	gi 21361321 ref NM_006087.2 Homo sapiens tubulin, beta, 5 (TUBB5), mRNA
TAGGAGCTGA	2 ribosomal protein S26
GCTGGGGTCA	9 defensin, alpha 1, myeloid-related sequence
TACCTCTGAC	2 ATPase, Class VI, type 11A
TGAGGCAGGG	2 syntaxin 5A
SOMA	15 Estrutura & Organização Intracelular
Contig1919	gi 181246 gb M92287.1 HUMCYCD3A Homo sapiens cyclin D3 (CCND3) mRNA, complete cds
AGGAGCTGCT	NADH dehydrogenase (ubiquinone) Fe-S protein 8 (23kD) (NADH-coenzyme Q)
GCCTGGCCAT	3 DKFZp564J157 protein
CCTCCAGCAG	3 retinoic acid receptor responder (tazarotene induced) 3
GGGCCTGGCC	2 midkine (neurite growth-promoting factor 2)
SOMA	10 Ciclo Celular
Contig1852	gi 565642 dbj D28877.1 HUMHNRNP Homo sapiens HNRPA2B1 gene for hnRNP protein A2 and B1, complete cds
TCACCTGAAA	2 DKFZP564K247 protein
SOMA	2 Metabolismo de RNA
Contig1705	gi 10047156 dbj AB046766.1 Homo sapiens mRNA for KIAA1546 protein, partial cds
TTTTAAAAA	4 sorting nexin 10
SOMA	4 Função Desconhecida
Contig1667	gi 31328 emb X62572.1 HSFPC23 H.sapiens RNA for Fc receptor, PC23
GGCACAGAAG	2 mutL homolog 3 (E. coli)
TGGCTGTGTG	6 supervillin
CTTCTGTGCC	3 serine/threonine kinase 11 (Peutz-Jeghers syndrome)
SOMA	11 Função Desconhecida
Contig1633	gi 21594898 gb BC032230.1 Homo sapiens, bactericidal/permeability-increasing protein, clone IMAGE:5211259,mRNA
TGTGTATTTT	2 Meis1, myeloid ecotropic viral integration site 1 homolog (mouse)
SOMA	2 Homeostase
Contig1372	gi 10435901 dbj AK023842.1 Homo sapiens cDNA FLJ13780 fis, clone PLACE4000450, highly similar to Homo sapiensBAZ2A mRNA for bromodomain adjacent to zinc finger domain 2A
GGAGGTGGGG	10 granulin
CTCCCTTGCC	2 capping protein (actin filament), gelsolin-like
SOMA	12 Função Desconhecida
Contig128	gi 16876941 gb BC016748.1 BC016748 Homo sapiens, ribosomal protein L37a, clone MGC:26772 IMAGE:4831278, mRNA,complete cds
TCGCCGCGAC	3 ribosomal protein L37a Homo sapiens cDNA FLJ13357 fis, clone PLACE1000061, weakly similar to
TCGCCGCGAC	3 similar to
AAGACAGTGG	48 ribosomal protein L37a
AAGACAGTGG	48 histone deacetylase 2
GCAAACCAA	2 spermidine/spermine N1-acetyltransferase Homo sapiens cDNA FLJ13357 fis, clone PLACE1000061, weakly similar to
AAGACAGTGG	48 similar to
SOMA	152 Metabolismo de Proteínas
Contig1043	gi 7019342 ref NM_013230.1 Homo sapiens CD24 antigen (small cell lung carcinoma cluster 4 antigen) (CD24),mRNA

GGAACAAACA	8 thyroid autoantigen 70kD (Ku antigen)
SOMA	8 Adesão
Contig8	12 gij14290497 gb BC009017.1 BC009017 Homo sapiens, S100 calcium-binding protein A6 (calcyclin), clone MGC:18233IMAGE:4157026, mRNA, complete cds
CCCCCTGGAT	20 S100 calcium binding protein A6 (calcyclin)
CCCCTGGATC	3 S100 calcium binding protein A6 (calcyclin)
SOMA	23 Ciclo Celular
Contig732	12 gij4557758 ref NM_000250.1 Homo sapiens myeloperoxidase (MPO), nuclear gene encoding mitochondrial protein, mRNA
TGCCATCTGT	3 cyclin B1
CTCAATAAAT	3 synaptogyrin 1
CTTTCTTGCC	2 ribosomal protein L14
TCAATAAATG	6 syntaxin 5A
SOMA	14 Resposta a Estimulos Externos
Contig388	12 gij16151096 gb AY047586.1 Homo sapiens calreticulin (CALR) mRNA, complete cds
GGTGATGAGG	2 putative breast adenocarcinoma marker (32kD)
SOMA	2 Regulação de Transcrição
Contig3408	12 gij7661625 ref NM_015640.1 Homo sapiens PAI-1 mRNA-binding protein (PAI-RBP1), mRNA
AGCACATTTG	8 fumarylacetoacetate hydrolase (fumarylacetoacetase)
AGCACATTTG	8 coactosin-like protein
AGGAGCAAGG	2 biliverdin reductase B (flavin reductase (NADPH))
SOMA	18 Metabolismo de RNA
Contig3401	12 gij14589948 ref NM_000937.2 Homo sapiens polymerase (RNA) II (DNA directed) polypeptide A, 220kDa (POLR2A), mRNA
TACCCACCC	9 MYC-associated zinc finger protein (purine-binding transcription factor)
SOMA	9 Metabolismo de DNA
Contig3211	12 gij22046152 ref XM_171150.1 Homo sapiens similar to KIAA1074 protein (LOC254027), mRNA
CTTTTCTTCT	6 insulin-like growth factor binding protein 7
CTTTTCTTCT	6 polymerase (RNA) II (DNA directed) polypeptide B (140kD)
CTTTTCTTCT	6 BRCA2 and CDKN1A interacting protein
GACTCTGAAA	5 RNA binding motif protein 3
CTTTTCTTCT	6 DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 32
GACTCTGAAA	5 ribosomal protein S15a
GACTCTGAAA	5 polynucleotide phosphorylase-like
SOMA	39 Função Desconhecida
Contig3182	12 gij11055671 gb AF304203.1 AF304203 Susicrofa breed Swedish wild boar mitochondrion, partial genome
TGGGGGTTTC	3 ferritin, heavy polypeptide 1
SOMA	3 Função Desconhecida
Contig3131	12 gij17386075 gb AF445025.1 AF445025 Homo sapiens clone 83 tumor rejection antigen mRNA, complete cds
TCCCCACACC	2 integrin-linked kinase
GAAACCCAG	Homo sapiens isolate donor N clone N168K immunoglobulin kappa light chain
SOMA	5 Função Desconhecida
Contig29	12 gij22059676 ref XM_166418.2 Homo sapiens similar to monoamine oxidase (LOC221656), mRNA
CTGCTTTAAA	2 annexin A1

SOMA	2	Metabolismo Geral
Contig295	12	gi 13543314 gb BC005823.1 BC005823 Homo sapiens, signal transducer and activator of transcription 6, interleukin-4 induced, clone MGC:1606 IMAGE:2958389, mRNA, complete cds
CAATGCTGCC	3	ribosomal protein L23
SOMA	3	Regulação de Transcrição
Contig2914	12	gi 3258613 gb AC004687.1 AC004687 Homo sapiens chromosome 17, clone hRPC.1171.1.10, complete sequence
TCCCTGGCTG	8	acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase)
SOMA	8	Função Desconhecida
Contig2354	12	gi 13543653 gb BC005976.1 BC005976 Homo sapiens, ras homolog gene family, member A, clone MGC:14663 IMAGE:4102976, mRNA, complete cds
TTAGTTACCT	7	ras homolog gene family, member A
SOMA	7	Ciclo Celular
Contig1184	12	gi 17105395 ref NM_000992.2 Homo sapiens ribosomal protein L29 (RPL29), mRNA
TTGGCAGAGA	2	ribosomal protein L29
GAGCCTGAGC	2	carcinoembryonic antigen-related cell adhesion molecule 4
SOMA	4	Metabolismo de Proteínas
Contig698	11	gi 15530329 gb BC013963.1 BC013963 Homo sapiens, Similar to adenylyl cyclase-associated protein, clone MGC:4492 IMAGE:2963922, mRNA, complete cds
CTGGTGTGTT	3	DKFZP586D0623 protein
ATGGAATGCT	2	epidermal growth factor receptor pathway substrate 15
SOMA	5	Transdução de Sinal
Contig557	11	gi 22478035 gb BC036658.1 Homo sapiens, Sjogren syndrome antigen A2 (60kD, ribonucleoprotein autoantigen SS-A/Ro), clone MGC:33170 IMAGE:4821776, mRNA, complete cds
AAGGCAATTT	4	ribosomal protein L31
AAGGCAATTT	4	vascular Rab-GAP/TBC-containing
SOMA	8	Estrutura & Organização Intracelular
Contig3148	11	gi 14602584 gb BC009806.1 BC009806 Homo sapiens, methylenetetrahydrofolate dehydrogenase (NADP+ dependent), methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolatesynthetase, clone MGC:1372 IMAGE:3508998, mRNA, complete cds
TGGATGGTGA	2	"aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid"
GCCGGGTGGG	3	basigin (OK blood group)
SOMA	5	Metabolismo Geral
Contig3132	11	gi 2103 emb X16513.1 SSREPDNA Piggender-neutral repetitive DNA
CCACTGCATT	9	SCAN domain-containing 2
ACTACCTTCA	2	px19-like protein
CCACTGCATT	9	inactivation escape 1
SOMA	20	Função Desconhecida
Contig2958	11	gi 398136 dbj D14466.1 HUMMPOAA Homo sapiens mpo gene for myeloperoxidase, partial cds, exons 9 and 10
TCACCGGTCA	4	gelsolin (amyloidosis, Finnish type)
SOMA	4	Resposta a Estimulos Externos
Contig2921	11	gi 4759139 ref NM_004252.1 Homo sapiens solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 1 (SLC9A3R1), mRNA

