

Universidade Federal de São Carlos
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética e Evolução

**"ESTUDOS TRANSCRICIONAIS EM LEUCÓCITOS DE PORTADOR
DA SÍNDROME DE DOWN, ATRAVÉS DA TÉCNICA DE ANÁLISE
SERIAL DA EXPRESSÃO GÊNICA (SAGE)".**

Wilson Malagó Junior

São Carlos – SP – 2004.

Universidade Federal de São Carlos
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética e Evolução

**"ESTUDOS TRANSCRICIONAIS EM LEUCÓCITOS DE PORTADOR
DA SÍNDROME DE DOWN, ATRAVÉS DA TÉCNICA DE ANÁLISE
SERIAL DA EXPRESSÃO GÊNICA (SAGE)".**

Wilson Malagó Junior

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Genética e Evolução do Centro de Ciências Biológicas e da Saúde da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Genética e Evolução; área de concentração: Genética e Evolução.

São Carlos – SP – 2004.

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar**

M236et

Malagó Junior, Wilson.

Estudos transcricionais em leucócitos de portador da síndrome de Down, através da técnica de análise serial da expressão gênica (SAGE) / Wilson Malagó Junior. -- São Carlos : UFSCar, 2004.

78 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2004.

1. Genética humana. 2 Down, Síndrome de. 3. SAGE. 4. Leucócitos. 5. Transcriptoma. 6. Humano. I. Título.

CDD: 573.21 (20^a)

Orientador:

Flávio Henrique da Silva

Dedicatória

Dedico minha dissertação aos meus pais, e agradeço pelo apoio, incentivo, pelo “porto seguro” que representam e pela beleza e nobreza do sentimento de orgulho que têm por mim.

Agradecimentos

Ao amigo, orientador e professor Dr. Flávio Henrique da Silva, pela amizade e oportunidade de orientação que abriram meus horizontes e que não só contribuíram para a minha formação acadêmica, como também para minha formação pessoal.

A todos os colegas do Laboratório de Biologia Molecular, pelo companheirismo e apoio na convivência desfrutada.

Aos colegas do Departamento de Genética e Evolução pelo incentivo e convivência.

Ao Programa de Pós-graduação em Genética e Evolução pela oportunidade.

À todos os professores pelo conhecimento transmitido.

As secretárias Leila, Rosemeire, Regiane e Tatiane pelos auxílios prestados.

Aos colegas do Instituto de Física da USP de São Carlos pelo companheirismo.

Ao Prof. Dr. Hamza El Dorry e seus orientados pelo apoio.

Aos demais colegas do Instituto de Química da USP de São Paulo pelo companheirismo e apoio.

À FAPESP, CAPES e CEPID pelo auxílio financeiro.

Resumo

A síndrome de Down (SD) é a anomalia genética mais freqüente em humanos e apresenta características como retardo mental, malformações cardíacas, defeitos do sistema imunológico, risco aumentado de desenvolvimento de leucemia e doença de Alzheimer, entre outras. Seus fenótipos são causados pela trissomia total ou parcial do cromossomo 21 (CH21) e provavelmente se originam devido a um desequilíbrio dos produtos gênicos deste cromossomo. Inserido em um contexto de "Genômica Funcional", o presente trabalho analisou o perfil de expressão gênica de leucócitos de portador da SD na busca de informações sobre o provável padrão molecular anômalo do transcriptoma do paciente. Para tanto, utilizou-se a técnica de Análise Serial da Expressão Gênica (SAGE). Foram listadas e analisadas diferenças entre os transcriptomas de leucócitos de indivíduos normais e de portador de SD. Verificou-se que a condição de SD comparada a não portadores apresenta baixa expressão relativa de transcritos envolvidos com o sistema imune e o metabolismo, além de transcritos ribossomais. Alguns transcritos diferencialmente expressos foram selecionados como possíveis candidatos ao envolvimento com a SD no contexto celular de leucócitos. Eles serviram também para validar os dados da SAGE. Assim, foi realizado o primeiro estudo de análise de expressão gênica em larga escala envolvendo transcriptoma de portadores de SD e utilizando a SAGE. Com isto foi disponibilizado um conjunto de dados importantes para estudos futuros que visem esclarecer a patogênese da SD.

Lista de figuras

Figura 1: Cariótipo de um portador de Trissomia do CH21.....	2
Figura 2: Idiograma do cromossomo humano 21 com regiões assinaladas.....	8
Figura 3: Esquema geral da obtenção e análise dos dados gerados pela técnica de SAGE (figura modificada do sítio “www.sagenet.org”).....	14
Figura 4: Esquema da abordagem experimental adotada pela técnica de SAGE.....	23
Figura 5: Representação dos adaptadores usados na técnica de SAGE.....	24
Figura 6: Trecho de cromatograma de seqüência da biblioteca de SAGE de portador de SD.....	25
Figura 7: Parte da tabela de contagem de “tags” de portador de SD gerada pelo “software” e-SAGE.....	26
Figura 8: Parte da tabela de contagem de “tags” de não portadores da SD gerada pelo “software” e-SAGE.....	27
Figura 9: Parte da tabela de comparação entre “tags” de portadores e não portadores de SD gerada pelo “software” e-SAGE.....	30
Figura 10: Parte da tabela de dados do mapeamento “CGAP SAGE Genie” (“Hs.best_gene.gz”).	31
Figura 11: Parte da tabela de dados final, comparativa entre “tags” (transcritos) de leucócitos de não portadores e de portadores de SD.....	33
Figura 12: RNA total obtido do sujeito de pesquisa (gel de agarose 1%).....	38
Figura 13: Amplificação dos “ditags” maiores (gel de poliacrilamida 12%).....	39
Figura 14: Extração dos “ditags” menores de 26pb (gel de poliacrilamida 12%).	40
Figura 15: Concatamerização (gel de poliacrilamida 8%).....	40
Figura 16: Amplificação dos clones de concatâmeros (gel de agarose 1%).....	41
Figura 17: Gráfico de aparecimento de “tags” novos (únicos) na biblioteca de SD.....	42
Figura 18: Distribuição de “tags” não identificados nas bibliotecas SD e	

Controle.....	46
Figura 19: Classificação funcional dos transcritos mais diferencialmente expressos nas bibliotecas controle e de paciente de SD.....	52
Figura 20: Classificação dos genes do CH21 pela função molecular.....	55
Figura 21: Fotos de gel de poliacrilamida 8% referentes às validações dos transcritos selecionados por “RT-PCR” semiquantitativa.....	58

Lista de tabelas

Tabela 1: Transcritos selecionados para validação.....	35
Tabela 2: Iniciadores usados nas validações.....	37
Tabela 3: Distribuição de “tags” únicos em cópias por célula das duas bibliotecas de acordo com suas abundâncias e porcentagem de massa de mRNA.....	43
Tabela 4: Os 40 transcritos mais superexpressos em leucócitos de portador de SD.....	48
Tabela 5: Os 40 transcritos mais sub-expressos em leucócitos de portador de SD.....	49
Tabela 6: Lista dos 10 genes do cromossomo 21 mais diferencialmente expressos entre as duas bibliotecas (controle e SD).....	54

Lista de abreviaturas

°C – Graus centígrados

μl – Microlitros

μM – MicroMolar

2D – Duas dimensões

APAE - Associação de Pais e Amigos dos Excepcionais

APP – “Amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)”

BCL2A1 – “BCL2- Related protein A1”

BsmF I – *Bacillus stearothermophilus* F I

C21orf43 (JAM2) – “junctional adhesion molecule 2 “

CAF1p60 – “chromatin assembly factor 1, subunit B (p60)”

CBR1 – “Homo sapiens carbonyl reductase 1 (CBR1)”

CBS – “Cystathionine-beta-synthase”

cDNA – DNA complementar

CEP-UFSCar – Comitê de Ética em Pesquisa da Universidade Federal de São Carlos

CGAP – “Cancer Genome Anatomy Project”

CH21 – Cromossomo humano 21

CXCR4 – “Chemokine C-X-C Motif Receptor 4”

D21S17 – “Human STS primer pGSH8 for locus D21S17 (21q22.2)”

D21S267 – “H. sapiens (D21S267) DNA segment containing (CA) repeat; clone AFM238wc3”

DCRA (DSCR3) – “Down syndrome critical region gene 3”

DCRB (DSCR4) – “Down syndrome critical region gene 4”

DNA – Ácido desoxirribonucléico

DNMT3L – “DNA (cytosine-5-)-methyltransferase-like”

dNTPs – desoxirribonucleosídeos trifosfatados

DSCAM – “Down syndrome cell adhesion molecule”

DSCR – Região Crítica da Síndrome de Down

DSCR1 – “Down syndrome critical region gene 1”

DSCR2 – “Down syndrome critical region gene 2”

DSCR5B – “Down syndrome critical region gene 5 “

DTT - Ditionitroto

DYRK1A - “Dual-specificity tyrosine - (Y)-phosphorylation regulated kinase 1A”

EDTA – “ethylenediaminetetraacetic acid”

ERG – “Homo sapiens genomic DNA of 21q22.2 Down Syndrome region, segment 12/13”

ESTs – “Expression Sequence Tags”

ETS2 – “Homo sapiens ETS2 intronic transcript 1 mRNA”

FTCD – “Formiminotransferase cyclodeaminase”

GART – “Phosphoribosylglycinamideformyltransferase”

GLGI – “Generation of longer cDNA fragments from SAGE tags for gene identification”

GPDH – “Glicerol Fosfato Desidrogenase”

GRIK1 – “Glutamate receptor, ionotropic, kainite 1”

GTF3A – “General Transcription Factor III A”

HCl – Ácido clorídrico

hnRNA – “Heterogeneous nuclear RNA”

HRMT1L1 – “HMT1 hnRNP methyltransferase-like 1”

IGSF5 – “immunoglobulin superfamily, member 5”

IL8 – “Interleukin 8”

ITGB2 – “integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit)”

Kb – Kilobases

KCl – Cloreto de potássio

KCNE1 – “potassium voltage-gated channel, Isk-related family, member 1”

KCNE2 – “potassium voltage-gated channel, Isk-related family, member 2”

KCNJ15 – “potassium inwardly-rectifying channel, subfamily J, member 15”

KCNJ6 – “potassium inwardly-rectifying channel, subfamily J, member 6”

Mb – Mega bases

MgCl₂ – Cloreto de magnésio

mL – Mililitros

mM – MiliMolar

M-MLV RT – “Moloney Murine Leukemia Virus Reverse Transcriptase”

MNBH (DYRK1A) – “dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A”

MX1 – “Homo sapiens myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse) (MX1)”

N6AMT1 – “Putative N6 DNA methyltransferase”

NCAM2 – “neural cell adhesion molecule 2”

NCBI – “National Center for Biotechnology Information”

NF-kappaB – “Nuclear factor kappaB”

ng – “Nanogramas”

Nla III – *Neisseria lactamica* III

pb – Pares de bases

PCP4 – “Purkinje cell protein 4”

PCR – “Polymerase Chain Reaction”

PDXK – “pyridoxal (pyridoxine, vitamin B6) kinase”

PFKL – “phosphofructokinase, liver”

pMol – PicoMol

PRSS7 – “protease, serine, 7 (enterokinase)”

pVal – Valor de probabilidade

RAST-PCR – “Rapid analysis of unknown SAGE tags”

RNA – Ácido ribonucléico

RNAm – Ácido ribonucléico mensageiro

RPL13A – “Ribosomal protein L13a”

RPL29 – Ribosomal protein L29

RPL37 – “Ribosomal protein L37”

RT-PCR – “reverse transcription-polymerase chain reaction”

S100B – “S100 calcium binding protein, beta (neural)”

SAGE – “Serial Analysis of Gene Expression”

SD – Síndrome de Down

SIM2 – “Single-minded homolog (Drosophila)”

SLC19A1 – “Solute carrier family 19 (folate transporter), member 1”

SMT3H1 – “SMT3 suppressor of mif two 3 homolog 1 (yeast)”

SNF1LK – “SNF1-like kinase”

SOD1 – “Superoxide dismutase 1, soluble (amyotrophic lateral sclerosis, adult)”

SRS – “Sequence Retrieval System”

TPTE – “Transmembrane Phosphatase with TEnsin homology”

TRPC7 – “transient receptor potential cation channel, subfamily M, member 2”

UBE2G2 – “ubiquitin-conjugating enzyme E2G 2 (UBC7 homolog, yeast)”

UniGene ID – “número de identificação de “clusters” do “UniGene””

USP16 – “ubiquitin specific protease 16”

USP25 – “ubiquitin specific protease 25”

µg – Microgramas

Índice

1 – Introdução.....	1
1.1 – As anomalias congênitas.....	1
1.2 - A síndrome de Down.....	1
1.3 – O cromossomo humano 21.....	4
1.4 - A Região Crítica da síndrome de Down.....	6
1.5 – Síndrome de Down e transcriptômica	10
1.6 – Proteômica.....	15
2 – Objetivos.....	17
2.1 –Objetivos.....	17
2.2 –Objetivos para alcançar os procedimetos.....	17
2.3 – Justificativa dos objetivos.....	18
3 – Materiais e Métodos.....	20
3.1 – Metodologia geral.....	20
3.2 – Obtenção de biblioteca de SAGE (Resumo da abordagem experimental). 21	
3.3 – Análise dos dados.....	24
3.4 – Validações.....	34
3.4.1 – Seleção de transcritos.....	34
3.4.2 – Reações de validação (“RT-PCR”).....	35

4 – Resultados e Discussão.....	38
4.1 – Biblioteca de SAGE.....	38
4.2 – Análise dos dados.....	41
4.3 – Validações da SAGE.....	52
4.3.1 – Seleção de transcritos.....	52
4.3.2 – Validações.....	56
5 – Conclusão.....	59
6 – Referências Bibliográficas.....	60
7 – Anexos.....	68
7.1 – Análise de riscos e benefícios do projeto.....	68
7.2 – Esclarecimentos prestados para submissão do projeto ao CEP-UFSCar.	70
7.3 – Modelo do consentimento esclarecido assinado pelos responsáveis do sujeito da pesquisa (portador de SD).....	72
7.4 – Carta de esclarecimento enviada aos responsáveis pelo sujeito de pesquisa.....	74

1 – Introdução

1.1 – As anomalias congênitas

As anomalias congênitas podem ocorrer devido a fatores externos que afetam o indivíduo em desenvolvimento, tais como uso de substâncias teratogênicas ou doenças adquiridas pela mãe durante a gestação. Elas também podem ser decorrentes de fatores genéticos presentes no zigoto (fatores internos) (Vogel e Motulsky, 1997). Nestas alterações um amplo espectro de características fenotípicas indesejáveis pode estar presente. No entanto, quadros clínicos semelhantes podem ter causas distintas e vice-versa. Dentre as alterações genéticas que um zigoto pode portar, as que se ressaltam são: Mutações pontuais, Deleções, Duplicações, Translocações cromossômicas e Aneuploidia. As síndromes são caracterizadas por alterações que ocorrem com certa frequência originando quadros clínicos típicos.

1.2 - A síndrome de Down

A síndrome de Down (SD) foi caracterizada por Langdon Down em 1866 (Down, 1866). Apesar do nome atual, ela foi primeiramente denominada de mongolismo porque os portadores apresentavam semelhanças fisionômicas aos habitantes da Mongólia. Segundo Down a anomalia seria resultado da influência desta raça. Esta conclusão reflete o preconceito que havia na época contra raças não européias, consideradas inferiores.

A SD só deixou de ser considerada influência de raças quando em 1959 Lejeune e colaboradores verificaram a presença de um pequeno cromossomo acrocêntrico adicional nos portadores (Lejeune *et. al.*, 1959) (Fig.1). Com isto se verificou que a SD resulta da trissomia do cromossomo humano 21 (CH21). De fato, todos estão sujeitos a ter um filho portador, e isto independe da raça ou condição sócio-econômica, (Christianson, 1996).

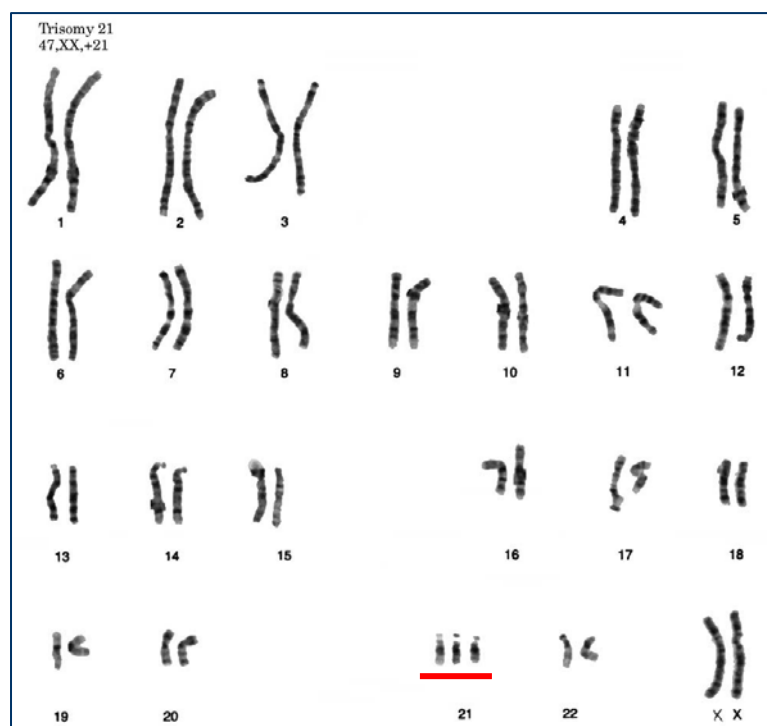


Figura 1: Cariótipo de um portador de Trissomia do CH21. Os cromossomos 21 sublinhados em vermelho refletem a condição de trissomia.

Sendo uma das anomalias genéticas mais comuns em humanos, a SD é a melhor estudada. Sua incidência é de aproximadamente um caso em cada setecentos nascimentos (Hassold e Jacobs, 1984). Isto significa dizer que nascem no Brasil cerca de 8 mil bebês com SD por ano. Se a idade materna é

elevada (mais de 35 anos), a incidência é aumentada para um caso a cada 370 nascimentos (Hook *et al.*, 1983).

A trissomia do CH21 é consequência de uma aneuploidia que acontece devido a não disjunção desse cromossomo. Estes erros têm origem principalmente maternal sendo que 67% dos casos ocorrem na meiose materna I; 20% na meiose materna II; 3% na meiose paterna I; 4% na meiose paterna II e 6% na mitose. Portadores de trissomia do CH21 oriunda de não disjunção mitótica possuem parte de suas células trissômicas e apresentam os fenótipos observados na SD. Estes casos são denominados mosaicismo e a não disjunção ocorre nas fases iniciais do desenvolvimento embrionário (Antonarakis, 1998). Um fenômeno raro, que também origina as principais características da síndrome, é a trissomia parcial do CH21 (ver item 1.4). Neste caso acontece a duplicação de uma pequena região do CH21, resultante de uma translocação (Korenberg *et al.*, 1994).

Mais de oitenta traços clínicos se manifestam devido a trissomia do cromossomo 21 (Reeves *et al.*, 2001). Dentre as manifestações fenotípicas, as principais são bem conhecidas e várias delas permitem o diagnóstico logo após o nascimento. Além disso, a diagnose intra-uterina da anomalia tem se tornado cada vez mais comum (Epstein *et al.*, 1996).

Algumas alterações marcantes são: hipotonia muscular, baixa estatura, mãos curtas, face achatada, fissura palpebral oblíqua ascendente, orelhas diplásticas, prega palmar única, cardiopatia, retardo mental, entre outras. Destas anormalidades, o retardo mental é presente em todos os pacientes (Epstein, 1995). Aparentemente, os períodos críticos no desenvolvimento cerebral anômalo são a proliferação, diferenciação e organização de neurônios

(Capone, 2001). Além da redução do número de neurônios, existe uma formação defeituosa e plasticidade sináptica (Nizetic, 2001). Também, com o avanço da idade, todos os portadores exibem alterações no sistema imunológico e endócrino e ainda apresentam risco aumentado de desenvolver doença de Alzheimer (Epstein, 1986). Crianças portadoras até os quatro anos de idade apresentam risco aumentado de desenvolver leucemia aguda (Fong *et al.*, 1987). Por outro lado, a trissomia do CH21 parece conferir proteção contra o desenvolvimento de tumores sólidos, sugerindo a existência de genes supressores de tumores nesse cromossomo (Hattori *et al.*, 2000).

