

Elisa Rennó Donnard Moreira

**DESENVOLVIMENTO E APLICAÇÃO DE MÉTODOS
BIOQUÍMICOS E BIOINFORMÁTICOS PARA BUSCAR
INTERAÇÕES REGULATÓRIAS:
Tead4 como modelo**

Orientador: J. Miguel Ortega

Co-orientador: R. Daniel Gietz

**Laboratório de Biodados, Biologia Celular e Desenvolvimento
Departamento de Bioquímica e Imunologia
Instituto de Ciências Biológicas - UFMG
2010**

Elisa Rennó Donnard Moreira

DESENVOLVIMENTO E APLICAÇÃO DE MÉTODOS BIOQUÍMICOS E
BIOINFORMÁTICOS PARA BUSCAR INTERAÇÕES REGULATÓRIAS:
Tead4 como modelo

Dissertação apresentada ao Departamento de
Bioquímica e Imunologia do Instituto de
Ciências Biológicas da Universidade Federal
de Minas Gerais como pré-requisito para a
obtenção do título de Mestre em Bioquímica e
Imunologia

Orientador: J. Miguel Ortega

Co-orientador: R. Daniel Gietz

Belo Horizonte

2010

AGRADECIMENTOS

Agradeço a todos que contribuíram para este trabalho, direta ou indiretamente. Ao meu orientador Miguel pelas excelentes idéias e toda a ajuda, aos amigos do Lab Biodados pela participação no trabalho e pela companhia mais que divertida. A Dan Gietz e Cheri por tornarem este trabalho possível, aos amigos de Bases por tornarem o trabalho mais agradável! Agradeço a minha família, aos meus amigos e ao Bê por todo o apoio.

LISTA DE FIGURAS

FIGURA 1: DESENVOLVIMENTO EMBRIONÁRIO PRÉ-IMPLANTAÇÃO	12
FIGURA 2: SISTEMA DE DUPLO-HÍBRIDO EM LEVEDURA.....	16
FIGURA 3: MAPAS DOS PLASMÍDIOS UTILIZADOS NOS EXPERIMENTOS DE DUPLO-HÍBRIDO.....	26
FIGURA 4 - PROTOCOLOS DE DILUIÇÃO DAS CULTURAS LÍQUIDAS	29
FIGURA 5: PROTOCOLO DE TRIAGEM EM CULTURA LÍQUIDA.....	31
FIGURA 6: PLATAFORMAS UTILIZADAS NOS EXPERIMENTOS DE MINERAÇÃO DE TEXTO.....	37
FIGURA 7: PROCEDIMENTO PARA CRIAÇÃO DA VIA DE INTERAÇÕES	39
FIGURA 8: PROCEDIMENTO PARA CRIAÇÃO DE AGRUPAMENTOS DE HOMÓLOGOS	41
FIGURA 9: BUSCA PELO ÚLTIMO ANCESTRAL COMUM (LCA).....	42
FIGURA 10: : ENSAIO DE <i>LACZ</i> DA TRIAGEM DE TRIM2	48
FIGURA 11: RESULTADO DO PESCADOR	50
FIGURA 12: VIA REGULATÓRIA DO DESENVOLVIMENTO EMBRIONÁRIO PRÉ-IMPLANTAÇÃO	52
FIGURA 13: CLASSIFICAÇÃO DOS HOMÓLOGOS AGRUPADOS PELO SEEDSERVER	61
FIGURA 14: VIA REGULATÓRIA DO DESENVOLVIMENTO EMBRIONÁRIO PRÉ-IMPLANTAÇÃO MOSTRANDO A ANCESTRALIDADE DOS GENES ENVOLVIDOS	64
FIGURA 15: DISTRIBUIÇÃO DA ORIGEM DOS GENES DA VIA DE PRÉ-IMPLANTAÇÃO	65
FIGURA 16: ANCESTRALIDADE OBTIDA PARA GENES DA PRÉ-IMPLANTAÇÃO - COMPARAÇÃO KO E SEEDSERVER.....	66
FIGURA 17: FUNÇÕES DOS GENES HOMÓLOGOS ENCONTRADOS EM <i>DROSOPHILA MELANOGASTER</i>	68
FIGURA 18: ENSAIO DE <i>LACZ</i> PARA RECONSTRUÇÃO DA INTERAÇÃO COM TEAD4	71

LISTA DE TABELAS

TABELA 1: ESPECIFICAÇÕES DAS LINHAGENS DE <i>SACCHAROMYCES CEREVISIAE</i> VALIDADAS PARA OS PROTOCOLOS DE DUPLO- HÍBRIDO COM TRIAGEM LÍQUIDA.....	24
TABELA 2: LEITURA DE OD600 DAS PLACAS DE TRIAGEM DA PROTEÍNA TRIM2 DESTACANDO POÇOS COM CRESCIMENTO ..	46
TABELA 3: CORRESPONDÊNCIA ENTRE OS NOMES DOS GENES HUMANOS E HOMÓLOGOS DE <i>DROSOPHILA MELANOGASTER</i> AGRUPADOS PELO SEEDSERVER	69

LISTA DE ABREVIATURAS E SIGLAS

3AT- 3-Aminotriazol

Amp- Ampicilina

cDNA- DNA complementar ao mRNA

DNA- Ácido Desoxirribonucléico

EPI- “Epiblast” / Epiblasto

Gal4_{DA}- Domínio de ativação transcricional do gene Gal4

Gal4_{DL}- Domínio de ligação ao DNA do gene Gal4

His- Histidina

HIS3- Gene codificador de uma enzima da síntese de histidina

ICM - “Inner Cell Mass” / Massa Celular Interna

KEGG – Kyoto Encyclopedia of Genes and Genomes

KO- KEGG Orthology

LacZ- gene da β -galactosidase

LB- Meio Luria Bertani

LCA- “Last Common Ancestor” / Último Ancestral Comum

Leu- Leucina

LEU2- Gene codificador de uma enzima da síntese do aminoácido leucina

LiAc- Acetato de Lítio

min- minuto

mRNA- RNA mensageiro

NCBI - National Center for Biotechnology Information

OD- absorvância medida a 600 nm (Optical Density)

PCR- Reação em cadeia da polimerase (Polimerase Chain Reaction)

PE- “Primitive Endoderm” / Endoderme Primitiva

PEG- Polietilenoglicol

PMID- “PubMed Unique Identifier”

RNA- Ácido Ribonucléico

SDs- Meio de cultura seletivo para leveduras

SD -W-L - Meio de cultura seletivo para leveduras sem os aminoácidos triptofano (W) e leucina (L)

SD -W-L-H - Meio de cultura seletivo para leveduras sem os aminoácidos triptofano (W), leucina (L) e histidina (H)

ssDNA- DNA de esperma de salmão desnaturado

TE- “Trophectoderm” / Trofectoderme

Tris- Tris-hidroximetil-aminometano

Trp- Triptofano

TRP1- Gene codificador de uma enzima da síntese do aminoácido triptofano

UEKO- UniRef Enriched KEGG Orthology

X-Gal- 5-bromo-4-cloro-3-indolil- β -d-galactopiranosídeo

RESUMO

O método de duplo-híbrido em leveduras é um sistema de seleção genético que permite a identificação de produtos de cDNA que interagem com uma proteína de interesse. Neste trabalho descrevemos dois novos protocolos para triagem em meio líquido que contribuem para aperfeiçoar este valioso método e eliminam a triagem laboriosa em placas de meio sólido. Para validar os novos protocolos utilizamos duas linhagens de levedura *Saccharomyces cerevisiae*, KGY37 e Y190, três iscas diferentes: m-RTel1, m-Trim2 ou h-TRIM32 e suas bibliotecas de cDNA respectivas: testículo de camundongo, cérebro de camundongo e músculo esquelético humano. Estas combinações já haviam sido previamente triadas com o protocolo tradicional, permitindo uma comparação com o protocolo novo. A nova triagem resultou na seleção de interações em todos os ensaios, o número obtido de clones positivos é comparável ao meio sólido e o meio líquido se mostrou uma abordagem mais rápida, econômica e eficiente. Atualmente, são gerados dados de interações entre proteínas e regulações gênicas em larga escala. Para uma visualização melhor de processos biológicos complexos, é interessante que estes dados sejam integrados e disponibilizados em bancos de dados como KEGG Pathway. Para o desenvolvimento embrionário pré-implantação, não existe uma via regulatória e nem todos os genes envolvidos no processo sequer possuem entradas no banco KEGG Orthology. Em vista disso, procuramos desenvolver esta via regulatória utilizando ferramentas de mineração de texto como Medline Ranker e PESCADOR para revelar biointerações entre genes envolvidos neste processo. Após a criação da via, os genes encontrados foram também utilizados como “seed” no software SeedServer desenvolvido no laboratório para gerar agrupamentos de homólogos. Os homólogos permitiram também a determinação do último ancestral comum de cada gene, revelando que a via do desenvolvimento pré-implantação é composta por uma porção ancestral e genes com origem moderna. Um dos fatores de transcrição essenciais para esta etapa do desenvolvimento é Tead4, envolvido em diversas regulações da pré-implantação. O único parceiro de interação conhecido para Tead4 é YAP que participa da regulação da diferenciação da trofotoderme. Em busca de novos parceiros que interagem com Tead4 utilizamos esta proteína como isca numa triagem de duplo-híbrido com o protocolo validado previamente. A biblioteca de cDNA utilizada foi a de embrião de sete dias (camundongo) e um dos clones recuperados no ensaio foi identificado como Tioredoxina1, cujo envolvimento em outras interações com fatores de transcrição é conhecido.

ABSTRACT

Yeast two-hybrid method consists of a genetic trap that selects for prey cDNA products within a library that interact with a bait protein of interest. Here we report two protocols of liquid screening that further improve this valuable method and eliminate the laborious classic screening in solid medium. To evaluate the accuracy of the new protocol we used two *Saccharomyces cerevisiae* yeast strains, KGY37 and Y190, three different baits: m-RTel1, m-Trim2 or h-TRIM32 and their respective cDNA libraries: mouse testis, mouse brain and human skeletal muscle. These combinations had been previously screened with the classic method allowing for a direct comparison to the novel method. The new screen resulted in the selection of interactions for all the bait proteins used, the number of positive clones obtained in the assays is comparable to that of the solid medium and the liquid screening is a faster, more economic and more efficient approach. Currently, the generation of protein interaction data and gene regulation occurs in a large scale. The integration of these data and generation of pathways and networks contained in databases such as KEGG Pathway is essential for the comprehension of complex biological processes. We noticed the absence of a chart or pathway describing the largely studied preimplantation development stages; furthermore, not all genes involved in the process have entries in KEGG Orthology. In this work we sought to develop this regulatory pathway using text-mining tools such as Medline Ranker and PESCADOR to reveal biointeractions among the genes involved in this process. The genes present in the resulting pathway were also used as seed in software developed by our group called SeedServer to create clusters of homologous genes. These homologs allowed the determination of the last common ancestor for each gene and revealed that the preimplantation development pathway consists of an ancient fraction with the addition of modern elements. One of the transcription factors essential for this developmental stage is Tead4, participating in several regulations during preimplantation. The only known interaction partner for Tead4 is YAP, co-activator of the differentiation of trophoctoderm lineage. In pursuit of new partners that interact with Tead4 we used this protein as bait in a two-hybrid screen with our previously reported protocol. The cDNA library used was mouse seven-day embryo and one of the recovered clones was characterized as Thioredoxin1, a protein with other reported associations to transcription factors.

ÍNDICE

AGRADECIMENTOS	II
LISTA DE FIGURAS.....	III
LISTA DE TABELAS.....	IV
LISTA DE ABREVIATURAS E SIGLAS	V
RESUMO	VII
ABSTRACT	VIII
1. INTRODUÇÃO.....	11
1.1 Desenvolvimento Embrionário Pré-implantação	11
1.2 O Fator de Transcrição Tead4.....	12
1.3 Proteômica Funcional e o Sistema de Duplo-Híbrido	13
1.4 Bioinformática na Integração do Conhecimento.....	18
1.4.1 Vias regulatórias	19
1.4.2 KEGG.....	19
1.4.3 Grupos de ortólogos	20
1.4.4 Identificando interações.....	21
2. JUSTIFICATIVA	22
3.OBJETIVOS.....	23
3.1 Objetivo Geral.....	23
3.2 Objetivos Específicos	23
4. MATERIAL E MÉTODOS	24
4.1 Duplo-híbrido em leveduras	24
4.1.1 Leveduras.....	24
4.1.2 Bactérias	25
4.1.3 Plasmídios.....	25
4.1.5 Protocolos	27
4.2 Bioinformática	33
4.2.1 PubMed	33
4.2.2 Medline Ranker	34
4.2.3 PESCADOR	34
4.2.5 Construção da via de regulações.....	38
4.2.6 SeedServer.....	40
4.2.7 Último Ancestral Comum	40
5. RESULTADOS	43
5.1 Validação da triagem líquida para o sistema de duplo-híbrido em leveduras	43
5.1.1 Escolha de iscas para triagens	43
5.1.2 Transformações e linhagens de leveduras	43
5.1.3 Triagem líquida com cRtel1 e Trim2.....	44
5.1.4 Monitoramento de leveduras com crescimento alterado	44
5.1.5 Triagem líquida com as iscas Trim2 e TRIM32, sob monitoramento	47
5.1.6 Ativação do gene repórter <i>LacZ</i>	47
5.1.7 Validação dos positivos.....	47
5.2 Criação de uma via de regulações utilizando ferramentas de mineração de texto ..	49
5.2.1 Busca na literatura.....	49
5.2.2 Seleção de trabalhos relevantes.....	49
5.2.3 Revelando biointerações com o software PESCADOR.....	49

5.2.4 Representação gráfica da via do desenvolvimento pré-implantação e revisão das regulações	51
5.3. Busca por genes homólogos	59
5.3.1 SeedServer.....	59
5.3.2 Análise da composição dos agrupamentos gerados.....	59
5.4. Ancestralidade dos genes da via do desenvolvimento pré-implantação	62
5.4.1 Determinação do último ancestral comum (“Last Common Ancestor”)	62
5.4.2 Funções de homólogos.....	67
5.5. Novos parceiros de interação para Tead4.....	70
5.5.1 Duplo-Híbrido em levedura	70
5.5.2 Triagem líquida dos transformantes obtidos.....	70
5.5.3 Ensaio de <i>LacZ</i>	70
5.5.4 Análise e confirmação das interações.....	70
6. DISCUSSÃO	72
7. CONCLUSÕES.....	77
8. PERSPECTIVAS.....	78
9. REFERÊNCIAS	79
LISTA DE ANEXOS	87

1. INTRODUÇÃO

1.1 Desenvolvimento Embrionário Pré-implantação

O período do desenvolvimento chamado de pré-implantação é uma fase com a duração de aproximadamente seis dias, a implantação normalmente ocorre até o sétimo dia embrionário (E7), período em que o útero está receptivo para o embrião (janela de receptividade uterina) (Wang e Dey, 2006). O desenvolvimento embrionário de mamíferos foi bastante estudado em camundongos e é diferente do desenvolvimento de outros animais que apresentam polarização de células resultando em blastômeros com potencial restrito (Johnson e McConnell, 2004). No camundongo, os blastômeros retêm o potencial de gerar qualquer tipo celular até o estágio de oito células. Durante os dois primeiros dias do desenvolvimento, o embrião passa por clivagens sucessivas até produzir um embrião de oito células (“eight-cell embryo”). A partir de então passa a ser chamado de mórula e os blastocistos aumentam o contato célula-célula para formar uma mórula compactada. As divisões subsequentes aumentam a complexidade da mórula e as células podem estar localizadas no interior, cercadas por outras células ou então na parte externa em contato com o ambiente. A identificação de células geradoras de cada linhagem mostrou que a trofotoderme (TE - “Trophectoderm”) é derivada principalmente de células externas, enquanto as células internas dão origem à massa celular interna (ICM - “Inner Cell Mass”). A ICM posteriormente se divide nas linhagens da endoderme primitiva (PE - “Primitive Endoderme”) e do epiblasto (EPI - “Epiblast”). Durante a formação da ICM e TE também ocorre a cavitação gerando a blastocele. O embrião é chamado de blastocisto no estágio em que apresenta ICM, TE e blastocele. 24 horas depois da formação do blastocisto ocorre a última fase de desenvolvimento antes da implantação do embrião na parede do útero e é formado o blastocisto tardio, no qual a linhagem da endoderme primitiva (PE, parte da endoderme extra-embrionária) já se separou da ICM. As três linhagens formadas no desenvolvimento pré-implantação têm destinos diferentes na formação do embrião. Enquanto o epiblasto dará origem ao feto no desenvolvimento posterior, a trofotoderme irá formar a porção fetal da placenta e a endoderme primitiva (posteriormente endoderme extra-embrionária) forma o saco vitelínico (Yamanaka, Ralston *et al.*, 2006). Os estágios do desenvolvimento durante a pré-implantação estão ilustrados na Figura 3.

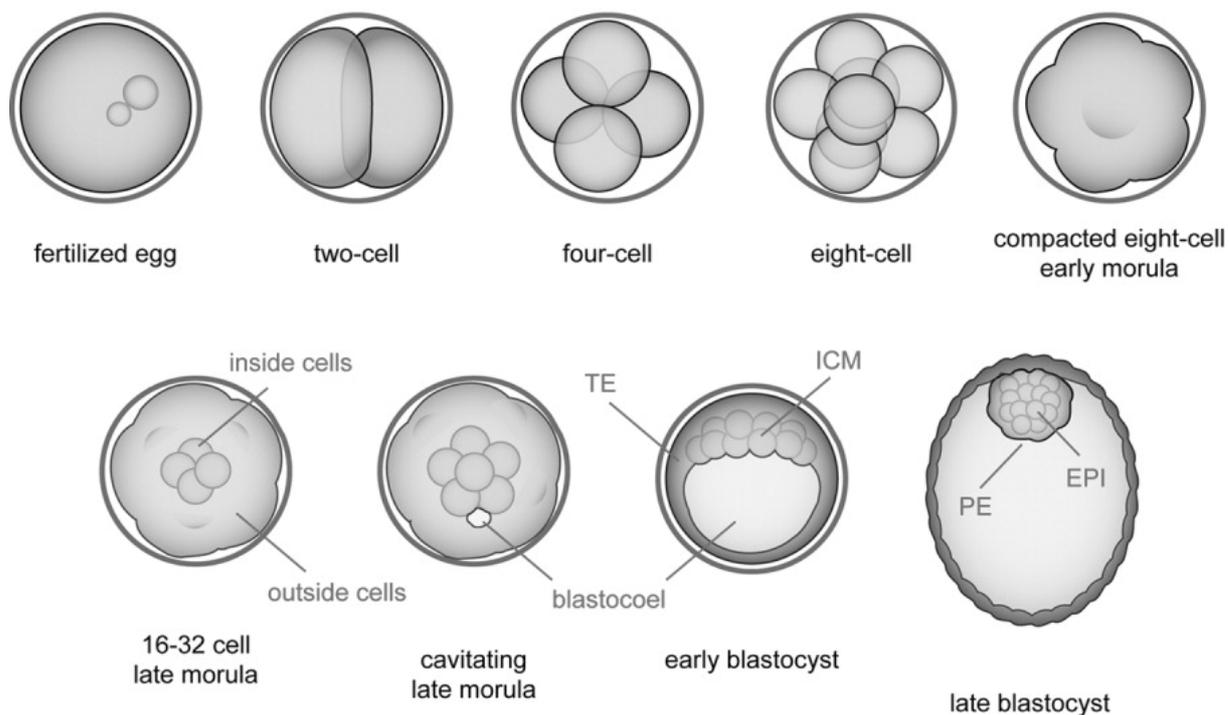


Figura 1: Desenvolvimento embrionário pré-implantação. (Yamanaka, Ralston *et al.*, 2006) Estágios embrionários no desenvolvimento pré-implantação, uma descrição detalhada pode ser encontrada no texto. TE=Trofotoderme; ICM=Massa Celular Interna; PE=Endoderme primitiva; EPI=Epiblasto.

Processos regulatórios complexos como o desenvolvimento de animais são resultantes da interação de diversos produtos gênicos e elementos que controlam a expressão destes genes. Resultados de experimentos tradicionais que determinam a função de um ou poucos genes não geram uma visão adequada para sistemas complexos. Uma rede regulatória complexa seria capaz de retratar aspectos específicos e gerais do desenvolvimento, como a maneira que determinadas células geram seus destinos no embrião e porque o processo caminha inexoravelmente adiante no tempo do desenvolvimento (Davidson, Rast *et al.*, 2002).

1.2 O Fator de Transcrição Tead4

Tead4 pertence a uma família de fatores de transcrição chamada TEA. As proteínas desta família contêm o domínio TEA de ligação ao DNA e são fatores de transcrição envolvidos em funções de diferenciação e desenvolvimento em uma variedade de animais e plantas (Andrianopoulos e Timberlake, 1991). A família recebeu este nome devido aos primeiros membros identificados: TEC1 de leveduras, TEF-1 de mamíferos (Tead1) e AbaA de *Aspergillus nidulans* (Burglin, 1991). Outro membro da família muito estudado é a proteína Scalloped de *Drosophila melanogaster*, envolvida no desenvolvimento da asa e sistema nervoso central do embrião (Bray, 1999).

O domínio TEA é constituído por 66-68 aminoácidos localizados na região N-terminal das proteínas e a resolução da estrutura de Tead1 mostrou que o domínio é composto por três alfa-hélices sendo que a H3 é a responsável pela ligação ao DNA juntamente com uma volta (“loop”) L1 (Anbanandam, Albarado *et al.*, 2006). A conservação do domínio entre os membros desta família é alta, Tead1 e Scalloped tem uma similaridade de 99% neste domínio e Tead1 pode inclusive ser utilizado para substituir Scalloped no desenvolvimento da asa de drosófila (Deshpande, Chopra *et al.*, 1997).

A expressão da proteína mTead2 de camundongo foi detectada em altos níveis em blastocistos e persiste durante todo o desenvolvimento pré-implantação, sendo que embriões sem esta proteína não sobrevivem após este período (Kaneko, Cullinan *et al.*, 1997). São conhecidas quatro genes Tead em camundongos e humanos, expressos em diversos tecidos adultos. As proteínas Tead de um mesmo organismo aparentemente não apresentam redundância funcional e estão envolvidas na ativação e repressão de diversos genes, esta ativação diferencial pode resultar de uma associação das proteínas Tead com outras proteínas (Kaneko e Depamphilis, 1998).

Atualmente, sabe-se que a proteína Tead4 associa-se com a proteína YAP (Yes-associated protein) e juntas as duas proteínas ativam a transcrição de outros genes. YAP possui um domínio de ativação mas não o domínio de ligação ao DNA e por isto precisa da interação com um fator de ligação ao DNA como Tead4 (Cao, Pfaff *et al.*, 2008). A interação entre estas proteínas também foi observada entre seus homólogos de *D. melanogaster*, revelando um mecanismo conservado de ativação da via de sinalização Hippo (Wu, Liu *et al.*, 2008). Embora esta interação de Tead4 já esteja bem descrita, existem evidências da participação de Tead4 na ativação de outros genes de forma independente de YAP, sugerindo a presença de outras proteínas que interagem com Tead4 no controle do desenvolvimento embrionário (Ralston, Cox *et al.*, 2010).

1.3 Proteômica Funcional e o Sistema de Duplo-Híbrido

A grande quantidade de dados gerados por pesquisas genômicas é acompanhada por uma necessidade cada vez maior de se obter dados funcionais com relação às proteínas. Embora análises de genes por comparação de sequências e identificação de motivos em comum indiquem possíveis funções para as proteínas, estudos de genômica funcional como nocaute e perfil metabólico são muitas vezes necessários para determinar uma função exata. Contudo, estes dados nem sempre são suficientes e necessitam de alto investimento e trabalho excessivo (Immink e Angenent, 2002).