CTTTGATCAG		ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin 2 A-inhibited)
SOMA	2	Transporte
Contig250	11	gij19845520 ref NM_002965.2 Homo sapiens S100 calcium binding protein A9 (calgranulin B) (S100A9), mRNA
CTGATGGCGA		16 S100 calcium binding protein A9 (calgranulin B)
TCTCAAGAAG		2 phosphodiesterase 8A
TCTCAAGAAG		2 DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (Myc-regulated)
GCCACCACCA		2 intercellular adhesion molecule 2
GTGGCCACGG		362 S100 calcium binding protein A9 (calgranulin B)
GCCACCACCA		2 sterol regulatory element binding transcription factor 1
SOMA	386	Ciclo Celular
Contig2395	11	gij10047078 ref NM_005745.3 Homo sapiens accessory protein BAP31 (DXS1357E), mRNA
TAGCAATCAG		3 interferon regulatory factor 2
GGCCTGGCAC		Homo sapiens, Similar to RIKEN cDNA 1300019N10 gene, clone 2 IMAGE:3889265,
SOMA	5	Função Desconhecida
Contig2170	11	gij18450370 ref NM_080841.1 Homo sapiens protein tyrosine phosphatase, receptor type, A (PTPRA), transcriptvariant 3, mRNA
GTGAAATCCT		proteasome (prosome, macropain) subunit, beta type, 9 (large 4 multifunctional)
SOMA	4	Transdução de Sinal
Contig1601	11	gij22041261 ref XM_086151.5 Homo sapiens phosphogluconate dehydrogenase (PGD), mRNA
GACGTTCACT		3 phosphogluconate dehydrogenase
SOMA	3	Metabolismo de Carboidratos
Contig1330	11	gij22052313 ref XM_043596.7 Homo sapiens KIAA1128 protein (KIAA1128), mRNA
TGCAGCAAAA		ESTs, Highly similar to R5HU22 ribosomal protein L17, cytosolic 3 [H.sapiens]
SOMA	3	Função Desconhecida
Contig1245	11	gij3153911 gb AF000670.1 AF000670 Homo sapiens elf-1 related protein (ELFR) mRNA, complete cds
CTACTCTTCT		2 glucose phosphate isomerase
SOMA	2	Regulação de Transcrição
Contig88	10	gij20552937 ref XM_007579.6 Homo sapiens cellular retinoic acid binding protein 1 (CRABP1), mRNA
TGGCTTGCTC		4 low density lipoprotein receptor defect B complementing
TGGCTTGCTC		4 acid-inducible phosphoprotein
GTGAAGGCAG		53 ribosomal protein S3A
TGGCTTGCTC		4 cisplatin resistance-associated overexpressed protein
SOMA	65	Resposta a Estimulos Externos
Contig876	10	gij16944077 emb AL512642.18 AL512642 Human DNA sequence from clone RP11-469L23 on chromosome 13, complete sequence [Homo sapiens]
TCCCTGGCTG		acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme 8 A thiolase)
SOMA	8	Função Desconhecida
Contig3515	10	gij16573330 emb X99133.1 HSNALGENH.sapiens NGAL gene
GGAAGGGGAG		4 chromodomain helicase DNA binding protein 2
SOMA	4	Organização Citoplasmática & Biogênese
Contig311	10	gij18043725 gb BC019816.1 BC019816 Homo sapiens, heat shock

		70kD protein 8, clone MGC:29929 IMAGE:2899894.mRNA, complete cds
CCAGGAGGAA		6 heat shock 70kD protein 8
AGAGGTTGAT		3 major histocompatibility complex, class I, A
SOMA	9	Função Desconhecida
Contig3068	10	gi 3411164 gb AF075590.1 AF075590 Homo sapiens MCF-7 peripheral-type benzodiazepine receptor (PBR) mRNA.partial cds
GGGCTCACCT		3 interleukin 1, delta
GGGGACGGC		2 CXYorf1 pseudoautosomal region gene
GGGGACGGC		2 Homo sapiens , Similar to RIKEN cDNA 1110049F14 gene, clone
SOMA	7	Transdução de Sinal
		gi 13477324 gb BC005137.1 BC005137 Homo sapiens, calmodulin 2 (phosphorylase kinase, delta), clone MGC:1447IMAGE:3504793, mRNA,
Contig3042	10	complete cds
CTGGGTCTCC		"ESTs, Highly similar to ribosomal protein L13; 60S ribosomal protein 2 L13;"
GGGGAAAAAA		4 complement component 1, q subcomponent binding protein
AAAAAACCCA		5 endosulfine alpha
CTGGGTCTCC		2 ribosomal protein L13
SOMA	13	Ciclo Celular
		gi 15929824 gb BC015332.1 BC015332 Homo sapiens, Similar to serine/threonine kinase 4, clone IMAGE:4400771,mRNA
Contig2891	10	complete cds
ACTCAGAAGA		2 NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2 (8kD, AGGG)
SOMA	2	Função Desconhecida
		gi 14550452 gb BC009474.1 BC009474 Homo sapiens, allograft inflammatory factor 1, clone MGC:13674IMAGE:4249306, mRNA,
Contig280	10	complete cds
CCTCAAATC		31 carbonic anhydrase I
SOMA	31	Ciclo Celular
		gi 11036541 gb AF237772.1 AF237772 Homo sapiens gamma-parvin (PARVG) mRNA, complete cds
Contig2409	10	complete cds
GCCCAGCTGG		2 associated molecule with the SH3 domain of STAM
GAGCCCCGAA		3 parvin, gamma
GCCCAGCTGG		eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange
GAGCCCCGAA		2 exchange
		3 timeless homolog (Drosophila)
SOMA	10	Função Desconhecida
		gi 22058218 ref XM_006748.6 Homo sapiens seven transmembrane protein TM7SF3 (TM7SF3), mRNA
Contig1704	10	complete cds
GAGGGAGTTT		77 ribosomal protein L27a
GAGGGAGTTT		77 ESTs, Highly similar to 2113200C ribosomal protein L27a [H.sapiens]
TTTTTGATT		20 thioredoxin interacting protein
ATTTTAGAAT		3 RAN binding protein 2
SOMA	177	Transdução de Sinal
		gi 21040335 ref NM_138931.1 Homo sapiens B-cell CLL/lymphoma 6 (zinc finger protein 51) (BCL6), transcriptvariant 2, mRNA
Contig1592	10	complete cds
TTCAAAAAGG		3 30 kDa protein
SOMA	3	Regulação de Transcrição
		gi 21359859 ref NM_002823.2 Homo sapiens prothymosin, alpha (gene sequence 28) (PTMA), mRNA
Contig1582	10	complete cds
AAAGTCAGAA		4 ubiquinol-cytochrome c reductase core protein II
GTCACACCAC		2 prothymosin, alpha (gene sequence 28)