1.3 – O cromossomo humano 21

Devido ao seu pequeno tamanho e envolvimento com a SD, o CH21 tem sido um dos alvos importantes dos estudos em genética humana; ele foi o primeiro a ser seqüenciado e possui genes amplamente estudados (Hattori *et al.*, 2000; Gosset *et al.*, 1999; Antonarakis, 2001). Este cromossomo é considerado de baixa densidade gênica e representa cerca de 1-1,5% do genoma humano. No manuscrito original, no qual o sequenciamento e a anotação de genes foram descritos, 225 genes ou genes preditos foram identificados (Hattori *et al.*, 2000). O conteúdo estimado de genes agora é de 329, incluindo 165 genes confirmados experimentalmente, 150 modelos de genes baseados em bancos de dados de “ESTs”, e 14 predições computacionais (Roizen e Patterson, 2003). Estes genes se localizam no braço 21q; O braço 21p possui apenas 1 gene, o *TPTE*, de função desconhecida.

Sabe-se que a trissomia do braço 21p não altera o fenótipo de seus portadores (Antonarakis, 1998).

Devido ao grande interesse no CH21, ele foi o primeiro autossomo a ter uma série de mapas desenvolvidos, entre eles: mapa de ponto de quebra, mapa de restrição, mapa de ligação e mapa de sequenciamento. Este último é fundamental para o fácil acesso a genes candidatos ao envolvimento com doenças e também para o aprofundamento no conhecimento dos genes presentes neste cromossomo (Hattori *et. al.*, 2000).

O CH21 apresenta grupos de genes que estão envolvidos na mesma via metabólica ou sistema biológico (Roizen e Patterson, 2003). Um grupo genes que exercem influência na função ou estrutura do sistema nervoso central pode ter um papel na neuropatogênese da SD, são eles: *SIM2*, *DYRK1A*, *GART*, *PCP4*, *DSCAM*, *GRIK1*, *APP*, *S100B*, e *SOD1*. Pelo menos seis genes do CH21 podem exercer função no metabolismo de folato ou grupos metil. São eles: *N6AMT1*, *CBS*, *DNMT3L*, *SLC19A1*, *FTCD* e *HRMT1L1* (Roizen e Patterson, 2003). O folato é necessário para o metabolismo de grupos metil e isto inclui a metilação do DNA, portanto, estes genes podem ter função no controle da expressão gênica. Outros genes se destacam: os que codificam kinases (*PRSS7*, *SNF1LK*, *PDXK* e *PFKL*); os envolvidos na via de “ubiquitinização” (*USP25*, *USP16*, *UBE2G2* e *SMT3H1*); moléculas de adesão celular (*NCAM2*, *IGSF5*, *C21orf43* e *ITGB2*); proteínas de canais iônicos (*KCNE2*, *KCNE1*, *KCNJ6*, *KCNJ15* e *TRPC7*) e vários fatores de transcrição.

1.4 - A Região Crítica da síndrome de Down

Estudos em pacientes portadores de trissomia parcial permitiram o estabelecimento de uma região mínima do cromossomo 21, situada no braço maior (21q–região distal), provavelmente responsável pelas principais características da SD e comum a todos os portadores. Esta região foi denominada Região Crítica da Síndrome de Down (DSCR) (Fuentes *et al.*, 1995 –1997; Epstein *et al.*, 1991; Korenberg *et al.*, 1988). Trabalhos realizados através de mapeamento com o uso de marcadores moleculares, que auxiliam na localização dos genes do cromossomo 21 e na determinação da DSCR (McInnis *et al.*, 1993), demonstram variações de tamanho desta região entre pacientes. Acredita-se que tais variações podem ser responsáveis pela origem de conjuntos distintos de fenótipos da síndrome. No entanto, existem outros fatores que provavelmente contribuem para a variabilidade fenotípica observada nos indivíduos com SD. São eles: combinações diferentes de alelos dos genes do CH21 e constituição genética do indivíduo no qual ocorre a trissomia e diferentes influências ambientais (Reeves *et al.*, 2001).

A delimitação da DSCR é um assunto que tem gerado controvérsias (Fig.2). Segundo McCormick *et al.*, (1989); Delabar *et al.*, (1993); Korenberg *et al.*, (1994), os estudos em pacientes com a trissomia parcial sugerem a delimitação da DSCR na região entre os marcadores moleculares D21S17 e ETS2, possuindo aproximadamente 4 Mb. Segundo Korenberg *et al.*, (1992), a região responsável pelo defeito no canal atrioventricular (característico da SD), foi mapeada entre os marcadores D21S267 e MX1 e, portanto, a DSCR apresentaria aproximadamente 5 a 6 Mb. Segundo Lucente *et al.*, (1995), a

DSCR possui 2,5 Mb e situa-se entre CBR1 e ERG. Mais recentemente a Região Crítica foi definida entre t(4;21) "break point" e ERG; com aproximadamente 1,6 Mb (Ohira *et al.*, 1996).

Segundo Korenberg *et al.*, (1994), a DSCR pode não ser a região exclusivamente responsável pela SD. Há casos em que indivíduos portadores de trissomia parcial, que possuem a DSCR maior, exibem fenótipos adicionais se comparados a trissômicos parciais com a DSCR menor. No entanto, indivíduos trissômicos totais não apresentam todas as características fenotípicas (Antonarakis, 1998). Também, existem três pacientes portadores de trissomia parcial na região proximal do CH21 (que não alcança a DSCR); que apresentam fenótipos da SD como: característica facial típica, baixa estatura, hipotonia muscular, e retardo mental (Korenberg *et al.*, 1994). Estes casos confrontam a afirmação de que existe uma só região responsável pelos fenótipos da SD. Para esclarecer a contribuição das regiões 21q aos fenótipos, mais casos se fazem necessários (Antonarakis, 1998). Apesar das discussões, existe um consenso entre os estudiosos: os genes presentes na DSCR (ou próximos a ela), podem exercer importantes papéis no desenvolvimento das alterações características da SD.

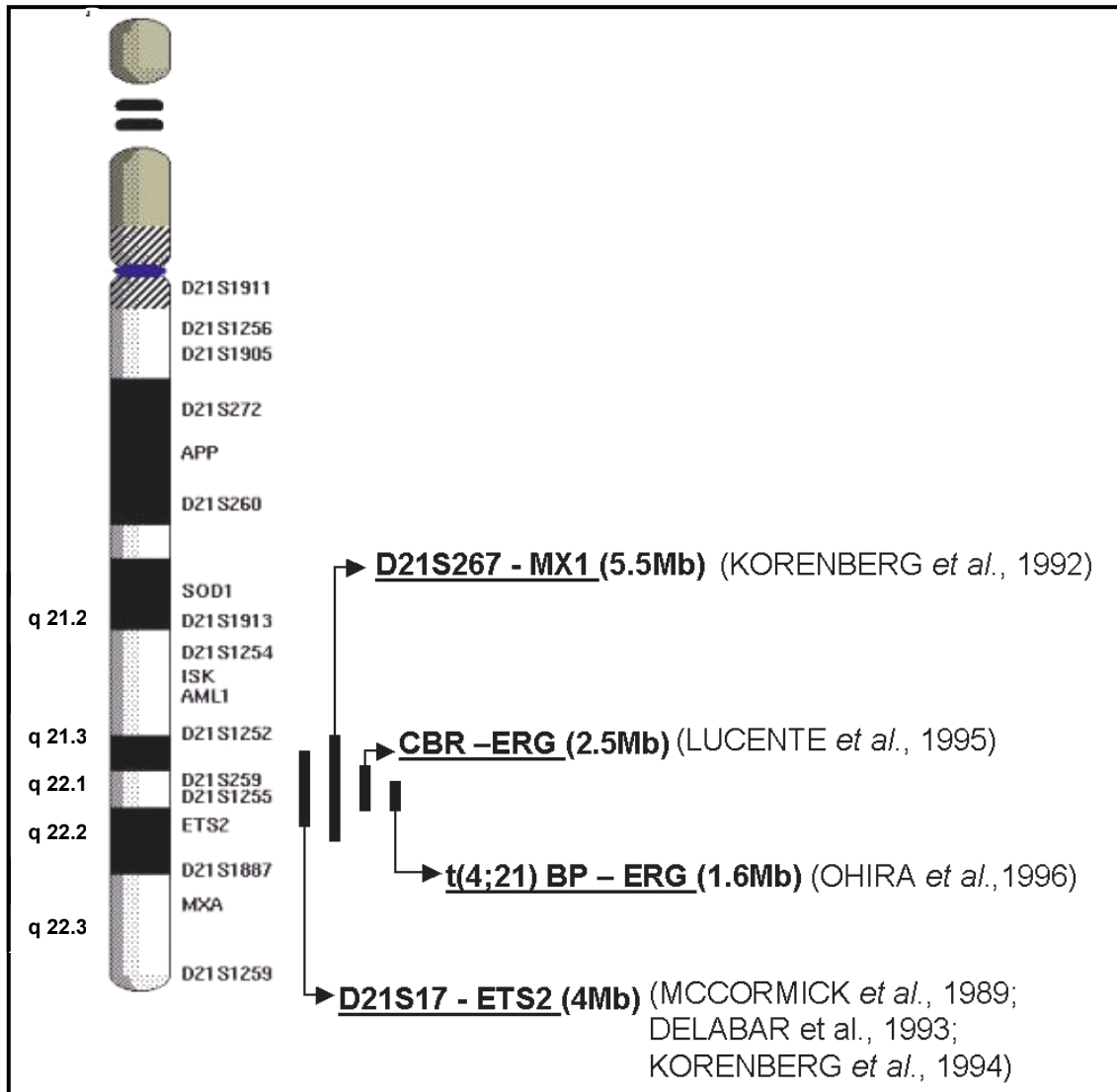


Figura 2: Idiograma do cromossomo humano 21 com regiões assinaladas.

Os diferentes autores defendem distintas delimitações para a DSCR.

Um fato relevante na aneuploidia do CH21 é a questão da contribuição de três cópias de genes normais para os fenótipos anômalos. Segundo Chrast (2000) a presença de três cópias de um certo número de genes do CH21, que contribuem para os fenótipos da SD, é uma hipótese já amplamente aceita (Chrast *et al.*, 2000). Isto já era previsto por Epstein (1995) que afirmava que os fenótipos seriam causados por cópias extra de genes que atuariam via

desequilíbrio bioquímico (Epstein, 1995). No entanto, Antonarakis (1998) reporta que não há ligação convincente entre a superexpressão dos genes do cromossomo 21 com os fenótipos característicos da SD (Antonarakis, 1998).

Recentemente, foi verificada por “Microarray” uma supra-regulação global dos genes do CH21 (comparado a outros cromossomos) em cérebros fetais humanos (Mao *et al.*, 2003). Estes autores também verificaram, em cérebros fetais, a existência de um balanço variável (que provavelmente ocorre em todos os tecidos) nos níveis de expressão em um grupo de genes supra-regulados do CH21. Em outras palavras, um grupo de genes supra-regulado possui representantes que se comportam distintamente, segundo seu nível de expressão, entre diferentes portadores. Este fato poderia ser responsável pelas variações das características clínicas cognitivas dos pacientes.

Dentre os genes situados na DSCR (entre D2S17 e ETS2), destacam-se: *CBR1* (Gonzales *et al.*, 1995), *SIM2* (Chen *et al.*, 1995), *CAF1p60* (Blouin *et al.*, 1996), *ERG* (Rao *et al.*, 1988), *MNBH* (Guimera *et al.*, 1997) e *ETS2* (Sumarsono *et al.*, 1996). Alguns deles são alvos de estudo em nosso laboratório: *DSCR1* (Fuentes *et al.*, 1995), *DCRA* e *DCRB* (Nakamura *et al.*, 1997), *DSCR2* (Vidal-Taboaba *et al.*, 1998) e *DSCR5B* (Shibuya *et al.*, 2000). Na tentativa de elucidar suas funções, realizamos, além do estudo presente, estudos de expressão heteróloga para produção de anticorpos e análises estruturais, interação proteína-proteína e localização subcelular.

1.5 – Síndrome de Down e transcriptômica

Existem duas hipóteses que tentam prever a influência das três cópias dos genes do cromossomo 21, em portadores da SD, no desenvolvimento dos fenótipos Down. Primeiro, a presença de cópias extra de certos (não todos) genes do CH21 contribuindo para os fenótipos. Segundo, os produtos destes genes agindo diretamente ou, mais possivelmente, indiretamente através de processos que afetam a expressão de outros genes ou produtos gênicos (Antonarakis *et al.*, 2001).

O nível de expressão dos genes do 21 é relativamente baixo sendo a maioria deles envolvida com regulação gênica. Este fato reforça a crença de que a influência destes genes na patogênese da SD ocorre de maneira indireta. Neste contexto, o perfil de ação dos produtos gênicos não só do cromossomo 21, como também do resto do genoma (genes “influenciáveis”), se mostraria alterado em portadores, nos diversos tecidos e tipos celulares. Portanto a condição global de desequilíbrio gênico ou bioquímico provavelmente reflete a condição fenotípica de SD. As análises em larga-escala, utilizadas em estudos de genômica funcional, geram informações sobre transcriptomas e proteomas. Estes estudos geram dados importantes que auxiliam na busca de diferenças no padrão de funcionamento gênico entre condições normais e anômalas.

O presente trabalho faz parte da implementação da técnica de SAGE (“Serial Analysis of Gene Expression”) em nosso laboratório. Projetos desta natureza geram perfis de transcriptomas, provendo informações importantes na busca do esclarecimento da patogênese da SD. Estes estudos podem ajudar a esclarecer a ação de genes candidatos ao envolvimento com a SD (genes do

CH21 e da DSCR) e sua provável influência na atividade de outros genes (outros candidatos presentes no resto do genoma), culminando nas características da SD. Isto pode ser alcançado porque a técnica permite a identificação de transcritos diferencialmente expressos em portadores, quando comparados a não-portadores da SD.

Considerando a importância do genoma, transcriptoma e proteoma na manutenção dos processos vitais, e com o fim do sequenciamento do genoma humano, os estudos de transcriptômica e proteômica têm se revelado cada vez mais importantes. Neste contexto de “Genômica Funcional”, a identificação de genes, transcritos e proteínas, bem como o esclarecimento de suas funções, tem crescido de forma significativa.

Os primeiros métodos de análise de transcritos, entre eles o “Diferencial Display” (Liang e Pardee, 1992), o “Representational Difference Analysis” (Hubank e Schatz, 1994) e o “Supression Subtractive Hibridization” (Diatchenko *et al.*, 1996), visavam eleger genes diferencialmente expressos. A identificação dos últimos os tornaria candidatos ao envolvimento com um certo fenótipo estudado, quando comparado a outro fenótipo controle. No entanto estas técnicas geram alto número de falsos positivos e são bastante laboriosas.

Os projetos de sequenciamento de cDNAs (Adams *et al.*, 1991) produziram informações cruciais que promoveram a identificação de genes e transcritos. Além de prover ESTs (“Expression Sequence Tags”), este trabalho de sequenciamento fornece aspectos estruturais de genes (Miller *et al.*, 1997; Jiang e Jacob, 1998; Wheeler *et al.*, 2001). Entretanto, o sequenciamento de cDNAs em larga escala com o intuito de se obter informações quantitativas

(Okubo *et al.*, 1992) depende de esforços exaustivos para a obtenção de amostras significativas.

O aumento de seqüências de cDNA acumuladas e os avanços tecnológicos possibilitaram o desenvolvimento da técnica de “Microarrays” (Lockhart *et al.*, 1996) (uma revisão pode ser obtida no suplemento especial V21 da “Nature Genetics”, 1999). Este método tem sido aplicado com êxito e boa aceitação, já que proporciona o estudo simultâneo de um grande número de transcritos. Entretanto existem algumas desvantagens. Painéis de hibridização montados em membrana requerem o uso de radioativos e não permitem ensaios simultâneos, além de apresentarem um número de “spots” limitados. Painéis montados em lâminas de vidro são mais caros e não reutilizáveis (Sambrook *et al.*, 1989). Ambos envolvem a deposição de cDNAs e isto requer coleção e processo trabalhoso de clones (Sambrook *et al.*, 1989). Já os “Chips”, que se caracterizam pela produção direta de DNAs no substrato, são proibitivamente caros e requerem conhecimento prévio da seqüência de DNA (Sambrook *et al.*, 1989). Além disso, independente do tipo de substrato, a análise dos dados é de uma complexidade acentuada.

A técnica de SAGE foi concebida por Veuculescu *et al.*, (1995) e tem se mostrado valiosa na análise global de transcritos, gerando padrões digitais de expressão gênica (Fig.3). Esta técnica apresenta um potencial de análise em larga escala, já que se baseia na geração de pequenos “ESTs de SAGE” chamados “tags” (aprox. 14 pb). Estes são específicos para cada cDNA do qual se originam e possuem tamanho suficiente para identificar os respectivos “clusters” de cDNA no banco de dados do “UniGene”. Os “tags” são obtidos a partir de uma população inicial de RNAm extraído de um órgão, tecido ou tipo

celular específico, em uma dada condição. Eles são depois ligados em seqüência (“concatamerizados”), clonados e sequenciados. Desta forma, são obtidas bibliotecas de SAGE que possuem seus “tags” contados e analisados por “softwares” apropriados (Fig.3). Como os “tags” (ou “ESTs de SAGE”) são pequenos, o esforço de sequenciamento gerado por técnicas triviais de obtenção de “ESTs” é reduzido em até 50 vezes (Bertelsen e Veuculescu, 1998; Veuculescu *et al.*, 2000).

Após a contagem dos “tags”, se obtém informações quantitativas e qualitativas a respeito dos transcritos. Duas bibliotecas, ou transcriptomas representativos do funcionamento gênico de um dado tipo celular, por exemplo, que potencialmente apresentariam padrões distintos de expressão gênica, podem então ter seus dados comparados (Fig. 3). Os transcritos diferencialmente expressos entre as bibliotecas são identificados através de seus respectivos “tags”. Com a submissão destes ao banco de dados de “Clusters de ESTs do UniGene”, no “NCBI” (<http://www.ncbi.nlm.nih.gov/>) eles apresentam, quando possuem, sua descrição disponível. Assim com os dados em mãos, e dependendo do contexto, são escolhidos alguns genes candidatos a estarem envolvidos com um certo fenótipo. Estes genes apresentam níveis de expressão distintos e significativos quando se comparam as duas condições nas quais os transcriptomas são estudados, ou seja, eles são diferencialmente expressos.

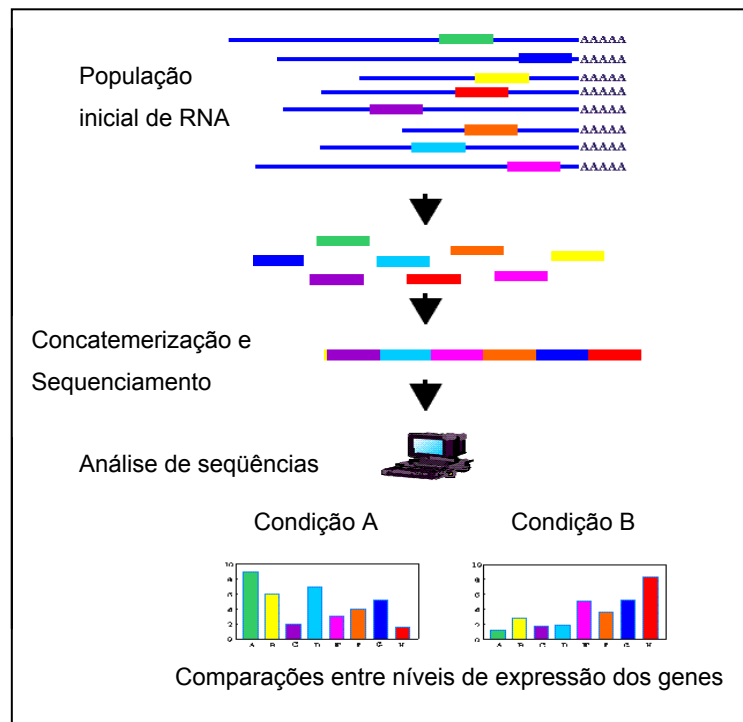


Figura 3: Esquema geral da obtenção e análise dos dados gerados pela técnica de SAGE (figura modificada do sítio “www.sagenet.org”).