Para a melhor compreensão de processos biológicos, a determinação funcional de proteínas é indispensável. Processos metabólicos como síntese de DNA, ativação

transcricional, tradução protéica, localização de proteínas e transdução de sinal são todos dependentes da interação de complexos protéicos (Gietz, Triggs-Raine *et al.*, 1997). Nas últimas décadas, uma grande variedade de métodos foram desenvolvidos para detectar, analisar e quantificar interações proteína-proteína, incluindo espectroscopia de ressonância, NMR (“Nuclear Magnetic Resonance”), ensaios de duplo-híbrido, “tagging” de peptídeos combinado com espectrometria de massa e tecnologias baseadas em fluorescência (Lalonde, Ehrhardt *et al.*, 2008).

O sistema de duplo-híbrido (Fields e Song, 1989) é um ensaio genético molecular que detecta interações proteína-proteína e tornou-se um procedimento padrão para a biologia molecular. O método é conduzido na levedura *Saccharomyces cerevisiae*. Embora existam adaptações a outros organismos, *S. cerevisiae* é um modelo melhor para o desenvolvimento do sistema pela rapidez do crescimento, simplicidade de transformação e por ser uma célula eucariota, proporcionando ambiente apropriado para o estudo de proteínas deste clado *in vivo* (Fields e Sternglanz, 1994). O sistema é baseado no produto do gene GAL4 de leveduras, uma proteína com dois domínios funcionais que é responsável pela ativação de genes envolvidos no metabolismo da galactose. O domínio de ligação ao DNA (GAL4_{DL}) interage com a seqüência UAS do promotor de GAL1 e o domínio de ativação (GAL4_{DA}) estimula a transcrição. O método de duplo-híbrido (Figura 1a) consiste na separação das seqüências codificadoras para esses domínios (GAL4_{DL} e GAL4_{DA}) e na inclusão destas em plasmídios diferentes, de modo a ficarem fusionadas a seqüências de DNA codificando proteínas cuja interação deseja-se verificar (GAL4_{DL}proteínaA e GAL4_{DA}proteínaB). Os plasmídios devem ser introduzidos numa linhagem de levedura contendo genes repórteres. Um exemplo é o gene *LacZ*, que codifica a enzima β -galactosidase de *Escherichia coli* fusionado ao promotor de GAL1. A interação das proteínas A e B será responsável reconstrução funcional do fator de transcrição GAL4 levando à ativação da transcrição do gene repórter *LacZ*. A atividade de β -galactosidase pode ser detectada por um ensaio com o substrato cromogênico X-Gal (5-bromo-4-cloro-3-indolil- β -d-galactopiranosídeo), identificando as leveduras que contém a interação pela cor azul resultante da metabolização do X-Gal.

Posteriormente, o desenvolvimento do sistema permitiu a utilização de uma proteína de interesse (isca) para encontrar proteínas desconhecidas (presas) que interajam com ela dentre diversas outras (Chien, Bartel *et al.*, 1991). Para isto, normalmente é utilizada uma biblioteca de cDNA de interesse como fonte de possíveis proteínas presa. Esta versão do método utiliza um segundo gene repórter, que codifica uma enzima da via de síntese do aminoácido histidina, GAL1-HIS3 (Figura 1b). A região codificadora da proteína de interesse é fusionada ao domínio de ligação (GAL4_{DL}), num plasmídio contendo o gene TRP1 e a biblioteca de cDNA é construída em um vetor GAL4_{DA} contendo o gene LEU2. Leveduras

deficientes para os genes TRP1, LEU2 e HIS3 (com uma cópia de HIS3 controlada pelo promotor GAL1) são transformadas com os dois plasmídios, e então submetidas a uma seleção em meio SD sem os aminoácidos Trp, Leu e His. A presença de uma interação entre a proteína isca e um produto da biblioteca de cDNA permitirá o crescimento da levedura. Como a maioria das leveduras utilizadas para este fim apresentam transcrição basal do gene repórter, é comum a adição de um inibidor competitivo da enzima produto de HIS3 denominado 3-Aminotriazol. O 3-Aminotriazol funciona como um controle para que somente leveduras contendo uma forte ativação do gene repórter resultante de uma interação entre proteínas de interesse sejam capazes de crescer no meio.

O sistema de duplo híbrido consiste numa melhor abordagem para o estudo de interações proteína-proteína em relação aos estudos *in vitro* devido à comum necessidade de algumas proteínas do ambiente celular para que ocorram as corretas modificações pós-traducionais, mesmo que muitas vezes o núcleo das leveduras não seja semelhante à situação real. Outra vantagem do sistema é a possibilidade de detectar mesmo interações fracas, embora isto se torne uma desvantagem pelo número elevado de falso-positivos (Van Crielinge e Beyaert, 1999). O ensaio também é considerado adequado para os estudos de interações protéicas por ser simples, barato e não depender de conhecimento prévio sobre as interações para se realizar uma triagem, além de poder ser automatizado (Tucker, Gera *et al.*, 2001).

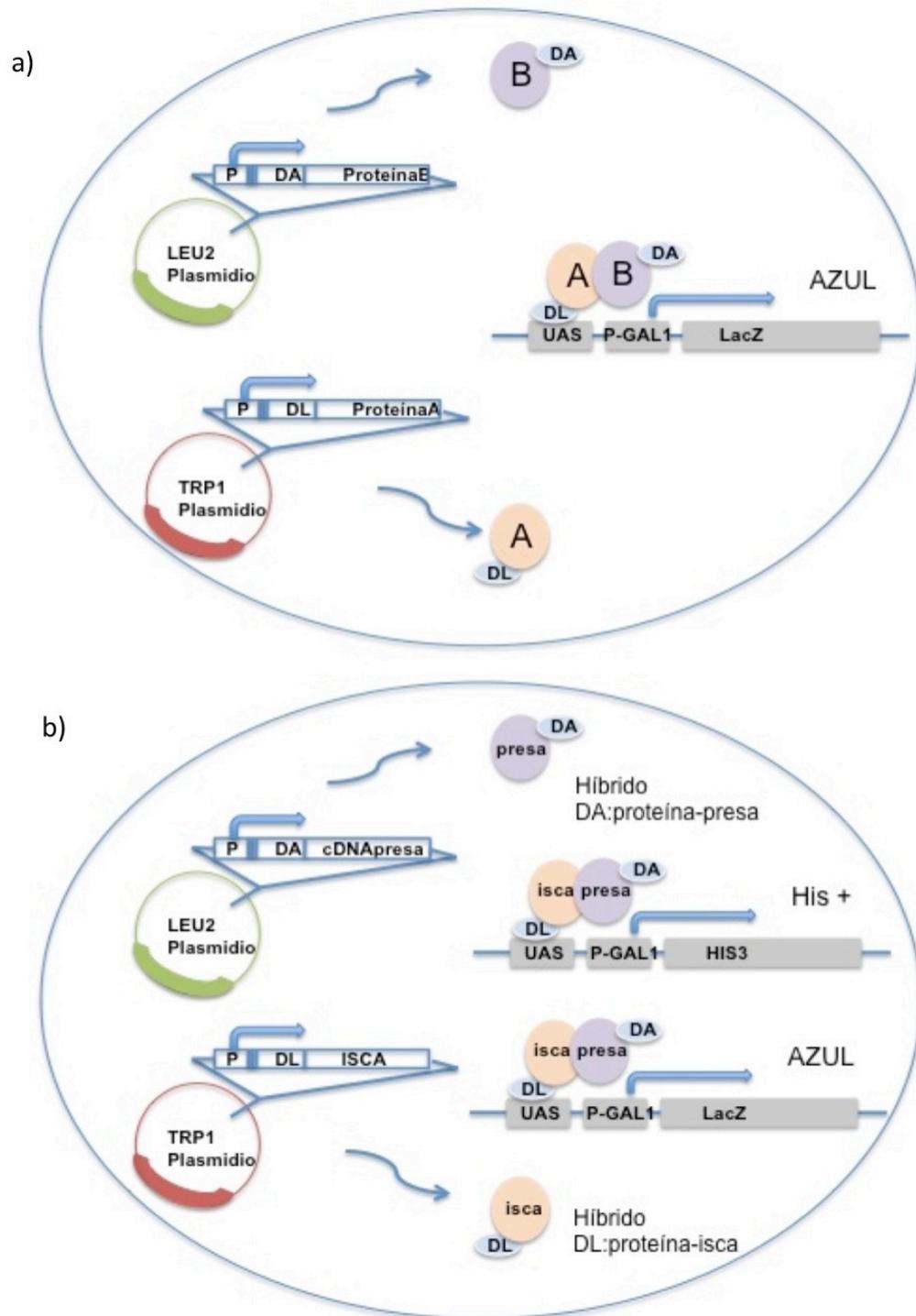


Figura 2: Sistema de Duplo-Híbrido em Leveduras. a) Representação esquemática do funcionamento do sistema de duplo-híbrido em leveduras para seleção de interações entre duas proteínas de interesse. DL: Domínio de ligação ao DNA; DA: Domínio de ativação da transcrição. Uma descrição mais detalhada pode ser encontrada no texto. b) Representação similar mostrando a seleção de interações entre uma proteína de interesse e um clone desconhecido de uma biblioteca de cDNA, são utilizados dois genes repórteres: HIS3 e *LacZ*.

O sistema também apresenta suas limitações. Uma grande desvantagem é a seleção de muitos falso-positivos, como já mencionado anteriormente, o que exige do pesquisador um esforço em vão, para analisar clones que inicialmente parecem interações relevantes e análises posteriores revelam ser espúrias. Refere-se aqui a falso-positivos do próprio sistema, não a interações reprodutíveis em leveduras e que porventura não venham a ocorrer no organismo em estudo. Esta seleção de falso-positivos mostra-se ainda mais elevada nos ensaios de triagem de bibliotecas (Chien, Bartel *et al.*, 1991). Embora a repetição dos ensaios possa distinguir corretamente as interações verdadeiras das falsas, fica clara a necessidade de análises laboriosas para avaliar a veracidade dos clones recuperados (Vidalain, Boxem *et al.*, 2004). A triagem de uma biblioteca é, além de tudo, muito trabalhosa, por ser realizada em um número elevado de placas de diâmetro grande. Uma mistura de transformação deve ser dividida em 100 placas de 15 cm de diâmetro (300 µL por placa) para uma triagem mais eficiente, reduzindo parcialmente o aparecimento de falso-positivos (Gietz, Triggs-Raine *et al.*, 1997).

Tradicionalmente, a seleção de leveduras contendo ativação do gene HIS3 no sistema de duplo-híbrido é feita em placas de meio SD -Trp -Leu -His suplementado com 3-Aminotriazol, um inibidor do produto de HIS3, como mencionado, que torna o crescimento possível somente para leveduras contendo uma ativação considerável do gene HIS3. Trabalhos reportando o uso de meio líquido foram descritos anteriormente como uma alternativa para a seleção de ativadores do gene HIS3 (Diaz-Camino, Risseeuw *et al.*, 2003), relatando maior eficiência quando comparado ao ensaio de *LacZ*. A possibilidade de automação deste ensaio foi relatado como também um grande atrativo (desenvolvimento em placas 96-deepwell). Para o sistema de tri-híbrido (Tirode, Malaguti *et al.*, 1997) também foi descrita uma abordagem em meio líquido que se mostrou capaz de distinguir interações verdadeiras de falsos positivos pela análise do crescimento da levedura no meio durante um intervalo fixo de 24 horas de crescimento (Srivastava e Lal, 2002).

Algumas proteínas podem tornar difícil a triagem pelo sistema por sua expressão ser tóxica para as leveduras tornando difícil ou até inviável o crescimento das leveduras (Van Criekinge e Beyaert, 1999). Fragmentos menores da proteína podem ser usados com o objetivo de eliminar a toxicidade (Gietz, Triggs-Raine *et al.*, 1997). Não há relatos de monitoramento paralelo do crescimento das leveduras sob efeito da isca, o que é normalmente feito nesses casos é esperar indefinidamente pelo crescimento das colônias nas placas.

Em nosso laboratório, Queiróz e colaboradores realizaram testes iniciais com as leveduras Y190 e HF7C para o desenvolvimento de uma metodologia de triagem em meio líquido. O objetivo da adaptação seria melhorar a seleção de interações de duplo-híbrido, diminuindo a incidência de falso-positivos. No trabalho, foi utilizada como modelo de

interação forte a proteína p53 e o antígeno Large T (Iwabuchi, Li *et al.*, 1998), enquanto as interações fracas foram representadas por um mutante conhecido de p53 (54.2) e o antígeno Large T. Os testes mostraram que a levedura Y190 em meio líquido era capaz de distinguir a interação forte (Gal4_{DL}-p53 + Gal4_{DA}-LargeT) e o controle negativo (Gal4_{DL}-p53 + Gal4_{DA}) em meio SDs -W-L-H líquido suplementado com 3-aminotriazol (3AT) após 24 horas de crescimento e após diluições. Concentrações de 3AT 2,5 mM foram suficientes para minimizar o crescimento devido a transcrição basal de HIS3 na levedura Y190 em procedimentos de 24 horas de cultivo e de diluições, não sendo necessárias concentrações da droga tão elevadas (35 mM) como normalmente utilizadas em meio SDs sólido (mais detalhes dos experimentos iniciais podem ser vistos em Material e Métodos).

A seleção em meio líquido também foi testada de forma robotizável, distribuindo a mistura de transformação em placas de 96 poços “deepwell”. Um cálculo da correspondência colônia-poço permitiu a comparação da seleção robotizável com a triagem em placas convencional mostrando uma equivalência dos métodos na recuperação de clones. O número de poços de uma placa “deepwell” para cada transformação deve ser maior se a incidência de clones é mais freqüente, enquanto que em triagens nas quais poucos clones são obtidos é possível diminuir o número de poços utilizados.

Outra metodologia testada foi a de crescimento em cultura única, com o intuito de selecionar clones com interações mais fortes através de uma competição. Esta metodologia só foi possível com a levedura Y190, pois testes iniciais mostraram que a linhagem HF7C não era capaz de distinguir interações fortes e fracas. Os testes de competição mostraram que a cultura contendo leveduras portando a interação forte, independentemente da presença de leveduras com interação fraca ou controle negativo, apresentou crescimento e as leveduras resultantes foram positivas no teste *LacZ*. Criando uma nova perspectiva para a triagem de interações de duplo-híbrido, de maneira mais eficiente e prática. Todavia, as simulações que desenvolveram o sistema precisavam ser testadas em um ambiente de seleção real e, se possível, desenvolvidas também com a colaboração de outros grupos experientes no sistema. Neste trabalho, buscamos o preenchimento destas perspectivas.

1.4 Bioinformática na Integração do Conhecimento

O campo da bioinformática possibilita atualmente um extenso número de pesquisas focadas na integração dos dados produzidos por técnicas de sequenciamento, análise de expressão diferencial, estudos de interação entre proteínas dentre outras técnicas atualmente adaptadas para coleta de dados em larga escala. A organização da informação em bases de dados secundárias, onde o conhecimento é organizado, é essencial num momento em que o poder de obtenção de dados em larga escala cresce aceleradamente. Assim, várias iniciativas organizam o conhecimento de interações gênicas e a subsequente

comparação de sequências permite uma rápida propagação desse conhecimento a outros genomas sequenciados.

O campo da Biologia de Sistemas propõe a integração destes dados para que a fisiologia de um organismo seja estudada por completo ao contrário de uma coleção de dados separados. A visão do sistema como um todo pode levar a soluções alternativas nas áreas de aplicação como biotecnologia e medicina (Hecker, Lambeck *et al.*, 2009).

1.4.1 Vias regulatórias

Diversas bases de dados compreendem vias metabólicas e regulatórias de processos biológicos importantes. O iPath (Letunic, Yamada *et al.*, 2008) do European Molecular Biology Laboratory permite a visualização e análise de um mapa metabólico completo de processos biossintéticos para o organismo de escolha e ainda manipulações de dados externos e sua integração com a via de interesse. Outro exemplo é a ferramenta BioCarta que procura oferecer acesso a vias de diversos processos armazenadas para consulta. O site BioCyc (<http://biocyc.org/>) possui uma coleção de bases de dados de vias, cada base de dados compreende as vias metabólicas de um organismo com genoma sequenciado (Karp, Ouzounis *et al.*, 2005). BioCyc é também uma fonte de informação em genômica comparativa e disponibiliza software que podem ser usados no estudo de novos organismos. Esta coleção todavia é focada apenas em vias metabólicas e em nenhum dos diversos outros processos regulatórios biológicos. Uma grande iniciativa para catalogar e permitir consulta a diversas vias metabólicas e regulatórias é a Kyoto Encyclopedia of Genes and Genomes, analisada mais detalhadamente a seguir.

1.4.2 KEGG

A base de dados KEGG (Kyoto Encyclopedia of Genes and Genomes) é um conjunto de recursos bioinformáticos que contém diversos pontos de entrada orientados pelos dados (<http://www.genome.jp/kegg/>). Por exemplo, a partir de uma consulta a uma via em KEGG Pathway como a glicólise podem ser acessados os grupos de ortólogos relacionados a cada gene que participa da via. KEGG Pathway é uma coleção de vias desenhadas manualmente representando o conhecimento da interação molecular e regulação para diversos processos. KEGG Orthology consiste em grupos de ortólogos definidos manualmente que correspondem a genes representados em KEGG Pathway. A aplicação de KEGG Orthology é ampliar a representação das vias para um número maior de organismos. KEGG Pathway e KEGG Orthology consistem portanto em ferramentas importantes para anotação de sequências de outros organismos.

1.4.3 Grupos de ortólogos

Dois genes são considerados ortólogos segundo sua história evolutiva (Fitch, 1970), quando estes genes divergiram após uma especiação. Atualmente o termo também é usado para se referir a genes com uma função conservada. Quando uma duplicação ocorre após um evento de especiação os dois genes contidos no mesmo organismo são considerados parálogos, mas com relação ao gene do organismo onde a duplicação não ocorreu são co-ortólogos e portanto, igualmente importantes em estudos evolutivos (Tatusov, Fedorova *et al.*, 2003). Como genes ortólogos normalmente ocupam nichos funcionais similares em organismos diferentes enquanto parálogos tendem a evoluir para uma diversificação funcional, podemos também ressaltar que a eficiência da anotação de um genoma depende muito da qualidade da identificação de genes ortólogos. A identificação de grupos de ortólogos é útil para estudos de evolução gene/proteína, genômica comparativa, identificação de sequências taxonomicamente restritas e anotação de genomas, por facilitar a identificação de proteínas e a atribuição de funções a elas (Altenhoff e Dessimoz, 2009).

KEGG Orthology consiste em uma excelente fonte de agrupamentos de ortólogos definidos manualmente e submetidos a rigorosa curadoria. As sequências agrupadas são provenientes de projetos genoma completos (Kanehisa e Goto, 2000). A base de dados COG do NCBI (Clusters of Orthologous Groups) é também curada manualmente e agrupa sequências do proteoma de diversos organismos procariotos e também de eucariotos unicelulares, as leveduras *Saccharomyces cerevisiae*, *S. pombe* e *C. albicans*. O sistema foi também expandido para organismos eucariotos complexos com genomas sequenciados, formando o KOG (Eukariotic Orthologous Groups) (Tatusov, Koonin *et al.*, 1997; Tatusov, Fedorova *et al.*, 2003). A base de dados EGGNOG possui agrupamentos de ortólogos gerados a partir de 630 genomas completos e atualizou os COGs e KOGs já mencionados provendo informação funcional para 88% das cerca de 2.200.000 proteínas anotadas (Muller, Szklarczyk *et al.*, 2010). Foi também desenvolvida a base de dados OrthoMCL como uma abordagem alternativa para a identificação automatizada de grupos de ortólogos em eucariotos, levando em consideração a similaridade entre parálogos “recentes” (Li, Stoeckert *et al.*, 2003).

Em nosso laboratório um procedimento de integração de dados enriquece a base de dados COG com agrupamentos UniRef50 da base UniProt (Suzek, Huang *et al.*, 2007), criando a base UECOG (Fernandes, Barbosa *et al.*, 2008). Recentemente, procedimento similar foi aplicado por Fernandes à base KEGG Orthology criando o banco de dados enriquecido UEKO (não publicado).

1.4.4 Identificando interações

Os diversos estudos genéticos e bioquímicos para detecção de interações proteína-proteína geraram uma grande quantidade de dados sobre interações que atualmente estão disponíveis em bases de dados curadas (Ooi, Schneider *et al.*, 2010). Estes dados tornaram-se uma excelente fonte para novas pesquisas. Dentre as bases de dados que contém informações sobre interações, destaca-se a iHOP (<http://www.ihop-net.org/>) que através de marcações de nomes de genes em diversas frases cria conexões entre elas e transforma a literatura contida na PubMed em um recurso navegável (Hoffmann e Valencia, 2004). Outra base de dados com interações é a STRING (<http://string-db.org/>) que contém interações físicas e associações funcionais entre proteínas e integra os dados provenientes de literatura (PubMed), contexto genômico, experimentos em larga escala e co-expressão conservada (Jensen, Kuhn *et al.*, 2009).

A mineração de texto tem, portanto, um papel fundamental nestas ferramentas e proporciona o acesso a interações dispersas na literatura. A extração de eventos biológicos da literatura por ferramentas de mineração de texto é essencial não só para abastecer as bases de dados de interações, mas também para a criação e anotação de vias (Ananiadou, Pyysalo *et al.*, 2010). Um software recente para extrair estas interações é o LAITOR (Barbosa-Silva, Soldatos *et al.*, 2010), capaz de identificar nomes de proteínas nas frases de um resumo e a palavra de biointeração que pode corresponder à associação entre eles.

2. JUSTIFICATIVA

O sistema de duplo-híbrido é um método fundamental na identificação de interações proteína-proteína, mas a triagem tradicional que utiliza placas de meio sólido seletivo apresenta muitas dificuldades para a aplicação deste método, especialmente em larga escala. O desenvolvimento de uma triagem em meio líquido é promissor para que este método se torne mais prático e também mais eficiente já que o crescimento de falsos positivos do sistema em estudos de culturas líquidas é menor. Contudo, em todas as aplicações de culturas líquidas para este sistema a abordagem não foi utilizada para triar bibliotecas complexas de cDNA, a qual torna-se mais favorável se utilizado o método de alta eficiência desenvolvido por Gietz e colaboradores.

O desenvolvimento embrionário pré-implantação é um processo controlado por diversas regulações que anteriormente só foram retratadas individualmente, mesmo em revisões da área. A ausência desta via complexa de regulações em importantes bancos de dados como KEGG Pathways é resultado da falta de uma publicação contendo todas as regulações envolvidas.

A importância de Tead4 no desenvolvimento embrionário é parcialmente compreendida. A associação com a proteína YAP parece ser insuficiente para resultar em todas as regulações na qual este fator está envolvido. Evidências como o fenótipo diferente apresentado pelo nocaute de Tead4 ou de YAP corroboram esta hipótese (Matthew J. Kohn - NYS Dept. of Health; comunicação pessoal). É interessante, portanto, a busca por mais interações que envolvem esta proteína.

Assim como a regulação completa do desenvolvimento não foi publicada, a genômica comparativa do processo não é conhecida. Sabe-se a função ancestral para alguns dos genes, mas a noção de qual porção do desenvolvimento tem origem antiga e quais mecanismos foram incorporados posteriormente ainda falta. Para conhecer a ancestralidade de genes os agrupamentos de homólogos são essenciais e precisam ser corretos. O laboratório de Biodados dispõe de técnicas capazes de melhorar os agrupamentos existentes em bases como KEGG Orthology, KOG e EGGNOG.