TTCATTATAA	16 prothymosin, alpha (gene sequence 28)
TGACAACAGA	4 ribonuclease, RNase A family, 3 (eosinophil cationic protein)
AATTTTATTT	5 poly(rC) binding protein 1
SOMA	31 Ciclo Celular
Contig1558	10 gil22047326 ref XM_173209.1 Homo sapiens LOC254058 (LOC254058), mRNA
ACTCCCTCCT	2 ribosomal protein S10
SOMA	2 Função Desconhecida
Contig1548	10 gil4503934 ref NM_000157.1 Homo sapiens glucosidase, beta: acid (includes glucosylceramidase) (GBA), mRNA"
TCCAGGCCTA	2 protein kinase (cAMP-dependent, catalytic) inhibitor gamma
AGGAAAGCTG	27 ribosomal protein L36
TCCAGGCCTA	2 adenosine deaminase
SOMA	31 Metabolismo de Lipídeos
Contig1263	10 gil4503080 ref NM_000760.1 Homo sapiens colony stimulating factor 3 receptor (granulocyte) (CSF3R), mRNA
TTCCCTTCTT	7 major histocompatibility complex, class II, DP beta 1
GGGCTGGGGT	70 ribosomal protein L29
TGGAGAAGAG	4 thioredoxin interacting protein
SOMA	81 Transdução de Sinal
Contig1227	10 gil31107 emb Z11692.1 HSEF2MR H.sapiens mRNA for elongation factor 2
TTTGTTAAAA	2 HIF-1 responsive RTP801
AAACCTCTCA	2 DKFZP586M1523 protein
SOMA	4 Regulação de Transcrição
Contig1157	10 gil6912575 ref NM_012387.1 Homo sapiens peptidyl arginine deiminase, type V (PADI5), mRNA
GAGCTCTTGT	2 Fanconi anemia, complementation group A
CAGCCCTCCC	2 calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma
SOMA	4 Metabolismo de Proteínas
Contig763	9 gil4557646 ref NM_001537.1 Homo sapiens heat shock factor binding protein 1 (HSBP1), mRNA
TGGAAAGTGA	6 v-fos FBJ murine osteosarcoma viral oncogene homolog
SOMA	6 Regulação de Transcrição
Contig569	9 gil31957 emb X54326.1 HSGTS H.sapiens mRNAfor glutaminyl-tRNA synthetase
ACCCTGCCAA	2 DKFZP434F1735 protein
SOMA	2 Metabolismo Geral
Contig3205	9 gil4506208 ref NM_002803.1 Homo sapiens proteasome (prosome, macropain) 26S subunit, ATPase, 2 (PSMC2), mRNA
AATATGTGGG	9 cytochrome c oxidase subunit VIc
GACCTTATTT	8 formin-like
CTGTGTCCAA	2 epidermal growth factor receptor pathway substrate 15
SOMA	19 Metabolismo de Proteínas
Contig2871	9 gil14150146 ref NM_032349.1 Homo sapiens hypothetical protein MGC11275 (SDOS), mRNA
GTGGCGCACA	3 adenosine deaminase, tRNA-specific 1
SOMA	3 Função Desconhecida
Contig2611	9 gil12652648 gb BC000072.1 BC000072 Homo sapiens, ribosomal protein L28, clone MGC:3230 IMAGE:3504015, mRNA,complete cds
GCAGCCATCC	22 triosephosphate isomerase 1

GCAGCCATCC	22 ribosomal protein L28
GCAGCCATCC	22 ESTs, Highly similar to S55915 ribosomal protein L28 [H.sapiens]
SOMA	66 Estrutura & Organização Intracelular
Contig2608	9 qil34475 emb Y00809.1 HSMBP Human mRNA foreosinophil granule major basic protein (MBP)
CACAGGGCCA	12 ankyrin repeat and BTB (POZ) domain containing 1
CCTGTGATCC	6 potassium channel, subfamily K, member 6 (TWIK-2)
SOMA	18 Resposta a Estimulos Externos
Contig2577	9 qil11055657 gb AF304202.1 AF304202 Susicrofa breed Landrace mitochondrion, partial genome
TGGGGGTTTC	3 ferritin, heavy polypeptide 1
SOMA	3 Função Desconhecida
Contig2484	9 qil22046239 ref XM_088239.3 Homo sapiens neutrophil cytosolic factor 1 (47kD, chronic granulomatous disease,autosomal 1) (NCF1), mRNA
CAGAGCTGTG	3 phorbolin-like protein MDS019
GTGTCCCTGT	4 lysozyme (renal amyloidosis)
CAGAGCTGTG	3 stromal cell-derived factor 2
GTGTCCCTGT	4 growth arrest-specific 11
SOMA	14 Metabolismo de Proteínas
Contig2363	9 qil11493668 gb AF201687.1 AF201687 Homo sapiens ETS1-associated protein 2 (EAP2) mRNA, complete cds
AAAGCATTTT	2 MO25 protein
SOMA	2 Transdução de Sinal
Contig2256	9 qil16950657 ref NM_001760.2 Homo sapiens cyclin D3 (CCND3), mRNA
AGGAGCTGCT	NADH dehydrogenase (ubiquinone) Fe-S protein 8 (23kD) (NADH-2 coenzyme Q
GCCTGGCCAT	3 DKFZp564J157 protein
CCTCCAGCAG	3 retinoic acid receptor responder (tazarotene induced) 3
GGGCCTGGCC	2 midkine (neurite growth-promoting factor 2)
SOMA	10 Ciclo Celular
Contig1977	9 qil22047416 ref XM_030771.5 Homo sapiens gp25L2 protein (HSGP25L2G), mRNA
CCCACCTGCC	2 protein tyrosine phosphatase, non-receptor type 23
GCCCCAGAAT	2 Homo sapiens clone TCCCIA00164 mRNA sequence
ACCCACCTGC	3 calpain 1, (mu/l) large subunit
GCAGCTCAGG	4 cathepsin D (lysosomal aspartyl protease)
SOMA	11 Função Desconhecida
Contig179	9 qil22045642 ref XM_005323.7 Homo sapiens exosome component 9 Rrp40 (RRP40), mRNA
TGTGAACACA	4 interferon regulatory factor 1
TGTGAACACA	4 ESTs, Weakly similar to R5HU7A ribosomal protein L7a, cytosolic
ACTCAGAAGA	2 NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2 (8kD, AGGG)
TCACCTGAAA	2 DKFZP564K247 protein
AAAAAACCCA	5 endosulfine alpha
SOMA	17 Metabolismo de RNA
Contig1744	9 qil460282 gb L29348.1 HUMGMCS Homo sapiens granulocyte-macrophage colony-stimulating factor receptoralpha-subunit soluble isoform 2 (GM-CSF-RAS2) mRNA, complete cds
GTGAAACCCC	92 "kangai 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2"
CAGGAGTTCA	3 enolase 1, (alpha)

GGAGGCTGAG	5 degenerative spermatocyte homolog, lipid desaturase (Drosophila)
CAGGAGTTCA	3 actin related protein 2/3 complex, subunit 2 (34 kD)
CGCCTGTAAT	5 MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)
GGAGGCTGAG	5 DKFZP434F1735 protein
AGCTGGGATG	2 2,4-dienoyl CoA reductase 2, peroxisomal
SOMA	115 Reconhecimento
Contig1727	9 gi 347313 gb L22009.1 HUM49KDA Human hnRNPH mRNA, complete cds
TTTGTTAATT	3 heterogeneous nuclear ribonucleoprotein H1 (H)
SOMA	3 Metabolismo de RNA
Contig1551	9 gi 11386146 ref NM_002778.1 Homo sapiens prosaposin (variant Gaucher disease and variant metachromaticleukodystrophy) (PSAP), mRNA
TCAACAGCCA	2 eukaryotic translation initiation factor 3, subunit 5 (epsilon, 47kD)
TCCCTGGCTG	8 leukodystrophy
TCCCTGGCTG	8 acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase)
SOMA	18 Metabolismo de Lipídeos
Contig1375	9 gi 18375627 ref NM_004638.2 Homo sapiens HLA-B associated transcript 2 (BAT2), transcript variant 2, mRNA
GAGGGAGTTC	3 ESTs, Highly similar to 2113200C ribosomal protein L27a [H.sapiens]
SOMA	3 Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig1316	9 gi 338056 gb M74816.1 HUMSGLY Human sulfated glycoprotein-2 mRNA, 3'end
CAGCTCATCT	5 jumping translocation breakpoint
GTGAAGCTCT	2 casein kinase 2, beta polypeptide
GGCTGGGGCC	21 elastase 2, neutrophil
CAGCTCATCT	5 adhesion regulating molecule 1
TCTCCAGGAA	18 CGI-69 protein
SOMA	51 Metabolismo de Lipídeos
Contig1293	9 gi 5174678 ref NM_006065.1 Homo sapiens signal-regulatory protein beta 1 (SIRPB1), mRNA
AATGCCAGCA	2 multiple endocrine neoplasia I
SOMA	2 Transdução de Sinal
Contig1194	9 gi 16753213 ref NM_005022.2 Homo sapiens profilin 1 (PFN1), mRNA
TGGGCTGGGG	3 adipose differentiation-related protein
CACACACAAA	2 myosin, light polypeptide, regulatory, non-sarcomeric (20kD)
TGGTTTTGGC	7 itchy homolog E3 ubiquitin protein ligase (mouse)
GGCTGGGGGC	40 profilin 1
CATTCATAA	4 GA binding protein transcription factor, alpha subunit (60kD)
GGCTGGGGGC	40 chromosome 1 amplified sequence 3
GGTTTGTGTG	2 protein associated with PRK1
GCAAAGAAGG	2 glycyl-tRNA synthetase
CATTCATAA	4 ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6
CTTCTTTGCC	6 hemoglobin, alpha 2
GCAAAGAAGG	2 hemoglobin, gamma A
SOMA	112 Metabolismo de Proteínas
Contig115	9 gi 5032058 ref NM_005621.1 Homo sapiens S100 calcium binding protein A12 (calgranulin C) (S100A12), mRNA
GCTCACCTT	12 lymphocyte antigen 117