Os genes candidatos passam por um processo de validação normalmente realizado em estudos de SAGE, antes de serem tidos como possíveis alvos de estudos em outros trabalhos. Esta validação confirma a condição de sub ou superexpressão relativa de um dado transcrito (gene) em uma das fontes amostrais de RNAm que originou uma das bibliotecas de SAGE. Os termos sub ou superexpressos costumam referir-se as condições apresentadas por transcritos oriundos de fontes anômalas estudadas; no presente trabalho isto também é aplicado. Técnicas de “Northern Blot” (Sambrook e Russell, 2001) e “Real Time-PCR” (Kang *et al.*, 2000; Higuchi *et al.*, 1993), que utilizam os RNAs obtidos das mesmas fontes biológicas em

questão, podem ser empregadas para a validação. A validação pode meramente confirmar a eficiência da SAGE e neste caso qualquer transcrito com diferença de expressão significativa pode ser usado. Porém, para além de validar a técnica, eleger candidatos ao envolvimento com a anomalia estudada, há necessidade de se conhecer a seqüência do transcrito, ou seja, seu respectivo “cluster” deve ter sido identificado. Assim é possível conhecer a função ou o papel biológico do transcrito em questão e inferir seu possível envolvimento com a anomalia estudada.

Além da geração de transcriptomas, que mostram padrões de expressão relativos e absolutos de transcritos, e da obtenção de números absolutos de transcritos por célula (através de normalizações), as informações da SAGE são usadas na geração de mapas genômicos de transcrição (Veuculescu *et al.*, 1997; Caron *et al.*, 2001). Outra utilidade é a identificação de novos transcritos, já que se trata de uma técnica sensível. Isto pode se fazer necessário quando surgem “tags” não associados a seqüências conhecidas (“unknown tags”), que apresentam contagens significativamente distintas entre as fontes biológicas em estudo. Para a identificação existem duas técnicas que se baseiam na obtenção de seqüências “3’” (correspondentes aos “tags”) dos cDNAs desconhecidos: “RAST-PCR” (Van Den Berg *et al.*, 1999) e “GLGI” (Chen *et al.*, 2000).

1.6 – Proteômica

As técnicas de análise de RNAs em larga escala são capazes de gerar importantes e abrangentes panoramas de funcionamento gênico. Com isto,

abrem possibilidade para uma série de estudos que certamente contribuem para o esclarecimento de diversos mecanismos biológicos estudados. Entretanto, apesar de muitas vezes os níveis de RNA representarem os de proteínas, não é possível prever exatamente os níveis das últimas em função dos primeiros, sejam eles determinados por “microarrays” (Le Naour *et al.*, 2001) ou SAGE (Gygi *et al.*, 1999).

Como os estudos que envolvem transcriptomas, a proteômica se insere nas análises funcionais de genomas, possibilitando a análise em massa de proteínas expressas. Nestes estudos, se utilizam técnicas de análise de proteínas em gel 2D, sequenciamento de proteínas, espectrofotometria de massa e “arrays” de proteínas. A proteômica também apresenta grande importância, porém, quando se trata de proteínas não se pode contar com a relativa facilidade de clonagem e manipulação que os ácidos nucleicos oferecem e isto é um fator complicador.

2 – Objetivos

2.1 –Objetivos

Analisar o padrão global de expressão gênica de leucócitos de portador da SD através da técnica de SAGE. Comparar os dados obtidos aos dados de expressão gênica, previamente gerados (também pela SAGE) e disponíveis na rede mundial de computadores, oriundos de leucócitos de não portadores da SD.

Identificar e analisar as prováveis diferenças (significativas) entre os padrões de expressão gênica do indivíduo normal e dos afetados.

Validar a técnica de SAGE procurando selecionar transcritos que possam estar envolvidos com anomalias dos portadores de SD, no contexto celular de leucócitos e com especial atenção aos problemas imunológicos e risco aumentado de desenvolver mal de Alzheimer.

2.2 –Procedimentos para alcançar os objetivos

- Obter sangue periférico de paciente portador da síndrome de Down.
- Isolar leucócitos e extrair RNA total.
- Construir biblioteca de SAGE representativa de leucócitos de portador da SD.
- Analisar os dados obtidos e compará-los aos dados oriundos de uma biblioteca de SAGE representativa de leucócitos de não portadores da SD.

- Realizar validações da SAGE selecionando genes com diferenças de expressão significativa entre os transcriptomas e possíveis candidatos ao envolvimento com fenótipos característicos da SD.

2.3 – Justificativa dos objetivos

A influencia para a escolha da técnica de SAGE para iniciar os estudos de transcriptômica em nosso laboratório se originou de dois fatores principais: O caráter quantitativo, qualitativo e de larga escala da técnica e a expectativa de obtenção de informações sobre os reflexos de desequilíbrio bioquímico oriundo de interações dos produtos gênicos da região DSCR e do CH21 com o genoma, transcriptoma e proteoma, no contexto celular de leucócitos de portador de SD. Além disso, chamaram a atenção características como acessibilidade, custo, facilidade na análise dados e a diversidade de usos da técnica. Pode-se destacar ainda o fato de que os dados gerados pela SAGE são duradouros e permitem comparações entre vários experimentos.

O envolvimento dos portadores da SD com problemas imunológicos e a predisposição para desenvolver doença de Alzheimer (Epstein, 1986) são características relevantes. Portanto, se mostram importantes alvos de estudo que visam contribuir para o entendimento dos mecanismos moleculares causadores do fenótipo de SD. Neste contexto, se justifica a decisão por analisar dados gerados a partir da construção de biblioteca de SAGE representativa de leucócitos de portadores, e compará-los a dados obtidos de uma biblioteca de SAGE representativa de leucócitos de não portadores da SD.

Também, a simplicidade da obtenção do sangue e da extração de RNA a partir deste, facilitaram a execução do presente projeto.

Com este trabalho, a primeira biblioteca de SAGE de humanos com SD se efetivou. Isto gerou, portanto, a inédita possibilidade de se eleger genes candidatos ao envolvimento com a síndrome. Além disso, seus dados ofereceram uma visão global do padrão de expressão gênica de leucócitos de portadores. Estes dados podem contribuir para o esclarecimento da influência dos genes trissômicos do CH21, no funcionamento do resto do genoma, do transcriptoma e proteoma; especialmente no que se refere à possível influência de tais genes na manifestação de fenótipos anômalos como problemas imunológicos e doença de Alzheimer, característicos dos pacientes.

3 – Materiais e Métodos

3.1 – Metodologia geral

Para a geração da biblioteca de SAGE, o RNA total foi obtido a partir de sangue periférico coletado de doador portador da síndrome de Down (sujeito da pesquisa), de sexo masculino e de idade de 28 anos. O referido indivíduo é matriculado na Associação de Pais e Amigos dos Excepcionais (APAE) da cidade de Passos – MG, e portador de trissomia total do CH21. Sendo portador da síndrome de Down, este indivíduo pertence a grupos vulneráveis e não possui plena autonomia, portanto a coleta de sangue foi realizada com autorização prévia dos responsáveis (pais), frente a um consentimento livre e esclarecido (ver itens 7.3 e 7.4). O sangue foi coletado em tubos contendo EDTA. Foi imediatamente realizado o isolamento do RNA total dos leucócitos com uso do “RNeasy Blood Mini Kit” (Qiagen), conforme recomendações do fabricante. Neste isolamento as hemácias foram desprezadas e os leucócitos sofreram lise, e do RNA total extraído, 10µg foram empregadas na geração da biblioteca de SAGE. Para isto foi utilizado o Kit I-SAGE (Invitrogen) conforme as instruções do fabricante (ver resumo da abordagem experimental na Fig. 4 do item 3.2).

Os cDNAs gerados (“tags”) foram clonados no vetor p-ZERO (Invitrogen) e os clones obtidos foram seqüenciados utilizando Kits de Terminadores ET-Dye (Amersham) no seqüenciador ABI-Prisma 377 (Perkin Elmer-Applied Biosystems). As seqüências foram analisadas através do “software” e-SAGE

(Margulies e Innis, 2000). A seleção de genes diferencialmente expressos para a posterior validação foi feita com base na descrição dos mesmos no banco de dados do “Uni Gene” (<http://www.ncbi.nlm.nih.gov/>). Para esta seleção os transcritos foram analisados quanto ao seu possível envolvimento com os fenótipos da SD (principalmente problemas imunológicos e doença de Alzheimer). As validações foram realizadas por “PCR semiquantitativa”. Para isso, o RNA total do sujeito da pesquisa (portador da SD) foi utilizado. As reações de controle para as validações foram realizadas com o emprego de RNA total de um indivíduo não portador de SD, também do sexo masculino e de 28 anos.

3.2 – Obtenção de biblioteca de SAGE (Resumo da abordagem experimental)

Uma visão geral da abordagem experimental da SAGE pode ser obtida na figura 4 que mostra os principais passos realizados, identificados por números:

- 1- O RNA total obtido foi usado para a síntese do cDNA correspondente, por “RT-PCR”, com o uso de oligo (dt) biotilado (Invitrogen).
- 2- Foi realizada a fixação da porção 3’ do cDNA às partículas magnéticas revestidas com Streptavidina, e a posterior clivagem do cDNA com a enzima de ancoragem *Nla* III (Invitrogen) (que reconhece o sítio CATG).

- 3- Após o descarte das porções 5' digeridas, a amostra foi dividida em duas e foi realizada a ligação de adaptadores "A" e "B" (que possuem sítios para os iniciadores "A" e "B" respectivamente, e o sítio para a enzima de etiquetagem *BsmF I* (Invitrogen)) aos cDNAs com o uso da enzima DNA ligase (Invitrogen) (ver figura 5).

- 4- Os cDNAs foram submetidos a uma nova clivagem com a enzima de etiquetagem *BsmF I* (Invitrogen), que possui sítio nos adaptadores "A" e "B", adjacentes à extremidade 5' do sítio da enzima de ancoragem *Nla III* (Fig. 5). Esta enzima de etiquetagem, por clivar 14 pb adiante de seu sítio ("down stream"), liberou e delimitou "tags" específicos de cada cDNA, que tiveram suas extremidades coesivas preenchidas pela enzima Klenow (Invitrogen).

- 5- Os cDNAs gerados (contendo os "tags" e os adaptadores) foram ligados com uso da enzima DNA ligase (Invitrogen) e formaram os "ditags" maiores (com adaptadores). Estes foram amplificados por "PCR", com uso da enzima *Platinum Taq* (Invitrogen) e dos iniciadores "A" e "B".

- 6- Após a purificação dos "ditags" maiores, os "ditags" menores foram isolados, com uso da enzima de ancoragem *Nla III*, e conseqüente retirada dos adaptadores. Eles foram então recuperados e ligados com uso da enzima DNA ligase (Invitrogen), para a formação dos "concatâmeros". A clonagem e o sequenciamento dos clones foram efetuados com uso do

plasmídeo pZERO (Invitrogen) e do seqüenciador “ABI PRISM 377 DNA Sequencer” (Perkin Elmer), respectivamente.

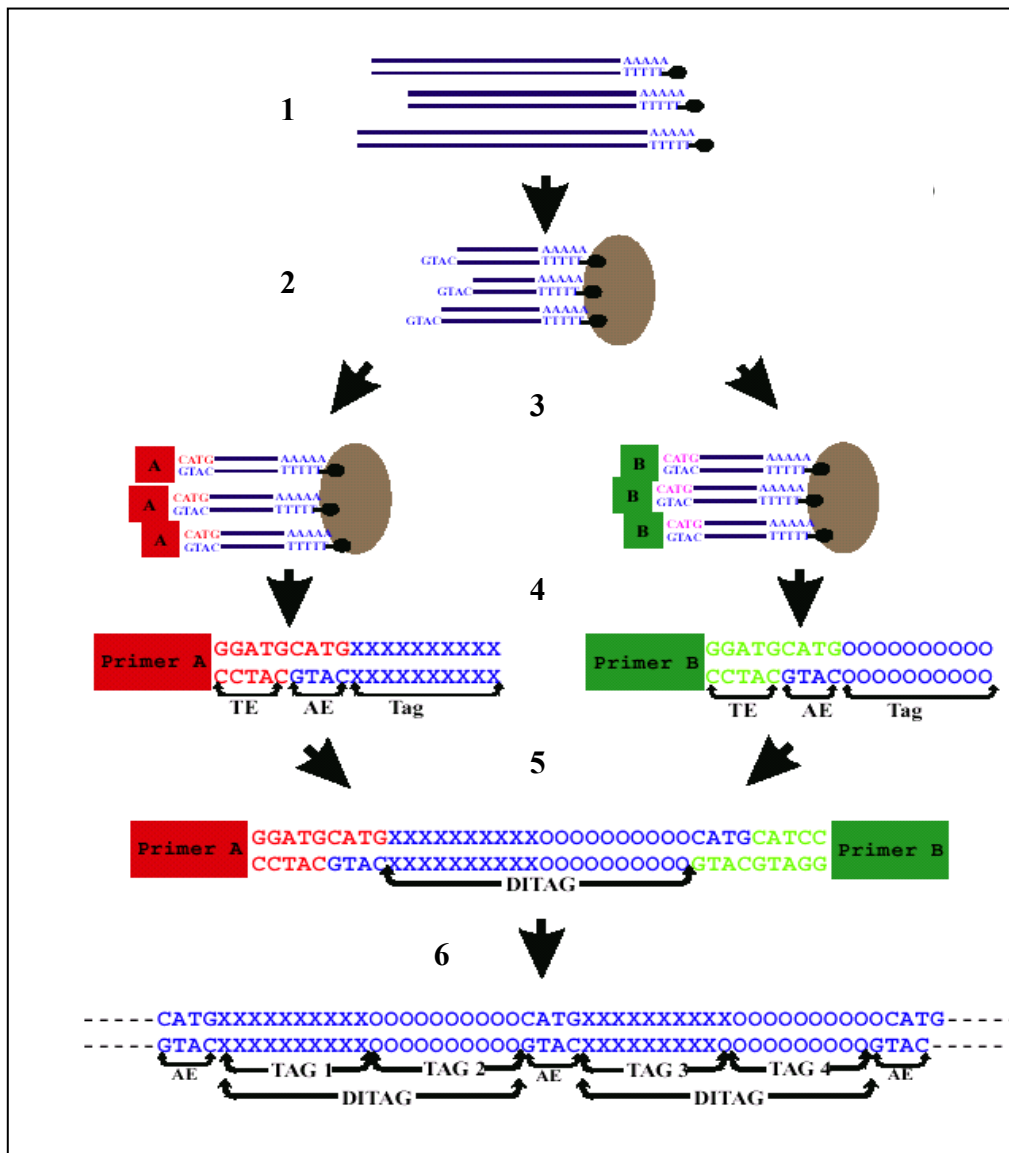


Figura 4: Esquema da abordagem experimental adotada pela técnica de SAGE. Figura modificada do protocolo original, disponível no sítio <http://www.sagenet.org/>.

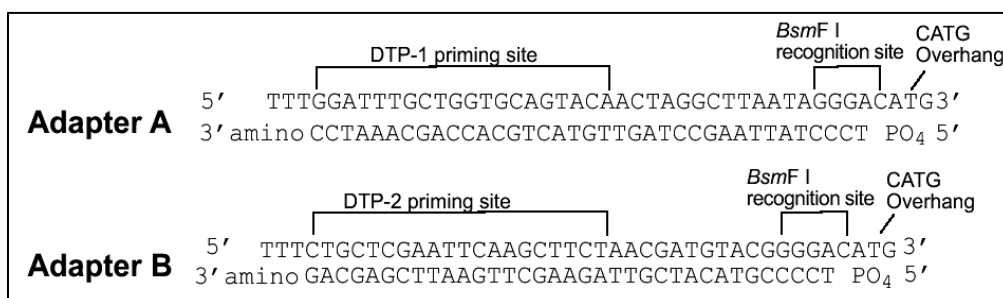


Figura 5: Representação dos adaptadores usados na técnica de SAGE. A diferença entre os adaptadores reside somente nos sítios dos iniciadores.

3.3 – Análise dos dados

Os clones da biblioteca de SAGE, de leucócitos de portador da SD foram seqüenciados até que 31.643 “tags” totais foram atingidos. Isto representou posteriormente 10.814 “tags” únicos. Em uma biblioteca de SAGE, entende-se por “tags” únicos o conjunto dos tipos de “tags” diferentes entre si; e por “tags” totais a soma das freqüências dos “tags” únicos.

Após o sequenciamento dos clones, e antes da contagem dos respectivos “tags”, houve uma seleção de seqüências (ou “tags”) de boa qualidade. Para isso, os arquivos de cromatogramas de seqüências (Fig. 6) foram convertidos para arquivos tipo “.PHD” com uso do “software” Phred Phrap versão 0.990722.g obtido do sítio <http://www.phrap.org/consed/consed.html#howToGet>. Nesta conversão os cromatogramas de seqüências foram avaliados quanto à qualidade e tiveram pontuações Phred atribuídas (“Phred score”) às bases. O “software” eSAGE (Margulies e Innis, 2000) foi utilizado para selecionar qualitativamente (baseado nas pontuações Phred), contar e catalogar as seqüências (ou “tags”). Este “software” trabalha juntamente com o “software” “Microsoft Access” do sistema

“windows”, que gera e compara tabelas de dados. Na seleção, as seqüências aceitas para contagem foram aquelas que apresentaram bases com pontuação Phred maior ou igual a 15 para os “tags” e maior ou igual a 12 para os sítios da enzima de ancoragem *Nla* III (sítios “CATG”). Após a seleção ou “filtragem” de seqüências, uma primeira tabela foi gerada contendo os “tags” (sem os sítios da enzima de ancoragem) e suas respectivas freqüências (Fig. 7).

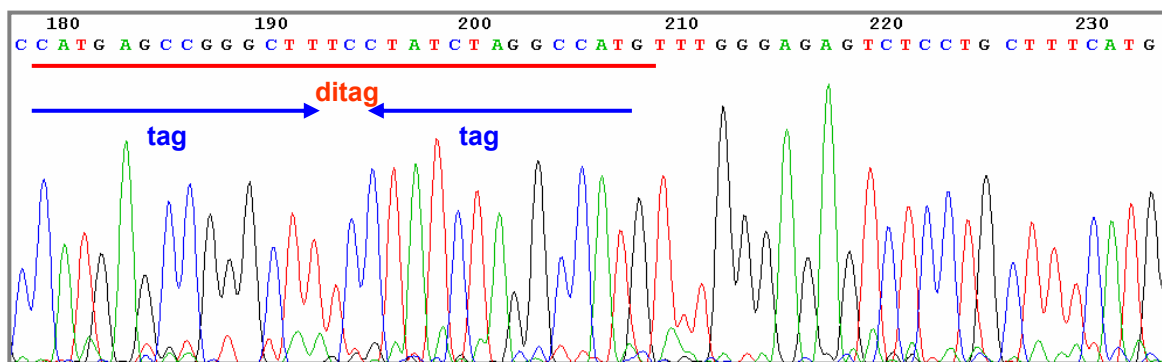


Figura 6: Trecho de cromatograma de seqüência da biblioteca de SAGE de portador de SD. A boa qualidade de sequenciamento é visualizada neste trecho devido à presença de “picos” estreitos, longos e com pouco “ruído”, referente às bases. O trecho sublinhado em vermelho indica um “ditag”; as setas azuis indicam os “tags”.

Frequency	Tag
273	TTTTTTTAGG
236	TGTA CTGGT
217	AGAGCAGAGA
203	TCCCTATTAA
193	GTTGTGGTTA
178	GGAGCGACGT
165	GATTTATGCC
119	GCCGAGCGAA
107	CCTCTGCCAG
106	TGTGTTGAGA
101	ATAATTCTTT
95	CACAAACGGT
93	CACACGCGTC
93	TGTGTATGTG

Figura 7: Parte da tabela de contagem de “tags” de portador de SD gerada pelo “software” e-SAGE. Foram catalogados os “tags” únicos (10.814) e suas respectivas freqüências.