Neste trabalho procuramos a validação de uma nova triagem para o sistema de duplo-híbrido em leveduras, utilizando o meio líquido. A aplicação do sistema com a nova triagem será também utilizada para buscar novas proteínas que interagem com o fator de transcrição Tead4 durante o desenvolvimento embrionário. Visando um maior entendimento do período pré-implantação do desenvolvimento, técnicas de mineração de texto serão utilizadas para a integração das regulações conhecidas deste processo em uma só via. Os genes compreendidos nesta via serão utilizados em experimentos de agrupamento de homólogos que possibilitam a determinação da ancestralidade destes genes.

3.OBJETIVOS

3.1 Objetivo Geral

Identificar interações proteína-proteína pelo sistema de duplo-híbrido com triagem líquida e criar uma via de regulações compreendidas no desenvolvimento embrionário pré-implantação.

3.2 Objetivos Específicos

- Validar a triagem líquida para o sistema de duplo-híbrido
- Agrupar as regulações conhecidas do desenvolvimento pré-implantação em uma via
- Criar agrupamentos de homólogos para genes da via de pré-implantação
- Utilizar o sistema de duplo-híbrido para buscar interações com Tead4

4. MATERIAL E MÉTODOS

4.1 Duplo-híbrido em leveduras

4.1.1 Leveduras

As leveduras utilizadas neste trabalho foram *Saccharomyces cerevisiae*, linhagens KGY37 (Hemming, Agatep *et al.*, 2001) e Y190 (Clontech). As especificações genotípicas, genes repórteres e marcadores genéticos de ambas as leveduras estão especificados na Tabela 1. A linhagem KGY37 foi escolhida por ser comumente utilizada em triagens do sistema de duplo-híbrido no laboratório do Dr. R. Daniel Gietz.

Tabela 1: Especificações das linhagens de *Saccharomyces cerevisiae* validadas para os protocolos de duplo-híbrido com triagem líquida.

Linhagem	Genótipo	Genes Repórteres	Marcadores Genéticos
KGY37	<i>MATα</i> , <i>ade2</i> , <i>trp1Δ-901</i> , <i>leu2Δ-inv::pUC18</i> , <i>his3Δ-200</i> , <i>gal4Δ</i> , <i>gal80Δ</i> , <i>ura3Δinv::GAL1-LacZ</i> , <i>lis2Δ-inv::GAL1-HIS3</i>	<i>HIS3</i> , <i>LacZ</i>	<i>trp1</i> , <i>leu2</i>
Y190	<i>MATα</i> , <i>ura3-52</i> , <i>his3-200</i> , <i>lys2-801</i> , <i>ade2-101</i> , <i>trp1-901</i> , <i>leu23</i> , <i>112</i> , <i>gal4Δ</i> , <i>gal80Δ</i> , <i>cyhr2</i> , <i>LYS2::GAL1_{UAS}-HIS3_{TATA}-HIS3</i> , <i>URA3::GAL1_{UAS}-GAL1_{TATA}-LacZ</i>	<i>HIS3</i> , <i>LacZ</i>	<i>trp1</i> , <i>leu2</i> , <i>cyhr2</i>

Manutenção e cultivo de leveduras

Leveduras utilizadas nos ensaios foram mantidas em meio de cultura sólido YPAD (leveduras sem plasmídeo) ou SDs (SD seletivo, leveduras com plasmídeo) a 4°C. Periodicamente, as leveduras eram semeadas em um novo meio para serem mantidas viáveis. As culturas líquidas de leveduras eram mantidas sob agitação e incubadas a 30°C. Em culturas líquidas era também adicionado o antibiótico Ampicilina na concentração final de 100 µg/ml.

4.1.2 Bactérias

A bactéria utilizada foi *Escherichia coli* da linhagem DH5 α , eletrocompetente (genótipo: supE44, DlacU169, f80, LacZDM15, hsdR17, recA1, 'endA1, gyrA69, thi-1, relA1).

Manutenção e cultivo de bactérias

As bactérias eram inoculadas em meio LB líquido suplementado com antibiótico e incubadas a 37°C durante 12-16 horas. As bactérias também eram inoculadas em placas de meio sólido LB suplementado com antibiótico e incubadas a 37°C por 12 horas, sendo depois mantidas a 4°C.

4.1.3 Plasmídios

Os plasmídios utilizados nas triagens aqui reportadas foram pAS1 e pAS2 (iscas) e pACT2 e pGAD10 (bibliotecas). Os mapas estão representados na Figura 4.

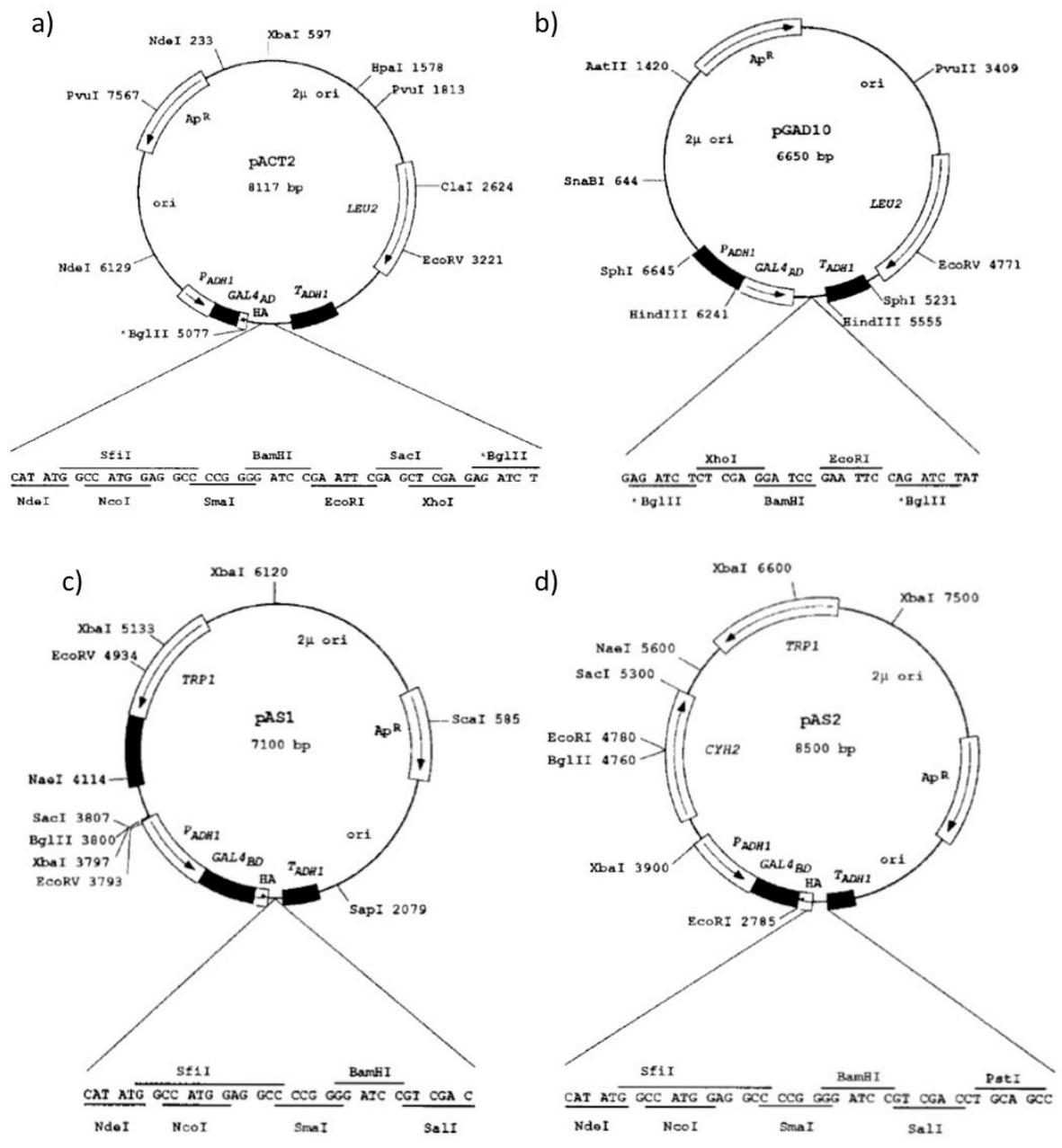


Figura 3: Mapas dos plasmídios utilizados nos experimentos de duplo-híbrido. a) pACT2 b) pGAD10 c) pAS1 d) pAS2

4.1.4 Bibliotecas de cDNA

Testículo de Camundongo (“Mouse Testis”) (CLONTECH)

Vetor: pACT2

Sítio de clonagem: *XhoI/EcoRI*

Total de clones independentes: 3,5 x 10⁶

cDNA faixa de tamanho: 0,4 - 4,0Kb

cDNA tamanho médio: 2,0Kb

Músculo Esquelético Humano (“Human Skeletal Muscle”) (CLONTECH)

Vetor: pGAD10

Sítio de clonagem: *EcoRI*

Total de clones independentes: $2,0 \times 10^6$

cDNA faixa de tamanho: 0,6 kb - 4,0 kb

cDNA tamanho médio: 1,3 kb

Cérebro de Camundongo (“Mouse Brain”) (CLONTECH)

Vetor: pACT2

Sítio de clonagem: *XhoI/EcoRI*

Total de clones independentes: $3,5 \times 10^6$

cDNA faixa de tamanho: 0,4 - 4,0 Kb

cDNA tamanho médio: 2 kb

Embrião de 7 dias (“7-day Embryo”) (CLONTECH)

Vetor: pGAD10

Sítio de clonagem: *EcoRI*

Total de clones independentes: $3,5 \times 10^6$

cDNA faixa de tamanho: 0,4 - 3,8 Kb

cDNA tamanho médio: 1,9 kb

4.1.5 Protocolos

Transformação de leveduras

Quick and Easy (Gietz e Schiestl, 2007b)

Este método pode ser utilizado para transformar células de *S. cerevisiae* em vários estágios de crescimento e armazenamento. O rendimento obtido é, contudo, menor e o método deve ser utilizado quando uma quantidade pequena de transformantes é suficiente. um volume semelhante a 50 μL de células é coletado de uma placa YPAD/ágar e dissolvido em 1 mL de água MilliQ em um tubo de microcentrifuga (1,5 mL). As células são então centrifugadas a 13000 g por 30 segundos e o sobrenadante é descartado. Ao pellet adiciona-se 240 μL de polietilenoglicol (PEG-3350 50% p/v), 36 μL de acetato de lítio (LiAc 1,0M), 50 μL de DNA carregador (ssDNA $2,0 \text{ mg mL}^{-1}$), e 34 μL de água MilliQ esterilizada e o DNA plasmidial (até 1 μg). O tubo é vortexado até que o sedimento seja ressuspenso e incubado por 30 minutos a 42°C. A transformação é centrifugada a 13000 g por 30 segundos a 20°C e o sobrenadante é retirado com uma micropipeta. Adiciona-se 1 mL de

água MilliQ esterilizada para ressuspender o sedimento com uma micropipeta. Amostras de 100 ou 200 µL da transformação são então semeadas em meio SD seletivo sólido e as placas são incubadas a 30°C por 3-4 dias.

Large Scale High Efficiency (Gietz e Schiestl, 2007a)

A eficiência deste método é alta o suficiente para triar bibliotecas complexas de cDNA. Uma reação em larga escala pode compreender 30, 60 ou até 120 receitas de transformação. Uma colônia da linhagem de levedura é inoculada em 50mL de meio YPAD e incubada durante a noite a 30°C sob agitação constante (aproximadamente 200 r.p.m.). Após o crescimento, determina-se a concentração da cultura utilizando um hemocitômetro e então, para 120 transformações é feito um novo inóculo com $2,5 \times 10^9$ células em 280 mL de meio YPAD duas vezes concentrado, a concentração final da cultura deve ser 5×10^6 células/mL. Uma nova incubação é então feita a 30°C por aproximadamente quatro horas (duas divisões celulares). Um mix de transformação é preparado de acordo com o volume necessário. Para 120 transformações adiciona-se 28,8 mL de PEG 50% (p/v); 4,32 mL de LiAc 1M; 6,0 mL de ssDNA 2 mg/mL e 4,08 mL de DNA plasmidial e água MilliQ esterilizada, totalizando 43,2 mL. As células da cultura são coletadas por centrifugação a 3000 g por 5 minutos e ressuspensas em água MilliQ esterilizada (1/2 do volume da cultura). Novamente as células são centrifugadas e ressuspensas em água MilliQ (1/5 do volume anterior), os *pellets* são então combinados e centrifugados da mesma forma. Adiciona-se sobre o *pellet* o mix de transformação, vortexando até que ele seja dissolvido. A transformação é incubada a 42°C por 50 minutos (linhagens Y190 e KGY37), misturando o tubo a cada 5 minutos. Após este período, centrifuga-se a transformação e adiciona-se 50 mL de meio SD^s deixando o *pellet* descansar por 10 minutos e depois ressuspendendo-o.

Em todas as transformações o DNA carregador deve ser fervido previamente por 5 minutos.

Triagem em cultura líquida

Experimentos anteriores conduzidos por Queiróz e outros membros do laboratório revelaram o melhor protocolo de diluições para a análise de duplo-híbrido em meio líquido. Os ensaios foram conduzidos por co-transformação de leveduras *S. cerevisiae* das linhagens HF7c e Y190 com 1 µg de pISCA (expressão de Gal4_{DL}-p53), 1 µg de pSEMPRESA (expressão de Gal4_{DA}) e 0; 0,1; 1; 10 ou 100 ng de pPRESA (expressão de Gal4_{DA}-largeT). Simulando assim a triagem de clones positivos em concentrações crescentes em meio a clones de controle negativo. Os protocolos, mostrados na Figura 4, consistiram em diluições (1:10 e 1:100) de culturas de leveduras co-transformadas, durante

um período de 120 horas, finalizando-se pela análise de ativação de gene repórter HIS3 por medida de densidade do meio SDs líquido em espectrofotômetro. A diferença principal entre os protocolos consiste na amplitude (1:10 ou 1:100) e intervalo (após 24, 48 ou 72 horas de cultivo) da primeira diluição. Os resultados mostraram que protocolo F (Figura 4) resultava em maior eficiência na recuperação dos clones positivos. A primeira diluição deste protocolo é de 1:10 após 72 horas de crescimento pós-transformação, seguida por uma diluição 1:100 24 horas depois da primeira. Enquanto em leveduras *S. cerevisiae* HF7c este protocolo apresentou sensibilidade semelhante à observada na seleção convencional em placas, em leveduras *S. cerevisiae* Y190 a sensibilidade foi um pouco melhor.

testes	0 h	24 h	48 h	72 h	96 h	120 h
A	Transformação	1:100	1:100	1:100	1:100	OD
B			1:100	1:100	1:100	OD
C				1:100	1:100	OD
D		1:10	1:100	1:100	1:100	OD
E			1:10	1:100	1:100	OD
F				1:10	1:100	OD

Figura 4 - Protocolos de diluição das culturas líquidas. Esquema de diluições utilizados na padronização dos testes de duplo-híbrido em cultura líquida realizados por Queiróz e colaboradores. Ver texto para mais detalhes.

A partir dos resultados destes experimentos iniciais foi estabelecido um protocolo padrão para os ensaios em meio líquido que foi adotado em todas as triagens do presente trabalho e está descrito a seguir.

Após a transformação as leveduras são ressuspensas em meio líquido SDs -W-L-H suplementado com 3AT 0,5 mM e 2,5 mM para KGY37 e Y190, respectivamente. Ao meio deve ser também acrescentado o antibiótico Ampicilina na concentração de 100 µg/mL para evitar contaminações. O volume de meio corresponde a 10 mL para cada receita de transformação.

As transformações podem ser divididas em frascos (Erlenmeyer 200 mL) ou em placas “deepwell” (placa de polietileno “Deep-Well” de 96 poços, BD Biosciences). As culturas são então incubadas por 72 horas a 30°C sob constante agitação (incubadora refrigerada TE-422, Tecnal). Após este período é feita uma diluição na razão de 1:10. Para as placas “deepwell”, a maneira mais simples de realizar esta diluição consiste em distribuir 1,25 mL de meio líquido (SDs -W-L-H +3AT +Amp) em cada poço de novas placas “deepwell” com uma pipeta repetidora e então acrescentar 150 µL da cultura anterior com o auxílio de uma pipeta multicanal. Após uma nova incubação de 24 horas a 30°C realiza-se uma diluição na razão de 1:100. As culturas são então incubadas por mais 24 horas a 30°C e em seguida analisadas quanto à presença de crescimento (ativação do gene repórter HIS3). O crescimento das culturas dos frascos pode ser determinado por contagem das células em um hemocítômetro. O crescimento das culturas dos poços foi avaliado transferindo 400 µL de cada poço para uma placa de ELISA com uma pipeta multicanal e medindo a densidade em leitor de placas de ELISA (Spectra Max Plus, Molecular Devices) com comprimento de onda de 600 nm (OD600). A metodologia está esquematizada na Figura 5.

Ensaio para verificação da ativação do gene repórter *LacZ*

As culturas que apresentam crescimento são plaqueadas em meio sólido SDs -W-L. Após o crescimento de colônias isoladas, um filtro de nitrocelulose é aplicado na superfície da placa e logo retirado. O filtro passa então por congelamento (com nitrogênio líquido ou freezer -80°C) e descongelamento à temperatura ambiente. Adiciona-se então a solução contendo X-Gal 20 mg/mL (5-bromo-4-cloro-3-indolil-β-d-galactopiranosídeo, Sigma) em tampão-Z (“Z-buffer”) e incubado a 30°C. Colônias capazes de ativarem o gene repórter *LacZ* adquirem coloração azul devido a metabolização do substrato X-Gal pela β-galactosidase, produto do gene repórter.

Soluções:

Tampão-Z Na₂HPO₄ 60 mM (Synth)
 (“Z- buffer”) Na₂H₂PO₄ 40 mM (Synth)
 KCl 10 mM (Synth)
 MgSO₄ 1 mM (Synth)
 SDS 0,025% (p/v)
 β-mercaptoetanol 0,27 mL (v/v)
 pH 7

X-Gal – 2% (p/v) em N,N dimetilformamida (DMF, Aldrich) ou Dimetil Sulfoxido (DMSO) e utilizado na concentração final de 20 mg/mL.

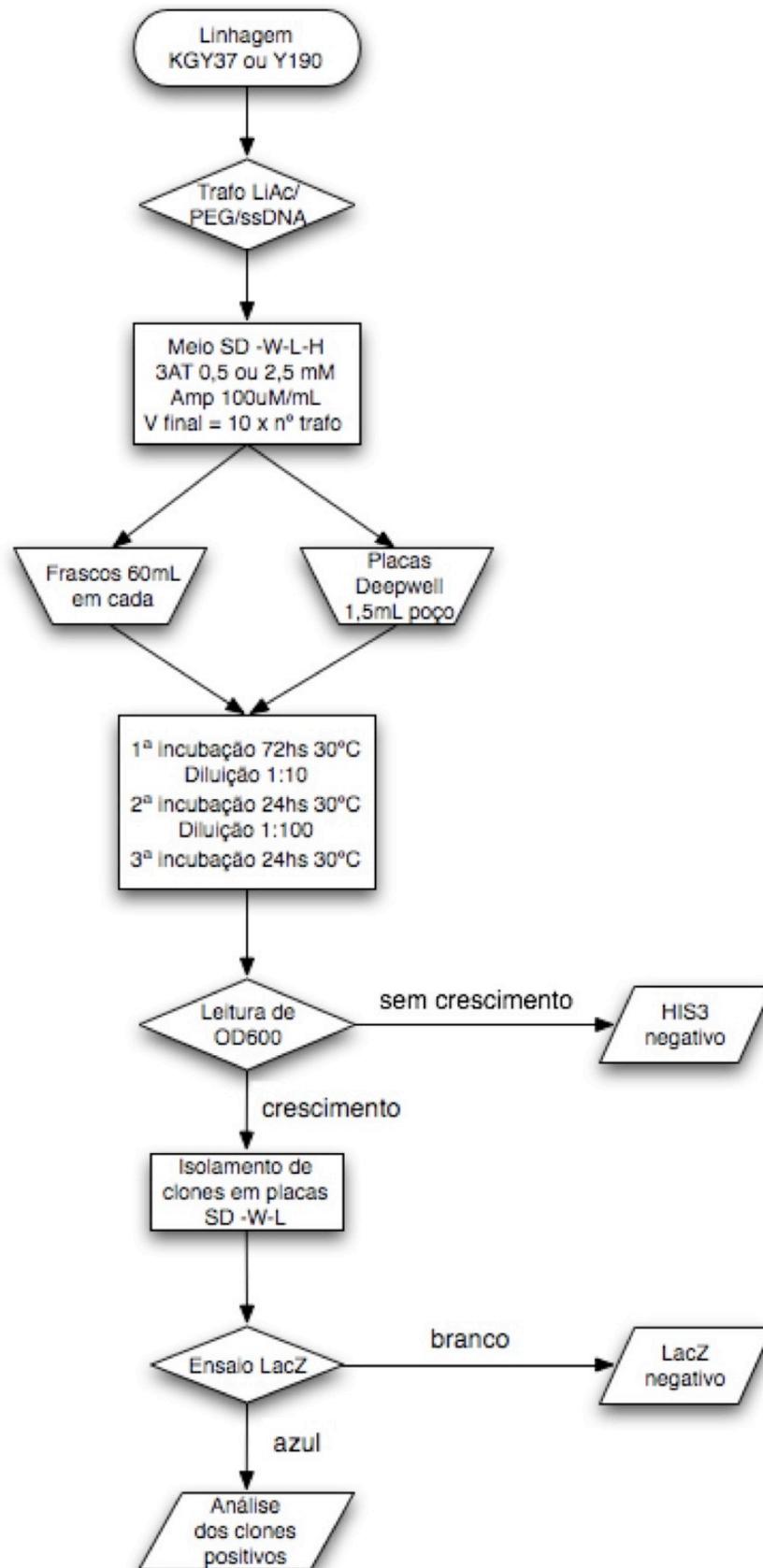


Figura 5: Protocolo de triagem em cultura líquida. Esquema ilustrando a triagem de duplo-híbrido em cultura líquida. Detalhes do método estão descritos no texto.

Minipreparação para extração de plasmídio de levedura

As leveduras são inoculadas em 5 ml de meio SD-W-L a 30°C sob agitação (incubadora refrigerada TE-422, Tecnal) durante 12 horas. Desta cultura, 1 mL é centrifugado e o sobrenadante descartado. As células depositadas eram ressuspensas no meio residual. O volume de 400 µL de solução de lise é adicionado e homogeneizado. Adiciona-se um volume de 400 µL da solução fenol:clorofórmio:álcool isopropílico (25:24:1) e 0,3 g de pérolas de vidro (Sigma, 425-600 *microns*, G-8772) lavadas com ácido e a cultura é homogeneizada por 2 minutos. A cultura é centrifugada durante 5 minutos. Retira-se a fase aquosa e seu volume é medido. Adiciona-se ao tubo um volume de solução 10 M de acetato de amônio correspondente a 0,25 do volume inicial, 1 µL de glicogênio e 2V de etanol absoluto gelado. O material é homogeneizado e incubado a -20°C durante 10 minutos. O sedimento é então lavado com etanol 80% e ressuspensado em 25 µL de água milliQ estéril. O DNA deve ser armazenado a -20°C.

Solução de lise para levedura:

- Triton x 100 (Sigma)
- SDS 10% (Sigma)
- NaCl 100 mM (MERK)
- Tris-HCl 10 mM, pH 8,0 (Sigma)
- EDTA 0,5 mM (Synth)
- Água MiliQ esterilizada

Eletroporação

Os plasmídios extraídos das leveduras são utilizados para transformar bactérias DH5α eletrocompetentes. Uma alíquota de 40 µL destas bactérias é descongelada por 15 minutos e incubada com 1 µL de um dos plasmídios por 5 minutos em gelo. As bactérias são então submetidas a uma descarga elétrica de 2500V (Eletroporador BIORAD). Por fim as bactérias são incubadas em meio SOC a 37°C durante 1 hora e semeadas em meio LB sólido suplementado com antibiótico (ampicilina 100 µg/mL).