CTGAAGGCTG	5 RAN binding protein 16
CTGAAGGCTG	5 platelet/endothelial cell adhesion molecule (CD31 antigen)
SOMA	22 Ciclo Celular
Contig1151	9 gi 7382490 ref NM_002298.2 Homo sapiens lymphocyte cytosolic protein 1 (L-plastin) (LCP1), mRNA
TGGCTCCTCC	12 calreticulin
SOMA	12 Metabolismo de Proteínas
Contig715	8 gi 3201999 gb AF068846.1 AF068846 Homo sapiens scaffold attachment factor A (SAF-A) mRNA, complete cds
TTTTTCTTAA	5 tumor protein, translationally-controlled 1
ACATTTTAA	3 NEFA-interacting nuclear protein NIP30
AGTCTTAATG	2 CUG triplet repeat, RNA binding protein 2
SOMA	10 Metabolismo Geral
Contig3457	8 gi 17511739 gb BC018726.1 BC018726 Homo sapiens, CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated), clone MGC:31825 IMAGE:4853578, mRNA, complete cds
CAGGAACGGG	7 mitogen-activated protein kinase kinase 2
TACCTGCAGG	2 S100 calcium binding protein A8 (calgranulin A)
SOMA	9 Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig3387	8 gi 1685050 gb U76764.1 HSU76764 Human CD97 mRNA, complete cds
CTTTTCTCT	4 tropomodulin 3 (ubiquitous)
TGTCTTTGCT	2 zinc finger protein 255
SOMA	6 Adesão
Contig3307	8 gi 14165438 ref NM_002140.2 Homo sapiens heterogeneous nuclear ribonucleoprotein K (HNRPK), transcript variant 1, mRNA
TTGTAATAA	3 discs, large (Drosophila) homolog 1
TAATTTCTCA	2 axotrophin
SOMA	5 Metabolismo de RNA
Contig3271	8 gi 12005923 gb AF261137.1 AF261137 Homo sapiens HT031 mRNA, complete cds
AGAGACTCTT	2 DKFZP547E1010 protein
GATGCATATA	2 CD164 antigen, sialomucin
SOMA	4 Função Desconhecida
Contig3218	8 gi 17512031 gb BC018990.1 BC018990 Homo sapiens, clone MGC:20176 IMAGE:3503710, mRNA, complete cds
CCTGTGGAGG	2 retinoid X receptor, gamma
CCCTGGGTTC	47 ferritin, light polypeptide
CCTGGGTTC	7 major histocompatibility complex, class II, DR alpha
CCCTGGGTTC	47 major histocompatibility complex, class II, DR alpha
CCTGGGTTC	7 ferritin, light polypeptide
TCTCCAGGAA	18 CGI-69 protein
SOMA	128 Função Desconhecida
Contig313	8 gi 340440 gb M80583.1 HUMZINFING Homo sapiens zinc finger protein mRNA
TACCTACTGA	3 thymopoietin
SOMA	3 Regulação de Transcrição
Contig3095	8 gi 13787202 ref NM_006386.2 Homo sapiens DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 17, 72kDa (DDX17), transcript variant 1, mRNA
ATCAGTGGCT	Homo sapiens serologically defined breast cancer antigen NY-BR-40 2 mRNA,

ATCAGTGGCT	2 proteasome (prosome, macropain) subunit, beta type, 4
GTGGCTGCTG	3 mitochondrial ribosomal protein S36
SOMA	7 Metabolismo de RNA
	gij14737457 ref XM_041507.1 Homo sapiens guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2 (GNAI2), mRNA
Contig3074	8 mRNA
GTTCTGGTTT	20 ATPase inhibitor precursor
CCAGAACAGA	55 ribosomal protein L30
TGGGCAAAGC	28 eukaryotic translation elongation factor 1 gamma
CTGAGTCTCC	4 guanine nucleotide binding protein (G protein), alpha inhibiting activity
SOMA	107 Transdução de Sinal
	gij21411332 gb BC031012.1 Homo sapiens, eukaryotic translation elongation factor 1 gamma, clone MGC:32765IMAGE:4654721, mRNA, complete cds
Contig3011	8 complete cds
ACTCCAGAAA	3 eukaryotic translation initiation factor 3, subunit 6 (48kD)
ACTCCAGAAA	3 ribosomal protein S15 ESTs, Highly similar to R3HU15 ribosomal protein S15, cytosolic
ACTCCAGAAA	3 [H.sapiens]
TGGACAAGCT	3 brain abundant, membrane attached signal protein 1 "ESTs, Highly similar to ribosomal protein L13; 60S ribosomal protein
CTGGGTCTCC	2 L13;"
TGGGCAAAGC	28 eukaryotic translation elongation factor 1 gamma
CTGGGTCTCC	2 ribosomal protein L13
SOMA	44 Metabolismo de Proteínas
	gij4322996 gb AF096810.1 AF096810 Diomedeachrysostoma microsatellite Dc31 sequence
Contig2939	8 microsatellite Dc31 sequence
TACCCCTGGA	3 high density lipoprotein binding protein (vigilin)
SOMA	3 Função Desconhecida
	gij6552332 ref NM_005252.2 Homo sapiens v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS), mRNA
Contig2908	8 osteosarcoma viral oncogene homolog (FOS), mRNA
TTGATGTTGA	2 macrophage erythroblast attacher
TTGATGTTGA	2 ATP binding protein associated with cell differentiation
SOMA	4 Regulação de Transcrição
	gij4505988 ref NM_000308.1 Homo sapiens protective protein for beta-galactosidase (galactosialidosis) (PPGB), mRNA
Contig2822	8 galactosidase (galactosialidosis) (PPGB), mRNA
ACATCCCAGA	2 coronin, actin binding protein, 1C
SOMA	2 Estrutura & Organização Intracelular
	gij20521697 dbj_AB023150.2 Homo sapiens mRNA for KIAA0933 protein, partial cds
Contig2780	8 protein, partial cds
GCTCACTGCA	3 peptidylprolyl isomerase (cyclophilin)-like 2
CGCCTGTAAT	5 MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)
SOMA	8 Função Desconhecida
	gij6649241 gb AF201422.1 AF201422 Homo sapiens splicing coactivator subunit SRm300 (SRM300) mRNA, completecds
Contig2745	8 coactivator subunit SRm300 (SRM300) mRNA, completecds
CAACTAATTC	5 clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2,
SOMA	5 Metabolismo de RNA
	gij20070800 gb BC026276.1 Homo sapiens, ribosomal protein L10, clone MGC:23823 IMAGE:4276627, mRNA, completecds
Contig2670	8 clone MGC:23823 IMAGE:4276627, mRNA, completecds
AAAGAAGTGG	2 HIV TAT specific factor 1
GTGGCTGAAA	2 ESTs, Highly similar to A42735 ribosomal protein L10, cytosolic
GTGAAGCCCC	3 WW45 protein
GTGGCTGAAA	2 ribosomal protein L10

TGACCCACA	2 ferritin, heavy polypeptide 1
SOMA	11 Função Desconhecida
	gij22046607 ref XM_170516.1 Homo sapiens similar to neutrophil cytosolic factor 1 (47kD, chronic granulomatous disease, autosomal 1)
Contig2468	8 (LOC256379), mRNA
CTCGAGGAGG	6 mitochondrial ribosomal protein L23
SOMA	6 Função Desconhecida
	gij14149884 ref NM_032193.1 Homo sapiens hypothetical protein
Contig2463	8 FLJ20974 (FLJ20974), mRNA
CTGGCCCGGA	4 vasodilator-stimulated phosphoprotein
GTGACAGAAG	8 eukaryotic translation initiation factor 4A, isoform 1
CTGGCCCGGA	4 high-mobility group 20B
CTGCCAACTT	3 cofilin 1 (non-muscle)
SOMA	19 Função Desconhecida
	gij11055671 gb AF304203.1 AF304203 Susicrofa breed Swedish wild boar mitochondrion, partial genome
Contig2312	8 boar mitochondrion, partial genome
TGGGGGTTTC	3 ferritin, heavy polypeptide 1
SOMA	3 Função Desconhecida
	gij6841193 gb AF161390.1 AF161390 Homo sapiens HSPC272 mRNA, partial cds
Contig2240	8 partial cds
GGGGAAAAA	4 complement component 1, q subcomponent binding protein
SOMA	4 Função Desconhecida
	gij21594898 gb BC032230.1 Homo sapiens, bactericidal/permeability-increasing protein, clone IMAGE:5211259, mRNA
Contig214	8 increasing protein, clone IMAGE:5211259, mRNA
TGTGTATTTT	2 Meis1, myeloid ecotropic viral integration site 1 homolog (mouse)
SOMA	2 Função Desconhecida
	gij23271193 gb BC036014.1 Homo sapiens, clone MGC:32916
Contig2074	8 IMAGE:5271668, mRNA, complete cds
CTGATCTCCA	7 D component of complement (adipsin)
SOMA	7 Metabolismo de RNA
	gij4505574 ref NM_002568.1 Homo sapiens poly(A) binding protein, cytoplasmic 1 (PABPC1), mRNA
Contig1978	8 cytoplasmic 1 (PABPC1), mRNA
AAATTTTGT	2 T-cell acute lymphocytic leukemia 1
SOMA	2 Metabolismo de RNA
	gij19923731 ref NM_001516.2 Homo sapiens general transcription factor IIH, polypeptide 3 (34kD subunit)(GTF2H3), mRNA
Contig1737	8 factor IIH, polypeptide 3 (34kD subunit)(GTF2H3), mRNA
CGCCTGTAAT	5 MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)
AGAGGTTGAT	3 major histocompatibility complex, class I, A
SOMA	8 Regulação de Transcrição
	gij22325355 ref NM_147150.1 Homo sapiens Akinase (PRKA) anchor protein 2 (AKAP2), transcript variant 2, mRNA
Contig1657	8 protein 2 (AKAP2), transcript variant 2, mRNA
TTTGTAATAT	2 grancalcin, EF-hand calcium binding protein
SOMA	2 Função Desconhecida
	gij3258613 gb AC004687.1 AC004687 Homo sapiens chromosome 17, clone hRPC.1171.1.10, complete sequence
Contig1570	8 clone hRPC.1171.1.10, complete sequence
TCCCTGGCTG	acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase)
SOMA	8 Função Desconhecida
	gij21752078 dbj AK093259.1 Homo sapiens cDNA FLJ35940 fis, clone TEST12011612, weakly similar to Aegilops squarrosa GAG56D gene for gamma-gliadin, accession Clae 24
Contig1483	8 gamma-gliadin, accession Clae 24
CTTTTTCTG	3 mitochondrial solute carrier
GCACAGAGCT	4 v-fos FBJ murine osteosarcoma viral oncogene homolog