Foram realizadas comparações entre os níveis de expressão gênica (freqüências dos “tags”) de leucócitos de portador de SD e de indivíduos não portadores. Para isto, além dos dados da biblioteca de SAGE de leucócitos de portador da SD, foram utilizados dados originados de uma biblioteca de SAGE de leucócitos de indivíduos não portadores da SD (biblioteca controle), dados estes, disponibilizados na rede de computadores pela Universidade de Duke, EUA, no sítio <ftp://ftp.ncbi.nih.gov/pub/sage/seq/>. Esta biblioteca foi construída a partir de RNAs de leucócitos de cinco mulheres adultas e não portadoras de SD.

Uma outra tabela, dos dados de não portadores, foi também gerada pelo “software” e-SAGE (Fig. 8). A “filtragem” de qualidade de seqüência não foi realizada para os dados da biblioteca controle, pois os arquivos de seqüência

disponibilizados são do tipo texto e, portanto não adequados para o uso do “software” Phred Phrap.

Frequency	Tag
1945	GCAAGAAAGT
1542	CCCAACGCGC
1109	CTTCTTGCCC
846	TTGGGGTTTC
609	GTTGTGGTTA
420	TAGGTTGTCT
407	CTAAGACTTC
402	AAAACATTCT
247	ATAATTCTTT
246	GAGGGAGTTT
234	TTGGTCCTCT
229	TTCATACACC
226	GTGGCCACGG
213	TTGGTGAAGG
212	GCATAATAGG

Figura 8: Parte da tabela de contagem de “tags” de não portadores da SD gerada pelo “software” e-SAGE. Foram catalogados os “tags” únicos (15.046) e suas respectivas frequências.

Estas duas tabelas foram então unidas por associação das colunas “tags” (coluna “chave primária”), que são comuns a ambas as tabelas, gerando uma tabela de comparação (Fig. 9) (as colunas chamadas “chave primária” são eleitas como referência para as comparações de tabelas através do “software” “Microsoft Access”). Esta nova tabela, além de apresentar as duas contagens de “tags”, correspondentes as bibliotecas Controle (SAGE1) e SD (SAGE2), possui duas pontuações (calculadas automaticamente pelo “software” e-SAGE)

que medem as diferenças de níveis de expressão encontradas entre os transcritos.

Uma medida é chamada “SAGE1toSAGE2” e realiza uma divisão normalizada entre os valores de expressão (frequência de “tags”) encontrados nos campos “SAGE1” e “SAGE2”. Como citado acima, o campo “SAGE1” representa a biblioteca controle (ou de não portadores) e o campo “SAGE2” representa a biblioteca de SD (Fig. 9). Para este cálculo os “tags” com contagem de valor zero são considerados como contagem de valor 1 (para não haver divisão por zero). A divisão efetuada considera sempre o maior valor de expressão (normalizado) como numerador e o menor como denominador, independentemente do campo onde se encontre. Se o valor de expressão é maior no campo “SAGE1”, ele ocupa o numerador da divisão e o valor anotado no campo “SAGE1toSAGE2” (resultado da divisão) é positivo. Entretanto se o valor do campo “SAGE2” é maior, este ocupa o numerador da divisão, mas o valor anotado no campo “SAGE1toSAGE2” é negativo. Este padrão facilita o pesquisador que, ao ler a tabela, percebe que se o valor citado for positivo trata-se de um transcrito mais expresso em condição normal, porém se o valor for negativo, trata-se de um transcrito mais expresso em condição de SD. Também, este valor indica quantas vezes um transcrito é mais ou menos expresso em relação a outro. Há pesquisadores que se referem a este valor como “fold”.

A normalização se faz necessária para o cálculo do “fold”, pois as bibliotecas apresentam tamanhos amostrais diferentes. A biblioteca de SD possui 10.814 “tags” únicos e 31.643 “tags” totais; a biblioteca normal possui 15.046 “tags” únicos e 48.169 “tags” totais. Sendo assim, após a normalização

(não mostrada na Fig. 9) o campo “SAGE2” tem seus valores multiplicados por aproximadamente 1,52.

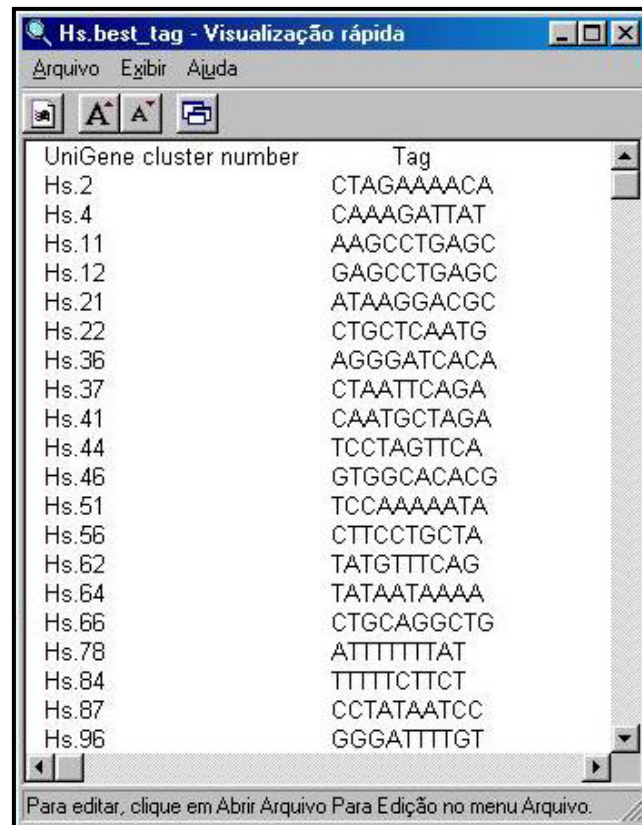
A outra medida é chamada “pVal” ou “valor de probabilidade”. Ela anota a probabilidade de a diferença de expressão encontrada entre transcritos ser creditada ao acaso. Esta medida além de ponderar a normalização supracitada, leva em consideração a contagem dos “tags” comparados. Por exemplo: se o campo “SAGE1” apresenta a contagem 10 e o campo “SAGE2” (já normalizado) apresenta a contagem 5, então o “fold” é 2. Do mesmo modo, se o campo “SAGE1” apresenta a contagem 100 e o campo “SAGE2” (já normalizado) apresenta a contagem 50, o “fold” também é 2. Neste exemplo o valor de “fold” não se altera do primeiro para o segundo caso, já o valor “pVal” se alteraria e no segundo caso ele seria menor, acentuando a condição de expressão diferencial entre os transcritos e, portanto, interpretando as contagens maiores como “mais confiáveis”.

Tag	SAGE1	SAGE2	PVal	SAGE1toSAGE2
AACCACATTG	13	0	0,0011177177	8,5
AACCACCACA	1	0	0,478562748	-1,5
AACCACCACG	1	0	0,478562748	-1,5
AACCACTACC	0	3	0,0494160269	-4,6
AACCAGAGGT	0	1	0,3143756573	-1,5
AACCAGCACC	0	1	0,3143756573	-1,5
AACCAGGAAA	0	5	0,007767598	-7,6
AACCAGGCTG	2	0	0,2888273569	1,3
AACCATAAAA	1	0	0,478562748	-1,5
AACCATAGCA	1	0	0,478562748	-1,5
AACCATCACC	1	0	0,478562748	-1,5
AACCCAGGAG	1	0	0,478562748	-1,5
AACCCAGTAA	1	0	0,478562748	-1,5
AACCCACAG	1	0	0,478562748	-1,5

Figura 9: Parte da tabela de comparação entre “tags” de portadores e não portadores de SD gerada pelo “software”e-SAGE. Observe que, por exemplo, na primeira linha o valor 8,5 é resultado da divisão: $13/1,52$ (“SAGE1”/“SAGE2”) e neste caso o valor do campo SAGE2, primeiramente interpretado como 1, após a normalização (sendo multiplicado por aproximadamente 1,52) vale 1,52. Também devido à normalização, por exemplo, na última linha o valor 1,5 é resultado da divisão: $1,52/1$ (“SAGE2”/“SAGE1”).

A tabela de comparação (Fig. 9) foi associada pela coluna de “tags” (coluna “chave primária”) a uma tabela de dados do mapeamento “CGAP SAGE Genie” (“Hs.best_gene.gz”) que foi obtida no sítio <ftp://ftp1.nci.nih.gov/pub/SAGE/>. Esta tabela possui além da coluna com os “tags”, uma outra chamada “UniGene cluster number” com os respectivos números dos clusters de “Expression Sequence Tags” (“ESTs”) do banco de dados (“UniGene” “IDs”) (Fig. 10). Estes “clusters” são conjuntos de seqüências

alinhas de “ESTs” que remontam transcritos. Assim, nesta tabela estão listados os transcritos humanos e seus “tags” representativos.



UniGene cluster number	Tag
Hs.2	CTAGAAAACA
Hs.4	CAAAGATTAT
Hs.11	AAGCCTGAGC
Hs.12	GAGCCTGAGC
Hs.21	ATAAGGACGC
Hs.22	CTGCTCAATG
Hs.36	AGGGATCACA
Hs.37	CTAATTCAGA
Hs.41	CAATGCTAGA
Hs.44	TCCTAGTTCA
Hs.46	GTGGCACACG
Hs.51	TCCAAAAATA
Hs.56	CTTCCTGCTA
Hs.62	TATGTTTCAG
Hs.64	TATAATAAAA
Hs.66	CTGCAGGCTG
Hs.78	ATTTTTTAT
Hs.84	TTTTTCTTCT
Hs.87	CCTATAATCC
Hs.96	GGGATTTTGT

Para editar, clique em Abrir Arquivo Para Edição no menu Arquivo.

Figura 10: Parte da tabela de dados do mapeamento “CGAP SAGE Genie” (“Hs.best_gene.gz”). Esta tabela lista os “tags” referentes a todos os “clusters” de humanos presentes no banco de dados do “UniGene”.

Na associação de tabelas, cada “tag” da tabela de comparação que possuía um “tag” correspondente (igual) na tabela do mapeamento, foi associado a um transcrito ou “cluster” por seu “UniGene ID” (número de identificação de “clusters” do “UniGene”). Os “tags” que não tinham correspondentes são “no matches” e não possuem “UniGene ID”, ou seja, representam transcritos não presentes no banco de dados do “UniGene”,

portanto não possuem “clusters” correspondentes e por isso são não identificados. A tabela resultante (não mostrada) ganhou então mais uma coluna: o “UniGene ID” do mapeamento.

Usando a nova coluna “UniGene ID” como “chave primária”, esta nova tabela foi então associada a outra obtida pelo serviço de “SRS” (“Sequence Retrieval System”) no site <http://srs.hgmp.mrc.ac.uk/>. Esta última continha além da coluna “UniGene ID” (“chave primária”), várias outras colunas de dados: “Description” (descrição dos genes correspondentes aos “clusters” do “UniGene”), “Gene” (Abreviação ou sigla, do gene em questão, comumente usada), “CHR” (cromossomo no qual o gene se localiza) , “Cyto” (posição cromossômica do gene), “LocusLink” (identificação disponível no sítio <http://www.ncbi.nlm.nih.gov> do “National Center of Biotechnology Information” - “NCBI” - que tenta reunir todo o tipo de informação sobre um gene, presente nas várias bases de dados existentes).

Após a incorporação de todos estes dados, outros dois campos ainda foram acrescentados: as colunas SAGE1n (não portador de SD) e SAGE2n (portadores de SD) que estimam o número de cópias por célula de cada transcrito em questão. Os valores destes campos foram calculados pela normalização de ambas as bibliotecas para um total de 300.000 “tags” (número médio de transcritos em células de mamíferos). Ou seja, o número da contagem de cada “tag” obtido em cada biblioteca é multiplicado por 300.000 e dividido pelo numero total de “tags” da sua biblioteca (48.169 “tags” totais para a biblioteca de não portadores e 31.643 para a biblioteca de portadores de SD). Uma tabela de dados final foi então gerada contendo todas as informações

citadas acima (Fig. 11). Esta tabela completa será disponibilizada no nosso sítio <http://www.ufscar.br/~dge/lbm.html>.

Tag	SAGE1	SAGE2	SAGE1n	SAGE2n	Pval	SAGE1toSAGE2	UniGene	DESCRIPTION	GENE	CHR	CYTO	LOCUSLINK
CCACTGCACACT	51	9	318	85	2,89606E-05	3,7	Hs.107003	enhancer of invasion	HEI10	14	14q11.2	57820
CTGTACAGAC	11	2	69	19	0,055290003	3,6	Hs.433615	tubulin beta 2	TUBB2			10393
AGGCTACGGGA	126	23	785	218	1,06527E-10	3,6	Hs.419535	ribosomal protein L RPL13A	RPL13A	19	19q13.3	23521; 282988
GTACTGTGGC	11	2	69	19	0,055290003	3,6	Hs.414565	chloride intracellular CLIC1	CLIC1	6	6p22.1-p21.2	1192
TGCTTGTCCC	11	2	69	19	0,055290003	3,6	Hs.286221	ADP-ribosylation fa ARF1	ARF1	1	1q42	375
TAAAGCTGTT	11	2	69	19	0,055290003	3,6	Hs.365986	ubiquitin-conjugatin UBE2B	UBE2B	5	5q23-q31	7320
CAATGTGTTA	11	2	69	19	0,055290003	3,6	Hs.74823	zinc finger protein 1 ZNF183	ZNF183		Xq25-q26	7737; 4694
GAAATGTAAG	11	2	69	19	0,055290003	3,6	Hs.132977	poly rC binding prot PCBP2	PCBP2	12	12q13.12-q13	5094; 286157
CACCTAATTG	155	28	965	265		3,6	Hs.136015	hypothetical proteir FLA0906	FLA0906	1	1p34.2	284654
AGGGTGTTTT	16	3	100	28	0,022649434	3,5	Hs.356139	heterogeneous nuc HNRPC	HNRPC	14	14q11.2	3183; 350871;
GGGTCAAAAAG	27	5	168	47	0,002927796	3,5	Hs.39911	hypothetical proteir LOC90673	LOC90673	14	14q11.2	90673
GGCTGGGGGC	48	9	299	85	8,24341E-05	3,5	Hs.408943	profilin 1	PFN1	17	17p13.3	5216
CTAAGACTTC	407	76	2535	721		3,5	Hs.462451	Homo sapiens tran				
GCTTGCAAAA	26	5	162	47	0,004160019	3,4	Hs.384944	superoxide dismutase SOD2	SOD2	6	6q25.3	6648
CGCTGGTTCC	130	25	810	237		3,4	Hs.388664	ribosomal protein L RPL11	RPL11	1	1p36.1-p35	6135
GGCAGCACAAA	5	0	31	0	0,063494577	3,3	Hs.446623	heterogeneous nuc HNRPL	HNRPL	19	19q13.2	3191; 145990
TACATTTCGTG	5	0	31	0	0,063494577	3,3	Hs.444324	protein tyrosine phs PTPRC	PTPRC	1	1q31-q32	5788

Figura 11: Parte da tabela de dados final, comparativa entre “tags” (transcritos) de leucócitos de não portadores e de portadores de SD.

3.4 – Validações

3.4.1 – Seleção de transcritos

Para validar os dados de SAGE encontrados e tentar eleger possíveis genes candidatos ao envolvimento com anomalias típicas de SD, ou de leucócitos de portadores de SD, alguns transcritos (genes) foram selecionados. Para isto, baseou-se em suas descrições e nível de expressão diferencial. Entende-se aqui por gene superexpresso, aquele que se apresenta significativamente mais expresso em portador de SD e por sub-expresso o gene que se apresenta significativamente menos expresso em portador de SD, ambos, nestas condições, quando comparados ao controle (não portadores).

Como o valor de probabilidade “pVal” analisa os dados de maneira mais eficaz, ele serviu de referencia na escolha dos transcritos. Portanto os transcritos escolhidos, que possuem um “fold” que poderia ser considerado relativamente baixo, apresentam um valor de “pVal” bastante significativo, como no caso do transcrito RPL29 (Tabela 1). Algumas informações acerca dos transcritos selecionados estão listadas na tabela 1.

Gene	Controle	SD	pVal	Fold	Descrição	UniGene	Crom.	Condição
IL8	198	0	0	130,1	Interleukin 8	624	4	Sub-exp.
RPL29	158	37	0	2,8	Ribosomal protein L29	430207	3	Sub-exp.
RPL37	177	14	0	8,3	Ribosomal protein L37	80545	5	Sub-exp.
GTF3A	0	42	1,06E-17	-63,9	General Transcription Factor III A	445977	13	Superexp.
RPL13A	126	23	1,07E-10	3,6	Ribosomal protein L13a	419535	19	Sub-exp.
CXCR4	43	0	2,95E-10	28,2	Chemokine C-X-C Motif Receptor 4	421986	2	Sub-exp.
BCL2A1	41	0	8,09E-10	26,9	BCL2- Related protein A1	227817	15	Sub-exp.

Tabela 1: Transcritos selecionados para validação. Os transcritos aparecem em ordem crescente de valor de probabilidade (“pVal”). Os campos “Controle” e “SD” representam a contagem absoluta de “tags” das bibliotecas de não portador e portador da SD respectivamente (antigos campos “SAGE1” e “SAGE2”). Com exceção do campo “Condição” (que caracteriza um transcrito como sub ou superexpresso) os demais campos já foram citados no item 3.3.

3.4.2 – Reações de validação (“RT-PCR”)

A validação se iniciou com as reações de síntese de cDNA, a partir de RNA, por ação da enzima transcriptase reversa. Cada par de reações de síntese de cDNA realizadas visou comparar os níveis de expressão de cada transcrito escolhido, entre as duas bibliotecas. Para cada par foram usados 400ng de RNA total oriundo do portador, e 400ng de RNA total do não portador de SD (controle), (indivíduos citados no item 3.1). Em cada reação do par, além dos iniciadores “anti-sense” específicos para o transcrito a validar, foi usado o iniciador “anti-sense” específico para o transcrito da enzima Glicerol Fosfato Desidrogenase (“GPDH”) como controle interno (Tabela 2). Desta maneira, para cada uma das duas amostras de RNA citadas acima, a síntese de cDNA

referente a cada transcrito escolhido foi realizada juntamente com a síntese de cDNA controle (“GPDH”). Para isto, foram adicionados 2pMoles dos iniciadores reversos (do transcrito em questão e do controle) aos dois moldes de RNA separadamente. As amostras perfizeram 15 μ l e foram incubadas a 70°C por 5 minutos. Em seguida a reação foi colocada no gelo e foram adicionados então, a cada amostra: Tampão apropriado “M-MLV Reverse Transcriptase Reaction Buffer” (Promega) com a composição: 50mM Tris-HCl (pH 8.3), 75mM KCl, 3mM MgCl₂ e 10mM DTT; “dNTPs” (Gibco BRL) a 500 μ M cada; 78 unidades de “RNA guard Rnase Inhibitor (Porcine)” (Pharmacia Biotech) e 200 unidades da enzima “M-MLV Reverse Transcriptase” (Promega). Esta nova mistura, com volume de 25 μ l foi incubada a 42°C por 60 minutos.

Ao término da síntese de cDNA, 2,5 μ l de cada reação foram usados para realizar as respectivas “PCRs” subseqüentes. Foram realizadas quatro reações de “PCR” para validar cada transcrito, ou seja, duas reações de “PCR” para cada molde de cDNA oriundo de um dos indivíduos. Isto ocorreu porque estes moldes continham cDNA do transcrito em questão e cDNA do controle de “GPDH”, que haviam sido sintetizados juntos. Desta forma, utilizando a mesma amostra como molde, uma das reações de “PCR” continha os iniciadores “sense” e “anti-sense” para o transcrito selecionado e a outra continha os iniciadores “sense” e “anti-sense” para a enzima “GPDH”. Isto para cada transcrito a validar de cada indivíduo. Todas as “PCRs” foram realizadas em volume final de 50 μ L e em micro-tubos de 0,2mL contendo, além do molde, “dNTPs” (Gibco BRL) a 200 μ M cada; Tampão apropriado “PCR buffer” (Gibco BRL) contendo 20 mM Tris-HCl (pH 8,4), 1.5 mM MgCl₂ e 50mM KCl; 25pMoles de cada iniciador; e 1 unidade da enzima “Taq DNA Polymerase”

(Gibco BRL). As reações foram iniciadas com um aquecimento de 94°C por 1 minuto, seguido de ciclos de 94°C por 15 segundos, 55°C por 30 segundos e 72°C por 5 minutos.