Minipreparação para extração de plasmídio de bactérias

As colônias de bactérias são inoculadas em 10 mL de meio LB líquido suplementado com ampicilina (100 µg/mL) e incubadas por uma noite a 37°C. Aproximadamente 5 mL da cultura são centrifugados e submetidos ao protocolo do kit NucleoSpin® Plasmid (Macherey-

Nagel). O DNA purificado obtido pode ser usado nas reações de sequenciamento ou nas transformações.

Sequenciamento

O DNA plasmidial purificado na etapa anterior é submetido a uma eletroforese em gel de agarose 1% contendo brometo de etídio na concentração final de 0,5 µg/ml (Life technologies) em tampão TAE 1x (Tris base 40 mM pH 7,2, Acetato de Sódio 20 mM, Sal tetrassódico de Etilenodiaminatetracético (EDTA) 1 mM), realizada a 60V por 40 minutos. A eletroforese permite determinar a concentração dos plasmídios pela comparação com um padrão de peso molecular. Os insertos são seqüenciados numa reação com 200 ng do plasmídio, 10 pmol dos iniciadores apropriados e 4 µL do Kit DYEnamic™ terminator e estas reações são realizadas em seqüenciador automático MegaBace 1000 DNA Analysis System (GE Healthcare).

Reconstrução da ativação do gene *LacZ*

Para eliminar possíveis falso-positivos selecionados no sistema, plasmídios Leu (biblioteca) recuperados dos clones positivos encontrados na triagem são utilizados para uma transformação de novas leveduras. As leveduras transformadas com o plasmídio Trp contendo a isca original e o plasmídio Leu recuperado são testadas para verificar a ativação do gene *LacZ*. A re-confirmação da ativação é uma evidência de que o plasmídio Leu responsável pela expressão do cDNA cujo produto apresenta interação com a proteína isca foi isolado e a isca não sofreu modificações ao longo da triagem.

4.2 Bioinformática

4.2.1 PubMed

A base de dados NCBI PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) foi utilizada nas pesquisas de literatura para os procedimentos de mineração de texto. Após realizar uma pesquisa com o termo de busca desejado (ex: “preimplantation development”) os números de identificação dos resumos (PMIDs) devem ser salvos como um arquivo de texto. A Figura 6a mostra a página da PubMed após uma pesquisa realizada e destaca em amarelo os PMIDs. No canto direito superior da Figura 6a pode ser visualizado o procedimento para criação de um arquivo contendo todos os PMIDs da busca.

A PubMed também deve ser utilizada para procurar resumos que contém muitas informações interessantes para a pesquisa desejada. Estes trabalhos (cerca de 10) serão usados na próxima etapa.

4.2.2 Medline Ranker

Para classificar os resumos obtidos na busca da PubMed quanto à importância das informações que eles contém para a pesquisa desejada, os PMIDs encontrados na busca PubMed são adicionados ao software Medline Ranker (Fontaine, Barbosa-Silva et al., 2009) (<http://cbdm.mdc-berlin.de/tools/medlineranker/>) no campo chamado “background set”, como mostra a Figura 6b. Os PMIDs dos resumos selecionados manualmente como informativos devem ser adicionados no campo chamado “test set”. A Figura 6b também mostra como deve ser selecionada a escolha de quais resumos devem ser classificados, a seleção é de “background set”. A Figura 6c mostra a página obtida após a classificação dos resumos; para o trabalho aqui descrito, selecionamos os 1000 resumos com a melhor classificação para continuar a análise.

4.2.3 PESCADOR

Os PMIDs dos 1000 resumos classificados anteriormente são inseridos na plataforma PESCADOR (ainda em publicação, esta pesquisa foi realizada como estudo de caso). O PESCADOR é uma aplicação web desenvolvida por Barbosa-Silva e colaboradores e será disponibilizada para uso público em breve. O software contido na plataforma que marca os nomes de proteínas (“terms”) e suas biointerações é chamado LAITOR (Barbosa-Silva, Soldatos *et al.*, 2010). As interações encontradas por este software podem ser de 4 tipos: a interação de tipo 1 é sinalizada quando são encontrados na mesma frase dois nomes de proteínas com uma palavra de biointeração entre eles; o tipo 2 consiste em dois nomes de proteínas com uma palavra de biointeração na mesma frase mas não entre eles; no tipo 3 encontra-se dois nomes de proteínas sem uma palavra de biointeração na mesma frase; no tipo 4 o software encontra os dois nomes de proteínas em frases distintas. Na Figura 6d pode ser observada a página inicial do PESCADOR, no campo superior devem ser inseridos os PMIDs dos resumos que se deseja marcar e também podem ser adicionados à pesquisa alguns conceitos desejados pelo usuário para melhor visualizar os resultados relacionados a estes (exemplo: Trophectoderm).

a)

NCBI Resources How To My NCBI Sign In

PubMed.gov U.S. National Library of Medicine National Institutes of Health

Search: PubMed
preimplantation development Search Clear

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to: Filter your results:

Results: 1 to 20 of 4114

- Sex-related physiology of the preimplantation embryo. Gardner DK, Larman MG, Thouas GA. Mol Hum Reprod. 2010 May 25. [Epub ahead of print] PMID 20501630
- Birth of a healthy infant following preimplantation PKHD1 haplotyping for autosomal recessive p... using multiple displacement amplification. Lau EC, Janson MM, Roesler MR, Avner ED, Strawn EY, Bick DP. J Assist Reprod Genet. 2010 May 20. [Epub ahead of print] PMID 20490649
- Brg1 is required for Cdx2-mediated repression of Oct4 expression in mouse blastocysts. Wang K, Sengupta S, Magnani L, Wilson CA, Henry RW, Knott JG. PLoS One. 2010 May 12;5(5):e10622. PMID 20485553

Choose Destination: File (selected), Clipboard, Collections, E-mail, Order

Download 4114 items. Format: PMID List (selected), Sort by: Recently Added (selected), Create File

Search details: preimplantation[All Fields] AND ("growth and development"[Subheading] OR ("growth"[All Fields] AND "development"[All Fields]) OR "growth and development"[All Fields] OR

b)

Medline Ranker	about	query page	tutorial page	contact	CBDM group page
Abstracts selection		Short documentation			
<p>The query topic (the training set) is defined by:</p> <ul style="list-style-type: none"> <input type="radio"/> the following PubMed query (NEW!) <input type="radio"/> all the following MeSH terms (tree top) <input checked="" type="radio"/> the following list of PMIDs <p>18160257 15880404 18035060</p> <p>one per line</p> 		<p>The training set is a set of scientific articles related to the same topic. You can write a PubMed query to build automatically a training set. Alternatively, you can use biomedical terms from the Medical Subject Headings (MeSH). Please input <u>only exact terms</u>. These terms can be found in the MeSH Browser or hierarchical tree. The procedure is detailed in the tutorial section. The training set can be also defined with your own list of articles (identified by PubMed identifiers, e.g. PMIDs). You can get a list of PMIDs in few clicks from the PubMed interface: go to the PubMed page, make a query, select 'UI List' in the Display menu, and Send to file. Please see the related tutorial.</p> <p>For demo: click on one of the following training set, and then on the 'Rank it' button:</p> <ul style="list-style-type: none"> Mitochondria Host-Pathogen Interactions Alternative Splicing Cardiovascular Diseases and Smoking Phosphorylation and Protein Binding Phosphorylation-dependent molecular processes Neoplasms and Stem Cells Up-Regulation of physiological processes in Alzheimer Disease Computational Biology 			
<p>The reference (the background set) is defined by:</p> <ul style="list-style-type: none"> <input type="radio"/> the whole Medline database <input checked="" type="radio"/> the following list of PMIDs <p>557421 860290 850212</p> <p>one per line</p>		<p>The background set should use the whole medline database or a random selection of articles. You can also provide your own list to take into account a bias in the test set.</p>			
<p>The abstracts to be ranked (the test set) are defined by:</p> <ul style="list-style-type: none"> <input type="radio"/> the training set <input checked="" type="radio"/> the background set <input type="radio"/> 10 000 randomly chosen recent abstracts <input type="radio"/> publications of the last 1 month(s) <input type="radio"/> the 1-year(s) old abstracts <input type="radio"/> the following list of PMIDs <p>one per line</p> <p>Rank it Reset</p>		<p>The test set defines abstracts which will be ranked by the MedlineRanker program. Ranking all the medline abstracts of the last months or years may be long. The processing speed is approximately 1 Million abstracts (~2 years old abstracts) per minute after initialization steps. The speed may vary depending on the server load.</p> <p>Please cite:</p> <ul style="list-style-type: none"> MedlineRanker: flexible ranking of biomedical literature. Fontaine JF, Barbosa-Silva A, Schaefer M, Huska MR, Muro EM, Andrade-Navarro MA. Nucleic Acids Research 2009 May 8; doi: 10.1093/nar/gkp353 <p>Please send me your comments and suggestions: JF's email</p>			

c)

The parameters

The training set: 7 / 7 abstracts.
 The background set: 3653 abstracts.
 The test set: 3653 abstracts from the background set.
 Initialisation: 25 seconds.

[top results](#) [cross validations](#) [discriminative word](#)

Significant results for the test set

Processing 3653 abstracts: 0% - 9% - 19% - 29% - 39% - 49% - 59% - 69% - 79% - 89% - 99% - done.

Ranked in less than 1 second.

colors: yellow for [training set pmids](#), green for [background set pmids](#), and brown for [discriminative words](#) (high  low weight).

[Click on the pmid to read the abstract with highlighted discriminative words](#)

Rank	PMID	Abstract Title	P-value
1	19732419	Promoter analysis of the rabbit POU5F1 gene and its expression in preimplantation stage embryos .	2.73e-04
2	17872502	Putative role of hyaluronan and its related genes , HAS2 and RHAMM , in human early preimplantation embryogenesis and embryonic stem cell characterization .	2.74e-04
3	15625129	Stably transfected human embryonic stem cell clones express OCT4 - specific green fluorescent protein and maintain self - renewal and pluripotency .	8.20e-04
4	20231280	Functional analysis of the Cdk7 , cyclin H , Mat1 complex in mouse embryonic stem cells and embryos .	1.09e-03
5	20018908	Cloning of complementary DNAs encoding structurally related homeoproteins from preimplantation mouse embryos : their involvement in the differentiation of embryonic stem cells .	1.37e-03
6	19915186	Involvement of a novel preimplantation - specific gene encoding the high mobility group box protein Hmgpi in early embryonic development .	1.64e-03
7	20485553	Brg1 is required for Cdx2 - mediated repression of Oct4 expression in mouse blastocysts .	1.91e-03
8	15880683	Eomesodermin is expressed in mouse oocytes and pre - implantation embryos .	2.19e-03
9	19582159	microRNA expression during trophectoderm specification .	2.46e-03
10	19564914	Transcription - independent heritability of induced histone modifications in the mouse preimplantation embryo .	2.73e-03
11	17287250	Mbd3 , a component of the NuRD co - repressor complex , is required for development of pluripotent cells .	3.01e-03
12	19960551	Epigenetic regulatory mechanisms during preimplantation development .	3.28e-03
13	19087264	RBM19 is essential for preimplantation development in the mouse .	3.55e-03
14	1821852	Specific expression of a retinoic acid - regulated , zinc - finger gene , Rex - 1 , in preimplantation embryos , trophoblast and spermatocytes .	3.83e-03
15	12514105	Multinotent cell lineages in early mouse development depend on SOX2 function .	4.10e-03

d)

PESCADOR
Platform for Exploration of Significant Concepts Associated to co-Occurrences Relationships.
[Input](#) | [Concepts](#) | [Retrieval](#) | [Development](#) | [Help](#)

cb dm **MDC**
Berlin-Buch

INPUT [Required]

Human user interface test
Choose a PubMed ID list file to upload:

Machine import interface
Use our programatic interface to deliver your PubMed IDs.
Data imported from:
Inactive

OR

Paste below your list of PubMed IDs (**one per line!**):

19732419
17872502
15625129
20231280
20018908
19915186
20485553
15880683
19582159

Delivered Pubmed IDs list:

[Example](#) [Clear](#) ["Start Analysis"](#) [Clear](#)

PLUS

Customized concepts

Load below the biological concepts to be checked for (**one per line!**)

Trophectoderm
TE
Inner Cell Mass
ICM
Differentiation
Pluripotency

[Example](#) [Clear](#)

RETRIEVAL

Load a previously analysed project:

Figura 6: Plataformas utilizadas nos experimentos de mineração de texto. a) Pesquisa PubMed com termos de interesse (ex: “preimplantation development”). Em amarelo está destacado o identificador do trabalho chamado de PubMed ID. **b)** Página inicial do software Medline Ranker demonstrando a inclusão de uma lista de PMIDs como “background set” que são classificadas quanto ao conteúdo em comparação com os resumos selecionados manualmente como informativos (cerca de 10) que formam o “test set”. **c)** Exemplo de resultado gerado pelo Medlline Ranker, lista dos resumos classificada em ordem de relevância. **d)** Página inicial da plataforma PESCADOR. O usuário pode adicionar a lista de resumos a serem marcados pelo software e também conceitos de interesse para a pesquisa que facilitam a visualização de biointerações relacionadas posteriormente.

4.2.5 Construção da via de regulações

Um esquema do procedimento para criação de uma via de regulações utilizando as ferramentas de mineração de texto aqui descritas está representado na Figura 7.

A via de regulações foi construída utilizando o programa Keynote iWORK 2009 (semelhante ao PowerPoint da Microsoft) e foi baseada na linguagem utilizada nas vias do KEGG Pathway, a KGML (“KEGG Markup Language” <http://www.genome.jp/kegg/docs/xml/>) que especifica como a via deve ser representada graficamente. Outra maneira de gerar a via de regulações é através de software como o CellDesigner (Noverre, Hucka *et al.*, 2009).

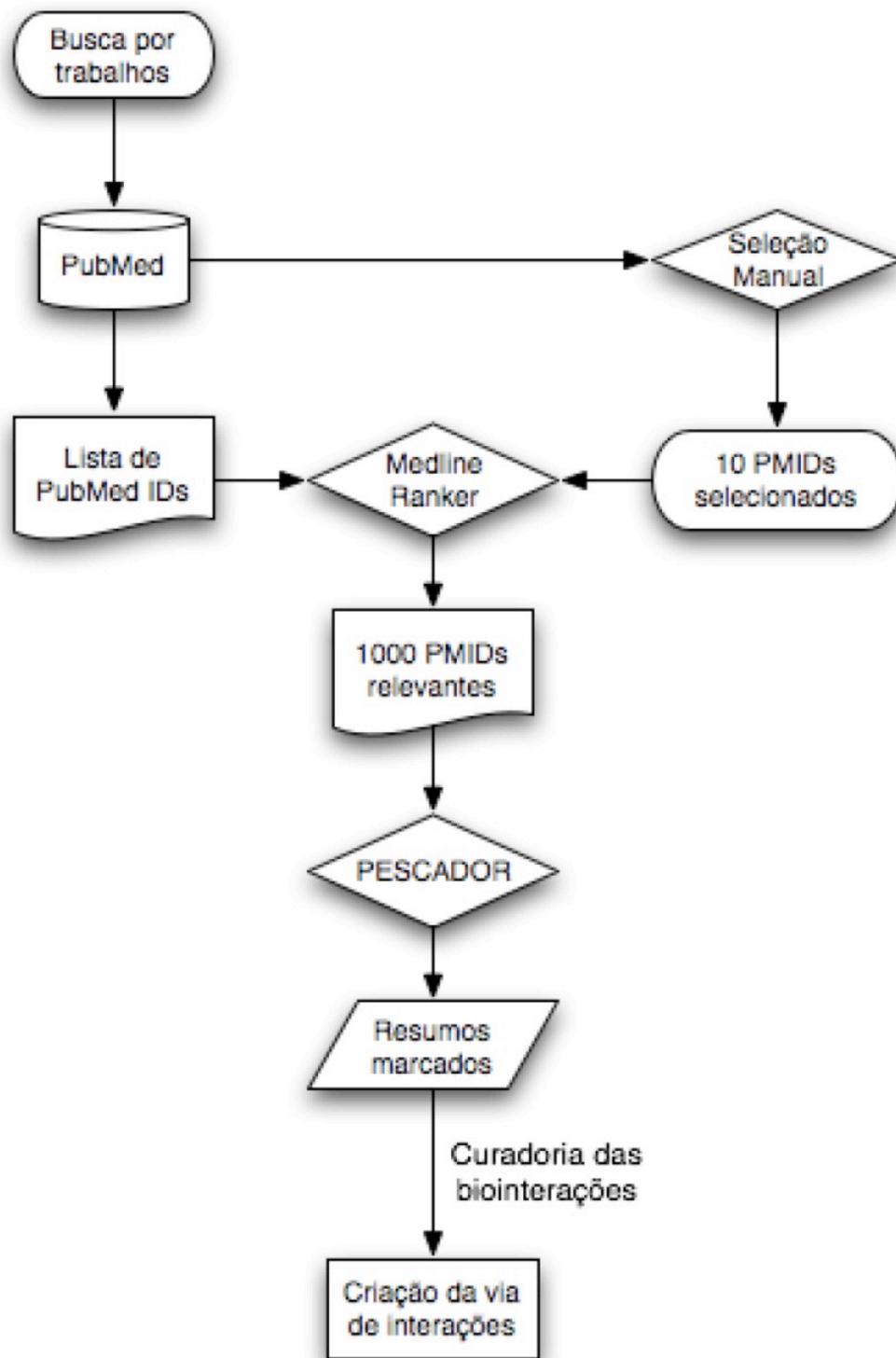


Figura 7: Procedimento para criação da via de interações. Inicialmente realiza-se uma pesquisa PubMed com os termos de interesse (ex: “preimplantation development”). A lista dos números de identificação dos trabalhos (PubMed IDs ou PMIDs) obtidos na pesquisa é introduzida no software Medline Ranker como “background set” junto com PMIDs de resumos selecionados manualmente como informativos (cerca de 10) que formam o “test set”. O software gera uma lista dos resumos classificada em ordem de relevância e cerca de 1000 resumos com a melhor classificação são escolhidos para alimentar o software PESCADOR. O PESCADOR identifica e marca símbolos de genes e biointerações. Com os resumos marcados é possível analisar as biointerações encontradas e gerar manualmente uma via para visualização gráfica destas informações.

4.2.6 SeedServer

Para a criação de agrupamentos de homólogos foi utilizado o software SeedServer. Este software foi desenvolvido por Guedes e colaboradores no laboratório Biodados e estará disponível em breve para uso público como um aplicativo web. Primeiramente, a lista de símbolos de genes que se deseja usar como semente (“seed”) é adicionada a uma tabela do MySQL e através de uma comparação com o banco de dados Gene (<http://www.ncbi.nlm.nih.gov/gene/>) podemos atribuir a cada símbolo um identificador, GeneID. Os identificadores são então utilizados juntamente com o banco de dados UniProt ID Mapping (<http://www.uniprot.org/>) para atribuir a cada GeneID um identificador UniProt das sequências, UniProtID. Para cada gene presente na via, duas sequências são utilizadas como “seed”, uma humana e a segunda de camundongo. O SeedServer realiza buscas através de dois componentes, o programa SeedLinkage (Barbosa-Silva, Satagopam *et al.*, 2008) e os bancos de dados KEGG Orthology (KO) e sua versão enriquecida UEKO (UniRef enriched KO) desenvolvida no laboratório Biodados por Fernandes e colaboradores (em publicação). Após a formação inicial dos agrupamentos, o SeedServer realiza uma comparação por alinhamento PSI-BLAST das sequências “seed” com todas as sequências recrutadas, eliminando aquelas que não se agrupam quando ocorre a convergência (as iterações param antes que a “seed” receba score menor que 50% de seu auto-score – módulo desenvolvido por Ribeiro, no laboratório). O procedimento para criação de grupos de homólogos pelo SeedServer está esquematizado na Figura 8.

4.2.7 Último Ancestral Comum

O último ancestral comum para os genes da via de regulações foi determinado com os dados obtidos dos agrupamentos feitos pelo SeedServer. Cada agrupamento de sequências possui uma tabela MySQL da qual podemos obter todos os identificadores taxonômicos para estas sequências, TxIDs; esta lista de TxIDs é adicionada à ferramenta Taxonomy CommonTree (<http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>) da base de dados Taxonomy do NCBI. A árvore taxonômica gerada apresenta em negrito os nomes das espécies correspondentes aos TxIDs inseridos e o grupo mais recente que contém todas estas espécies representadas no cluster é considerado o último ancestral comum (LCA). Um esquema da determinação do LCA e um exemplo de árvore gerada pelo NCBI podem ser vistos na Figura 9. Para números maiores de genes é utilizado um programa desenvolvido no laboratório por Velloso, que consulta a árvore taxonômica do NCBI alocada em memória e gera o LCA para cada agrupamento de sequências submetido.

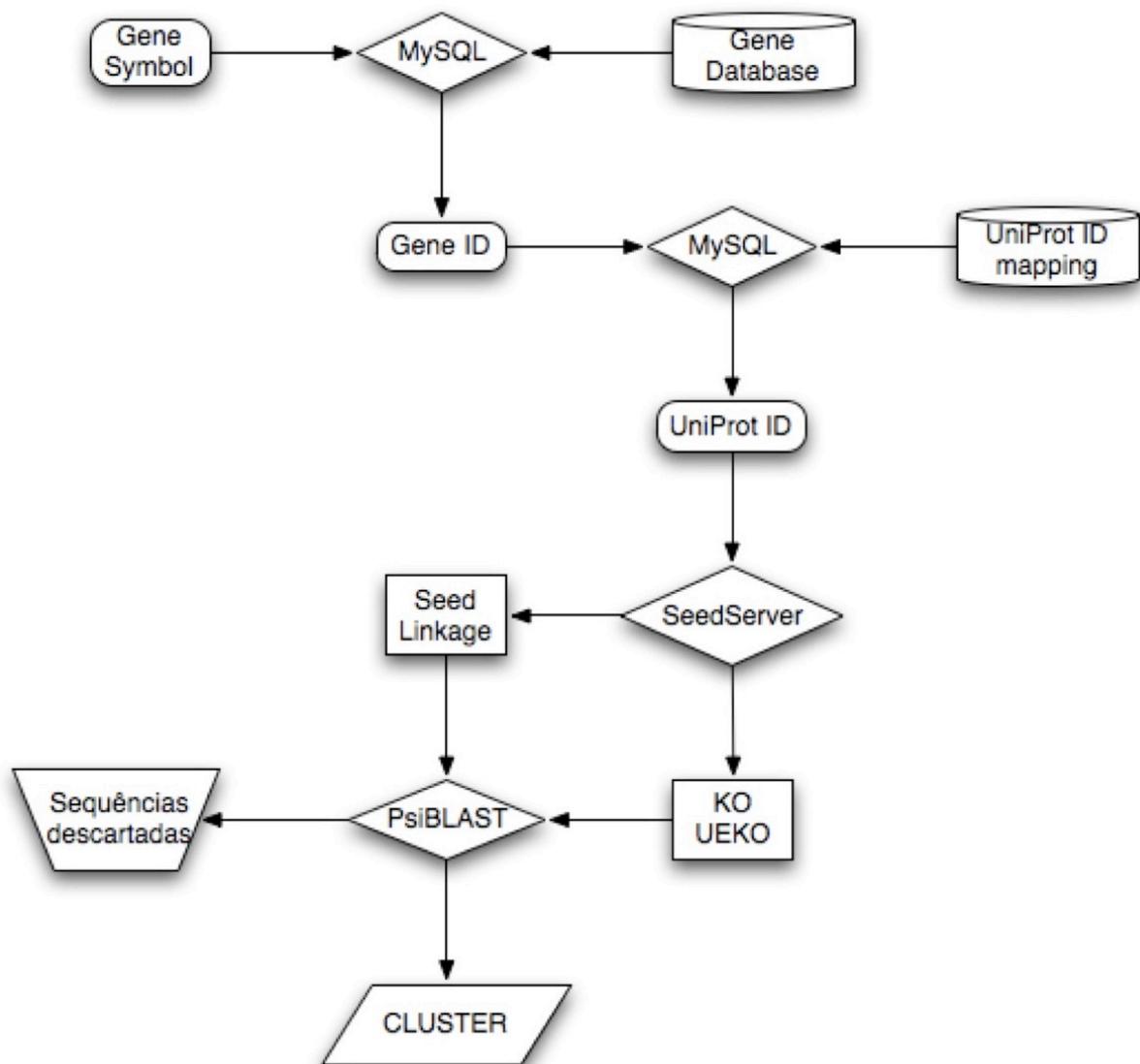


Figura 8: Procedimento para criação de agrupamentos de homólogos. Utilizando os símbolos dos genes (“gene symbols”), que participam da via de regulação gerada com o PESCADOR, e duas bases de dados, Gene e UniProt ID Mapping, obtêm-se os identificadores UniProt das sequências (“UniProt ID”). Estes são usados como entrada do software SeedServer, que contém dois componentes: o programa SeedLinkage, que realiza buscas de homólogos em BBH, e os bancos de dados Kegg Orthology (KO) e sua versão enriquecida UEKO (“UniRef enriched KO”). Um outro módulo do SeedServer compara as sequências de entrada com as recrutadas pelos componentes com PsiBLAST gerando um agrupamento de homólogos validado.