GCACAGAGCT	4 "kangai 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2"
SOMA	11 Função Desconhecida
Contig142	8 gij536843 gb_L33075.1 HUMIQGA Homo sapiens ras GTPase-activating-like protein (IQGAP1) mRNA, complete cds
TGTGTTGAGG	2 eukaryotic translation elongation factor 1 alpha 1
AACTCTGCTC	2 heme-regulated initiation factor 2-alpha kinase
SOMA	4 Transdução de Sinal
Contig1310	8 gij21757410 dbj_AK097592.1 Homo sapiens cDNA FLJ40273 fis, clone TESTI2027165, highly similar to Oryctolagus cuniculus RNA-binding protein lark (lark) mRNA
CCTTTCCTTT	3 small inducible cytokine subfamily B (Cys-X-Cys), member 14 (BRAK)
GCCTGCTGGG	5 actinin, alpha 1
GCCTGCTGGG	5 glutathione peroxidase 4 (phospholipid hydroperoxidase)
GTGCTGTCTC	59 hemoglobin, alpha 1
CCTTTCCTTT	3 GDP dissociation inhibitor 1
TCTGTAGCTT	4 endothelial-derived gene 1
SOMA	79 Função Desconhecida
Contig881	7 gij15011935 ref NM_001029.2 Homo sapiens ribosomal protein S26 (RPS26), mRNA
ACAGCAAAGT	11 Rhesus blood group, D antigen
GCCCAAGGAC	3 filamin A, alpha (actin binding protein 280)
ACAGCAAAGT	11 small membrane protein 1 "phosphorylase, glycogen; liver (Hers disease, glycogen storage disease 2 type VI)"
TCGCTGCACT	2 type VI"
GCCCAAGGAC	3 19A24 protein
ACTAACTGTG	4 retinoblastoma binding protein 4
ACTAACTGTG	4 intermediate filament protein syncoilin
SOMA	38 Estrutura & Organização Intracelular
Contig880	7 gij15011937 ref NM_001030.2 Homo sapiens ribosomal protein S27 (metallopanstimulin 1) (RPS27), mRNA
ACACACAAAA	3 step II splicing factor SLU7
CACAAACGGT	98 ESTs, Highly similar to A48045 ribosomal protein S27, cytosolic
CATTTTGGAT	2 glutamate-cysteine ligase, catalytic subunit
CACAAACGGT	98 ribosomal protein S27 (metallopanstimulin 1)
SOMA	201 Estrutura & Organização Intracelular
Contig727	7 gij5454161 ref NM_006472.1 Homo sapiens thioredoxin interacting protein (TXNIP), mRNA
ACCCCCTTCC	4 transketolase (Wernicke-Korsakoff syndrome)
TGTGATCAGA	26 intercellular adhesion molecule 3
SOMA	30 Diferenciação Celular e Tissular
Contig717	7 gij485363 gb_M59819.1 HUMGCSFR2 Human granulocyte colony-stimulating factor receptor (G-CSFR-2) mRNA, complete cds
TTCCCTTCTT	7 major histocompatibility complex, class II, DP beta 1
GGGCTGGGGT	70 ribosomal protein L29
TGGAGAAGAG	4 thioredoxin interacting protein
SOMA	81 Transdução de Sinal
Contig6	7 gij19773576 gb_AF276934.1 Sus scrofa domestica mitochondrial D-loop, complete sequence
ACCCCAAAAA	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic 4 thread
SOMA	4 Função Desconhecida

Contig465	7	gi 16740549 gb BC016148.1 BC016148 Homo sapiens. Similar to integral membrane protein 2B, clone MGC:10219IMAGE:3912066, mRNA, complete cds
TAGAGAATGA	2	peptidylprolyl isomerase B (cyclophilin B)
ACTACAAATA	5	ESTs, Highly similar to S42106 ribosomal protein L9 homolog [H.sapiens]
ACTACAAATA	5	ribosomal protein L9
SOMA	12	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig437	7	gi 184231 gb M21339.1 HUMHMG14A Human non-histone chromosomal protein HMG-14 gene, complete cds
GGAGGCTGAG	5	DKFZP434F1735 protein
SOMA	5	Regulação de Transcrição
Contig41	7	gi 5453749 ref NM_006317.1 Homo sapiens brain abundant, membrane attached signal protein 1 (BASP1), mRNA
CTTTTCTTCT	6	insulin-like growth factor binding protein 7
CTTTTCTTCT	6	polymerase (RNA) II (DNA directed) polypeptide B (140kD)
CTTTTCTTCT	6	BRCA2 and CDKN1A interacting protein
TGGACAAGCT	3	brain abundant, membrane attached signal protein 1
CAGAAGAAAA	2	protein phosphatase 3 (formerly 2B), regulatory subunit B (19kD), alpha
CTTTTCTTCT	6	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 32
SOMA	29	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig415	7	gi 13540472 ref NM_030752.1 Homo sapiens t-complex 1 (TCP1), mRNA
CATCTGTA	2	major histocompatibility complex, class II, DR beta 5
TCTTTTAAA	2	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer)
TGGTCCTCTG	2	NIMA (never in mitosis gene a)-related kinase 4
TTGTGATGTA	3	glucose phosphate isomerase
CATCTGTA	2	protocadherin 20
TTGGTCCTCT	103	ribosomal protein L41
SOMA	114	Organização Citoplasmática & Biogênese
Contig381	7	gi 22095361 ref NM_022343.2 Homo sapiens chromosome 9 open reading frame 19 (C9orf19), mRNA
TCTGTCAAGA	4	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, O subunit prosaposin (variant Gaucher disease and variant metachromatic)
AAGTTGCTAT	14	leukodystrophy)
CCCCCTGCC	3	E3 ubiquitin ligase SMURF1
SOMA	21	Função Desconhecida
Contig3488	7	gi 6900101 emb AJ271684.1 HSA271684 Homo sapiens mRNA for myeloid DAP12-associating lectin (MDL-1 gene)
GCCTTAAAA	6	pre-B-cell colony-enhancing factor
ACATTTTAA	3	NEFA-interacting nuclear protein NIP30
GTTGTGGCCA	2	DnaJ (Hsp40) homolog, subfamily C, member 5
SOMA	11	Transdução de Sinal
Contig3259	7	gi 4502662 ref NM_001774.1 Homo sapiens CD37 antigen (CD37), mRNA
GAAGTGGCTG	2	FK506 binding protein 8 (38kD)
TGCTGCCTGT	4	HCGIV-6 protein
TGCTGCCTGT	4	bone marrow stromal cell antigen 2
SOMA	10	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig3200	7	gi 14591915 ref NM_001023.2 Homo sapiens ribosomal protein S20 (RPS20), mRNA
AGAACCTTCA	7	major histocompatibility complex, class I, A