Para a posterior análise dos produtos de “PCR” foram retiradas alíquotas das “PCRs” dos transcritos a validar ao fim de 28, 30 e 35 ciclos. Já para as “PCRs” dos controles de “GPDH” foram retiradas amostras ao fim de 22, 25 e 28 ciclos. Desta maneira as “PCRs” tornaram-se semiquantitativas, facilitando as análises. Os produtos amplificados foram submetidos à eletroforese em gel de poliacrilamida 8% que foi posteriormente corado com nitrato de prata.

As seqüências dos iniciadores usados na validação e os tamanhos esperados dos produtos de “PCR” estão listados na Tabela 2 abaixo:

Gene	Iniciador sense	Iniciador anti-sense	“PCR” (pb)
IL8	ATGACTTCCAAGCTGGCCGTG	TTATGAATTCTCAGCCCTCTTCAA	299
RPL29	CTTTCTCTTCCGGTTCTAGG	ACAAATAGCACAGGAGGACC	630
RPL37	CTGCTATATCTTTCACCACC	TTTGTCCAGTAAGTACAGGG	579
GTF3A	GCGCCAATTACAGCAAAGCC	GACATACATCCCTTTCTGGG	596
RPL13A	GAAGGCATCAACATTTCTGG	TAAGACCCTTTCCTTGCTCC	643
CXCR4	ATGTCCATTCTTTGCCTCTTTTGC	TTAGCTGGAGTGAAAACCTGAAGAC	1070
BCL2A1	ATGACAGACTGTGAATTTGGATATATT	TCAACAGTATTGCTTCAGGAGAG	527
GPDH	TGAAGGTCGGAGTCAACGGATTTGGT	CATGTGGGCCATGAGGTCCACCAC	930

Tabela 2: Iniciadores usados nas validações.

4 – Resultados e Discussão

4.1 – Biblioteca de SAGE

Os principais passos da construção da biblioteca de SAGE de leucócitos de portador de SD foram documentados por fotografias de gel de eletroforese e serão descritos a seguir:

Após a extração do RNA total, este pôde ser observado em gel de agarose 1%. Foram visualizadas as bandas referentes aos RNAs ribossômicos 18s (1.9 Kb) e 28s (4.5 Kb) (Fig. 12).

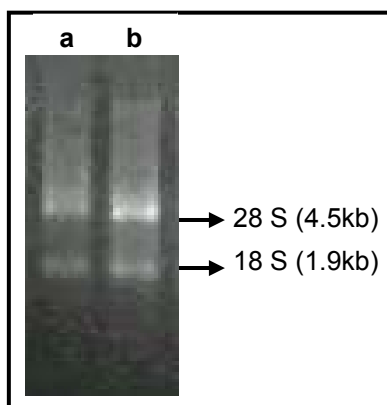


Figura 12: RNA total obtido do sujeito de pesquisa (gel de agarose 1%). As setas indicam as bandas referentes aos RNAs ribossômicos. A coluna “a” contém metade da amostra aplicada na coluna “b”.

A ocorrência destas bandas indica a qualidade adequada do RNA total que foi utilizado. Após a produção de cDNA e sua primeira digestão com a enzima de ancoragem *Nla* III; a ligação dos adaptadores “A” e “B”; a clivagem com a enzima de etiquetagem *BsmF* I e a ligação para a formação dos “ditags” maiores, estes últimos foram amplificados e a reação foi submetida à

eletroforese em gel de poliacrilamida 12%, para a purificação das bandas de 100pb. Estas bandas continham os “ditags” maiores, ou seja, o “ditag” menor mais os dois adaptadores. As bandas de 80pb observadas continham somente os dois adaptadores e foram desprezadas (Fig. 13).

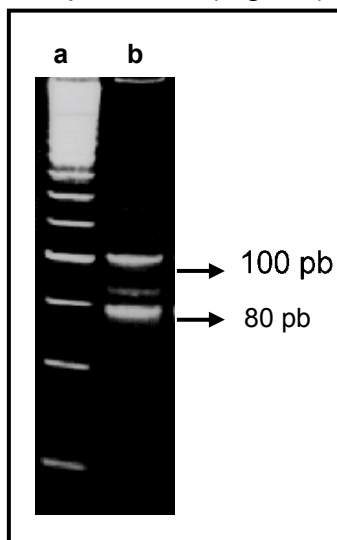


Figura 13: Amplificação dos “ditags” maiores (gel de poliacrilamida 12%).

A coluna “a” contém o marcador de peso molecular. Na coluna “b” as setas indicam: banda de 100pb (referente aos “ditags” maiores) e banda de 80pb referente aos adaptadores.

A banda de 100pb purificada teve seus “ditags” menores (26pb) extraídos e assim livres dos adaptadores. Para isso as amostras foram digeridas com a enzima de ancoragem *N/a* III e depois submetidas à eletroforese em gel de poliacrilamida 12%. Quatro bandas foram observadas: 100pb que refletem digestão parcial (“ditag” menor + dois adaptadores); 60pb também referente à digestão parcial (“ditag” menor + um adaptador); 40pb referente aos adaptadores e 26pb referente aos “ditags” menores que foram purificados posteriormente (Fig. 14).

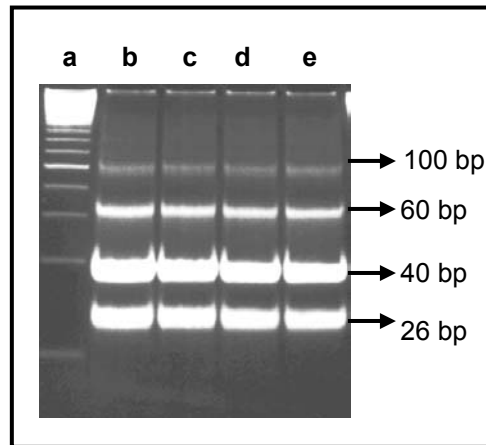


Figura 14: Extração dos “ditags” menores de 26pb (gel de poliacrilamida 12%). A coluna “a” contém o marcador de peso molecular. As demais colunas possuem as amostras apenas divididas e as setas indicam: banda de 26pb (“ditags” menores); banda de 40pb (adaptadores); bandas de 60 e 100pb (digestão parcial).

A banda de 26pb foi então isolada e os “ditags” menores foram ligados para a formação de concatâmeros. A reação de ligação foi submetida à eletroforese em gel de poliacrilamida 8% e a região contendo cDNAs de tamanho entre aproximadamente 25 a 55 “tags” (0.4 a 0.9Kb) foi purificada para clonagem no vetor pZERO (Invitrogen) (Fig. 15).

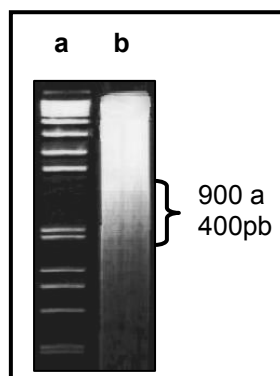


Figura 15: Concatamerização (gel de poliacrilamida 8%). A coluna “a” contém o marcador de peso molecular. Na coluna “b” a região assinalada indica as frações purificadas de concatâmeros com 25 a 55 “tags”.

Os vetores tiveram seus insertos amplificados com o uso de iniciadores específicos para o vetor pZERO (Invitrogen) (“M13 (-40) Forward” e “M13 Reverse”) para a checagem do tamanho dos concatâmeros que variaram entre 0.4 e 0.9Kb (Fig. 16).

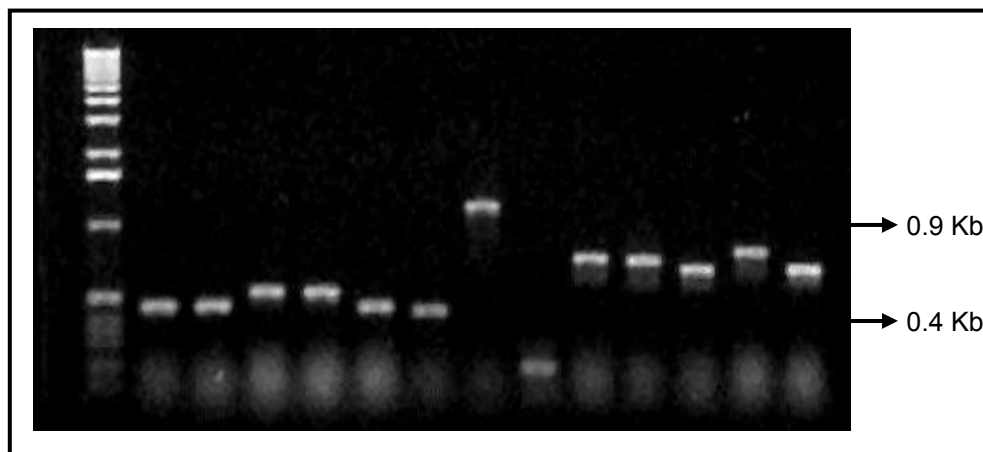


Figura 16: Amplificação dos clones de concatâmeros (gel de agarose 1%).

A primeira coluna contém o marcador de peso molecular. As demais colunas possuem clones diversos, as setas indicam aproximadamente os tamanhos verificados, que foram adequados para o sequenciamento.

4.2 – Análise dos dados

Após o sequenciamento, a contagem de “tags” da biblioteca de leucócitos de portador de SD foi de 31.643 “tags” totais e 10.814 “tags” únicos. Os números da biblioteca controle são 48.169 “tags” totais e 15.046 “tags” únicos. Naturalmente, à medida em que se avançou no sequenciamento, o número de “tags” totais aumentou proporcionalmente, ou seja, cada novo “tag” seqüenciado contava como mais um para o montante de “tags” totais. Já para

os “tags” únicos, isto não aconteceu pois os transcritos mais abundantes “coibiram” o aparecimento de transcritos mais raros, de modo que ficou cada vez mais difícil o aparecimento de algum transcrito diferente, ao se obter mais seqüências (Fig. 17).

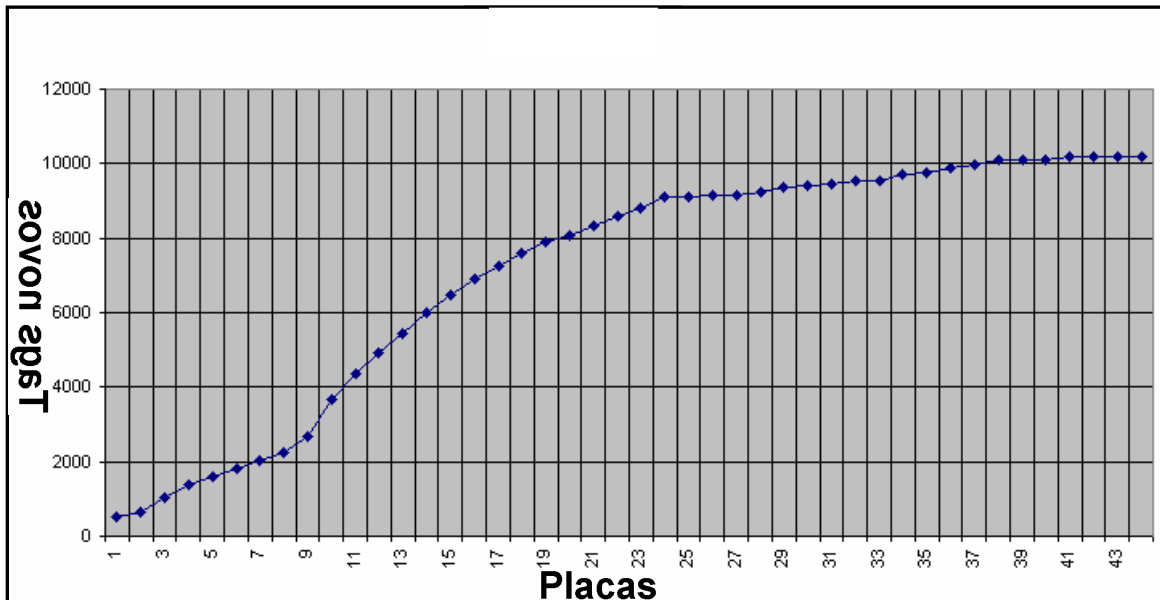


Figura 17: Gráfico de aparecimento de “tags” novos (únicos) na biblioteca de SD. Na ortogonal representa-se os “tags” únicos, na abscissa representa-se as seqüências por placa do seqüenciador “ABI PRISM 377 DNA Sequencer” (Perkin Elmer). Observa-se o aumento de “tags” únicos na medida em que se prossegue o sequenciamento, que cessou ao se atingir 10.814 “tags” únicos.

Após a normalização para a estimativa de número de cópias por célula, os “tags” únicos da biblioteca de SD se distribuíram da seguinte maneira: 0.36% apresentaram mais de 500 cópias por célula, perfazendo 14.2% da massa de mRNA; 8.86% apresentaram de 51 a 500 cópias por célula perfazendo 36.9% da massa de mRNA; 31.13% apresentaram de 11 a 50 cópias por célula perfazendo 29.5% da massa de mRNA e 59.65%

apresentaram menos de 10 cópias por célula perfazendo 19.4% da massa de mRNA (Tabela 3).

A distribuição de “tags” únicos da biblioteca controle foi: 0.42% apresentaram mais de 500 cópias por célula, perfazendo 32.2% da massa de mRNA; 3.53% apresentaram de 51 a 500 cópias por célula perfazendo 24.0% da massa de mRNA; 22.15% apresentaram de 11 a 50 cópias por célula perfazendo 21.4% da massa de mRNA e 73.9% apresentaram menos de 10 cópias por célula perfazendo 22.4% da massa de mRNA (Tabela 3). Nas duas bibliotecas a maior parte dos transcritos apresentou baixo nível de expressão. Isto fica mais evidente na biblioteca controle, o que pode ser decorrente da presença de um grande número de “tags” com contagem absoluta “1” nesta biblioteca, oriundos de possíveis erros de sequenciamento (comentado a seguir). No entanto, para ambas as bibliotecas, as categorias de nível de expressão que possuem mais de 10 cópias por célula compreendem mais de 75% da massa de mRNA (Tabela 3).

Frequência	“tags” únicos Controle			“tags” únicos SD		
	Contagem de “tags”	%	Massa de mRNA (%)	Contagem de “tags”	%	Massa de mRNA (%)
Mais de 500	63	0.42	32.2	39	0.36	14.2
51 a 500	532	3.53	24	958	8.86	36.9
11 a 50	3 332	22.15	21.4	3 366	31.13	29.5
10 ou menos	11 119	73.9	22.4	6 451	59.65	19.4
Total	15 046	100.0	100.0	10 814	100.0	100.0

Tabela 3: Distribuição de “tags” únicos em cópias por célula das duas bibliotecas de acordo com suas abundâncias e porcentagem de massa de mRNA.

Aproximadamente 45% dos “tags” únicos da biblioteca de SD e por volta de 27 % dos “tags” únicos da biblioteca controle não apresentaram descrição no banco de dados do “UniGene”, ou seja, não foram encontrados “clusters” com os respectivos genes para os “tags”. Estes “tags” são denominados “no matches” ou “tags” não identificados. Frente a esta diferença significativa encontrada entre as contagens de “tags” absolutos desconhecidos das bibliotecas, buscou-se esclarecimentos a respeito:

Uma vez que os “tags” foram selecionados quanto à qualidade (ver item 3.3), a grande contagem de “tags” não identificados da biblioteca do paciente de SD, provavelmente não aconteceu devido a erros de sequenciamento. Além disso, uma comparação entre a distribuição de “tags” não identificados das duas bibliotecas mostrou que uma fração significativa dos “no matches” de SD é representada por “tags” que apareceram duas vezes ou mais na contagem absoluta. Isto diminuiu ainda mais a possibilidade de que o grande número de “tags” não identificados, na contagem de SD, tivesse ocorrido devido a erros de sequenciamento (Fig. 18). Frente a estes dados, pode-se inferir que talvez o grupo de “tags” não identificados da biblioteca de SD, por se distribuir em classes de maior frequência, pode apresentar uma maior fração de “tags” reais. Por outro lado, a fração de “no matches” de contagem 1 (Fig. 18) da biblioteca controle pode representar, na sua maioria, “tags” irreais, frutos de erros de seqüência. Isto poderia ocorrer mesmo com a diferença aparentemente grande verificada entre os números de “no matches” das duas bibliotecas.

De fato, um grande número de “tags” únicos não identificados parece ocorrer normalmente em leucócitos. Um grupo de pesquisadores aplicou a técnica de SAGE em uma variedade de tipos de células brancas de pessoas

normais; eles reportaram que aproximadamente 47% dos “tags” únicos contados eram “no match” (Hashimoto *et al.*, 2003). Este resultado foi verificado mesmo depois da omissão dos “tags” que apareceram somente uma vez na contagem absoluta (o que evita “tags” “no matches” devido a erros de sequenciamento). Estes autores atribuíram o grande número de “no matches” encontrado a várias possibilidades: seqüências derivadas de hnRNA, erro por amplificação de “PCR”, variante de “splicing” (junção exon-intron) desconhecida, ligação inespecífica ou ao acaso de “oligo-dT” ao mRNA durante a construção da biblioteca e ocorrência de mRNA raro (Hashimoto *et al.*, 2003).

É difícil explicar se os “tags” não identificados são reais ou não, mas o potencial da SAGE de identificar novos genes (Veuculescu *et al.*, 1997) nos fez suspeitar que existe uma fração de “tags” “no match” que refletem transcritos reais e isto se aplica também, obviamente, à biblioteca construída neste trabalho. Neste contexto isto poderia ocorrer porque transcritos raros poderiam ser expressos em leucócitos de pacientes de SD. Deste modo, a supra-regulação de tais transcritos poderia ser reflexo de um “desequilíbrio” de expressão gênica nos pacientes. Assim, abre-se aqui a perspectiva de se identificar estes transcritos ainda não conhecidos, pois existem técnicas (item 1.5) que levam à identificação de transcritos correspondentes a um “tags” “no matches”, são elas: “RAST-PCR” (Van Den Berg *et al.*, 1999) e “GLGI” (Chen *et al.*, 2000). Nelas a seqüência do “tag” é usada como iniciador juntamente com iniciadores “dt”, numa reação de amplificação por “PCR” para amplificar a região “3’” do cDNA desconhecido.

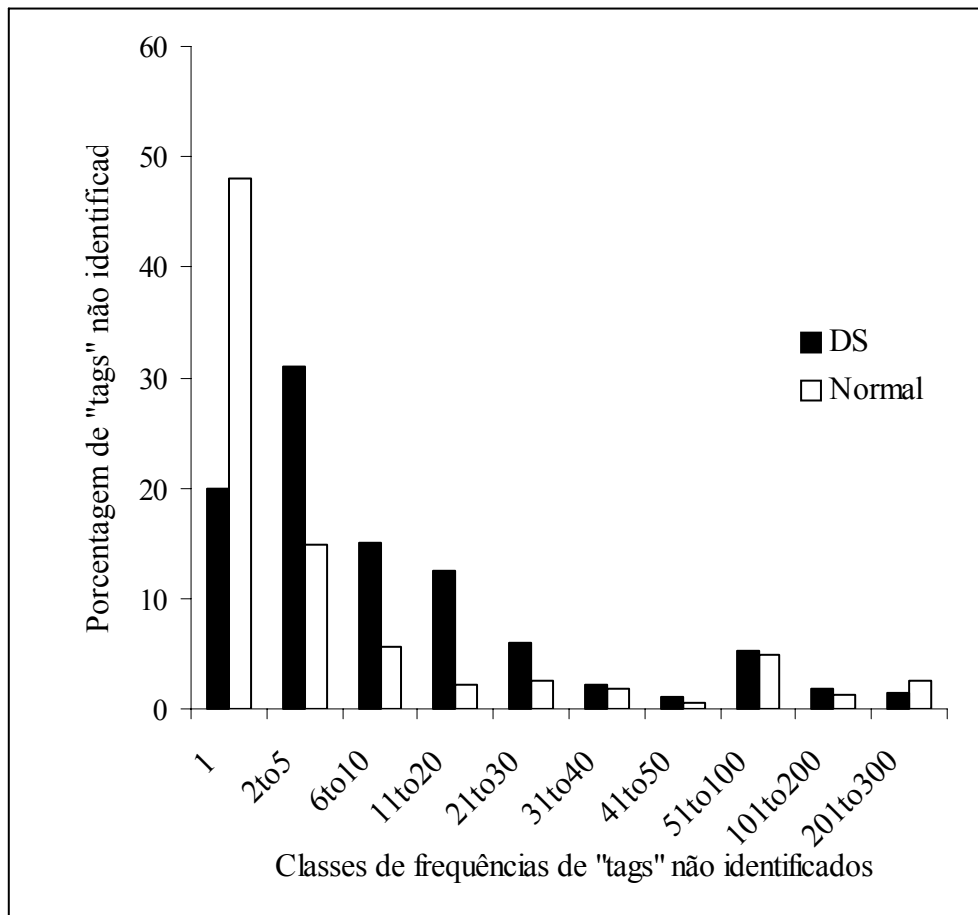


Figura 18: Distribuição de “tags” não identificados nas bibliotecas SD e Controle. As barras pretas representam as porcentagens de “tags” não identificados da biblioteca SD em cada classe de contagens absolutas de “tags”. As barras brancas representam as porcentagens de “tags” não identificados da biblioteca controle em cada classe de contagens absolutas de “tags”.