5. RESULTADOS

5.1 Validação da triagem líquida para o sistema de duplo-híbrido em leveduras

O sistema de triagem de duplo-híbrido é uma importante metodologia para o estudo de interações entre proteínas. A adaptação para triagem líquida e a validação da utilização de duas linhagens de leveduras, KGY37 e Y190 são descritas a seguir.

5.1.1 Escolha de iscas para triagens

Para validar o novo protocolo de triagem desenvolvido, selecionamos três iscas que já haviam sido previamente triadas por colaboradores com o protocolo convencional em placas de meio sólido. As iscas Trim2 e cRTel1 foram cedidas pelo Dr. Hao Ding com as respectivas bibliotecas de cDNA (cérebro e testículo de camundongo). A isca TRIM32 foi cedida pelo Dr. Klaus Wrogemann com a biblioteca de cDNA de músculo esquelético humano.

Trim2 é uma proteína de camundongo (*Mus musculus*) mediadora de ubiquitinação, sua função exata ainda não é conhecida mas parece envolver a regulação da polarização de neurônios durante o desenvolvimento. RTel1 é uma proteína também de camundongo (*Mus musculus*) cuja função envolve a manutenção e regulação de telômeros. TRIM32 é uma proteína humana (*Homo sapiens*) mediadora de ubiquitinação, com função exata ainda não definida.

5.1.2 Transformações e linhagens de leveduras

Para a triagem da isca cRTel1 (porção c-terminal da proteína RTel1) foi utilizada a levedura KGY37, que foi transformada pelo método “Quick and Easy” somente com o DNA isca, um plasmídio de expressão em levedura que produz cRTel1 em fusão com o domínio de ligação em DNA de Gal4 (pRTEL1; 1 µg). Em sequência, a levedura transformada com o DNA isca foi usada na transformação pelo método “High-Efficiency” com 10 µg da biblioteca de cDNA de testículo de camundongo. As 120 receitas de transformação resultaram em uma eficiência de 8×10^6 transformantes, uma cobertura de aproximadamente 2x da biblioteca.

Para a triagem da proteína Trim2 utilizamos as duas linhagens de levedura, KGY37 e Y190. Cada transformação foi de acordo com o método “High-Efficiency” e utilizamos 5 µg de DNA isca e 10 µg de biblioteca de cDNA de cérebro de camundongo. As 120 receitas de transformação com KGY37 resultaram em $4,6 \times 10^6$ transformantes e com Y190 resultaram em $5,4 \times 10^6$ transformantes. Cobrindo a biblioteca de estudo de 2-3 vezes aproximadamente.

Com a proteína TRIM32 a triagem foi feita na levedura KGY37. A levedura foi transformada de acordo com o método “High-Efficiency” e para cada receita utilizamos 5 µg de DNA isca e 15 µg de biblioteca de cDNA de músculo esquelético humano. As 120 receitas de transformação resultaram em uma eficiência de 4×10^7 transformantes, gerando uma cobertura de 1,15 vezes da biblioteca.

5.1.3 Triagem líquida com cRtel1 e Trim2

A transformação com a isca cRtel1 e biblioteca de testículo foi ressuspensa em 1200 mL de meio líquido SD -W-L-H +3AT +Amp (ver “Métodos” para detalhes) e distribuída em 10 frascos (60 mL em cada) e 4 placas “deepwell” (1,5 mL por poço). As culturas foram incubadas sob agitação por 72 horas e diluídas de acordo com o protocolo de triagem líquida descrito em “Métodos”. Após todo o processo verificamos o crescimento de 13 poços através da leitura de OD600 para as placas e nos frascos houve crescimento de 7 culturas.

A transformação com a isca Trim2 e biblioteca de cérebro de camundongo na levedura KGY37 foi igualmente ressuspensa em 1200 mL de meio líquido SD -W-L-H +3AT +Amp e distribuída em 10 frascos e 4 placas “deepwell”. Após o protocolo de incubações e diluições as culturas não apresentaram crescimento. Re-incubamos as placas iniciais (por dois dias adicionais aos três previamente utilizados) e prosseguimos novamente com as diluições para dar mais tempo de crescimento seletivo aos clones positivos, visto que a isca Trim2 interfere significativamente no crescimento da levedura (não mostrado). Após este procedimento, 21 poços mostraram crescimento como pode ser observado na Tabela 2. As culturas foram semeadas em meio sólido seletivo (10^3 células por placa) para o isolamento de colônias. A re-incubação das placas permitiu o isolamento de clones de interesse mas este procedimento pode resultar na inclusão de clones contendo interações fracas. Sendo assim, procuramos desenvolver um procedimento auxiliar padrão para monitorar o crescimento das leveduras sob interferência das iscas que será descrito a seguir. A triagem da isca Trim2 com a levedura Y190 foi feita já de acordo com este procedimento e será descrita no item 5.1.5.

5.1.4 Monitoramento de leveduras com crescimento alterado

As iscas que afetam o crescimento das leveduras apresentam um obstáculo para a triagem de duplo-híbrido. O procedimento para verificar o crescimento das leveduras consiste em diluir um volume de 0,5 mL (dos 50 mL de células inicialmente ressuspensas após as 120 transformações), correspondente a uma receita de transformação, em 9,5 mL de meio líquido SD -W-L (para selecionar apenas leveduras contendo o plasmídeo isca e o da biblioteca) e este tubo é chamado de A. O tubo B consiste em 1 mL do tubo A diluído em 9,0 mL de meio SD -W-L (diluição 1:10). Foram feitas também diluições 1:10 e 1:100 do tubo

B (tubos C e D). Os tubos foram incubados paralelamente à incubação das placas e frascos das triagens e o crescimento das culturas dos tubos era verificado por contagem de células no hemocitômetro nos momentos indicados pelo protocolo, indicando se a diluição subsequente poderia ser feita ou não. Após a incubação inicial de 72 horas, a contagem de células mostrou que o tubo B é o melhor para acompanhar o crescimento das culturas e determinamos que quando a concentração deste tubo fosse $2,0 \times 10^7$ células/mL a diluição 1:10 poderia ser feita. O tubo B também é diluído 1:10 (B1) e 1:100 (B2), a diluição maior 1:100 é feita para simular um evento mais raro correspondente à presença da interação na levedura. Para as iscas utilizadas que tornaram o crescimento da levedura lento, o tubo B mostrou que são necessárias de 12 a 24 horas adicionais antes da diluição 1:100 e outras 12 horas adicionais antes da leitura de OD600.

Tabela 2: Leitura de OD600 das placas de triagem da proteína Trim2 destacando poços com crescimento.

Plate 1	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0433	0.0417	0.0415	0.0444	0.0418	0.0432	0.0605	0.5331	0.0452	0.041	0.0505	0.0477
B	0.0487	0.0643	0.0444	0.0446	0.0475	0.0457	0.7664	0.8543	0.0456	0.045	0.3784	0.0467
C	0.0501	0.0465	0.0458	0.1089	0.0443	0.0469	0.0909	0.1998	0.0484	0.048	0.0631	0.0435
D	0.0483	0.0438	0.0603	0.0446	0.0442	0.0463	0.0538	0.0431	0.044	0.0458	0.0462	0.0474
E	0.0461	0.047	0.0448	0.0443	0.0476	0.0399	0.0458	0.0458	0.0457	0.046	0.0452	0.0583
F	0.0424	0.042	0.0458	0.0493	0.0413	0.042	0.043	0.0417	0.0487	0.0432	0.0421	0.0455
G	0.048	0.0763	0.0429	0.0444	0.0451	0.0438	0.0442	0.1932	0.0443	0.0442	0.046	0.0471
H	0.0564	0.0493	0.0537	0.0412	0.0523	0.0504	0.3851	0.2753	0.0451	0.0528	0.0504	0.2217
Plate 2	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0448	0.0485	0.0478	0.0499	0.0449	0.0414	0.047	0.0444	0.0425	0.047	0.0481	0.0459
B	0.0462	0.0469	0.045	0.0489	0.0437	0.0446	0.0445	0.0465	0.0574	0.0456	0.0455	0.0446
C	0.0488	0.0492	0.0439	0.0485	0.0449	0.044	0.0449	0.0512	0.044	0.0551	0.0482	0.0466
D	0.0454	0.0466	0.0445	0.049	0.0479	0.0506	0.048	0.0468	0.0465	0.6923	0.047	0.0445
E	0.0459	0.0499	0.0503	0.0461	0.0484	0.0449	0.0659	0.0596	0.0486	1.0925	0.0448	0.0532
F	0.039	0.044	0.0412	0.0567	0.043	0.0413	0.0414	0.0423	0.0418	0.103	0.0431	0.0431
G	0.053	0.052	0.0447	0.0458	0.0471	0.0979	0.0434	0.045	0.0575	0.0443	0.0459	0.045
H	0.051	0.0553	0.0501	0.055	0.0446	0.0544	0.0488	0.0577	0.0515	0.0504	0.0578	0.0601
Plate 3	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0441	0.0431	0.0405	0.0453	0.0419	0.0458	0.071	0.0431	0.0418	0.0418	0.0435	0.047
B	0.045	0.0543	0.1029	0.153	0.0798	0.0553	0.0481	0.0479	0.0469	0.0439	0.047	0.0461
C	0.0474	0.0997	0.0602	0.0386	0.1105	1.2777	0.0405	0.0447	0.0459	0.0451	0.0493	0.0441
D	0.0505	0.061	0.0473	0.0378	0.0397	0.0345	0.0415	0.0432	0.0458	0.0485	0.0454	0.2784
E	0.045	0.0409	0.0364	0.0459	0.0412	0.0569	0.0465	0.0459	0.0477	0.0457	0.0462	0.0463
F	0.0421	0.0424	0.0602	0.0369	0.043	0.0636	0.0386	0.0425	0.043	0.0416	0.0485	0.0422
G	0.0475	0.0794	0.1522	0.04	0.1261	0.0408	0.0372	0.0485	0.0624	0.0455	0.0494	0.0459
H	0.0485	0.04	0.0435	0.0442	0.0464	0.0412	0.059	0.0442	0.3701	0.0495	0.0524	0.0502
Plate 4	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0456	0.0553	0.4756	0.0419	0.0416	0.0378	0.0413	0.0407	0.0454	0.0465	0.0432	0.0456
B	0.0453	0.0448	0.0431	0.0466	0.0457	0.0428	0.0435	0.0465	0.0511	0.0455	0.049	0.0504
C	0.0434	0.0442	0.0444	0.0456	0.044	0.0443	0.0457	0.0456	0.0482	0.0472	0.0453	0.0433
D	0.0473	0.0444	0.0437	0.0443	0.045	0.0455	0.0448	0.0456	0.0806	0.0444	0.0453	0.0453
E	0.0573	0.0453	0.0446	0.0443	0.0454	0.045	0.0452	0.0518	0.0758	0.0475	0.0444	0.0465
F	0.0415	0.0409	0.0409	0.041	0.0418	0.0416	0.0401	0.0434	0.0659	0.0402	0.0414	0.0455
G	0.0438	0.0448	0.0431	0.044	0.047	0.0448	0.05	0.0437	0.0463	0.0457	0.0457	0.0444
H	0.0625	0.0599	0.0569	0.0655	0.0576	0.0553	0.0505	0.0486	0.0567	0.0492	0.0555	0.0494

5.1.5 Triagem líquida com as iscas Trim2 e TRIM32, sob monitoramento

Uma nova triagem com a proteína Trim2 foi feita utilizando a linhagem de levedura Y190. As 120 receitas de transformação foram ressuspensas em 1200 mL de meio SD -W-L-H +3AT +Amp e distribuídas em 4 placas “deepwell” e 10 frascos. O protocolo de diluições foi realizado de acordo com o monitoramento do crescimento da levedura indicado pelo tubo B. Em seguida, o crescimento foi determinado pela leitura de OD600 e foi observado crescimento ($OD \geq 0.1$) em todos os frascos e em 35 poços. As culturas com crescimento foram plaqueadas para isolamento de colônias.

A triagem com a proteína TRIM32, que também afetava o crescimento das leveduras (não mostrado), e a biblioteca de músculo esquelético humano foi realizada na levedura KGY37 e as transformações foram ressuspensas em 1200 mL de meio SD -W-L-H +3AT +Amp e distribuídas em 4 placas “deepwell” e 10 frascos. Após o protocolo de diluições acompanhado do monitoramento do crescimento, a leitura de OD600 indicou crescimento em 25 poços e 9 frascos. As culturas com crescimento foram plaqueadas para isolamento de colônias.

5.1.6 Ativação do gene repórter *LacZ*

As colônias obtidas para cada triagem foram transferidas para filtros de nitrocelulose e submetidas a um ensaio de ativação do gene repórter *LacZ*. Para a triagem com a isca cRTel1, foram positivos todos os 13 poços e todos os 7 frascos. Na triagem de Trim2 com KGY37 ficaram azuis (positivos) 8 dos 21 poços, parte dos filtros deste ensaio estão mostrados na Figura 10. Esta baixa correlação pode ser um reflexo do crescimento de fracas interações ou falso-positivos devido ao maior tempo de crescimento proporcionado nessa triagem, que não foi monitorada. Já na triagem de Trim2 com Y190 todos os 35 clones de poços e 10 frascos foram positivos para ativação de *LacZ*. Finalmente, na triagem de TRIM32 os 9 frascos resultaram em clones positivos para ativação de *LacZ* assim como 17 dos 25 poços. Estes resultados do ensaio de *LacZ* mostram uma melhora importante no protocolo de duplo-híbrido quando a triagem é realizada em meio líquido, já que nas triagens em placas de meio sólido é bastante comum que a maioria das colônias crescidas sejam negativas para a ativação de *LacZ*.

5.1.7 Validação dos positivos

Os clones positivos resultantes das triagens foram posteriormente sequenciados e submetidos a uma transformação com a isca original para reconstruir a interação. Os resultados obtidos foram similares e comparáveis em rendimento de clones positivos aos resultados obtidos de triagens em placas. Como exemplo serão descritos os clones confirmados para a triagem de Trim2 e levedura Y190 com a permissão de nosso

colaborador, Dr Hao Ding. Dos 10 clones positivos de frascos, a interação foi reconstruída com sucesso para seis. No sequenciamento verificou-se que cinco destes correspondiam à proteína EFS e o sexto à proteína Trim3. Os 35 clones de poços resultaram na reconstrução de 14 interações, sete destas correspondiam à proteína EFS, um deles correspondia à Trim3, e os demais a quatro proteínas diversas (não mostrado). A obtenção de clones repetidos provenientes de frascos diferentes é, portanto, um indicativo da intensidade da interação, reflexo da capacidade da levedura contendo este clone de dominar a cultura.



Figura 10: : Ensaio de *LacZ* da triagem de Trim2. Fotos de alguns resultados de ensaio de *LacZ* com clones isolados da triagem em meio líquido como isca Trim2 na levedura KGY37 (5 positivos e um negativo - ao centro, abaixo).

5.2 Criação de uma via de regulações utilizando ferramentas de mineração de texto

Ao lado de investigações de interações proteína-proteína, para as quais a metodologia de triagem líquida contribui, a literatura científica e dados depositados em bases públicas permitem a elaboração de uma via de regulação resultante de interações funcionais entre proteínas e genes por elas regulados. Uma lacuna em bases de dados é a regulação do desenvolvimento embrionário pré-implantação, tratado a seguir.

5.2.1 Busca na literatura

Utilizamos a plataforma PubMed do NCBI para buscar trabalhos relacionados com o desenvolvimento embrionário pré-implantação. O termo de busca utilizado ("preimplantation development") selecionou 3524 resumos.

5.2.2 Seleção de trabalhos relevantes

Para obter um resultado mais eficiente do que o retorno de uma pesquisa tradicional na PubMed, esta mesma pesquisa foi submetida a uma classificação da relevância de cada um dos trabalhos obtidos utilizando o software Medline Ranker (ver "Métodos"). O software atribuiu a cada um dos 3524 resumos um valor "p" (exemplo mostrado na figura 6c), indicando quanto este resumo seria informativo para nossa pesquisa. Esta atribuição se baseia no caráter informativo de 10 trabalhos selecionados manualmente listados a seguir: (Ambrosetti, Schöler *et al.*, 2000; Scaffidi e Bianchi, 2001; Plusa, Frankenberg *et al.*, 2005; Strumpf, Mao *et al.*, 2005; Mizuno, Sono *et al.*, 2006; Zhang, Tam *et al.*, 2006; Tam, Lim *et al.*, 2008; Cauffman, De Rycke *et al.*, 2009; Nishioka, Inoue *et al.*, 2009). Os 1000 trabalhos melhor ranqueados foram selecionados.

5.2.3 Revelando biointerações com o software PESCADOR

Com o resultado da classificação do Medline Ranker, o software PESCADOR foi capaz de marcar nomes de proteínas ("terms") e biointerações entre eles nos 1000 trabalhos que inserimos inicialmente. O PESCADOR marcou 722 nomes e selecionou 223 biointerações de tipo 1 entre eles, além de outras biointerações (tipos 2 a 4) também informativas (detalhes do software em "Métodos"). A partir deste ponto, a interpretação deste resultado é manual e as informações retiradas dos resumos foram colocadas de uma forma visual, agrupando regulações do desenvolvimento embrionário pré-implantação. Como exemplo, na Figura 11 está mostrado um resumo marcado pelo PESCADOR e também a interpretação da informação descrita no resumo. O fator de transcrição Tead4 é o ativador da transcrição do gene Cdx2, levando à diferenciação da trofotoderme, e para esta regulação é necessário que Tead4 interaja com Yap. Esta via de ativação de Cdx2 não acontece na massa celular interna ("inner cell mass") devido à presença nestas células de

um outro fator da via Hippo chamado Lats, uma cinase capaz de fosforilar Yap, que fica restrito ao citosol e não pode interagir com Tead4 localizada no núcleo. Esta importante regulação do destino das células embrionárias é apenas uma das contidas na via final obtida para a pré-implantação.

PESCADOR
 Platform for Exploration of Significant Concepts AssociateD to co-Occurrences Relationships.
 Results for ID 1274991988: [Summary](#) | [Terms](#) | [Concepts](#) | [Abstracts](#) | [Network](#) | [Pathways](#) | [Help](#) | [Home](#) |

Selected PubMed ID: 19289085

Sentence 1: The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass.

Sentence 2: Outside cells of the preimplantation mouse embryo form the trophectoderm (TE), a process requiring the transcription factor Tead4.

Sentence 3: Here, we show that transcriptionally active Tead4 can induce Cdx2 and other trophoblast genes in parallel in embryonic stem cells.

Sentence 4: In embryos, the Tead4 coactivator protein Yap localizes to nuclei of outside cells, and modulation of Tead4 or Yap activity leads to changes in Cdx2 expression.

Sentence 5: In inside cells, Yap is phosphorylated and cytoplasmic, and this involves the Hippo signaling pathway component Lats.

Sentence 6: We propose that active Tead4 promotes TE development in outside cells, whereas Tead4 activity is suppressed in inside cells by cell contact- and Lats-mediated inhibition of nuclear Yap localization.

Sentence 7: Thus, differential signaling between inside and outside cell populations leads to changes in cell fate specification during TE formation.

Legend of colors:
 ■ Target sentence(s);
 ■ Biological terms;
 ■ Biointeraction terms;
 ■ Biological concept(s).

All loaded PMIDs
 [Show Top 500]
 1: [20231280](#)
 2: [20018908](#)
 3: [19915186](#)
 4: [20485553](#)
 5: [20100249](#)
 6: [19949895](#)
 7: [20353723](#)
 8: [20308546](#)
 9: [20479880](#)
 10: [19696013](#)
 11: [20046880](#)
 12: [19776387](#)
 13: [20034412](#)
 14: [20132010](#)
 15: [20336607](#)
 16: [20047725](#)
 17: [20362968](#)
 18: [20021719](#)
 19: [19967554](#)
 20: [20179092](#)
 21: [20422011](#)
 22: [20047732](#)
 23: [20116247](#)
 24: [20300987](#)
 25: [20176746](#)
 26: [20047729](#)
 27: [20136652](#)
 28: [19596312](#)

Biological terms co-occurrences with concepts				
Type	Sentence	Pair 1	Pair 2	Biointeractions
1	3	CDX2	TEAD4	induce
1	4	CDX2	TEAD4	localizes activity
2	3	CDX2	TEAD4	
2	4	CDX2	TEAD4	
3	3	CDX2	TEAD4	active induce

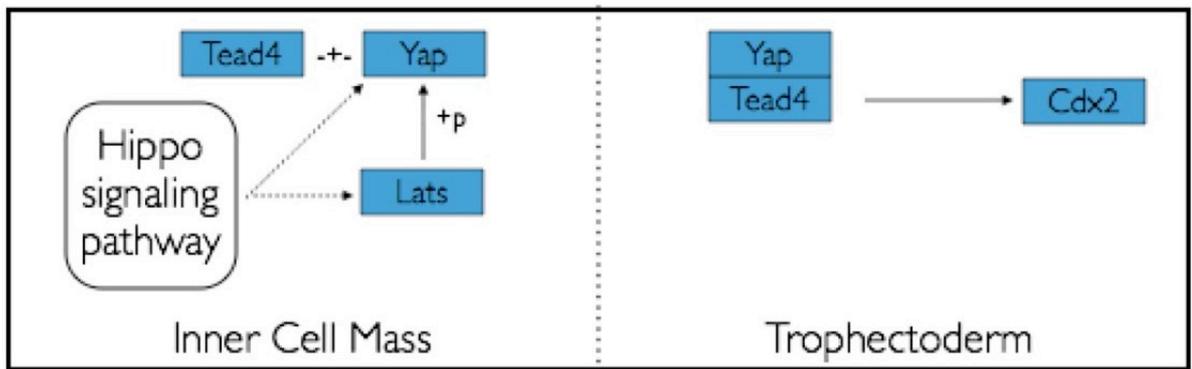


Figura 11: Resultado do PESCADOR. Exemplo de resumo marcado com o software PESCADOR, os nomes de genes reconhecidos são marcados de vermelho (“terms”) e as palavras de biointeração em amarelo. A plataforma permite que o usuário procure pelas interações de interesse na lista de genes encontrados (“terms”), na lista de resumos (“abstracts”) ou pelos conceitos de interesse indicados inicialmente. Abaixo está mostrado um exemplo de curadoria manual da informação extraída do resumo e sua representação gráfica, iniciando a construção da via de interações.