SOMA	7	Estrutura & Organização Intracelular
Contig3029	7	gi 4758937 ref NM_004573.1 Homo sapiens phospholipase C, beta 2 (PLCB2), mRNA
CCAGGATCTG		2 protein phosphatase 2, regulatory subunit B (B56), gamma isoform
TCCCTGGCTG		8 prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)
TCCCTGGCTG		8 acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase)
SOMA	18	Transdução de Sinal
Contig3002	7	gi 4504164 ref NM_000177.1 Homo sapiens gelsolin (amyloidosis, Finnish type) (GSN), mRNA
CACCACGGTG		5 RNB6
SOMA	5	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig2961	7	gi 4104352 gb AF035191.1 AF035191 Homo sapiens nuclear autoantigenic sperm protein autosomal variant mRNA, partial cds
GCCCAGCTGG		2 associated molecule with the SH3 domain of STAM
GCCCAGCTGG		2 eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange)
SOMA	4	Metabolismo de DNA
Contig2959	7	gi 5031976 ref NM_005746.1 Homo sapiens pre-B-cell colony-enhancing factor (PBEF), mRNA
GCCTTAACAA		13 pre-B-cell colony-enhancing factor
SOMA	13	Sinalização Intercelular
Contig284	7	gi 183684 gb M20681.1 HUMGTLPA Human glucosetransporter-like protein-III (GLUT3), complete cds
CGCCTGTAAT		5 MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)
GTGAAACCCT		42 CAP-binding protein complex interacting protein 2
SOMA	47	Transporte
Contig2789	7	gi 20270336 ref NM_138792.1 Homo sapiens hypothetical protein BC018147 (LOC123169), mRNA
CAAATAAAA		2 translocase of outer mitochondrial membrane 70 homolog A (yeast)
TCTTCTGCTT		3 hepatocellular carcinoma-associated antigen 59
SOMA	5	Função Desconhecida
Contig2440	7	gi 11055671 gb AF304203.1 AF304203 Suscrofa breed Swedish wild boar mitochondrion, partial genome
TGGGGGTTTC		3 WAS protein family, member 1
SOMA	3	Função Desconhecida
Contig239	7	gi 22507396 ref NM_002473.2 Homo sapiens myosin, heavy polypeptide 9, non-muscle (MYH9), mRNA
ATCAGAGCT		45 glycyl-tRNA synthetase
SOMA	45	Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig2356	7	gi 8886010 gb AF157510.1 AF157510 Homo sapiens beta endosulfine mRNA, complete cds
GTCCCAAAT		13 nuclear receptor coactivator 4
GGCTGCTATT		3 defensin, alpha 3, neutrophil-specific
SOMA	16	Metabolismo de Carboidratos
Contig2310	7	gi 22052384 ref XM_056963.4 Homo sapiens cytoplasmic FMR1 interacting protein 2 (CYFIP2), mRNA
TCTTCTGCTT		3 hepatocellular carcinoma-associated antigen 59
SOMA	3	Função Desconhecida
Contig2280	7	gi 19401710 gb AF468111.1 Homo sapiens DNA-directed RNA polymerase II subunit 11 mRNA, complete cds
TGGGCCTGTG		2 similar to envelope protein

SOMA	2	Metabolismo de RNA
Contig2179	7	gi 5901911 ref NM_006888.1 Homo sapiens calmodulin 1 (phosphorylase kinase, delta) (CALM1), mRNA
TGGGTGAAGG	3	thymosin, beta 4, X chromosome
TGGGTGAAGG	3	sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like
TACAAGAGGA	9	ribosomal protein L6
SOMA	15	Ciclo Celular
Contig2023	7	gi 7661879 ref NM_014716.1 Homo sapiens centaurin, beta 1 (CENTB1), mRNA
GTAGCAGGGT	3	synuclein, alpha (non A4 component of amyloid precursor)
AGGGTCCCTG	2	immunoglobulin kappa constant
SOMA	5	Transdução de Sinal
Contig1983	7	gi 488096 gb L32767.1 HUMF513 Human coagulation factor V gene, exon 13
AAGGAGATGG	24	ESTs, Highly similar to R5HU31 ribosomal protein L31 [H.sapiens]
AAGGAGATGG	24	ribosomal protein L31
AAGGAGATGG	24	vascular Rab-GAP/TBC-containing
SOMA	72	Homeostase
Contig1847	7	gi 8574037 emb AJ132695.5 HSA132695 Homo sapiens rac1 gene
TGGGTGAGCC	4	cathepsin B
SOMA	4	Ciclo Celular
Contig1719	7	gi 22066514 ref XM_171902.1 Homo sapiens similar to Hkr1p (LOC254967), mRNA
GGAAGGACAG	2	ATPase, H+ transporting, lysosomal 38kD, V0 subunit d isoform 1
SOMA	2	Função Desconhecida
Contig149	7	gi 5729811 ref NM_006532.1 Homo sapiens ELL gene (11-19 lysine-rich leukemia gene) (ELL), mRNA
GACAGCTGAG	5	adenylate kinase 1
GACCCTGCCC	4	fibroblast growth factor receptor 4
GACCCTGCCC	4	FK506 binding protein 8 (38kD)
SOMA	13	Metabolismo de RNA
Contig147	7	gi 4503302 ref NM_001925.1 Homo sapiens defensin, alpha 4, corticostatin (DEFA4), mRNA
TTTCCTTTGC	2	peripheral benzodiazepine receptor-associated protein 1
ATACTGCTGC	3	cullin 2
SOMA	5	Homeostase
Contig1454	7	gi 18027315 gb AF289556.1 AF289556 Homo sapiens clone pp4491 unknown mRNA
CCTTCCAAAT	3	malate dehydrogenase 2, NAD (mitochondrial)
SOMA	3	Função Desconhecida
Contig1436	7	gi 4557556 ref NM_001429.1 Homo sapiens E1A binding protein p300 (EP300), mRNA
CACACACAAA	2	myosin, light polypeptide, regulatory, non-sarcomeric (20kD)
SOMA	2	Regulação de Transcrição
Contig1381	7	gi 23238212 ref NM_005717.2 Homo sapiens actin related protein 2/3 complex, subunit 5, 16kDa (ARPC5), mRNA
GCTTATAAAA	3	histidine triad nucleotide binding protein 1
SOMA	3	Mobilidade
Contig1329	7	gi 17402908 ref NM_012101.2 Homo sapiens tripartite motif-containing 29 (TRIM29), transcript variant 1, mRNA
CTGAGGGTGG	3	BCL2/adenovirus E1B 19kD interacting protein 3

SOMA	3	Regulação de Transcrição
Contig1162	7	gi 6358506 gb AF043723.1 AF043723 Homo sapiens guanine exchange factor MCG7 isoform 2 mRNA, complete cds
CACCCCCAGG	2	G protein pathway suppressor 2
GCCAGGGCGG	3	histocompatibility (minor) 13
GCCAGGGCGG	3	RuvB-like 2 (E. coli)
GCCTCCAGAG	4	alpha glucosidase II alpha subunit
GGAGGCTGGG	3	general transcription factor IIIC, polypeptide 3 (102kD)
SOMA	15	Transdução de Sinal
Contig1141	7	gi 22042371 ref XM_043678.7 Homo sapiens KIAA1096 protein (KIAA1096), mRNA
AAGTTTCCAA	2	protein phosphatase 4 (formerly X), catalytic subunit
SOMA	2	Função Desconhecida
Contig1102	7	gi 3258613 gb AC004687.1 AC004687 Homo sapiens chromosome 17, clone hRPC.1171.1.10, complete sequence
TCCCTGGCTG	8	acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase)
SOMA	8	Função Desconhecida
Contig1015	7	gi 6912575 ref NM_012387.1 Homo sapiens peptidyl arginine deiminase, type V (PADI5), mRNA
GAGCTCTTGT	2	Fanconi anemia, complementation group A
CAGCCCTCCC	2	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma
SOMA	4	Metabolismo de Proteínas
Contig965	6	gi 4730928 gb AF039942.1 AF039942 Homo sapiens HCF-binding transcription factor Zhangfei (ZF) mRNA, complete cds
ATGTAACAAA	109	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains),
ATGTAACAAA	109	DNA segment, single copy probe LNS-CAI/LNS-CAII (deleted in polyposis)
SOMA	218	Regulação de Transcrição
Contig963	6	gi 3219696 gb AF046785.1 AF046785 Homo sapiens liver glycogen phosphorylase (PYGL) mRNA, complete cds
GACTCACTTT	5	peptidylprolyl isomerase B (cyclophilin B)
GATGTGGCTG	4	"ESTs, Highly similar to eukaryotic translation elongation factor 1 beta 2;"
GATGTGGCTG	4	eukaryotic translation elongation factor 1 beta 2
AATCCAAAGG	2	"phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI)"
SOMA	15	Metabolismo de Carboidratos
Contig953	6	gi 19263676 gb BC025275.1 Homo sapiens, clone MGC:39180 IMAGE:4646010, mRNA, complete cds
GTCTGGGGCT	32	thiopurine S-methyltransferase
SOMA	32	Função Desconhecida
Contig893	6	gi 19923408 ref NM_014445.2 Homo sapiens stress-associated endoplasmic reticulum protein 1; ribosome associated membrane protein 4 (SERP1), mRNA"
TGTGCTAAAT	39	ribosomal protein L34
ATGTAACAAA	109	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains),
TTTTTAATGT	26	H3 histone, family 3A
TTCAAAAAGG	3	30 kDa protein
SOMA	177	Metabolismo de Proteínas
Contig884	6	gi 20357526 ref NM_002074.2 Homo sapiens guanine nucleotide binding protein (G protein), beta polypeptide 1 (GNB1), mRNA