Os 40 transcritos de maior superexpressão e os 40 de maior sub-expressão, de portador de SD em relação ao controle, foram listados nas tabelas 4 e 5 respectivamente. Eles foram assim classificados conforme o “fold”, ou seja, foram catalogados conforme os valores crescentes (Tabela 4) ou decrescentes (Tabela 5) do campo “SAGE1toSAGE2”.

Apesar dos “tags” “no match” geralmente não apresentarem nível de expressão alta, a tabela 4 chama a atenção por listar um grande número de “tags” que apresentam alto nível de expressão, mas não possuem identificação no “UniGene” (já discutido acima). Já a tabela 5 revela um alto número de transcritos de eritrócito que provavelmente ocorreu devido a uma contaminação por restos de hemácias durante a construção da biblioteca controle. Este fato foi considerado nas análises e na escolha dos transcritos para serem validados e eleitos como candidatos ao envolvimento com a SD.

“Tag”	Controle	SD	Descrição
TTTTTTTAGG	0	273	M-phase phosphoprotein 1
TGTA CT TGGT	0	236	No match
AGAGCAGAGA	0	217	zinc finger protein, subfamily 1A, 4 (Eos)
GGAGCGACGT	0	178	No match
GATTTATGCC	0	165	ESTs, Weakly similar to adenylate translocator (brittle-1)-like protein [Arabidopsis thaliana] [A.thaliana]
GCCGAGCGAA	0	119	No match
CCTCTGCCAG	0	107	hypothetical protein FLJ10713
TGTGTATGTG	0	93	hypothetical protein FLJ13511
TGTGTATGTG	0	93	lethal (3) malignant brain tumor l(3)mbt protein (Drosophila) homolog
CACACGCGTC	0	93	No match
CTCAGGGACT	0	80	No match
CACGACTGTG	0	79	No match
GAACAGAAGA	0	79	acyl-Coenzyme A dehydrogenase, short/branched chain
GAACAGAAGA	0	79	EST
AAGAATCTTG	0	79	No match
AAGAGGCGCT	0	79	myosin light chain kinase 2, skeletal muscle
CACCCGCGTA	0	78	No match
TGTGTTAGCG	0	72	No match
TGGTAAACAA	0	61	No match
CCGTTCAACG	0	59	No match
ATGGCGATGG	0	58	No match
GGGCGCCGGA	0	57	paired immunoglobulin-like receptor beta
GGCGGAGCAC	0	55	No match
TGGTGTGTGG	0	52	No match
GAGGAACGAA	1	52	chorionic somatomammotropin hormone 1 (placental lactogen)
GAGGAACGAA	1	52	ninein (GSK3B interacting protein)
CGCATTGCAC	0	51	No match
ATTTGAAGCT	0	48	No match
AAAAAAGACA	0	47	hypothetical protein MGC11335
CAAACATCCA	0	46	No match
TGATGGATGC	0	44	No match
AAAATCAACA	0	42	ESTs
AAAATCAACA	0	42	general transcription factor IIIA
TACGCCTGCC	0	41	No match
CTTCGGCTTT	0	40	No match
CGCAAAAACA	0	39	No match
TGTCGTGGAG	0	38	ribosomal protein L4
CCACGCAGAG	0	37	No match
CAGAGAGAGA	0	37	No match
TGGAGGACGA	0	36	phosphoglycerate kinase1

Tabela 4: Os 40 transcritos de maior superexpressão em leucócitos de portador de SD. Os transcritos grafados em vermelho foram validados.

“Tag”	Controle	SD	Descrição
GCAAGAAAGT	1945	1	hemoglobin, beta
CCCAACGCGC	1542	0	hemoglobin, alpha 2
CTTCTTGCCC	1109	1	hemoglobin, alpha 2
TGGAAGCACT	198	0	interleukin 8
TCGTCCACGT	169	0	hemoglobin, alpha 2
TCTCCATACC	140	0	No match
GTGGGTTCTC	124	0	hemoglobin, alpha 2
GTGCGCTGAG	98	1	major histocompatibility complex, class I, A
GTGCGCTGAG	98	1	major histocompatibility complex, class I, C
AGCCTTCACC	88	0	hemoglobin, delta
AGCCTTCACC	88	0	hemoglobin, beta
TAACAGCCAG	86	0	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
TAACAGCCAG	86	0	hypothetical protein FLJ14075
CCCTGGGTTT	164	2	ferritin, light polypeptide
CCCTGGGTTT	164	2	ESTs, Weakly similar to S71880 ferritin 1 precursor
TCAAGAAAGT	81	0	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1
TCCAACGCGC	76	0	No match
GCCAACCTCC	74	0	ADG-90 protein
GCTGAACGCG	73	1	CCAAT/enhancer binding protein (C/EBP), beta
GACAAAAAAA	69	1	ESTs, Moderately similar to RS1A ARATH 40S ribosomal protein S15A [A.thaliana]
GACAAAAAAA	69	1	ribosomal protein S15a
GTGCTGTCTC	62	0	hemoglobin, alpha 2
GTGCTGTCTC	62	0	CASP2 and RIPK1 domain containing adaptor with death domain
GTGTCTGTTT	57	0	ESTs
GTGTCTGTTT	57	0	Homo sapiens cDNA FLJ31102 fis, clone IMR322000010
TTTGTTAAAA	53	1	HIF-1 responsive RTP801
TTTGTTAAAA	53	1	EST
ATGGTGGGGG	51	0	zinc finger protein 36, C3H type, homolog (mouse)
CAATAAACTG	49	0	putative translation initiation factor
CAATAAACTG	49	0	ESTs
TTAAACTTAA	43	0	chemokine (C-X-C motif), receptor 4 (fusin)
AATGAACAAT	42	1	ninjurin 1
TAATGAATAA	41	0	BCL2-related protein A1
ACCCACGTCA	41	0	jun B proto-oncogene
ACCCACGTCA	41	0	potassium voltage-gated channel, shaker-related subfamily beta member 2
GTAAGAAAGT	39	0	hemoglobin, beta
GTAAGAAAGT	39	0	KIAA0077 protein
TTTAACGGCC	154	4	No match
AGAAATAAAG	38	1	adenosine monophosphate deaminase 2 (isoform L)
CTTGACATAC	37	0	dual specificity phosphatase 1

Tabela 5: Os 40 transcritos de maior sub-expressão em leucócitos de portador de SD. Os transcritos grafados em vermelho foram validados.

Se a contagem total de transcritos de eritrócitos (Tabela 5) diluiu a contagem total dos transcritos restantes, então uma determinada contagem de um “tag” da biblioteca controle pode ser considerada como sub representada. Isto equivale a dizer que ao observar tal contagem imagina-se que ela seria maior, não fosse a contaminação citada. Desta forma uma expressão diferencial encontrada entre um transcrito presente nas duas bibliotecas pode apresentar maior ou menor significância. Em outras palavras, um dado transcrito sub-expresso no paciente de SD pode ter esta condição mais acentuada devido a provável contagem maior de seu correspondente da biblioteca controle. Porém, ao classificar um transcrito como superexpresso no paciente de SD, sua condição de expresso diferencialmente terá significância menor pelo mesmo motivo. O transcrito superexpresso GTF3A (Tabelas 1, 2 e 4) foi escolhido por ter sido considerado como não sujeito à perda de significância na sua expressão diferencial, devido a sua contagem alta no portador de SD.

Uma segunda classificação foi realizada gerando uma tabela de dados com uma listagem de valores de probabilidade crescentes. Isto fez com que os transcritos de maior expressão diferencial (super ou sub-expressos) fossem observados no “topo” da tabela. Foi possível então comparar o padrão de expressão gênica diferencial entre as duas bibliotecas, ou seja, foi possível checar quais tipos de transcritos se mostraram seletivamente mais expressos em cada biblioteca. Para isto foram selecionados e classificados quanto à função os primeiros 239 transcritos mais diferencialmente expressos da lista. Dentre eles, 120 se apresentaram mais expressos na biblioteca controle e 119 se apresentaram mais expressos na biblioteca de SD. As classes de funções

(catorze categorias – Fig. 19) foram então analisadas. Genes envolvidos com sistema imune, metabolismo e proteínas ribossômicas se mostraram preferencialmente expressos na biblioteca controle. Já na biblioteca de paciente de SD os “tags” não identificados se mostraram preferencialmente expressos (Fig. 19). Este dado reflete, em parte, a significativa quantidade de “tags” não identificados na referida biblioteca (já discutido acima).

Ao analisar as contagens de classificações da biblioteca controle, pode-se inferir que existe um certo padrão de níveis baixos de expressão de genes relacionados ao metabolismo e sistema imune em portadores de SD. Isto pode refletir nos fenótipos anômalos dos pacientes, já que estes apresentam problemas que envolvem estas classes de funções de genes. Desta forma, estudos futuros de tais transcritos podem ser promissores. Os baixos níveis de expressão de transcritos ribossômicos em pacientes de SD serão discutidos no item 4.3.2.

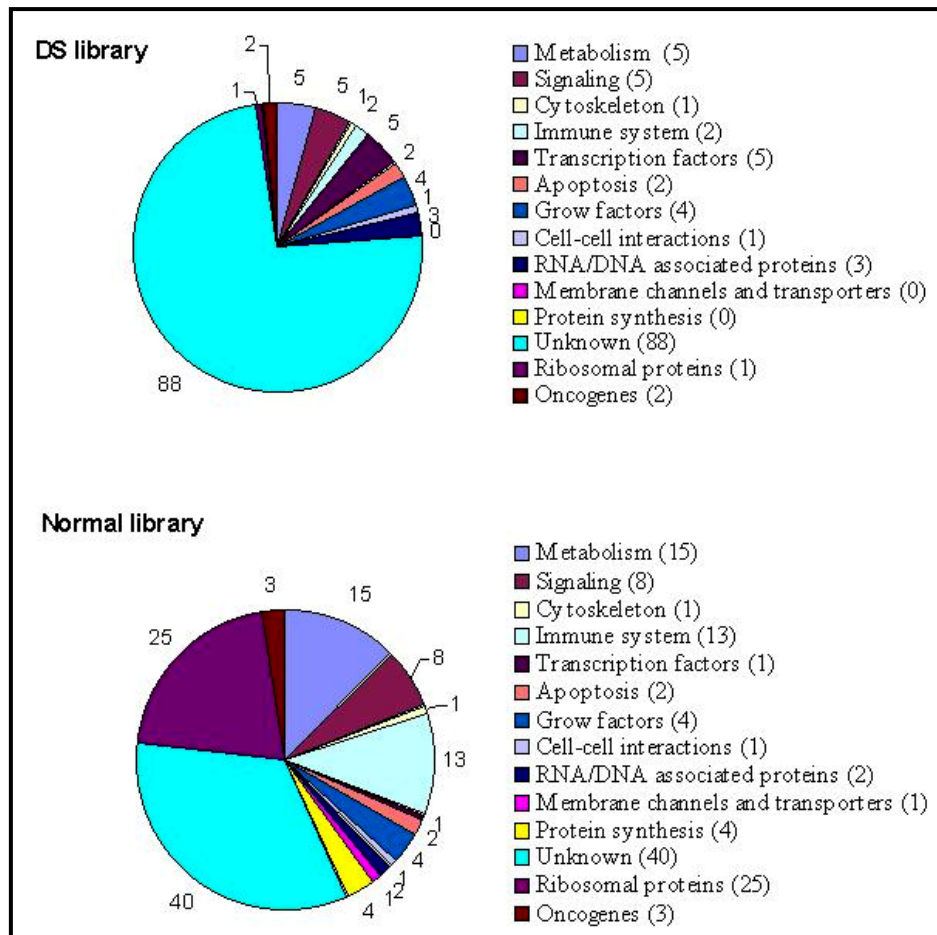


Figura 19: Classificação funcional dos transcritos mais diferencialmente expressos nas bibliotecas controle e de paciente de SD. O número próximo a cada categoria indica o número total de genes da classe.

4.3 – Validações da SAGE

4.3.1 – Seleção de transcritos

Para a validação da técnica de SAGE, há autores que aleatoriamente “pinçam” alguns transcritos com expressão diferencial, normalmente com “fold” maior que 10 e com isto confirmam o estabelecimento dos perfis de expressão gênica.

No presente trabalho, foram escolhidos transcritos (Tabela 1 – item 3.4.1) que além de validar a SAGE, talvez pudessem estar envolvidos com os fenótipos da SD no contexto celular de leucócitos. A primeira questão/hipótese que foi levada em consideração foi a eventual possibilidade de o cromossomo 21 possuir genes diferencialmente expressos que estariam envolvidos diretamente na patogênese da SD. Porém os genes deste cromossomo não apresentaram diferenças significativas entre os níveis de expressão, que justificassem sua escolha (Tabela 6). Além disto, os genes do CH21 são comumente pouco expressos, pois eles são em geral, fatores de regulação e envolvidos com sinalização (Fig. 20). Portanto, as diferenças entre os perfis de expressão gênica entre portador e não portadores de SD são mais evidentes entre os genes de outros cromossomos. Isto não significa afirmar que não haja diferenças de expressão dos genes do CH21 entre pacientes e não pacientes de SD, e muito menos dizer que elas não estejam envolvidas com a SD. Pelo contrário este fato reforça a hipótese de que os genes deste cromossomo agem na patogênese da SD indiretamente via outros genes e produtos gênicos e afetando o genoma, transcriptoma e proteoma.

Gene	Controle	SD	pVal	Fold	Descrição	UniGene	Crom.
C21orf7	15	0	0,0004	9,9	chromosome 21 open reading frame 7	Hs.41267	21
NDUFV3	0	6	0,0031	-9,1	NADH dehydrogenase ubiquinone flavoprotein 3 10kDa	Hs.199471	21
OLIG2	0	6	0,0031	-9,1	oligodendrocyte lineage transcription factor 2	Hs.176977	21
COL6A2	0	5	0,0078	-7,6	collagen type VI alpha 2	Hs.420269	21
C21orf18	0	5	0,0078	-7,6	chromosome 21 open reading frame 18	Hs.446259	21
CNN2	9	0	0,0084	5,9	calponin 2	Hs.169718	21
U2AF1	8	0	0,0140	5,3	U2 RNU2 small nuclear RNA auxillary factor 1	Hs.365116	21
PTTG1P	0	4	0,0196	-6,1	pituitary tumor-transforming 1 interacting protein	Hs.369026	21
ITGB2	10	1	0,0273	6,6	integrin beta 2 antigen CD18 p95 lymphocyte function-associated 1 macrophage mac-1 subunit	Hs.375957	21
C21orf80	0	3	0,0494	-4,6	chromosome 21 open reading frame 80	Hs.22982	21

Tabela 6: Lista dos 10 genes do cromossomo 21 de maior expressão diferencial entre as duas bibliotecas (controle e SD). Os transcritos aparecem em ordem crescente de valor de probabilidade (“pVal”). Os “Folds” são relativamente baixos e, principalmente, os valores de probabilidade são altos, se comparados aos valores dos transcritos da tabela 1, selecionados para validação.



Figura 20: Classificação dos genes do CH21 pela função molecular. A tabela foi gerada utilizando a coluna “UniGene” como “chave primária”, através do sítio <http://fatigo.bioinfo.cnio.es/>. Os genes classificados possuíam, no mínimo, a contagem de “tags” 1 em pelo menos uma das bibliotecas de SAGE analisadas neste trabalho.

4.3.2 – Validações

Ao serem realizadas as validações por “RT-PCR” semiquantitativa, os dados de comparação entre as bibliotecas de SAGE (controle e SD) referentes aos transcritos escolhidos (Tabela 1) puderam ser confirmados (Fig. 21).

A alta suscetibilidade a infecções apresentada por portadores de SD influenciou na escolha dos transcritos IL8, CXCR4 e BCL2A1 para validação.

As interleucinas são peptídeos e proteínas secretados que mediam interações locais entre leucócitos. Geralmente estas mediadoras possuem várias fontes, alvos e modos de ação (Alberts *et al.*, 1994). Macrófagos secretam a interleucina-8 que tem capacidade de atrair neutrófilos presentes na circulação. A atividade de neutrófilos poderia ser comprometida frente à baixa expressão desta interleucina em portadores de SD, confirmada nos experimentos de validação (Fig. 21).

As quimiocinas são peptídeos pequenos com potente atividade de atração química de leucócitos (Raport *et al.*, 1996.). Elas são produzidas por células endoteliais de tecidos inflamados e órgãos linfóides e junto com seus receptores apresentam importante papel na migração sítio-dirigida de resposta inflamatória e na migração para ativação de leucócitos, respectivamente (Salentin *et al.*, 2003). O CXCR4 é um dos principais receptores de quimiocina que provêem sinais para as células B entrarem no linfonodo (Okada *et al.*, 2002). A sub-expressão deste receptor em pacientes com SD, verificada na validação (Fig. 21), poderia ser responsável por uma possível não migração e não ativação de linfócitos B, o que afetaria o funcionamento do sistema imune.

As proteínas da família BCL2 são reguladoras fundamentais de resposta de apoptose. A expressão do gene pro-sobrevivência Bcl-2 homólogo Bfl-1/A1, ou BCL2A1 (alvo transcricional direto do fator nuclear NF-kappaB), apresenta função de proteção antiapoptose no sistema imune (Zong *et al.*, 1999). Um grupo de pesquisadores demonstrou que, em camundongos, a expressão do BCL2A1 é crítica para o controle de sobrevivência de células “B” e “T” (Grumont *et al.*, 1999). Neste contexto, a condição de sub-expressão deste transcrito validado (fig. 21) sugere seu envolvimento no impedimento da proliferação de linfócitos “T”, característico de portadores de SD (Ugazio, 1981).

O fator de transcrição 3A (GTF3A) é uma proteína ligante de ácidos nucléicos, possui motivos de ligação tipo “Zinc Finger” e provavelmente se liga a RNAs “5S” (Drew *et al.*, 1995). Por ser um fator de transcrição, talvez um excesso de GTF3A em pacientes de SD confirmado pela validação (Fig. 21) ou até mesmo um aumento de RNAs “5S” possa de alguma maneira intervir de maneira nociva em mecanismos bioquímicos de regulação, por influência direta ou indireta.

Um grupo de pesquisadores encontrou uma variedade de transcritos de proteínas ribossômicas que mostraram condição de sub-expressão em cérebros de camundongos portadores de SD (Chrast *et al.*, 2000). No presente trabalho, além de alguns transcritos ribossômicos terem sido selecionados, como sub-expressos, para validação (Tabela 1 e Fig 21), os dados mostrados na figura 19 revelam um padrão de expressão baixa destes transcritos em leucócitos de paciente de SD. Apesar de se haver estudado aqui leucócitos de humanos, não se pode descartar a hipótese da existência de um padrão de

expressão baixa de transcritos ribossômicos, talvez nos mais diversos tecidos ou tipos celulares de portadores de SD. De fato, as validações dos transcritos RPL29, RPL37 e RPL13A confirmam suas condições de sub-expressão (Fig. 21).