5.2.4 Representação gráfica da via do desenvolvimento pré-implantação e revisão das regulações

A extensa via obtida após a análise de todos os trabalhos marcados pelo PESCADOR está representada na Figura 12. A via compreende 86 genes (ou seus produtos) e diversas regulações entre eles. Uma revisão destas regulações será descrita a seguir.

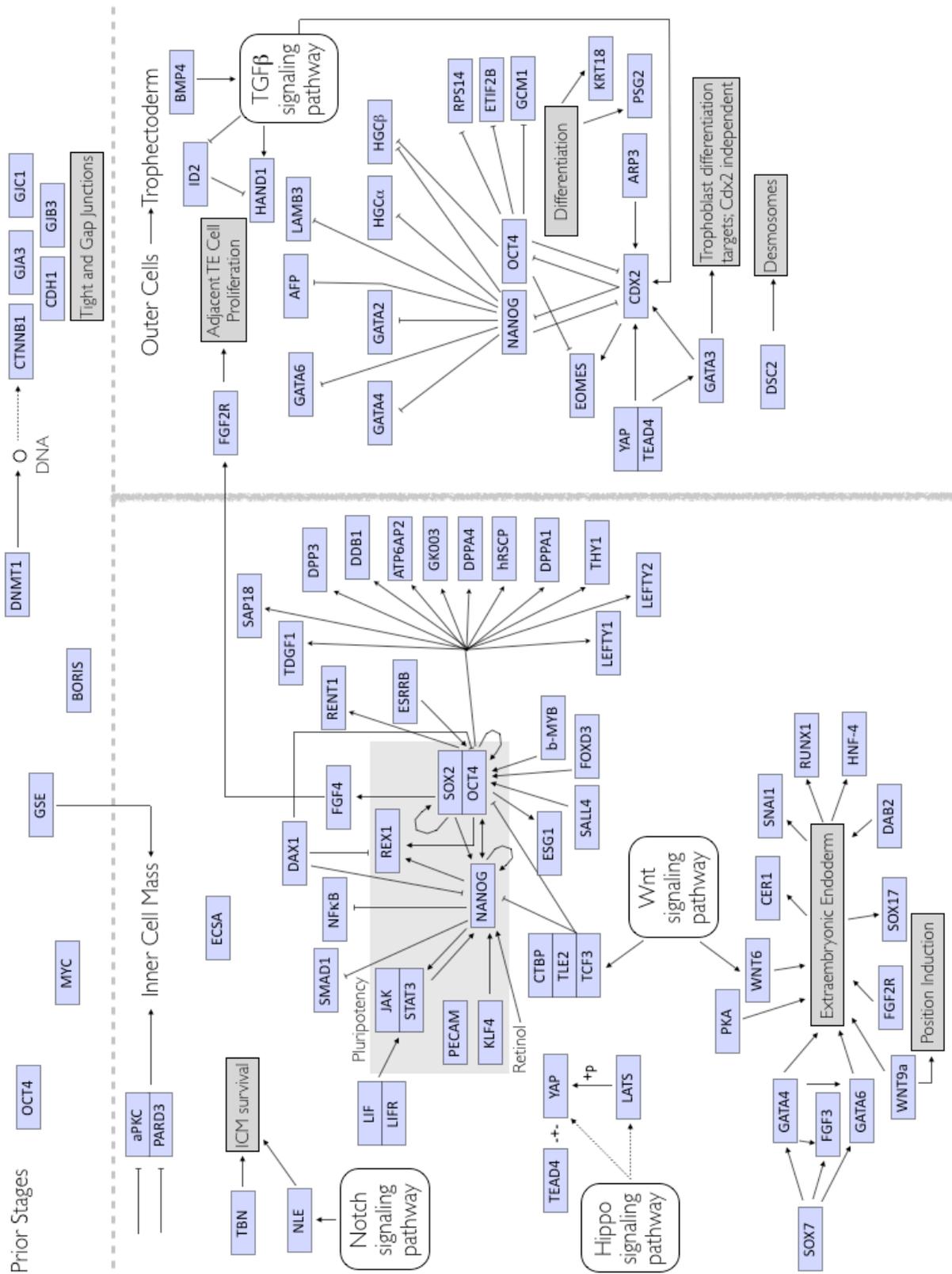


Figura 12: Via regulatória do desenvolvimento embrionário pré-implantação. A figura mostra os genes envolvidos no desenvolvimento embrionário no período pré-implantação e as regulações existentes entre eles. Algumas funções estão também representadas nos retângulos. Algumas das principais regulações estão descritas no texto. Como modelo de sinalização foi seguida a linguagem KEGG (ver texto para detalhes).

Primeiras clivagens embrionárias

O oncogene c-MYC é um importante regulador da transcrição e sua expressão é observada nos estágios iniciais do desenvolvimento, podendo ser encontrado nas células até a fase de mórula, sendo reprimido a partir de então (Suzuki, Abe *et al.*, 2009). Dois genes foram associados recentemente com o desenvolvimento, um deles envolvido nos estágios logo após a fecundação que depois passa a ser reprimido, chamado BORIS e o segundo chamado ECSA que começa a ser expresso já no blastocisto exclusivamente nas células da massa celular interna. A presença destes genes foi comparada com o padrão de expressão do fator de transcrição Oct4, que já está presente nas primeiras clivagens, sendo posteriormente reprimido e voltando a ser expresso no blastocisto (Monk, Hitchins *et al.*, 2008). A expressão de um gene associado com gametogênese, Gse foi recentemente identificada em células do embrião inicial, esta proteína posteriormente tem sua expressão restrita em células da massa celular interna, sugerindo uma função na especificação da linhagem celular (Mizuno, Sono *et al.*, 2006).

Padrões de metilação e o correto desenvolvimento pré-implantação

Desde o início do desenvolvimento embrionário pré-implantação padrões genômicos de metilação em células humanas dependem da proteína DNMT1 (DNA methyltransferase-1) enquanto em células de camundongo a variante correspondente é chamada Dnmt1o. No estágio de oito células esta proteína é realocada para o núcleo e mantém padrões de metilação essenciais, permitindo que o embrião complete eventos iniciais do desenvolvimento (Chung, Ratnam *et al.*, 2003). Recentemente foi demonstrado que a perda da localização correta de Dnmt1o não só resulta em uma interrupção do desenvolvimento em embriões de 5-7 células como também é responsável pela inibição da transcrição de cinco genes (GJA3, GJC1, GJB3, CDH1 e CTNNB1) envolvidos na formação de junções do tipo “gap” e “tight” (Yu, Xue *et al.*, 2009). Estas junções são cruciais para processos iniciais como a compactação do embrião de 8-células e a cavitação da blastocele.

Dicotomia Trofotoderme (TE) e Massa Celular Interna (ICM)

Nas células destinadas a formar a massa celular interna, marcadas pela repressão dos genes aPKC e Pard3 (Plusa, Frankenberg *et al.*, 2005) e também pela ativação de Sox2 (Guo, Huss *et al.*, 2010), os principais fatores de transcrição da pluripotência, incluindo Nanog e Oct4, permanecem ativos devido à expressão de um importante membro da via de sinalização Hippo: Lats. Esta serina/treonina proteína cinase é responsável por fosforilar Yap, levando ao seqüestro desta proteína pelo citosol e impedindo sua associação com o fator de transcrição Tead4.

Ativando a diferenciação da TE: Tead4/Yap ativam Cdx2 para reprimir Nanog e Oct4

Inversamente, nas células externas que vão se diferenciar e formar a trofotoderme, Yap não está fosforilado, permanece no núcleo e se associa com Tead4, levando à ativação de Cdx2, um repressor chave de Nanog e Oct4 (Ralston e Rossant, 2008). Esta repressão então suspende a inibição que Nanog e Oct4 estavam exercendo em vários genes, que por sua vez se tornarão ativos (Strumpf, Mao *et al.*, 2005; Nishioka, Inoue *et al.*, 2009). A ativação de Cdx2 requer um escape da repressão basal: Nanog (Hyslop, Stojkovic *et al.*, 2005) e Oct4 (Zhang, Tam *et al.*, 2006) reprimem níveis basais de transcrição de Cdx2; uma indução de níveis mais altos de transcrição por Tead4/Yap supera esta repressão e Cdx2 pode exercer a sua função (Nishioka, Inoue *et al.*, 2009). Tead4 também foi associada recentemente com a ativação de outro fator de diferenciação da trofotoderme, GATA3 (Ralston, Cox *et al.*, 2010), que atua em conjunto com Cdx2 e também em genes independentes do controle deste fator, esta ativação de GATA3 parece ser independente de Yap sugerindo a interação de Tead4 com algum outro fator. Também necessária para o alto nível de expressão de Cdx2 é a proteína Arp3. Experimentos com camundongos “knockout” para este gene apresentam células trofoblastos incapazes de se desenvolver apropriadamente, possivelmente entrando em apoptose como resultado da ausência de Cdx2 (Vauti, Prochnow *et al.*, 2007). Outra via importante para a diferenciação da trofotoderme é a via de TGFbeta estimulada pelo fator BMP4 que ativa proteínas Smad, estas proteínas estimulam a transcrição de Cdx2 (Hayashi, Furue *et al.*, 2010). BMP4 também é responsável pela inibição de Id2, inibidor da diferenciação (Kondo, Cubillo *et al.*, 2004), e pela ativação de Hand1, proteína específica da diferenciação de trofoblastos (Riley, Anson-Cartwright *et al.*, 1998).

Na ausência de Oct4 e Nanog

A repressão de Oct4 leva à ativação de um regulador positivo de destino de célula da TE, Eomes (“T-box protein eomesodermin”) (Strumpf, Mao *et al.*, 2005; Babaie, Herwig *et al.*, 2007), também apontado como alvo do controle de Cdx2 (Marikawa e Alarcón, 2009). A diferenciação subsequente destas células é acompanhada pela expressão de diversos genes, como a glicoproteína PSG2 (Adjaye, Herwig *et al.*, 2007) e o marcador KRT18, um dos primeiros sinais de que a célula perdeu sua capacidade totipotente antes de qualquer diferenciação visível (Vauti, Prochnow *et al.*, 2007; Cauffman, De Rycke *et al.*, 2009). A remoção da repressão exercida por Oct4 também resulta na ativação dos genes ETIF2B e Rps14 (Shin, Cui *et al.*, 2005) permitindo que estas células iniciem uma rotina de tradução intensa. Experimentos de “knockdown” de Oct4 mostraram também que a retirada deste fator permite a expressão de Gcm1, específico de placenta (Yamada, Ogawa *et al.*, 2000), e da cadeia beta do hormônio hCG (Matin, Walsh *et al.*, 2004).

Simultaneamente, a repressão de Nanog permite a expressão de vários genes tanto associados à trofotoderme (GATA2, hCG-alfa e hCG-beta) quanto associados à endoderme extra-embrionária (GATA4, GATA6, LAMB3 e AFP) (Hyslop, Stojkovic *et al.*, 2005). A endoderme extra-embrionária se diferencia posteriormente a partir da ICM também mas dá origem ao saco vitelínico, enquanto a trofotoderme origina a porção embrionária da placenta. Desde o início da formação do blastocisto, desmossomos são formados na trofotoderme e a proteína responsável é a desmocolina (DSC2), que não é expressa na ICM (Collins, Lorimer *et al.*, 1995).

Sob a contínua ativação de Oct4 e Nanog

Na massa celular interna, os genes responsáveis pela pluripotência permanecem ativos e formam uma complexa via de regulações. Recentemente foi descrito que a transcrição de Nanog é estimulada pela presença de compostos como retinol (Chen, Yang *et al.*, 2007). Exercendo um papel redundante na ativação de Nanog, as proteínas Klf2, Klf4 e Klf5 (“krüppel-like factor”) foram descritas como essenciais para a manutenção da pluripotência. Klf4 já era um fator conhecido por este papel e faz parte dos fatores de transcrição utilizados na reprogramação em laboratório de células diferenciadas. Contudo, somente a depleção simultânea de Klf4, 2 e 5 resulta na diferenciação de células tronco, indicando a redundância funcional (Jiang, Chan *et al.*, 2008). Outras proteínas são ativadoras já bem descritas de Nanog, incluindo os dois outros fatores principais reguladores da pluripotência, Oct4 (Pan, Li *et al.*, 2006; Babaie, Herwig *et al.*, 2007) e Sox2 (Van Den Berg, Zhang *et al.*, 2008). O receptor de estrogênio ESRRB também foi descrito como envolvido na ativação de Nanog por Oct4 e Sox2 (Van Den Berg, Zhang *et al.*, 2008). Inversamente, Nanog também ativa Oct4 (Pan, Li *et al.*, 2006), e o receptor ESRRB também é necessário para manter a atividade do promotor de Oct4 (Zhang, Zhang *et al.*, 2008).

Cada um destes três fatores também atua como um auto-ativador, por exemplo Oct4 e Sox2 se associam e ativam a transcrição do próprio Oct4 (Okumura-Nakanishi, Saito *et al.*, 2005). Outro fator de transcrição principal envolvido na manutenção do estado de pluripotência das células é Sall4 (Cauffman, De Rycke *et al.*, 2009). Sall4 se liga a uma região regulatória conservada no acentuador (“enhancer”) distal de Oct4 e ativa a transcrição do gene (Zhang, Tam *et al.*, 2006). Estudos com interferência de microRNA mostram que a perda de Sall4 leva a uma redução do nível de mRNA de Oct4 e um aumento significativo da expressão de Cdx2 na ICM (Zhang, Tam *et al.*, 2006). B-MYB, um gene expresso em células proliferativas, é também um regulador positivo de Oct4 e estudos descrevem diferenciação imprópria da ICM na ausência de b-MYB (Tarasov, Tarasova *et al.*, 2008).

A via de sinalização Notch é uma via conservada evolutivamente, envolvida em processos de comunicação celular e decisões celulares corretas e está envolvida no desenvolvimento da massa celular interna (Adjaye, Huntriss *et al.*, 2005). A proteína Nle, uma reguladora direta desta via, é essencial para a sobrevivência das células da massa celular interna (Cormier, Le Bras *et al.*, 2006). Outra proteína associada com o desenvolvimento e essencial para a sobrevivência da massa celular interna é Tbn (Taube nuss) sem a qual as células da ICM entram em apoptose (Voss, Thomas *et al.*, 2000).

A expressão da molécula de adesão de plaquetas e células endoteliais (PECAM1 ou CD31) foi detectada por microscopia confocal de imunofluorescência no blastocisto e restrita às células da ICM. Posteriormente PECAM1 permanece somente nas células do epiblasto pluripotentes, desaparecendo no momento em que estas sofrem diferenciação (Robson, Stein *et al.*, 2001). Estes resultados indicam um novo papel para esta molécula no desenvolvimento embrionário.

Ativação sob controle

Outras vias de controle mantêm a expressão dos fatores de transcrição principais gerando uma concentração celular adequada e equilibram estes diversos mecanismos de ativação e superexpressão da transcrição. Uma regulação complexa por retroinibição (“feedback”) é composta por FOXD3, Nanog e Oct4 (Pan, Li *et al.*, 2006). Para manter a concentração desejada de Oct4 e Nanog, estes três genes interagem da seguinte maneira: A expressão de Nanog é ativada por FOXD3 e Oct4 mas não acima do nível desejado por repressão exercida pelo Oct4; FOXD3 e Nanog ativam a expressão de Oct4 mas não acima do necessário pois Oct4 atua como auto repressor.

Dax1 é um receptor nuclear de hormônio órfão recentemente identificado como repressor da transcrição de Oct4 (Sun, Nakatake *et al.*, 2009). A expressão de Dax1 foi também capaz de reduzir a expressão de Nanog e Rex1. Ensaios mostraram que Dax1 se liga a Oct4 e elimina sua atividade de ligação a DNA, diminuindo assim a transcrição de Nanog e Rex1, alvos de ativação do Oct4.

Outro destes repressores na ICM é Tcf3, um efetor da via de sinalização Wnt. TLE2 (uma proteína da família Groucho) e CtBP (“C-terminal binding protein”) foram descritos como parceiros principais de Tcf3 na mediação de seu efeito repressor. Tcf3 se liga ao promotor de Oct4 e o reprime, este efeito repressor requer ambos os domínios de Tcf3 de interação com Groucho e com CtBP (Tam, Lim *et al.*, 2008). Tcf3 também limita a concentração celular do mRNA e da proteína Nanog, além de controlar a atividade do promotor de Nanog em células tronco embrionárias (ESCs); o domínio de interação com Groucho de Tcf3 está envolvido nesta repressão (Pereira, Yi *et al.*, 2006). Tcf3 é crítico para a manutenção dos níveis apropriados de Oct4 e Nanog em ESCs. Experimentos mostram

que a perda de Tcf3 por interferência de RNA (RNAi) bloqueia a habilidade destas células de se diferenciar posteriormente (Tam, Lim *et al.*, 2008).

Alvos de Oct4 e Sox2

Oct4 ativa o gene específico de células tronco-1 (Esg1), que codifica uma proteína de ligação a RNA presente na ICM responsável por regular diversos transcritos específicos (Tanaka, Lopez De Silanes *et al.*, 2006). Oct4 e Sox2 são também responsáveis por ativar a transcrição do fator de crescimento de fibroblastos-4 (FGF4) (Okumura-Nakanishi, Saito *et al.*, 2005). A expressão de FGF4 requer portanto a atividade combinada destes dois fatores de transcrição que se ligam em sítios adjacentes na região do DNA contendo o acentuador do gene FGF4 (Ambrosetti, Schöler *et al.*, 2000). Um vez transcrita, a proteína FGF4 pode interagir com seu receptor FGFR2 e ativar a proliferação de células da ICM e TE, posteriormente ativando também as células da endoderme extra-embriônica.

Diversos outros genes são também alvos dependentes da ativação de Oct4. Estes incluem o fator de crescimento TDGF1, Inibidor de crescimento SAP18, regulador de transcritos sem-sentido (“nonsense”) RENT1, duas proteínas envolvidas na auto-renovação de células tronco DPPA4 e DPPA1, marcadores da endoderme anterior visceral LEFTY1 e 2, antígeno de superfície THY1 e outros genes codificadores de proteínas envolvidas em processos celulares especializados (DPP3, ATP6AP2, DDB1) ou proteínas hipotéticas como GK003 e hRscp (Shin, Cui *et al.*, 2005; Babaie, Herwig *et al.*, 2007).

Acredita-se que esta ampla regulação exercida por Sox2 e Oct4 durante a embriogênese de mamíferos opera por meio de uma ligação ao DNA cooperativa em regiões regulatórias compostas de motivos HMG (Sox2) e POU (Oct4) adjacentes (cassetes HMG/POU) (Torres-Padilla, Richardson *et al.*, 2007). DPPA4 é um dos genes com a presença de um cassete HMG/POU em sua região promotora (Chakravarthy, Boer *et al.*, 2008).

Alvos de Nanog e STAT3

Uma importante contribuição para a pluripotência é exercida pela ativação da via JAK/STAT. A via LIF/STAT3 (Saito, Liu *et al.*, 2004; Chen, Yang *et al.*, 2007; De Felici, Farini *et al.*, 2009) para a manutenção da pluripotência celular consiste de LIF e o receptor de LIF, que desencadeiam um sinal celular através de STAT3. Este transdutor de sinal e ativador de transcrição é ativado pela cinase JAK1 e se liga a vários promotores induzindo a transcrição de genes relacionados à pluripotência. Nanog e STAT3 ligam-se a promotores dependentes de STAT3 e causam sua ativação sinergicamente. Nanog também funciona como um inibidor transcricional de NF κ B, um fator conhecido por sua atividade pró-diferenciação (Torres e Watt, 2008). Nanog também é responsável pela repressão de SMAD1, impedindo

a diferenciação induzida por BMP4 pela via de sinalização TGF β , da qual SMAD1 é um transdutor de sinal (Suzuki, Raya *et al.*, 2006).

Diferenciação da endoderme extra-embriônica a partir de células da ICM

Ainda antes da implantação embrionária, outra diferenciação acontece. Certas células da massa celular interna dão origem à endoderme primitiva, o primeiro tipo celular distinto da endoderme extra-embriônica. Os componentes da endoderme extra-embriônica são as endodermes primitiva, parietal e visceral. Este tecido vai desenvolver o saco vitelínico durante estágios posteriores da embriogênese.

Wnt6 foi recentemente identificado como um indutor da endoderme primitiva e esta indução é acompanhada pela translocação de beta-catenina (CTNNB1) e Snai1 para o núcleo (Krawetz e Kelly, 2008). Este estudo também mostrou que o aumento da expressão da proteína cinase A (PKA) induz marcadores da endoderme parietal. Outro membro da família Wnt, Wnt9a, tem sua expressão localizada em células da ICM que margeiam a blastocle (Kemp, Willems *et al.*, 2005) e induz o reposicionamento de células que expressam GATA6 necessário para a formação da endoderme primitiva (Meilhac, Adams *et al.*, 2009).

Sox7 tem papel fundamental na diferenciação da endoderme parietal. Foi mostrado por estudos utilizando moléculas curtas de RNA de interferência que Sox7 é responsável pela indução da transcrição de GATA4 e GATA6 (Futaki, Hayashi *et al.*, 2004). O silenciamento individual ou combinado de Sox7, GATA4 e GATA6 resulta em supressão das mudanças morfológicas celulares e também da produção de laminina1 (LAMB1), alterações características da diferenciação da endoderme parietal. GATA4 foi previamente identificado como um fator de transcrição responsável pela ativação de FGF3 (Murakami, Thurlow *et al.*, 1999). Sox7 também ativa o promotor de FGF3, já Sox2, embora da mesma família, pode modular negativamente a ativação de FGF3 mediada por GATA4 esta evidência é fortalecida pelo papel deste fator na pluripotência da ICM (Murakami, Shen *et al.*, 2004). Sox17, outro membro da família Sox, é responsável pela diferenciação da endoderme extra-embriônica em estágios posteriores (Shimoda, Kanai-Azuma *et al.*, 2007). Associada à expressão de Sox17 foi encontrado o fator de transcrição Runx1, também específico para a endoderme extra-embriônica (Kurimoto, Yabuta *et al.*, 2006). HNF-4 é um fator de transcrição específico da endoderme extra-embriônica com papel no desenvolvimento pós-implantação e na organogênese (Duncan, Manova *et al.*, 1994) sua expressão pode ser resultante da diferenciação induzida por BMP4 (Coucouvanis e Martin, 1999). Finalmente, para o desenvolvimento da endoderme visceral a proteína Dab2 é indispensável, embora seu papel não esteja bem descrito, talvez relacionado ao posicionamento celular correto (Morris, Tallquist *et al.*, 2002; Yang, Smith *et al.*, 2002). A expressão de Cer1, um marcador

da endoderme anterior visceral (AVE) começa antes da implantação do embrião no subconjunto de células que compõe a endoderme primitiva. Parte desta população ancestral de células expressando Cer1 junta-se a células que só começam a expressar Cer1 após a implantação e formam a endoderme anterior visceral (Torres-Padilla, Richardson *et al.*, 2007).