TCCAGGCCTA	2 adenosine deaminase
SOMA	2 Transdução de Sinal
Contig776	6 gi 16579887 ref NM_000507.2 Homo sapiens fructose-1,6-bisphosphatase 1 (FBP1), mRNA
AGAGCAGAAA	2 adaptor-related protein complex 3, sigma 2 subunit
CAAATGCTGT	9 signal transducer and activator of transcription 1, 91kD
SOMA	11 Metabolismo de Carboidratos
Contig729	6 gi 23346413 ref NM_002376.3 Homo sapiens MAP/microtubule affinity-regulating kinase 3 (MARK3), mRNA
TGGAAGCATC	5 annexin A2
SOMA	5 Ciclo Celular
Contig613	6 gi 7021909 dbj AK000934.1 Homo sapiens cDNAFLJ10072 fis. clone HEMBA1001714, highly similar to Homo sapiens mRNA:cDNA DKFZp564G0422"
TCTTCTGCTT	3 hepatocellular carcinoma-associated antigen 59
AGCCACCGTG	3 DKFZP434D146 protein
TCTTCTGCTT	3 ring finger protein 20
TGGCAGCTTT	2 CGI-48 protein
SOMA	11 Metabolismo de Carboidratos
Contig585	6 gi 18585370 ref XM_083939.1 Homo sapiens STIP1 homology and U-Box containing protein 1 (STUB1), mRNA
GGCCGGGGGC	5 pre-B-cell leukemia transcription factor 2
CAGAGATGAA	2 heat shock 70kD protein 1A
CAGAGATGAA	2 TAR DNA binding protein
GTTCTGGCC	"Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed (fox derived);"
GGGGCTTCTG	3 legumain
SOMA	15 Estrutura & Organização Intracelular
Contig544	6 gi 9966912 ref NM_020445.1 Homo sapiens actin-related protein 3-beta (ARP3BETA), mRNA
TGGCAAAAAA	membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase
GTTTTTAATT	2 DKFZP727G051 protein
TTTAATTGTG	2 SET translocation (myeloid leukemia-associated)
SOMA	7 Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig3383	6 gi 20070168 ref NM_002520.2 Homo sapiens nucleophosmin (nucleolar phosphoprotein B23, numatrin) (NPM1), mRNA
GTTTGATAAA	2 conserved gene amplified in osteosarcoma
TTTGTTAAAA	2 HIF-1 responsive RTP801
SOMA	4 Estrutura & Organização Intracelular
Contig3338	6 gi 4557678 ref NM_000214.1 Homo sapiens jagged 1 (Alagille syndrome) (JAG1), mRNA
CCAGAACAGA	55 ribosomal protein L30
SOMA	55 Transdução de Sinal
Contig3238	6 gi 4505574 ref NM_002568.1 Homo sapiens poly(A) binding protein, cytoplasmic 1 (PABPC1), mRNA
AAATTTTGT	2 T-cell acute lymphocytic leukemia 1
SOMA	2 Regulação de Transcrição
Contig3206	6 gi 10437938 dbj AK025429.1 Homo sapiens cDNA: FLJ21776 fis. clone HEP00171
GATGGATTTT	2 lactate dehydrogenase B
SOMA	2 Função Desconhecida

Contig3170	gi 509046 emb X70941.1 SSRPT45 S.scrofa DNAfor centromere specific motif
CCACTGCATT	9 SCAN domain-containing 2
ACTACCTTCA	2 px19-like protein
CCACTGCATT	9 inactivation escape 1
SOMA	20 Função Desconhecida
Contig3157	gi 2463576 dbj AB007510.1 Homo sapiens mRNA for PRP8 protein, complete cds
GCCAAGGAAA	27 solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane)
SOMA	27 Metabolismo de RNA
Contig3057	gi 9956025 gb AY007127.1 Homo sapiens clone CDABP0035 mRNA sequence
AGATGTGTGG	2 hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A
TGACTGTGCT	21 neurogranin (protein kinase C substrate, RC3)
SOMA	23 Função Desconhecida
Contig3025	gi 4506400 ref NM_002880.1 Homo sapiens v-raf-1 murine leukemia viral oncogene homolog 1 (RAF1), mRNA
TCTTCCAGGA	6 ribosomal protein L15
SOMA	6 Apoptose
Contig2990	gi 20070234 ref NM_006495.2 Homo sapiens ecotropic viral integration site 2B (EVI2B), mRNA
ACATATACTT	2 glycerol kinase
SOMA	2 Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig2947	gi 21732579 emb AL832039.1 HSM803346 Homo sapiens mRNA; cDNA DKFZp313B0110 (from clone DKFZp313B0110)"
TTACTAAATG	5 DKFZP564K247 protein
SOMA	5 Função Desconhecida
Contig2861	gi 7661885 ref NM_014764.1 Homo sapiens DAZ associated protein 2 (DAZAP2), mRNA
GATGAACTGA	3 splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)
GATGAACTGA	3 ATPase, Na ⁺ /K ⁺ transporting, alpha 1 polypeptide
TAGACTGGGT	2 adenyl cyclase-associated protein
CTGGGTTAAT	92 ribosomal protein S19
SOMA	100 Função Desconhecida
Contig2835	gi 4557800 ref NM_000270.1 Homo sapiens nucleoside phosphorylase (NP), mRNA
TGCCTCTGTC	6 nucleoside phosphorylase
TTTATTTTGT	5 CDC-like kinase 1
AGATGTGTGG	2 hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A
TAAAGCTGTT	10 ubiquitin-conjugating enzyme E2B (RAD6 homolog)
SOMA	23 Metabolismo de DNA
Contig2797	gi 21704257 gb AF314059.2 Homo sapiens chromosome 8 clone GS1-P1-861 map 8p22-p21, complete sequence
CATATTTTTT	3 calmodulin 2 (phosphorylase kinase, delta)
CACAGAACAC	3 beta-site APP-cleaving enzyme
SOMA	6 Função Desconhecida
Contig2763	gi 10047172 dbj AB046774.1 Homo sapiens mRNA for KIAA1554 protein, partial cds
TATCACTCTG	4 male-enhanced antigen

GGCCGGGGGC	5 pre-B-cell leukemia transcription factor 2
SOMA	9 Função Desconhecida
Contig2752	6 gi 5441365 emb AJ012506.1 HSA012506 Homo sapiens mRNA activated in tumor suppression, clone TSAP21 extended
ATTATTTTTC	14 ribosomal protein L7
ACATTTTAA	3 NEFA-interacting nuclear protein NIP30
ATTATTTTTC	14 eukaryotic translation initiation factor 3, subunit 10 (theta, 150/170kD ESTs, Highly similar to R5HU7 ribosomal protein L7, cytosolic [H.sapiens
ATTATTTTTC	14]
SOMA	45 Transporte
Contig2698	6 gi 13543892 gb BC006093.1 BC006093 Homo sapiens, matrix metalloproteinase 9 (gelatinase B, 92kD gelatinase, 92kD type IV collagenase), clone MGC:12688 IMAGE:4054882, mRNA, complete cds
CCAGAACAGA	55 ribosomal protein L30
SOMA	55 Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig2695	6 gi 5803186 ref NM_006755.1 Homo sapiens transaldolase 1 (TALDO1), mRNA
TTGTAATAAAA	2 taxol resistance associated gene 3
SOMA	2 Metabolismo de Carboidratos
Contig2598	6 gi 16552800 dbj AK057201.1 Homo sapiens cDNA FLJ32639 fis, clone SYNOV2001014, weakly similar to Mus musculusZfp228 (Znf228) mRNA
GGCCAGGACT	3 formyl peptide receptor 1
SOMA	3 Função Desconhecida
Contig258	6 gi 4502100 ref NM_000700.1 Homo sapiens annexin A1 (ANXA1), mRNA
CTATATTTT	2 cathepsin C
SOMA	2 Metabolismo de Lipídeos
Contig2497	6 gi 21237758 ref NM_001780.2 Homo sapiens CD63 antigen (melanoma 1 antigen) (CD63), mRNA
CAAAGACAAT	2 HSPC038 protein
AAGATTGGGG	2 fibronectin 1
TTGCTTTTGT	3 ADP-ribosylation factor 4
AAGATTGGGG	2 Homo sapiens CD44 isoform RC (CD44) mRNA, complete cds
AAGATTGGGG	2 CD44 antigen (homing function and Indian blood group system)
TTGCTTTTGT	3 Fanconi anemia, complementation group F
AAGATTGGGG	2 secreted protein, acidic, cysteine-rich (osteonectin)
ATTGCTTTTG	2 similar to S. cerevisiae RER1
SOMA	18 Ciclo Celular
Contig23	6 gi 4505640 ref NM_002592.1 Homo sapiens proliferating cell nuclear antigen (PCNA), mRNA
AGTTTTATTT	2 RAB10, member RAS oncogene family
SOMA	2 Ciclo Celular
Contig2349	6 gi 20070236 ref NM_006597.2 Homo sapiens heat shock 70kDa protein 8 (HSPA8), mRNA
AGAGGTTGAT	3 major histocompatibility complex, class I, A
SOMA	3 Resposta a Estimulos Externos
Contig2328	6 gi 11055671 gb AF304203.1 AF304203 Suscrofa breed Swedish wild boar mitochondrion, partial genome
TGGGGGTTTC	3 ferritin, heavy polypeptide 1
SOMA	3 Função Desconhecida
Contig2268	6 gi 5902089 ref NM_006931.1 Homo sapiens solute carrier family 2 (facilitated glucose transporter), member 3(SLC2A3), mRNA