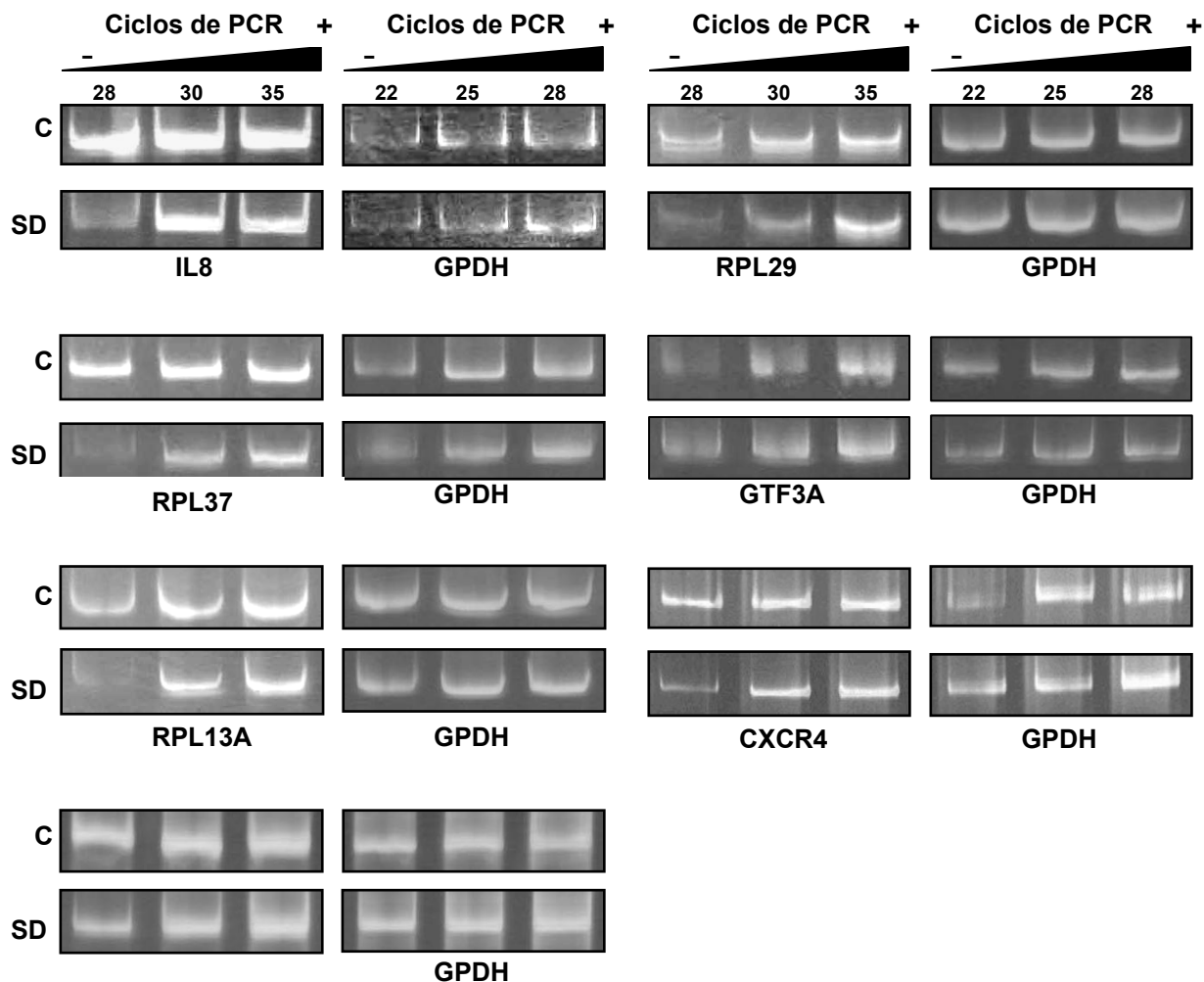


Figura 21: Fotos de gel de poliacrilamida 8% referentes às validações dos transcritos selecionados por “RT-PCR” semiquantitativa. Os transcritos selecionados e seus respectivos controles de GPDH estão identificados abaixo de cada grupo de fotos (4 fotos) de validação. Os números de ciclos em que amostras de “PCR” foram retiradas estão indicados na parte superior. **C**= “RT-PCR” do indivíduo não portador de SD (controle). **SD** = “RT-PCR” do indivíduo portador de SD.

5 – Conclusão

A técnica de SAGE foi implantada e validada com sucesso, marcando o início dos estudos envolvendo genômica funcional, em nosso laboratório.

O perfil do transcriptoma de leucócitos de portadores de SD, quando comparado ao controle, revelou sub-expressão de transcritos ribossômicos, transcritos envolvidos com metabolismo e com o sistema imune.

A expressão diferencial dos transcritos RPL13A, RPL29, RPL37, GTF3A, IL8, BCL2A1 e CXCR4 foi validada através da técnica de RT-PCR semiquantitativa.

Há dados disponibilizados para se buscar por transcritos ainda não identificados que seriam expressos em leucócitos de pacientes portadores de SD.

O trabalho realizado gerou novos dados sobre diferenças transcricionais entre portadores e não portadores de SD, abrindo perspectivas para estudos futuros que visem elucidar os mecanismos moleculares da patogênese da SD.

No que se refere a estudos de análise em massa de expressão gênica, neste trabalho, foi reportado pela primeira vez um perfil de transcriptoma de humano portador de SD.

6 – Referências Bibliográficas

- ADAMS MD, KELLEY JM, GOCAYNE JD, DUBNICK M, POLYMEROPOULOS MH, XIAO H, MERRIL CR, WU A, OLDE B, MORENO RF, KERLAVAGE AR, MCCOMBIE WR, VENTER JC. Complementary DNA- sequencing – Expressed sequence tags and human genome project. **SCIENCE** 252: (5013) 1651-1656 JUN 21 1991.
- ALBERTS B, BRAY D, LEWIS J, RAFF M, ROBERTS K, WATSON J D. Molecular Biology of the Cell 3rd ed. **NEW YORK AND LONDON: GARLAND PUBLISHING;** c1994.
- ANTONARAKIS SE. 10 years of *Genomics*, Chromosome 21, and Down Syndrome. **GENOMICS** 51 1 - 16 1998.
- ANTONARAKIS SE. Chromosome 21: from sequence to applications. **CURRENT OPINION IN GENETICS & DEVELOPMENT** 11: (3) 241-246 JUN 2001.
- BERTELSEN AH, VELCULESCU VE. High-throughput gene expression analysis using SAGE. **DRUG DISCOVERY TODAY** 3: (4) 152-159 APR 1998.
- BLOUIN JL, DURIAUX-SAIL G, CHEN H, GOS A, MORRIS MA, ROSSIER C, ANTONARAKIS SE. Mapping of the gene for the p60 subunit of the human chromatin assembly factor (CAF1A) to the Down syndrome region of chromosome 21. **GENOMICS**. 15 ;33(2):309-12 Apr 1996.
- CAPONE GT. Down syndrome: advances in molecular biology and the neurosciences. **J DEV BEHAV PEDIATR** 22 (1): 40-59. Feb 2001.
- CARON H, VAN SCHAIK B, VAN DER MEE M, BAAS F, RIGGINS G, VAN SLUIS P, HERMUS MC, VAN ASPEREN R, BOON K, VOUTE PA, HEISTERKAMP S, VAN KAMPEN A, VERSTEEG R. The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. **SCIENCE** 291: (5507) 1289+ FEB 16 2001.
- CHEN JJ, ROWLEY JD, WANG SM. Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. **PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA** 97: (1) 349-353 JAN 4 2000.
- CHRAST R, SCOTT HS, PAPASAVVAS MP, ROSSIER C, ANTONARAKIS ES, BARRAS C, DAVISSON MT, SCHMIDT C, ESTIVILL X, DIERSSEN M, PRITCHARD M,

ANTONARAKIS SE. The mouse brain transcriptome by SAGE: Differences in gene expression between P30 brains of the partial trisomy 16 mouse model of Down syndrome (Ts65Dn) and normals. **GENOME RESEARCH** 10: (12) 2006-2021 DEC 2000.

- CHRISTIANSON AL. Down syndrome in sub-Saharan Africa. **JOURNAL OF MEDICAL GENETICS** 33: (2) 89-92 FEB 1996.
- DELABAR JM, THEOPHILE DIZ, CHETTOUH Z, BLOUIN JL, PRIEUR M, NOEL B, SINET PM. Molecular mapping of 24 features of Down Syndrome on chromosome 21. **EUROPE JOURNAL OF HUMAN GENETICS** 1: 114-124 1993.
- DIATCHENKO L, LAU YFC, CAMPBELL AP, CHENCHIK A, MOQADAM F, HUANG B, LUKYANOV S, LUKYANOV K, GURSKAYA N, SVERDLOV ED, SIEBERT PD. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. **PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA** 93: (12) 6025-6030 JUN 11 1996.
- DOWN JL: Observations on an ethnic classification of idiots. 1866. **MENT RETARD** 33: 54-56 1995.
- DREW PD, NAGLE JW, CANNING RD, OZATO K, BIDDISON WE, BECKER KG. Cloning and expression analysis of a human cDNA homologous to *Xenopus TFIIIA*. **GENE**, 159, 215-218 1995.
- EPSTEIN C, Down syndrome (trisomy 21). In "The metabolic and molecular bases of inherited disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly and D. Valle Eds.), pp. 749-794, **MC GRAW HILL, NEW YORK** 1995.
- EPSTEIN C, HASSOLD T, LOTT I, NADEL L, PATTERSON D. Etiology and Pathogenesis of Down Syndrome. **WILEY LISS, NEW YORK** 1996.
- EPSTEIN C, KORENBERG J, ANNERÉN G, ANTONARAKIS S, COURCHESNE E, EPSTEIN I, FOWLER A, GRONER Y, HURET J, KEMPER T, LOTT I, LUBIN B, MAGENIS E, OPITZ E, PATTERSON D, PRIEST J, PUESCHEL S, RAPOPORT S, SINET PM, TANZI R, DE LA CRUZ F. Protocols to Establish Genotype-Fenotype Correlations in Down Syndrome. **AMERICAN JOURNAL OF HUMAN GENETICS** 49, 207-235 1991.

- EPSTEIN C. The Consequences of Chromosome Imbalance: Principles, Mechanisms and Models. **CAMBRIDGE UNIVERSITY PRESS, NEW YORK** 1986.
- FONG CT, BRODEUR GM. Down's syndrome and leukemia: epidemiology, genetics, cytogenetics and mechanisms of leukemogenesis. **CANCER GENETICS AND CYTOGENETICS**. 28: 55–76. 1987.
- FUENTES JJ, PRITCHARD MA, ESTIVILL X. Genomic Organization, Alternative Splicing, and Expression Patterns of the *dscr1* (Down Syndrome Candidate Region 1) Gene. **GENOMICS** 44, n.3, 358-361 1997.
- FUENTES JJ, PRITCHARD MA, PLANAS AA, BOSCH A, FERRER I, ESTIVILL X. A New Human Gene from the Down Syndrome Critical Region Encodes a Proline-Rich Protein highly Expressed in Fetal Brain and Heart. **HUMAN MOLECULAR GENETICS** 4, 1935-1944 1995
- GOSSET P, AIT-GHEZALA G, SINET PM, CREAU N. Isolation and analysis of chromosome 21 genes potentially involved in Down Syndrome. **JOURNAL OF NEURAL TRANSMISSION-SUPPLEMENT** (57) 197-209 1999.
- GRUMONT RJ, ROURKE IJ, GERONDAKIS S. Rel-dependent induction of A1 transcription is required to protect B cells from antigen receptor ligation-induced apoptosis. **GENES DEVELOPMENT**. 13, 400–411 1999.
- GUIMERA J, PRITCHARD M, NADAL M, ESTIVILL X. Minibrain (MNBH) is a single copy gene mapping to human chromosome 21q22.2. **CYTOGENET CELL GENET** 77(3-4) :182-4 1997.
- GYGI SP, ROCHON Y, FRANZA BR, AEBERSOLD R. Correlation between protein and mRNA abundance in yeast. **MOLECULAR AND CELLULAR BIOLOGY** 19: (3) 1720-1730 MAR 1999.
- HASHIMOTO S, NAGAI S, SESE J, SUZUKI T, OBATA A, SATO T, TOYODA N, DONG HY, KURACHI M, NAGAHATA T, SHIZUNO K, MORISHITA S, MATSUSHIMA, K. Gene expression profiles in human leukocytes. **BLOOD** 101(9), 3509-3513 2003.
- HASSOLD T, JACOBS P. Trissomy in Man. **ANNU. REV. GENET.** v. 18, p. 69-97 1984.
- HATTORI M, FUJIYAMA A, TAYLOR TD, WATANABE H, YADA T, PARK HS, TOYODA A, ISHII K, TOTOKI Y, CHOI DK, SOEDA E, OHKI M, TAKAGI T, SAKAKI Y,

TAUDIEN S, BLECHSCHMIDT K, POLLEY A, MENZEL U, DELABAR J, KUMPF K, LEHMANN R, PATTERSON D, REICHWALD K, RUMP A, SCHILLHABEL M, SCHUDY A, ZIMMERMANN W, ROSENTHAL A, KUDOH J, SHIBUYA K, KAWASAKI K, ASAKAWA S, SHINTANI A, SASAKI T, NAGAMINE K, MITSUYAMA S, ANTONARAKIS SE, MINOSHIMA S, SHIMIZU N, NORDSIEK G, HORNISCHER K, BRANDT P, SCHARFE M, SCHON O, DESARIO A, REICHELT J, KAUER G, BLOCKER H, RAMSER J, BECK A, KLAGES S, HENNIG S, RIESSELMANN L, DAGAND E, HAAF T, WEHRMEYER S, BORZYM K, GARDINER K, NIZETIC D, FRANCIS F, LEHRACH H, REINHARDT R, YASPO ML. The DNA sequence of human chromosome 21. **NATURE** 405: (6784) 311-319 MAY 18 2000.

- HIGUCHI R, FOCKLER C, DOLLINGER G, WATSON R. Kinetic PCR Analysis - Real-Time Monitoring of DNA amplification reactions. **BIO-TECHNOLOGY** 11: (9) 1026-1030 SEP 1993.
- HOOK EB, CROSS PK, SCHREINEMACHERS DM: Chromosomal abnormality rates at amniocentesis and in live-born infants. **JAMA**. 249: 2034-2038 1983.
- HUBANK M, SCHATZ DG. Identifying differences in messenger-RNA expression by representational difference analysis of cDNA. **NUCLEIC ACIDS RESEARCH** 22: (25) 5640-5648 DEC 25 1994.
- JIANG J, JACOB HJ. EbEST: An automated tool using expressed sequence tags to delineate gene structure. **GENOME RESEARCH** 8: (3) 268-275 MAR 1998.
- KANG JJ, WATSON RM, FISHER ME, HIGUCHI R, GELFAND DH, HOLLAND MJ. Transcript quantitation in total yeast cellular RNA using kinetic PCR. **NUCLEIC ACIDS RESEARCH**. 15;28(2):e2. Jan 2000.
- KORENBERG J, CHEN XN, SCHIPPER R, SUN Z, GONSKY R, GERWEHR S, CARPENTER N, DAUMER C, DIGNAN P, DISTECHE C, GRAHAM JMJ, HUGDINS L, MCGILLIVRAY B, MIYAZAKI K, OGASAWARA N, PARK JP, PAGON R, PUESCHEL S, SACK G, SAY B, SCHUFFENAUHER S, SOUKUP S, YAMANAKA T. The Consequence of Chromosome Imbalance. **PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA**. 91, 4997-5001 1994.

- KORENBERG J, PULST S, KAWASHIMA H, IKEUCHI T, YAMAMOTO K, OGASAWARA N, SCHONBERG S, WEST R, KOJIS T, EPSTEIN C. Familial Down Syndrome with Normal Karyotype: Definition of the Region. **AMERICAN JOURNAL OF HUMAN GENETICS** 43, A 110 1988.
- KORENBERG JR, BRADLEY C, DISTECHE CM. Down Syndrome: Molecular Mapping of the Congenital Heart Disease and Duodenal Stenosis. **AMERICAN JOURNAL OF HUMAN GENETICS** 50, 294-302 1992.
- LEJEUNE J, TURPIN R, GAUTIER M: Le Mongolisme. Premier Exemple Daberration Autosomique Humanie. **ANNALES DE GENETIQUE** 1: (2) 41-49 1959.
- LE NAOUR F, HOHENKIRK L, GROLLEAU A, MISEK DE, LESCURE P, GEIGER JD, HANASH S, BERETTA L. Profiling changes in gene expression during differentiation and maturation of monocyte-derived dendritic cells using both oligonucleotide microarrays and proteomics. **J BIOL CHEM.** 25;276(21):17920-31 May 2001.
- LIANG P, PARDEE AB. DIFFERENTIAL DISPLAY OF EUKARYOTIC MESSENGER-RNA BY MEANS OF THE POLYMERASE CHAIN-REACTION. **SCIENCE** 257: (5072) 967-971 AUG 14 1992.
- LOCKHART DJ, DONG HL, BYRNE MC, FOLLETTIE MT, GALLO MV, CHEE MS, MITTMANN M, WANG CW, KOBAYASHI M, HORTON H, BROWN EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. **NATURE BIOTECHNOLOGY** 14: (13) 1675-1680 DEC 1996.
- LUCENTE D, CHEN HM, SHEA D, SAMEC SN, RUTTER M, CHRAST R, ROSSIER C, BUCKLER A, ANTONARAKIS SE, MCCORMICK MK. Localization of 102 exons to a 2,5 Mb region involved in Down Syndrome. **HUMAN MOLECULAR GENETICS** 4: 1305-1311 1995.
- MAO R, ZIELKE CL, ZIELKE HR, PEVSNER J. Global up-regulation of chromosome 21 gene expression in the developing down syndrome brain. **GENOMICS** 81, 457-467. 2003.
- MARGULIES EH, INNIS JW. eSAGE: managing and analysing data generated with serial analysis of gene expression (SAGE). **BIOINFORMATICS**;16(7):650-1 Jul 2000.
- MC CORMIK MK, SCHINZEL A, PETERSEN MB, STETTEN G, DRISCOLL DJ, CANTU ES, TRANEBJAERG L, MIKKELSEN M, WATKINS PC, ANTONARAKIS SE.

Molecular genetic approach to the characterization of the "Down Syndrome" region on chromosome 21. **GENOMICS** 5: 325-331 1989.

- MCINNIS MG, CHAKRAVARTI A, BLASCHAK J, PERTERSEN MB, SHARMA V, AVRAMOPOULOS D, BLOUIN JL, KONIG U, BRAHE C, MATISE TC, WARREN A, TALBOT CC, VAN BROECKHOVEN C, LITT M, ANTONARAKIS SE. A linkage map of human chromosome 21: 43 PCR markers at average intervals of 2.5 cM. **GENOMICS** 16:562-571 1993.

- MILLER G, FUCHS R, LAI E. IMAGE cDNA clones, UniGene clustering, and ACeDB: An integrated resource for expressed sequence information. **GENOME RESEARCH** 7: (10) 1027-1032 OCT 1997.

- NAKAMURA A, HATTORI M, SAKAKI Y. Isolation of a novel human gene from the Down syndrome critical region of chromosome 21q22.2. **JOURNAL OF BIOCHEMISTRY** 122: (4) 872-877 OCT 1997.

- **NATURE GENETICS**. Volume 21: pp 1 - 60, Supplement JAN 1999

- NIZETIC D. Functional genomics of the Down syndrome. **CROAT MED J.**; 42(4): 421-7 Aug 2001.

- OHIRA M, OOTSUYAMA A, SUZUKI E, ICHIKAWA H, SEKI N, NAGASE T, NOMURA N, OHKI M. Identification of a novel human gene containing the tetratricopeptide repeat domain from the Down Syndrome region of chromosome 21. **DNA RESEARCH** 3: 9-16 1996.

- OKADA T, NGO VN, EKLAND EH, FÖRSTER R, LIPP M, LITTMAN DR, CYSTER JG. Chemokine requirements for B cell entry to lymph nodes and Peyer's patches. **JOURNAL OF EXPERIMENTAL MEDICINE** 196(1), 65–75 2002.