5.3. Busca por genes homólogos

5.3.1 SeedServer

Para estabelecer uma base de dados de ortólogos e providenciar informação de sequência para os genes que compõe a via de desenvolvimento pré-implantação, sequências de aminoácidos correspondentes aos genes humanos e de camundongo foram utilizadas como sementes (“seed”) no software SeedServer (ver “Métodos”). Na verdade apenas o identificador UniProt dessas proteínas foi necessário para execução do SeedServer – símbolos dos genes foram verificados no banco de dados Gene, NCBI, e convertidos para Gene-ID e subsequentemente na base UniProt obtiveram-se os identificadores desejados. Para aumentar a precisão, somente entradas SwissProt foram utilizadas. Para cada gene foi gerado um agrupamento composto por 2 até 260 sequências e o conteúdo destes será analisado a seguir.

Todos os genes presentes na via foram submetidos ao SeedServer para gerar agrupamentos de homólogos exceto hRSCP e DPPA1 (não possuem sequência com anotação SwissProt).

5.3.2 Análise da composição dos agrupamentos gerados

O software SeedServer gerou agrupamentos para cada “seed” contendo sequências que podem ter anotações Swiss-Prot ou não (Boutet, Lieberherr *et al.*, 2007). Quando uma sequência é anotada manualmente e passa por uma curadoria, esta sequência entra no banco de dados UniProt como Swiss-Prot. As sequências adicionadas ao banco automaticamente são sequências TrEMBL, não revisadas. De todos os agrupamentos gerados pelo SeedServer, somente cerca de 25% das sequências recrutadas possuem anotações Swiss-Prot (Figura 13a). Sendo assim, a busca por homólogos utilizando o SeedServer gera uma grande quantidade de sequências candidatas à curadoria manual para tornarem-se anotações Swiss-Prot. Os dados da quantidade de sequências Swiss-Prot ou não Swiss-Prot para cada um dos agrupamentos estão mostrados no gráfico da Figura 13a.

Devido ao fato de que o SeedServer recruta as sequências para os agrupamentos por três vias diferentes – KO, UEKO e SeedLinkage – (ver “Métodos” para detalhes) analisamos também a porcentagem de sequências com a qual cada uma destas buscas contribui. Considerando todos os agrupamentos, 28% das sequências são agrupadas com a “seed” pelo banco de dados KO, outros 51% são agrupados pelo UEKO e 20% são sequências recrutadas diretamente pelo SeedLinkage. Os dados mostram como a contribuição do SeedLinkage é importante para o enriquecimento dos agrupamentos de ortólogos já existentes, assim como a versão enriquecida do KO que contribui com metade das sequências agrupadas. Os dados da contribuição de cada uma das buscas feitas pelo SeedServer para cada agrupamento estão representados no gráfico da Figura 13b.

Além disso, as sequências recrutadas para os agrupamentos de homólogos podem ser classificadas quanto ao andamento de projetos genoma associados aos organismos dos quais estas sequências pertencem. Para todos os agrupamentos gerados, 48% das sequências são provenientes de organismos com projetos genoma em montagem, 27% das sequências são provenientes de organismos com projetos genoma completos, 2% das sequências são provenientes de organismos com projetos genoma em andamento e 24% das sequências são depósitos provenientes de organismos sem projetos genoma associados. O recrutamento de sequências depositadas que não pertencem a genomas completos denota a importância das buscas realizadas pelo SeedServer através do SeedLinkage e UEKO, já que o banco de dados KO somente agrupa sequências provenientes de projetos genoma completos. A associação com projetos genoma das sequências contidas em cada um dos clusters está representada no gráfico da figura 13c.

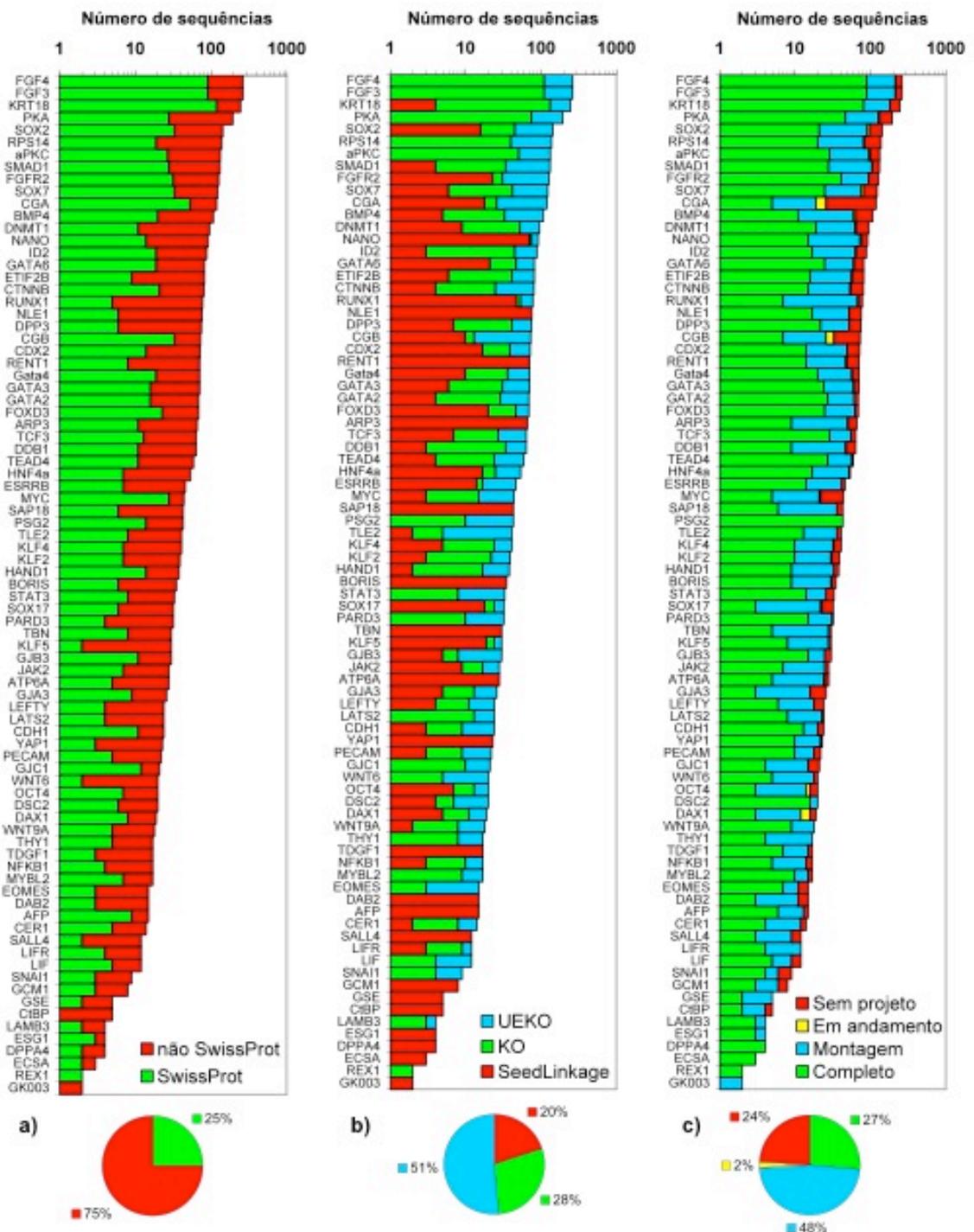


Figura 13: Classificação dos homólogos agrupados pelo SeedServer. a) O gráfico mostra para cada gene (eixo y) a quantidade de seqüências (eixo x) que possuem (verde) ou não (vermelho) anotação SwissProt. Note que o eixo x mostra números em escala logarítmica. O gráfico menor representa a fração total de seqüências recrutadas que tem ou não anotação SwissProt. b) Representação similar para as seqüências recrutadas exclusivamente pelo SeedLinkage (vermelho), pela base de dados KO (verde) ou pela sua versão estendida UEKO (azul). c) Representação similar para mostrar o andamento dos projetos genoma dos quais surgiu o depósito das seqüências recrutadas. Organismos com genoma completo (“Complete”), organismos com projetos genoma em montagem (“Draft”), organismos com projetos genoma em andamento (“In Progress”) e seqüências de organismos sem nenhum projeto genoma em andamento (“No Project”).

5.4. Ancestralidade dos genes da via do desenvolvimento pré-implantação

5.4.1 Determinação do último ancestral comum (“Last Common Ancestor”)

A geração dos agrupamentos contendo sequências homólogas para os genes envolvidos no desenvolvimento pré-implantação possibilitou a determinação do último ancestral comum (LCA) destes genes (ver “Métodos” para procedimento). Na Figura 14 estão representados os genes da via de pré-implantação de acordo com sua ancestralidade. Genes com origem considerada antiga são aqueles cujo LCA encontrado é anterior ao clado Euteleostomi e estão mostrados em cinza. Genes cujo LCA encontrado pertence ao clado Euteleostomi ou clados mais recentes são considerados genes de origem recente e estão marcados em azul. Genes de origem antiga que possuem homólogos em *Drosophila melanogaster* estão marcados com um asterisco. A Tabela S1 (anexo) mostra todos os genes usados como “seed” e os resultados da busca de LCA para cada um. Com esta pesquisa sobre ancestralidade podemos observar que a via de desenvolvimento pré-implantação tem origem antiga com adição de elementos novos principalmente em Euteleostomi e Eutheria. Mesmo funções essenciais estão divididas entre genes recentes e antigos, como podemos observar no caso de Nanog (antigo) e Oct4 (recente), dado que os dois genes são os principais fatores na manutenção da pluripotência das células embrionárias. O gráfico da Figura 15 mostra a distribuição de todos os genes da via de acordo com sua origem em clados da linhagem humana. Pode se observar que uma grande quantidade de genes se origina em certos períodos como visto em Eumetazoa, Coelomata, Euteleostomi e Eutheria. A razão para esta origem em ondas precisa ser melhor estudada, todavia aparentemente a origem de estruturas complexas, que caracterizou os descendentes a partir de um dado momento da evolução, deve ter ocorrido simultaneamente à especialização de grupos de genes.

O enriquecimento dos agrupamentos pelo procedimento utilizado pelo SeedServer pode ser observado na Figura 16 na qual está representada a ancestralidade obtida quando utilizamos os agrupamentos gerados pelo KO ou pelo SeedServer (o qual além do KO faz uso de busca com software Seed Linkage e do enriquecimento do KO, denominado UEKO). A ancestralidade está representada por números relativos à profundidade do grupo taxonômico; para visualizar qual grupo taxonômico corresponde a qual profundidade pode ser consultada a Tabela S1 em anexo contendo o LCA para cada um dos genes pesquisados neste trabalho. Brevemente, a profundidade 1 representa Cellular Organisms e 28, *Homo sapiens*. Podemos observar no gráfico que quando a geração de agrupamentos pelo SeedServer resulta em uma diminuição da profundidade do grupo taxonômico identificado como LCA, o ponto correspondente ao gene encontra-se abaixo da linha pontilhada (na linha diagonal, KO=SeedServer). Pontos que se encontram no eixo Y não

possuíam agrupamentos gerados pelo KO e portanto sua representação depende da busca feita pelo SeedServer. Para exemplificar destacamos os pontos correspondentes aos genes Nanog e Oct4. Os agrupamentos de KO para estes genes indicam que o último ancestral comum para eles pertence ao grupo taxonômico de profundidade 20 (Eutheria), porém nossos dados dos agrupamentos gerados pelo SeedServer indicam que a profundidade para Nanog é 4 (Metazoa) e para Oct4 é 14 (Euteleostomi). O resultado do SeedServer possibilitou portando a identificação de Nanog como um gene de origem antiga, para o qual inclusive observam-se homólogos em drosófila. O último ancestral comum de Oct4 teve também sua profundidade modificada, porém consideramos a origem em Euteleostomi como mais recente em relação a Nanog e concomitante com a ocorrência de um grande grupo de genes (Figura 15, Euteleostomi). Os agrupamentos contribuíram portanto para uma visualização mais correta da real composição da via e sua porção ancestral e moderna.

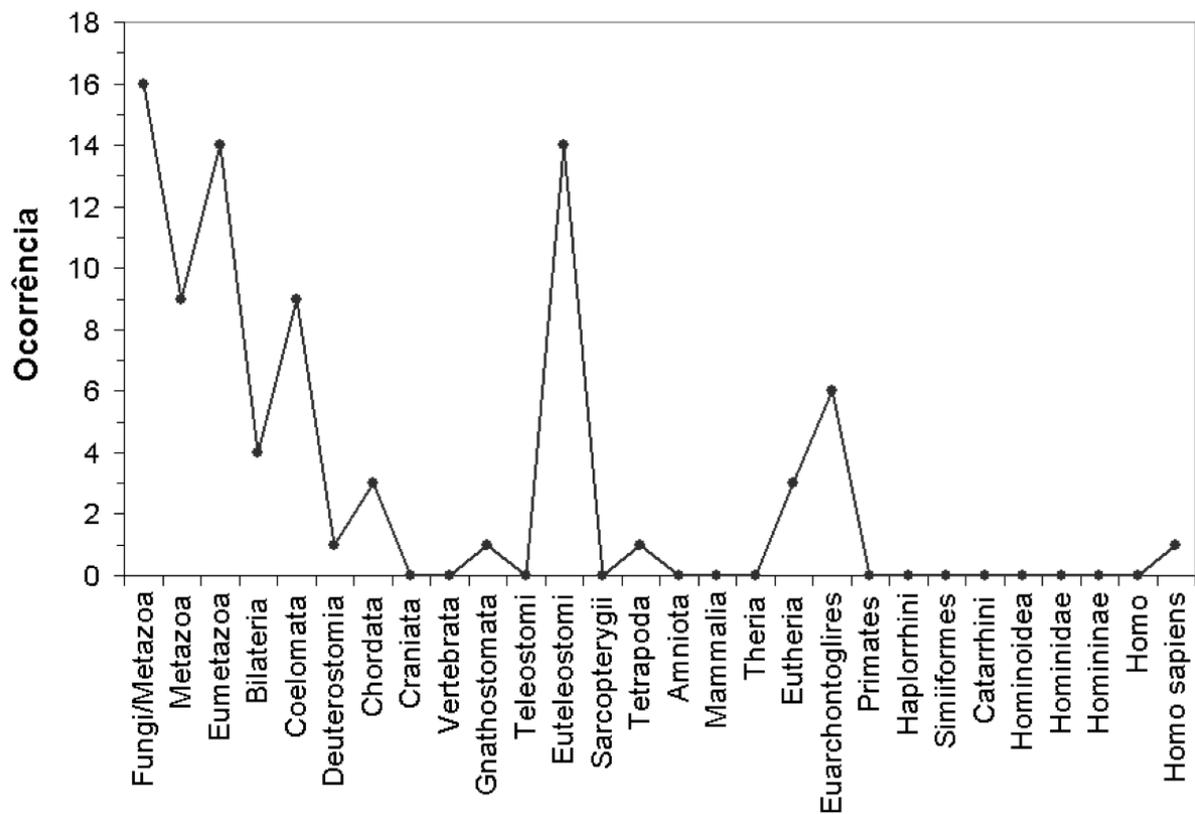


Figura 15: Distribuição da origem dos genes da via de pré-implantação. O gráfico mostra o surgimento dos genes envolvidos na via de desenvolvimento pré-implantação em momentos da evolução humana, segundo o último ancestral comum encontrado para os agrupamentos gerados pelo SeedServer. No eixo Y está representada a quantidade de genes e no eixo X localizam-se os grupos taxonômicos nos quais estes genes se originaram.

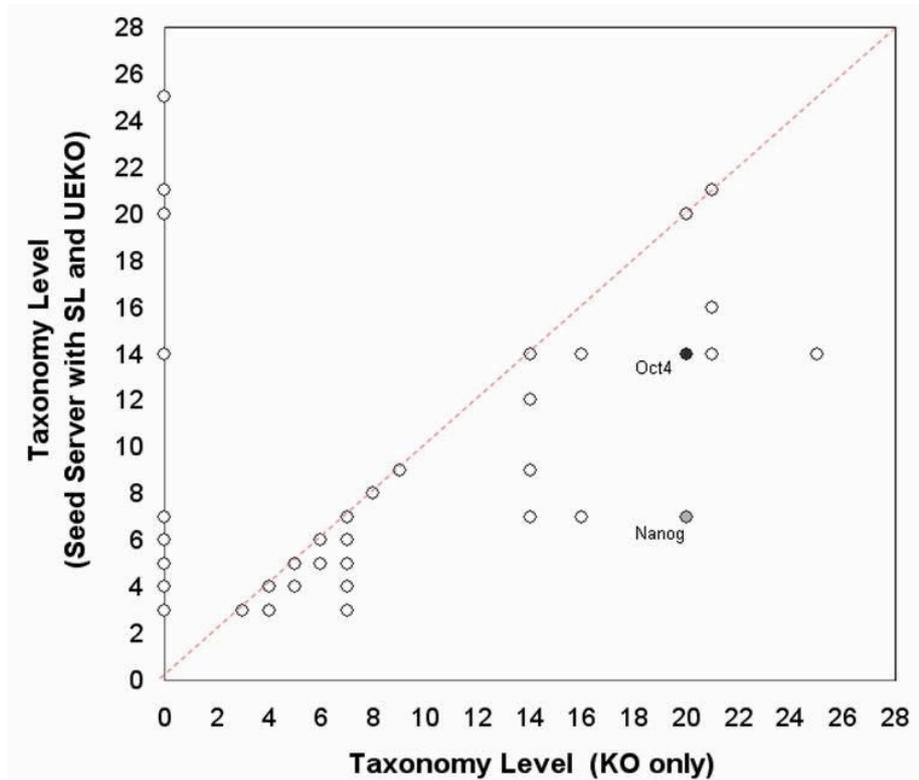


Figura 16: Ancestralidade obtida para genes da pré-implantação - comparação KO e SeedServer. O gráfico mostra o nível de profundidade do grupo taxonômico identificado como último ancestral comum dos genes da pré-implantação pelo KO (eixo X) e pelo SeedServer (eixo Y). Cada ponto no gráfico representa um gene para o qual determinamos o LCA. Dois genes estão destacados para exemplificar: Nanog e Oct4.

5.4.2 Funções de homólogos

Para analisar se os homólogos dos genes com origem antiga possuem funções relacionadas às exercidas no desenvolvimento pré-implantação, realizamos um segundo trabalho de mineração de texto para identificar estes genes em trabalhos descrevendo o desenvolvimento embrionário em *Drosophila melanogaster*. Trabalhos selecionados na PubMed com o termo de busca “Drosophila AND development” foram adicionados ao PESCADOR e os símbolos de genes marcados pelo software permitiram a análise dos homólogos encontrados anteriormente. Uma segunda via de regulações foi gerada a partir da análise destes trabalhos com o objetivo de mostrar as funções e as regulações das quais participam estes genes homólogos. A via mostrada na Figura 17 não tem como objetivo ilustrar todo o desenvolvimento de drosófila mas sim as regulações nas quais os homólogos têm participação. Uma correlação entre os símbolos dos genes humanos e o correspondente em drosófila pode ser encontrada na Tabela 3. Podemos observar várias regulações interessantes no desenvolvimento de *D. melanogaster* como por exemplo a via de sinalização Hippo que é extremamente conservada, já que Wts (homólogo de Lats) tem como função fosforilar Yki (homólogo de Yap) que no estado não fosforilado interage com Sd (homólogo de Tead4) e ativam juntos vias relacionadas à proliferação e desenvolvimento da asa no embrião (Huang, Wu *et al.*, 2005). A proteína Caudal (homóloga de Cdx2) está envolvida nos mesmos processos (Hwang, Kim *et al.*, 2002; Lengyel e Iwaki, 2002) mas em drosófila sua ativação não é feita pela parceria de Yki e Sd mas sim pela diminuição da concentração da proteína Bicoid que inibe a tradução do seu mRNA (Niessing, Blanke *et al.*, 2002). Outras regulações, embora não igualmente conservadas, mostram o envolvimento dos genes em funções homólogas às exercidas no desenvolvimento pré-implantação, como Grn (homólogo de GATA2) e Sox7 atuando na diferenciação de células em processos de neurogênese (Brown e Castelli-Gair Hombria, 2000; Kerner, Simionato *et al.*, 2009). Outro exemplo pode ser visto na repressão gênica exercida por Gro (homólogo de TLE2), Vnd e Scro (homólogos de Nanog), que na pré-implantação são também responsáveis pela repressão de diversos genes e são essenciais para o desenvolvimento inicial do embrião (Zaffran, Das *et al.*, 2000; Uhler, Zhang *et al.*, 2007). Outras regulações podem ser observadas na figura 17 e as referências correspondentes a cada uma delas podem ser encontradas na Tabela 3.

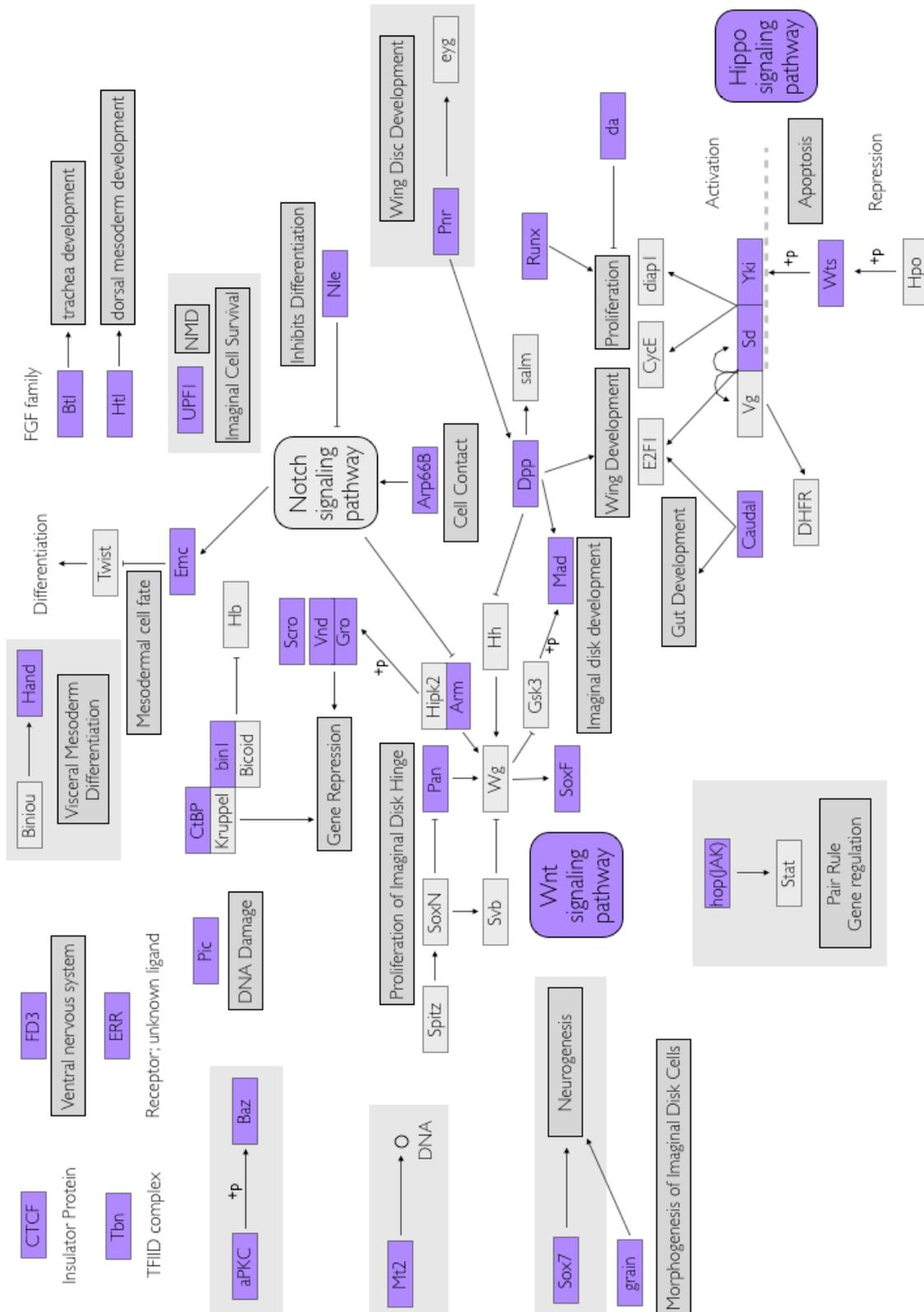


Figura 17: Funções dos genes homólogos encontrados em *Drosophila melanogaster*. A figura mostra as respectivas funções da maioria dos genes homólogos presentes em *D. melanogaster* encontrados nos agrupamentos feitos pelo SeedServer. Observa-se que estes homólogos estão envolvidos em processos relacionados ao desenvolvimento do embrião de *D. melanogaster*. Para uma correspondência entre o nome dos genes representados nesta figura e o nome do gene homólogo da via de preimplantação ver Tabela 3.