CGCTGTAAT	5 MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)
GTGAAACCCT	42 CAP-binding protein complex interacting protein 2
SOMA	47 Transporte
	qil15559521 gb BC014125.1 BC014125 Homo sapiens, Similar to splicing factor 3b, subunit 2, 145kD, cloneIMAGE:4562562, mRNA,
Contig2234	6 partial cds
CCTTATATTT	3 ribosomal protein L8
TGCTGCCCCC	2 ribosomal protein L10
TGCTGCCCCC	ESTs, Highly similar to A42735 ribosomal protein L10, cytosolic 2 [H.sapiens]
GAGAAGGAGG	2 CasL interacting molecule
GGTGTGGAAG	4 egl nine homolog 2 (C. elegans)
TGCTGCCCCC	2 placenta-specific 1
GAGAAGGAGG	2 peptidylprolyl isomerase A (cyclophilin A)
SOMA	17 Metabolismo de RNA
	qil14764461 ref XM_006867.3 Homo sapiens selectin P ligand
Contig2166	6 (SELPLG), mRNA
GTCACCAAAC	2 selectin P ligand
SOMA	2 Adesão
	qil4506516 ref NM_002923.1 Homo sapiens regulator of G-protein signalling 2, 24kDa (RGS2), mRNA
Contig1956	6 signalling 2, 24kDa (RGS2), mRNA
TAATGCTAAA	5 regulator of G-protein signalling 2, 24kD
SOMA	5 Transdução de Sinal
	qil14149862 ref NM_032179.1 Homo sapiens hypothetical protein
Contig1883	6 FLJ20542 (FLJ20542), mRNA
CTGGGGGAGG	6 erythrocyte membrane protein band 4.9 (dematin)
CTCACATCTC	2 ribosomal protein, large, P0
SOMA	8 Função Desconhecida
	qil14149687 ref NM_015449.1 Homo sapiens NICE-3 protein (NICE-3),
Contig1702	6 mRNA
AAAAAGCAGA	3 superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult))
AGGAAAGCTG	27 ribosomal protein L36
CTGGCAGATT	2 splicing factor 3a, subunit 3, 60kD
CAAGGATCTA	3 NICE-3 protein
ATATAATCTG	6 lectin, galactoside-binding, soluble, 3 (galectin 3)
AGTGTGGAAT	3 keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner) fibroblast growth factor receptor 2 (bacteria-expressed kinase,
CAAGGATCTA	3 keratinocyte
SOMA	47 Função Desconhecida
	qil15431300 ref NM_000971.2 Homo sapiens ribosomal protein L7
Contig1593	6 (RPL7), mRNA
TCTGAAGTCA	3 inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
SOMA	3 Metabolismo de Proteínas
	qil20482396 ref XM_166227.1 Homo sapiens similar to FLJ00239
Contig1588	6 protein (LOC219972), mRNA
AGAAAGATGT	9 annexin A1
SOMA	9 Função Desconhecida
	qil22748758 ref NM_152350.1 Homo sapiens hypothetical protein
Contig1491	6 MGC40157 (MGC40157), mRNA
AACTGCCCA	4 ribosomal protein L27
GCACAGGCCA	3 NEU1 protein
GGCAACATAG	2 zinc finger protein 274

TTGCCCAGGC	13 novel protein
CAAGTTCTTT	8 sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like
AACTTGCCCA	4 aminopeptidase puromycin sensitive
CAGCCTGGGC	2 protease, serine, 21 (testisin)
SOMA	36 Função Desconhecida
	gij2342716 gb AC002039.1 HUAC002039 Homo sapiens Chromosome 16 BAC clone CIT987-SKA-13F4 ~complete genomicsequence,
Contig1433	6 complete sequence
CTCCCTTGCC	2 capping protein (actin filament), gelsolin-like
SOMA	2 Matriz Extracelular, Membrana Celular, Citoesqueleto
	gij12698056 dbj AB051543.1 Homo sapiens mRNA for KIAA1756
Contig1432	6 protein, partial cds
TGGTGACAGC	2 putative secreted protein
TGGTGACAGC	2 nuclear receptor co-repressor/HDAC3 complex subunit
SOMA	4 Matriz Extracelular, Membrana Celular, Citoesqueleto
	gij14250560 gb BC008734.1 BC008734 Homo sapiens, Fc fragment of IgG, receptor, transporter, alpha, cloneMGC:1506 IMAGE:3163446,
Contig1306	6 mRNA, complete cds
CCACTCCTCA	5 defender against cell death 1
TGCTGATTTG	2 C-type (calcium dependent, carbohydrate-recognition domain) lectin,
CACGAAGGGA	3 immunoglobulin lambda-like polypeptide 1
CTGCTAAAAG	2 cystatin A (stefin A)
CCTTCAAATC	31 carbonic anhydrase I
CTGTGAGACC	8 Fc fragment of IgG, receptor, transporter, alpha
SOMA	51 Resposta a Estimulos Externos
	gij14591907 ref NM_000978.2 Homo sapiens ribosomal protein L23
Contig127	6 (RPL23), mRNA
CTTTTTTCTG	3 mitochondrial solute carrier
GCCAAAAACC	10 peptidoglycan recognition protein
GCTGCTGGTG	6 jagged 2
AAAATAAAGA	2 APEX nuclease (multifunctional DNA repair enzyme)
AAAACCTGTA	2 solute carrier family 5 (sodium iodide symporter), member 5
TCACCCACAC	12 ribosomal protein L23
GGGATCAAGG	2 mitochondrial ribosomal protein L24
GCCAAAAACC	10 collagen, type I, alpha 1
SOMA	47 Metabolismo de Proteínas
	gij20073232 gb BC026955.1 Homo sapiens, hypothetical protein from EUROIMAGE 1977056, clone MGC:34262IMAGE:5106605, mRNA,
Contig1267	6 complete cds
	proteinase 3 (serine proteinase, neutrophil, Wegener granulomatosis
GGCTGGGGCC	21 autoantigen)
GGCTGGGGCC	21 elastase 2, neutrophil
SOMA	42 Função Desconhecida
	gij11640575 gb AF113214.1 Homo sapiens MSTP034 mRNA, complete
Contig1253	6 cds
TAAGAAAAAA	Homo sapiens , Similar to chordin-like, clone MGC:11264
ACCCCCTTCC	4 IMAGE:3942645,
	4 transketolase (Wernicke-Korsakoff syndrome)
SOMA	8 Função Desconhecida
	gij123510386 ref NM_153050.1 Homo sapiens myotubularin related
Contig1190	6 protein 3 (MTMR3), transcript variant 1, mRNA
ATATTTTCT	3 putative ATPase

TGCACGACTA	3 RING1 and YY1 binding protein
SOMA	6 Matriz Extracelular, Membrana Celular, Citoesqueleto
Contig984	5 gi 6808282 emb AL137572.1 HSM802317 Homo sapiens mRNA: cDNA DKFZp586O0221 (from clone DKFZp586O0221)"
TTCCCTTCTT	7 major histocompatibility complex, class II, DP beta 1
ACAAAAGTGA	2 DEK oncogene (DNA binding)
TTCATTATAA	16 prothymosin, alpha (gene sequence 28)
SOMA	25 Função Desconhecida
Contig902	5 gi 20538904 ref XM_029830.7 Homo sapiens hypothetical protein AL133206 (LOC64744), mRNA
AGAGGGTGGG	2 DnaJ (Hsp40) homolog, subfamily B, member 1
SOMA	2 Função Desconhecida
Contig895	5 gi 18482312 gb AC073517.6 Homo sapiens BACclone RP11-163E9 from 7, complete sequence
ACCCACCTGC	3 calpain 1, (mu/l) large subunit
SOMA	3 Função Desconhecida
Contig878	5 gi 17391507 gb BC018694.1 BC018694 Homo sapiens, clone MGC:3256 IMAGE:3506191, mRNA, complete cds
CTGTACATAC	3 NS1-binding protein
TTGGGGTTTT	2 ferritin, heavy polypeptide 1
AGTTGTCCCG	4 retinoblastoma-associated factor 600
CAGGCCCTTT	2 glutathione S-transferase M1
TTGGGGTTTT	2 WAS protein family, member 1
SOMA	13 Função Desconhecida
Contig758	5 gi 188668 gb M19508.1 HUMMPOA Human myeloperoxidase gene, exons 1-4
GCCCTGGAGA	2 ferritin, light polypeptide
CCCCAGCCAG	50 ribosomal protein S3
GCGGCTGCGC	2 basigin (OK blood group)
GCCCTGGAGA	2 major histocompatibility complex, class II, DR alpha
SOMA	56 Resposta a Estímulos Externos
Contig68	5 gi 5032056 ref NM_005620.1 Homo sapiens S100calcium binding protein A11 (calgizzarin) (S100A11), mRNA
GCCAGGGCCA	2 MYC-associated zinc finger protein (purine-binding transcription factor)
GCCAGGGCCA	2 aldolase C, fructose-bisphosphate
GTGAAGGCAG	53 ribosomal protein S3A
SOMA	57 Ciclo Celular
Contig62	5 gi 23398606 gb BC038241.1 Homo sapiens, Similar to RIKEN cDNA 4921501M20 gene, clone MGC:46306 IMAGE:5180790,mRNA, complete cds
GACTGGA AAA	11 "Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide"
GGAGGTGGGG	10 granulin
GCTCACTGCA	3 peptidylprolyl isomerase (cyclophilin)-like 2
TAGGAGCTGA	2 ribosomal protein S26
CAGGGCACAG	2 immunoglobulin heavy constant mu
SOMA	28 Função Desconhecida
Contig589	5 gi 21756105 dbj AK096577.1 Homo sapiens cDNA FLJ39258 fis, clone OCBBF2009242
TGAAACTCAT	4 neurotrophic tyrosine kinase, receptor, type 1
SOMA	4 Função Desconhecida