- OKUBO K, HORI N, MATOBA R, NIIYAMA T, FUKUSHIMA A, KOJIMA Y, MATSUBARA K. LargeScale cDNA sequencing for Analysis of quantitative and qualitative aspects of gene-expression. **NATURE GENETICS** 2: (3) 173-179 NOV 1992.

- RAO VN, MODI WS, DRABKIN HD, PATTERSON D, O'BRIEN SJ, PAPAS TS, REDDY ES. The human erg gene maps to chromosome 21, band q22: relationship to the 8; 21 translocation of acute myelogenous leukemia. **ONCOGENE** 3(5):497-500 Nov 1988.

- RAPORT CJ, SCHWEICKART VL, CHANTRY D, EDDY RL JR, SHOWS TB, GODISKA R, GRAY PW. New members of the chemokine receptor gene family. **JOURNAL OF LEUKOCYTE BIOLOGY**. 59(1):18-23. Review Jan 1996.
- REEVES RH, BAXTER LL, RICHTSMEIER JT. Too much of a good thing: mechanisms of gene action in Down syndrome. **TRENDS IN GENETICS**. 17: (2) 83-88. 2001.
- ROIZEN NJ, PATTERSON D. Down's syndrome. **LANCET** 361:1281-1289 2003.
- SALENTIN R, GEMSA D, SPRENGER H, KAUFMANN A. Chemokine receptor expression and chemotactic responsiveness of human monocytes after influenza A virus infection. **JOURNAL OF LEUKOCYTE BIOLOGY**. 74, 252-259 2003.
- SAMBROOK J, FRITSCH EF, MANIATIS T. Molecular Cloning: A Laboratory Manual. 2nd ed. **COLD SPRING HARBOR LABORATORY PRESS**, Cold Spring Harbor, N.Y. 1989.
- SAMBROOK J, RUSSELL DW. Molecular cloning: A laboratory manual, 3rd ed. **COLD SPRING HARBOR LABORATORY PRESS**, Cold Spring Harbor, NY. 2001.
- SHIBUYA K, KUDOH J, MOSHIMA S, KAWASAKI K, ASAKAWA S, SHIMIZU N. Isolation of two novel genes, DSCR5 and DSCR6, from Down Syndrome Critical Region on Human Chromosome 21q22.2. **BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS** 271, 693-698 2000.
- SUMARSONO SH, WILSON TJ, TYMMS MJ, VENTER DJ, CORRICK CM, KOLA R, LAHOUD MH, PAPAS TS, SETH A, KOLA I. Down's syndrome-like skeletal abnormalities in ETS2 transgenic mice. **NATURE** 8;379(6565):534-7 Feb 1996..
- UGAZIO AG. Down's syndrome: problems of immunodeficiency. **HUM GENET SUPPL.**; 2:33-9 1981.
- VAN DEN BERG A, VAN DER LEIJ J, POPPEMA S. Serial analysis of gene expression: rapid RT-PCR analysis of unknown SAGE tags. **NUCLEIC ACIDS RESEARCH**. 1;27(17):e17. Sep 1999.
- VELCULESCU VE, VOGELSTEIN B, KINZLER KW. Analysing uncharted transcriptomes with SAGE. **TRENDS IN GENETICS** 16: (10) 423-425 OCT 2000.
- VELCULESCU VE, ZHANG L, VOGELSTEIN B, KINZLER KW. SERIAL ANALYSIS OF GENE-EXPRESSION. **SCIENCE** 270: (5235) 484-487 OCT 20 1995.

- VELCULESCU VE, ZHANG L, ZHOU W, VOGELSTEIN J, BASRAI MA, BASSETT DE, HIETER P, VOGELSTEIN B, KINZLER KW. Characterization of the yeast transcriptome. **CELL** 88: (2) 243-251 JAN 24 1997.
- VIDAL-TABOADA JM, SANZ S, EGEO A, SCARTEZZINI P, OLIVA R. Identification and characterization of a new gene from human chromosome 21 between markers D21S343 and D21S268 encoding a leucine-rich protein. **BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS** 250: (3) 547-554 SEP 29 1998.
- VOGEL F, MOTULSKY AG: Human Genetics: Problems and approaches, 3rd ed, **SPRINGER**. 851pg 1997.
- WHEELER DL, CHURCH DM, LASH AE, LEIPE DD, MADDEN TL, PONTIUS JU, SCHULER GD, SCHRIML LM, TATUSOVA TA, WAGNER L, RAPP BA. Database resources of the National Center for Biotechnology Information. **NUCLEIC ACIDS RESEARCH** 29: (1) 11-16 JAN 1 2001.
- ZONG WX, EDELSTEIN LC, CHEN C, BASH J, GELINAS C. The prosurvival Bcl-2 homolog Bfl-1/A1 is a direct transcriptional target of NF-kappaB that blocks TNFalpha-induced apoptosis. **GENES DEVELOPMENT**. 13(4), 382-387 1999.

7 – Anexos

7.1 – Análise de riscos e benefícios do projeto.

Quanto aos sujeitos da pesquisa:

O indivíduo que estará representando o sujeito da pesquisa (portador de síndrome de Down) terá uma amostra de sangue periférico coletada, da qual será extraído o RNA. A coleta, simples e trivial como as de laboratórios de análise clínica, não representa riscos significativos ou causa danos ao sujeito da pesquisa. Ela será realizada com uso de material estéril e descartável, em condições adequadas de assepsia, e por profissional de saúde treinado (biomédico). O RNA obtido do sangue (precisamente das células brancas do sangue) servirá como molde para a obtenção do cDNA (DNA complementar ao RNA). Desta forma, a coleta de sangue é uma etapa essencial para que o projeto seja iniciado. Após a extração do RNA e os estudos realizados, pode haver eventualmente uma coleta adicional de sangue do indivíduo. Parte do RNA que se isola na primeira etapa do estudo é reservada para utilização na parte final de confirmação dos dados. A coleta adicional pode se fazer necessária já que o RNA é muito suscetível à degradação. Nós garantimos, ao sujeito da pesquisa, que assumiremos total responsabilidade caso algum dano ou prejuízo venha a ocorrer com o mesmo devido à sua participação na pesquisa.

Também será garantido ao sujeito, sigilo absoluto quanto à sua identificação, não sendo este dado e/ou outros, que possam comprometê-lo e

acarretar algum tipo de problema ou desconforto, citados em nenhum momento.

Por outro lado, a realização deste projeto pode contribuir de forma muito significativa para um entendimento maior das alterações fenotípicas resultantes da síndrome de Down, e conseqüentemente trazer benefícios futuros como, por exemplo, gerar conhecimento para o desenvolvimento de tratamentos que possam vir a amenizar os problemas característicos dos pacientes.

Quanto aos benefícios científicos:

A síndrome de Down (SD), uma das maiores causas de retardo mental, é a alteração cromossômica de maior incidência em humanos (um caso a cada setecentos nascimentos) e a melhor estudada. Suas características típicas são bem familiares, porém anômalas e, portanto, indesejáveis. O cromossomo 21, que em triplicata causa a síndrome, foi o primeiro a ser seqüenciado e tem seus genes amplamente estudados.

Neste contexto, destaca-se a importância do estudo desta anomalia. O presente projeto propõe uma análise ampla da expressão gênica de leucócitos de portador da SD e sua comparação com a expressão gênica de leucócitos de não portadores. Desta forma, utilizando a técnica de Análise Serial da Expressão Gênica (SAGE) busca-se identificar genes diferencialmente expressos, em leucócitos de portadores, que representariam importantes alvos de estudo para o esclarecimento de mecanismos moleculares responsáveis por anomalias. Objetiva-se, por se tratar de leucócitos, identificar genes envolvidos: com as disfunções imunológicas e com o risco aumentado de se desenvolver doença de Alzheimer (problemas característicos dos portadores da SD).

A técnica de SAGE tem se mostrado de grande valia na obtenção de padrões digitais de expressão gênica. Esta técnica tem um potencial de larga escala por se basear na geração de pequenos “tags” ou cDNAs de aproximadamente 14pb específicos e representativos de cada cDNA oriundo da população inicial de RNAm. Os tags são “concatamerizados” (ligados), clonados e seqüenciados. Depois são contados e analisados. Isto gera dados comparativos entre a condição normal e anômala. Estas informações são quantitativas (mostram a abundância relativa dos genes) e qualitativas (fornecem suas identificações).

Assim, a utilização da SAGE será de grande importância na obtenção de resultados significativos para o entendimento global do padrão de expressão gênica de leucócitos de portador de SD. Padrão este, envolvido com prováveis processos bioquímicos alterados na síndrome. Também, a implantação da SAGE, ainda pouco explorada no Brasil, trará uma série de conhecimentos e novas possibilidades para a análise funcional dos genomas, abrindo portas para a nova era da Genômica Funcional.

7.2 – Esclarecimentos prestados para submissão do projeto ao CEP-UFSCar

É de responsabilidade do pesquisador responsável conduzir a pesquisa até que resultados positivos, nulos ou negativos, sejam publicados e conseqüentemente auxiliem a comunidade científica nas pesquisas sobre a síndrome de Down (SD). O projeto será realizado no Laboratório de Biologia Molecular (LBM), instalado no Departamento de Genética e Evolução (DGE) da

Universidade Federal de São Carlos (UFSCar). Este laboratório possui equipamentos especializados para pesquisa em Biologia Molecular e conta com apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). A UFSCar prove a infra-estrutura para as instalações do LBM.

Uma vez atingidos os objetivos principais explicitados neste projeto, e obtendo-se resultados esclarecedores, os mesmos serão publicados em revista científica apropriada. Este seria o critério para o encerramento da pesquisa, já que não existem riscos que tornem plausível a suspensão da realização do trabalho. O término da pesquisa está previsto para novembro do ano de 2003.

Quanto ao sujeito da pesquisa, este não será submetido a riscos significativos e não sofrerá danos pelos métodos utilizados. Serão realizadas apenas simples e cuidadosas coletas de sangue periférico, por profissional da área de saúde treinado, para que o projeto seja iniciado. É importante ressaltar que a participação no projeto não acarretará gastos nem causará danos ao sujeito da pesquisa e que o pesquisador responsável garante o ressarcimento de eventuais despesas.

A Análise Serial da Expressão Gênica (SAGE) é um método de estudo de padrão de expressão gênica, que compara: dados de seqüência de DNA obtidos a partir de uma amostra de RNA em questão, a dados de seqüência de DNA obtidos de uma amostra de RNA padrão. Neste projeto, a amostra em questão, o RNA total de leucócitos (células brancas do sangue) de portador da SD, representará os pacientes. Para as comparações serão utilizados os dados

de seqüência da amostra controle, obtidos de RNA de leucócitos de pessoas normais, que já foram gerados por outros pesquisadores e se encontram disponíveis publicamente na rede mundial de computadores.

Deste modo a inclusão de indivíduo portador da SD, que contribuirá como sujeito da pesquisa, se faz necessária sendo a obtenção de seu sangue periférico crucial para a realização do projeto. Sendo portador de síndrome de Down, este indivíduo pertence a grupos vulneráveis, no entanto, a obtenção do conhecimento desejado pelo presente projeto, não pode ser conseguida por sujeitos com plena autonomia. Dada a importância da síndrome de Down, devido a sua gravidade e alta incidência, e os promissores conhecimentos novos que serão gerados, a escolha do sujeito da pesquisa se justifica. Esta escolha se baseará na idade e características clínicas do portador de SD (adequadas para comparações ao controle) matriculado na Associação de Pais e Amigos dos Excepcionais (APAE) da cidade de Passos, Minas Gerais. O consentimento livre e esclarecido dos responsáveis pelos doadores, bem como os esclarecimentos prestados ao paciente (no limite de suas capacidades) são condições imprescindíveis para a coleta de sangue.

7.3 – Modelo do consentimento esclarecido assinado pelos responsáveis do sujeito da pesquisa (portador de SD).

O laboratório de Biologia Molecular (LBM) do Departamento de Genética e Evolução (DGE) da UFSCar realizará pesquisa envolvendo seres humanos e em prol dos mesmos. O projeto de pesquisa refere-se à identificação e ao estudo de genes que provavelmente atuam de forma distinta em portadores de síndrome de Down (SD), quando comparados a não portadores. Para a

realização dos estudos será necessária a utilização do RNA (material genético) de leucócitos (células brancas do sangue) de portador de SD envolvido na pesquisa (sujeito da pesquisa). Este material será obtido de sangue periférico coletado do indivíduo, e utilizado para iniciar os estudos. De fato, esta coleta é indispensável e insubstituível para a pesquisa, pois não há disponível no mercado, ou em outras fontes, amostras representativas de material genético de portadores desta síndrome que possam ser utilizadas com tal finalidade. Desta forma se torna imprescindível neste projeto, a participação de um doador de sangue paciente de SD (alvo da pesquisa).

O pesquisador responsável abaixo assinado é ciente das responsabilidades descritas na Resolução nº 196 de 10 de outubro de 1996, do Conselho Nacional de Saúde, e se compromete a cumprir inteiramente os termos desta resolução. O mesmo esclarece que haverá o devido ressarcimento de eventuais despesas decorrentes de toda e qualquer participação do sujeito da pesquisa doador de sangue; ainda, garante indenização diante de eventuais danos. É também de responsabilidade do pesquisador responsável a garantia de acompanhamento e esclarecimentos antes e durante os estudos, a garantia da liberdade do sujeito de retirar seu consentimento sem nenhum tipo de restrição em qualquer fase da pesquisa e a garantia do sigilo quanto à sua identidade, quanto à sua privacidade e quanto aos dados confidenciais envolvidos na pesquisa. O sujeito portador de síndrome de Down participante da pesquisa será acompanhado e orientado por seus representantes legais, tendo absoluto direito de ser informado sobre a pesquisa, no limite de sua capacidade. Os responsáveis abaixo assinados pelo sujeito da pesquisa, cientes dos estudos a serem realizados, autorizam neste

documento, a coleta de sangue e a publicação, em revista científica apropriada, dos resultados obtidos no término da pesquisa. Quaisquer outras atribuições ao pesquisador, e direitos do sujeito da pesquisa, descritos na referida Resolução, serão seguidas no desenvolvimento deste projeto, sem nenhuma restrição.

7.4 – Carta de esclarecimento enviada aos responsáveis pelo sujeito da pesquisa.

A síndrome de Down (SD) é uma das anomalias genéticas mais freqüentes em humanos ocorrendo um caso em cada setecentos nascimentos. Suas características mais comuns são bastante conhecidas, entre elas estão: baixa estatura, mãos curtas, face achatada, fissura palpebral oblíqua ascendente, orelhas diplásticas, prega palmar única e outras. No entanto, as principais anomalias que ocorrem em 100% dos casos são: o retardo mental, os defeitos do sistema imunológico, e o risco aumentado de desenvolvimento de leucemia e doença de Alzheimer; ocorrendo em 50% dos casos estão os problemas cardíacos. Neste contexto, a pesquisa científica que investiga a SD é de grande importância considerando a alta incidência e as características indesejáveis desta anomalia. Tais estudos se justificam por gerar conhecimento necessário para que surjam novos tratamentos que um dia, talvez no futuro, possam ajudar a neutralizar ou pelo menos amenizar os efeitos da síndrome.

Nosso Laboratório, o Laboratório de Biologia Molecular (LBM), se encontra no Departamento de Genética e Evolução (DGE) da Universidade Federal de São Carlos (UFSCar) e está sob coordenação do pesquisador responsável Prof. Dr. Flávio Henrique da Silva. Este professor realizou seu

doutoramento no Instituto de Química (IQ) da Universidade de São Paulo (USP) adquirindo vasta experiência em técnicas de Biologia Molecular aplicadas à pesquisa científica.

Para os projetos de pesquisa desenvolvidos no LBM e devido aos apoios financeiros da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), nosso laboratório conta com uma infra-estrutura completa para a pesquisa em Biologia Molecular, incluindo seqüenciador de DNA, purificador de proteínas e diversos outros equipamentos e materiais. Nossa equipe de pesquisadores é treinada em técnicas de genética molecular e constituída por alunos, de Pós-Graduação em Genética e Evolução (Mestrado e Doutorado).

Desde 1996, o LBM estuda detalhadamente a síndrome de Down utilizando técnicas de engenharia genética. Alguns dos genes potencialmente envolvidos com a síndrome já foram isolados em nosso laboratório e vêm sendo cuidadosamente investigados quanto as suas funções, são eles: *dscr-1*, *dcra*, *dcrb*, *dscr-2* e *dscr-5b*. Acredita-se que com o esclarecimento das atividades destes genes, ou seja, o entendimento dos seus mecanismos moleculares de atuação, se possa compreender os processos biológicos envolvidos no desenvolvimento anômalo dos portadores da SD. Alguns destes “genes–alvo” por nós estudados, já tiveram certas funções elucidadas que permitiram novas especulações sobre suas prováveis atuações que resultariam nas anormalidades dos pacientes.

O presente trabalho de investigação da SD realizado atualmente em nosso laboratório marca o início de uma nova abordagem de estudos. Neste

trabalho é utilizada a técnica de Estudo Serial da Expressão Gênica (SAGE) que é uma metodologia de genética molecular bastante recente e ainda pouquíssimo utilizada no Brasil. Ela é valiosa, pois permite a análise do funcionamento de milhares de genes simultaneamente. Com ela se pode verificar o funcionamento padrão do conjunto de genes ativos de um determinado órgão, tecido ou tipo celular normal, e compará-lo ao funcionamento do conjunto de genes ativos de um órgão, tecido ou tipo celular anormal (no caso, de pacientes com SD). Outras técnicas disponíveis que também permitem o estudo comparativo do funcionamento (ou expressão) de genes, tais como, “RT-PCR” e “Northern Blot”, não o fazem em larga escala como a SAGE, abrangendo apenas dezenas de genes.

Já é bem aceito na comunidade científica que as mudanças fisiológicas de qualquer natureza (inclusive patológica) são acompanhadas de mudanças no padrão de expressão gênica. Neste contexto, com a técnica de SAGE se podem buscar diferenças na expressão gênica (relacionadas com as anomalias dos pacientes) comparando indivíduos normais a indivíduos portadores de SD.

Para iniciar o uso da SAGE decidiu-se analisar o tecido sanguíneo humano, mais precisamente as células brancas do sangue (leucócitos), comparando então, os leucócitos de indivíduos normais aos de portadores da síndrome. Daí a necessidade de haver um portador da SD sujeito da pesquisa e, portanto doador de sangue. Ressalta-se que a este indivíduo será garantido o sigilo, quanto às informações envolvidas no trabalho, que assegura sua privacidade. A escolha por leucócitos se justifica porque tais células estão diretamente envolvidas com o funcionamento do sistema imunológico (anormal

em portadores da SD) e com o risco de desenvolver leucemia e doença de Alzheimer (aumentado em portadores da SD).

Com o presente trabalho, se pretende identificar diversos novos genes potencialmente envolvidos com a SD. Estes genes estariam envolvidos com as doenças acima citadas e os problemas imunes (característicos dos portadores da SD), pois de algum modo seus desempenhos irregulares (diferenças no nível de expressão), nos pacientes, estariam relacionados às anormalidades.

A técnica de SAGE gera tabelas de dados que comparam os genes atuantes nos leucócitos de pessoas normais aos genes atuantes nos leucócitos de portadores da SD. Daí são obtidas informações sobre quais genes se encontram relativamente ativos ou inativos nas células de um ou outro sujeito, quando comparados entre si, e qual a intensidade desta atividade. Assim, as intensidades de expressão gênica são comparadas e se pode então verificar, com auxílio de métodos estatísticos, diferenças significativas entre os níveis de atividade gênica de cada indivíduo.

A seleção de novos genes potencialmente envolvidos com a SD se dá devido a duas condições simultâneas: quando os genes apresentam níveis de atividade bastante distintos (relativamente aumentados ou reduzidos) entre os indivíduos analisados e quando eles exercem funções relacionadas com a manutenção do funcionamento dos leucócitos. Tais características tornam os genes candidatos ao envolvimento com as anormalidades dos pacientes, por eles poderem estar causando um desequilíbrio nos mecanismos moleculares (funcionamento de outros genes e proteínas) necessários para o desempenho normal das células brancas. Este desequilíbrio seria então a causa molecular dos problemas que ocorrem nos leucócitos dos pacientes. Desta forma os

genes candidatos escolhidos são importantes alvos novos para futuras investigações moleculares sobre a patologia da síndrome de Down.