Tabela 3: Correspondência entre os nomes dos genes humanos e homólogos de *Drosophila melanogaster* agrupados pelo SeedServer.

Gene	Gene de <i>D. melanogaster</i>	Referência função (PMID)
BORIS	CTCF	19210964; 17805343
PARD3	Baz	19472188
RENT1	Upf1	17196039
GATA6	Pnr	12900462; 11731463
CDX2	Cad	11846473; 12368268
SOX7	Sox100B; Q9va16	19754708
LATS2	Wts	16096061; 18328423
NLE1	Nle	16785210; 9857191
DPP3	DppIII	12846841; 9006064; 8700205
ATP6AP2	CG8444	não encontrada
PKA	PkaC1; PkaC2; PkaC3	18646611
TCF3	Da	12070099; 15024404
aPKC	Pkc53e; Wnt6	19472188
GATA2	Grn	11044401
DNMT1	Mt2	12944428; 15533947
CtBP	Rbbp8	19049982
YAP1	Yki	16096061; 18331708
ARP3	Arp66B	18976918; 19543274
RPS14	RpS14a	não encontrada
CTNNB1	Arm	19668359; 19232093; 19088090
SAP18	Bin1	19049982; 15649476; 16823614
TLE2	Gro	17070676; 15802274
GATA4	Pnr	12900462; 11731463
ETIF2B	Eif2B	12466526; 18313299
TEAD4	Sd	18258485; 18313299
DDB1	Pic	15514059
NANOG	Scro; Vnd	17070676; 10842079
JAK2	Hop	11804783; 12441298
TBN	Taf8	11438666
HAND1	Hand	17904115; 17511863; 16467358
BMP4	Dpp	9006064; 9012523
GATA3	Grn; GATAc grain	11044401
ID2	Emc	11840322; 15128668; 16026970
RUNX1	CG34145	19447650; 18758811
FOXD3	FD3	1356269
SOX17	Sox15; Sox50e	19176582
ESRRB	ERR	12767224
KLF5	Nucleic acid binding (sem nome)	não encontrada
FGFR2	Btl; Htl	12175485; 15084280
HNF4A	Hnf4	19254568
SMAD1	Mad	19657393; 19896409; 18997322

5.5. Novos parceiros de interação para Tead4

5.5.1 Duplo-Híbrido em levedura

Para procurar novas proteínas que interagem com Tead4 no desenvolvimento embrionário e que, conseqüentemente, podem estar envolvidas em outras regulações dependentes de Tead4 ainda não bem compreendidas, realizamos uma triagem pelo sistema de duplo-híbrido em leveduras. Utilizando Tead4 como isca, realizamos uma triagem de biblioteca de embrião de sete dias (Clontech) pelo sistema de duplo-híbrido em meio líquido previamente estabelecido na levedura KGY37. Foram feitas 60 receitas de transformação pelo método “High Efficiency” (Gietz e Schiestl, 2007a) cada uma utilizando 10 µg de DNA da biblioteca e 10 µg de pAS1-Tead4. O rendimento obtido foi de aproximadamente 5×10^5 células, suficiente para a triagem da biblioteca utilizada.

5.5.2 Triagem líquida dos transformantes obtidos

A transformação foi distribuída em 600 mL de meio líquido SD -W-L-H suplementado com 3AT 0,5 mM e distribuída em quatro placas “deepwell” com 96 poços cada. As incubações e diluições foram realizadas de acordo com o protocolo descrito em “Métodos” e após este período as placas foram examinadas para verificar a presença de crescimento nos poços. Os poços com crescimento sinalizam a ativação do gene repórter HIS3, que acontece apenas em leveduras contendo uma interação entre Tead4 e um dos clones da biblioteca, permitindo seu crescimento no meio sem histidina. A leitura de OD600 mostrou que 62 poços apresentavam crescimento e portanto estes foram plaqueados em meio SD -W-L para o isolamento de colônias.

5.5.3 Ensaio de *LacZ*

Para confirmar a presença da interação as colônias isoladas de cada um dos poços com crescimento foram submetidas a um teste de ativação do segundo gene repórter, *LacZ*. O ensaio de *LacZ* (descrição em “Métodos”) mostrou a ativação de todos os clones com crescimento (todos ficaram azuis; dados não mostrados). Alguns dos muitos clones foram utilizados para demonstrar que a triagem produzia interações fortes restauráveis.

5.5.4 Análise e confirmação das interações

Os plasmídios contidos nas leveduras positivas foram extraídos e o plasmídio pGAD10 contendo o inserto da biblioteca foi seqüenciado para revelar a identidade da proteína que mostrou interação com Tead4. Foi possível realizar esta etapa da análise dos clones com poucas amostras sendo que dois clones foram seqüenciados com sucesso. Um deles mostrou que o inserto correspondia à proteína Tioredoxina1 (Trx1, acesso RefSeq NP_035790) e o outro mostrou correspondência com a fita contrária à codificante para a

proteína Smg8. Existem registros de uma proteína hipotética (LOC74133) no local segundo o banco de dados Genome do NCBI.

Realizamos também a reconstrução da interação observada através de uma nova transformação com a levedura KGY37. A levedura foi transformada com 1 µg de DNA isca (pAS1-Tead4) e 1 µg do pGAD10-Tioredoxina purificado ou com 1 µg de DNA isca e 1 µg de pGAD10-ProtHipotética. As transformações foram plaqueadas em meio SD -W-L e as colônias obtidas foram submetidas a um ensaio de ativação do gene *LacZ* no qual se mostraram novamente positivas. A Figura 17 mostra as colônias azuis no ensaio *LacZ* reconfirmando a interação dos clones isolados da biblioteca com a isca original. Esta reconstrução é necessária para verificar se os resultados positivos são corretos e não devido a alguma mutação sofrida pela isca durante o processo de triagem ou devido a uma ativação independente da isca (somente a proteína presa ativa a transcrição).

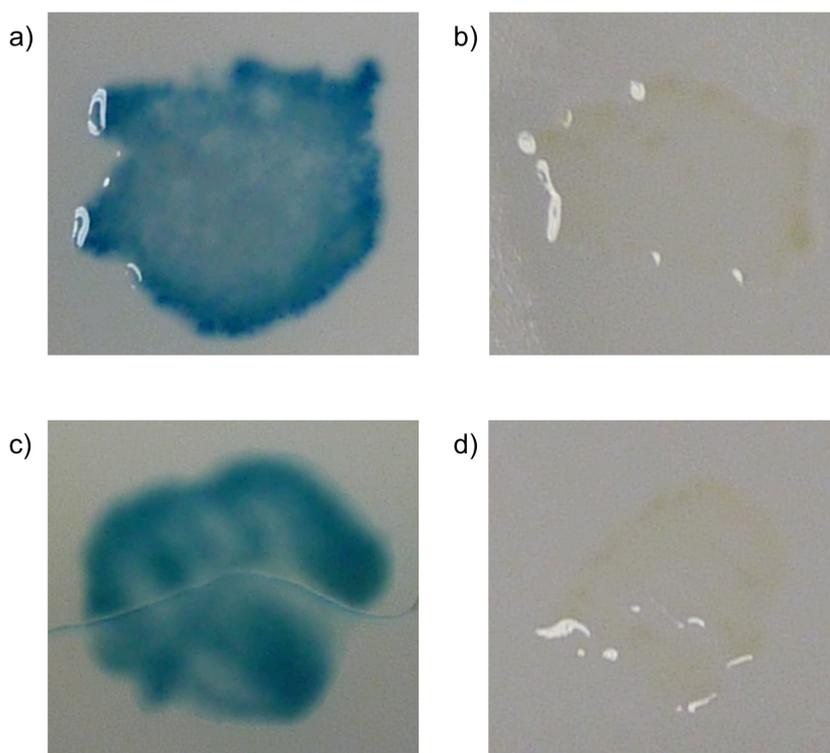


Figura 18: Ensaio de LacZ para reconstrução da interação com Tead4. **a)** Ensaio de LacZ com a levedura contendo o plasmídeo isolado da triagem cujo inserto foi identificado como Tioredoxina1 e o plasmídeo contendo a isca Tead4 original, a interação ativa a transcrição de LacZ. **b)** Levedura contendo somente o plasmídeo da Tioredoxina1, o resultado do ensaio LacZ é negativo neste caso. **c)** Levedura contendo o plasmídeo isolado da triagem cujo inserto foi identificado como uma proteína hipotética e o plasmídeo contendo a isca original, a interação ativa a transcrição de LacZ. **d)** Levedura contendo somente o plasmídeo da proteína hipotética, o resultado do ensaio LacZ é negativo neste caso.

6. DISCUSSÃO

O sistema de duplo-híbrido em leveduras é um dos métodos frequentemente utilizados para a busca de interações proteína-proteína e o aperfeiçoamento deste método, cujo protocolo de seleção é considerado laborioso, é desejável. A geração de dados genômicos em larga escala deve ser acompanhada também pela determinação funcional, que depende de métodos como o sistema de duplo-híbrido. O protocolo de triagem em meio líquido relatado neste trabalho resulta na redução do crescimento de clones de interação fraca e a abordagem descrita em placas “deepwell” é facilmente adaptada para uma aplicação robotizada, aumentando a eficiência do método. A triagem de bibliotecas complexas com ótima cobertura torna-se facilmente atingível, mas para que esta triagem seja eficiente é necessário obter uma alta eficiência de transformação. Verificamos que o protocolo desenvolvido por Gietz e Schiestl é atualmente o melhor em relação à eficiência de transformação, assim como as condições de cultivo das leveduras descritas neste trabalho, pois existe uma correlação entre o rendimento da transformação e o bom crescimento das leveduras no meio de cultivo antes das transformações (Gietz, University of Manitoba; comunicação pessoal). Além disso, a utilização dos frascos na triagem também se mostrou interessante para selecionar clones com interações mais fortes, medida pela capacidade das leveduras contendo estas interações de dominar as culturas. Isto foi observado na triagem relatada para validação do sistema, com a isca Trim2, na qual das culturas crescidas de frascos quase todas eram dominadas por somente uma interação e esta mesma interação também foi selecionada individualmente no ensaio simultâneo com placas “deepwell”. O protocolo de triagem líquida provavelmente irá substituir a triagem em placas convencional nos laboratórios que aplicam este método, esta mudança já ocorreu nos laboratórios de Gietz e colaboradores da Universidade de Manitoba, Canadá.

A utilização de ferramentas de mineração de texto para a geração de vias regulatórias é uma abordagem eficaz e importante para o interesse atual em reunir dados sobre um organismo ou processo biológico. Outros trabalhos que buscam também esta integração de dados regulatórios podem ser encontrados, um deles é considerado o melhor exemplo atualmente de via regulatória gerada para um processo biológico, este trabalho é a geração da via regulatória que controla o desenvolvimento embrionário do ouriço-do-mar (Davidson, Rast *et al.*, 2002). A via compreende inicialmente 40 genes, mas foi posteriormente atualizada (Peter e Davidson, 2010). Um aspecto interessante deste trabalho é a inclusão de elementos regulatórios em *cis* na representação, ilustrando melhor a atuação dos fatores de transcrição envolvidos no processo em diferentes promotores. Esta

representação é interessante para evitar a má compreensão dos dados ilustrados nas vias tradicionais, nas quais muitas vezes um nódulo representa um gene e ao mesmo tempo sua proteína. Outros trabalhos dos autores citados também foram publicados para completar a via regulatória gerada com novas referências e para enfatizar a importância da geração de vias regulatórias que auxiliam na compreensão dos dados biológicos (Li e Davidson, 2009). Todavia, devido à universalidade de uso da base de dados KEGG Pathway e o interesse atual em estabelecer padrões para os dados gerados, a linguagem de marcação da mesma (KGML) foi utilizada aqui (Figuras 14 e 16).

A busca de trabalhos relacionados com um conceito específico como desenvolvimento pré-implantação resultou na seleção de dados relativos somente a este processo. Quando são utilizadas outras ferramentas como iHOP (Hoffmann e Valencia, 2004) e STRING (Jensen, Kuhn *et al.*, 2009) para a busca de biointerações é necessário saber o nome do gene que se tem interesse e as informações são então recuperadas. No caso do iHOP, estas últimas consistem em uma extensa lista de trabalhos relacionados àquele gene que devem ser manualmente analisados para se extrair a informação relativa ao processo de interesse. Já no STRING, as associações com outros genes podem ser diretas como ativações ou repressões ou podem ser desconhecidas e em todos os casos não é possível realizar a pesquisa restringindo a busca a um processo específico no qual se deseja verificar a participação de um gene. A abordagem utilizada neste trabalho (PubMed/MedlineRanker/PESCADOR) permite iniciar a pesquisa sem se conhecer os genes envolvidos, somente selecionando os trabalhos relacionados ao processo de interesse. A curadoria manual exigida para criar uma via desta maneira é consideravelmente menor. Contudo, a verificação de todas as interações destacadas pela ferramenta é indispensável. A mineração de texto não é capaz de eliminar a seleção de falsos pares de interação; no caso do LAITOR (contido na plataforma PESCADOR) as interações do tipo 3 e 4 podem conter genes que não possuem associação especificada no texto. Os dados de mineração de texto contribuem ao mesmo tempo para a descrição da via na forma de revisão da literatura, uma etapa necessária para a validação das regulações representadas e para a inclusão da via em uma base de dados específica, tal como KEGG Pathway (Kanehisa - Kyoto University Bioinformatics Center; comunicação pessoal). O domínio deste procedimento para criação de vias possibilita a ampliação de temas atualmente não abordados, como vias regulatórias de diversos tipos de câncer, mecanismos de resistência a patógenos em plantas, resposta de plantas ao estresse abiótico e seca, dentre outros temas de interesse do laboratório e de futuros colaboradores.

A inserção desta nova via no banco de dados KEGG irá permitir a anotação automatizada para novos organismos, como normalmente é feito neste banco. Dos 86 genes que participam da via, 20 não possuem entradas em KEGG Orthology e serão

adições importantes (Figura 13b). Considerando que a contribuição do KEGG para o recrutamento de sequências nos agrupamentos do SeedServer é de somente 25%, alguns organismos ocasionalmente divergentes dos que estão representados no KEGG passam a ter um papel importante para uma anotação mais eficiente de novas sequências. É importante ressaltar que somente os componentes SeedLinkage e UEKO do SeedServer são capazes de agrupar sequências provenientes de organismos sem projeto genoma completo.

Outra contribuição importante do SeedServer é a identificação de candidatos a anotação SwissProt nos agrupamentos (Figura 13a). A anotação SwissProt depende da correta associação de sequências a famílias gênicas e proteínas com função já conhecida, utilizando para tanto a literatura disponível. A anotação pode ser auxiliada pelo fato de que cada um dos genes da via encontra-se associado a PMIDs, que são referências importantes para os homólogos respectivos. Estas referências ficam armazenadas permanentemente na ferramenta PESCADOR. As referências de publicação são utilizadas também pelo banco de dados Gene como fontes de informação funcional para os genes chamadas de GeneRIFs (“Gene reference into function”) e podem ser facilmente associadas com o gene respectivo por usuários da ferramenta.

A pesquisa de informações funcionais para os homólogos de drosófila revelou a participação dos genes em processos relacionados de desenvolvimento embrionário e foi também uma boa validação para o agrupamento de sequências homólogas pelo SeedServer, já que nenhuma sequência de *D. melanogaster* agrupada aos genes iniciais possui uma função divergente do esperado.

A geração de agrupamentos corretos é essencial para uma determinação correta da ancestralidade dos genes, mas não é o único fator limitante. O sequenciamento de novos organismos pertencentes a grupos taxonômicos irmãos dos que congregam os organismos com genomas já sequenciados será uma importante fonte de sequências que permitirão uma reavaliação da ancestralidade real dos genes. Um exemplo seria o sequenciamento do genoma da barata (*Periplaneta americana*), possivelmente auxiliando na compreensão da evolução de insetos. Todavia, foi importante a análise da ancestralidade da via de Desenvolvimento Embrionário Pré-implantação, por ser uma via com componentes esperadamente mais contemporâneos. Este estudo inicial inspirou a análise global de ocorrência de homólogos para os genes de *Homo sapiens*, trabalho em andamento no laboratório. Esta determinação do último ancestral comum pode ser aplicada também a diversas vias representadas no banco de dados KEGG Pathway, como por exemplo vias de reparo de DNA ou de patógenos e suas doenças, para descrever a evolução destes sistemas.

Um dos principais fatores envolvidos no desenvolvimento pré-implantação é a proteína Tead4. A interação do fator de transcrição Tead4 com a proteína YAP é bem descrita (Ralston e Rossant, 2008), porém esta interação somente não é capaz de explicar todas as regulações das quais Tead4 participa como observado recentemente por Ralston e colaboradores. A triagem realizada utilizando Tead4 como isca no sistema de duplo-híbrido em levedura resultou na confirmação da ativação dos genes repórteres para dois clones, um deles identificado por sequenciamento como sendo a proteína Tioredoxina1. A Tioredoxina1 é uma proteína com papel fundamental na manutenção do estado redox das células (Holmgren, 1985). Esta proteína também está envolvida em processos celulares como proliferação, apoptose e expressão gênica. Foram encontradas diversas evidências na literatura da interação de Tioredoxina com uma variedade de fatores de transcrição. Um dos trabalhos iniciais a respeito da interação de Tioredoxina com um fator de transcrição (Qin, Clore *et al.*, 1995) mostra que esta proteína reduz certos resíduos de cisteína presentes no fator NF κ B e com isso ativa as propriedades de ligação ao DNA do mesmo; a resolução da estrutura da proteína complexada com um peptídeo mostra também a elevada especificidade da ligação, estabilizada por muitas interações. Outra interação foi bem descrita na resposta a situações de estresse como a exposição das células à radiação ionizante (Wei, Botero *et al.*, 2000). A Tioredoxina1 interage com Ref-1 e juntas promovem a ativação transcricional de AP-1, complexo transcricional formado pelas proteínas Fos/Jun, mas a Tioredoxina deve ser previamente transportada para o núcleo por mecanismos ainda não bem descritos. A interação resulta na ativação da ligação ao DNA de AP-1 como previamente descrito para NF κ B. A interação de Tioredoxina com Ref-1 foi confirmada em estudos de duplo-híbrido em mamíferos e por estudos *in vitro* com ligação cruzada (Hirota, Matsui *et al.*, 1997). O complexo AP-1 possui um coativador chamado Jab1, esta proteína se liga ao domínio de ativação de Jun e auxilia na ativação da transcrição induzida por este complexo. Jab1 é responsável por promover a degradação do inibidor de uma cinase dependente de ciclina p27Kip1. Estudos de co-imunoprecipitação e FRET identificaram a interação de Tioredoxina com Jab1 que resulta na modulação da função exercida por Jab1. Esta modulação também resulta do fato de Tioredoxina competir com Jab1 pela ligação de p27Kip1, indicando um papel para Tioredoxina no controle da proliferação (Hwang, Ryu *et al.*, 2004).

Outros estudos mostram que Tioredoxina também é capaz de aumentar a atividade de ligação a DNA de p53, e ainda atuar em conjunto com Ref-1 (assim como na ativação de Fos/Jun) para a potencialização do efeito de Ref-1 sobre p53. A expressão de p21 dependente de p53 foi aumentada em ensaios de transfecção com Tioredoxina e novamente foi observado um efeito melhor da ativação quando o experimento incluía um composto que induz a translocação da Tioredoxina para o núcleo (Ueno, Masutani *et al.*, 1999). A respeito

da localização celular da Tioredoxina, foi observado em estudos com culturas de células que quando a cultura encontra-se confluenta a localização de Tioredoxina é predominantemente citoplasmática enquanto nas culturas cujas células estão espalhadas a localização de Tioredoxina é nuclear, sugerindo que os ambientes redox nuclear e citoplasmáticos e os requerimentos celulares são diferentes durante as duas condições (Spielberger, Moody *et al.*, 2008). Por último, a interação de Tioredoxina com o receptor de glucocorticoide (pertencente a uma família de fatores de transcrição) e a consequente ativação de seu domínio de ligação ao DNA foi também observada em estudos de duplo-híbrido em mamífero e a localização deste complexo de interação é nuclear durante condições de estresse oxidativo (Makino, Yoshikawa *et al.*, 1999).

No desenvolvimento embrionário, uma possibilidade seria o envolvimento da Thioredoxina1 no controle do estado redox das células que se encontram em uma fase altamente proliferativa. Talvez a identificação de Thioredoxina como parceira de interação de Tead4 signifique que este fator de transcrição precisa de um mecanismo de controle que garanta a redução de resíduos de seu sítio de ligação ao DNA, garantindo uma ativação máxima quando o fator se associa a YAP.

Todas estas referências corroboram com os resultados obtidos no ensaio de duplo-híbrido em leveduras com Tead4 e indicam que a interação com Tioredoxina é de fato verdadeira, pelo menos em leveduras, e pode levar ao desvendamento de um novo controle transcricional envolvendo estas duas proteínas no desenvolvimento embrionário pré-implantação. Evidentemente a interação funcional das duas proteínas precisa de uma confirmação através de ensaios *in vivo* com células de mamífero e vetores de expressão para que seja possível afirmar com clareza a existência desta interação e também para determinar as consequências resultantes deste mecanismo no controle transcricional. Estes estudos serão realizados em continuidade pelo laboratório.

7. CONCLUSÕES

- A triagem líquida para o sistema de duplo-híbrido em leveduras torna a seleção de interações entre proteínas por este método mais eficiente e menos laboriosa, possibilitando duas abordagens diferentes dependendo do interesse da pesquisa. A abordagem em placas “deepwell” pode ser utilizada em estações robóticas para experimentos em larga escala ou manualmente quando o interesse é a seleção de diversas interações. A abordagem em frascos gera uma competição entre as leveduras e possibilita a seleção de clones contendo interações mais fortes. As linhagens Y190 e KGY37 foram certificadas para o protocolo.
- A geração de vias regulatórias através de ferramentas de mineração de texto permite a integração de dados gerados em estudos prévios para que a visualização de um processo biológico seja mais completa. Quando os genes presentes nesta via são associados a grupos de homólogos acrescenta-se esta informação à via regulatória permitindo a visualização do mesmo processo para diferentes organismos. Os homólogos também possibilitam a determinação da ancestralidade dos genes envolvidos no processo levando a uma melhor compreensão do mesmo.
- A triagem pelo sistema de duplo-híbrido utilizando a proteína Tead4 como isca permitiu a seleção eficiente de novas proteínas que podem estar envolvidas nos processos de regulação do desenvolvimento embrionário dos quais o fator de transcrição Tead4 participa.

8. PERSPECTIVAS

- Descrever a via regulatória gerada em um artigo de revisão para a validação da mesma e a inclusão desta via no banco de dados KEGG Pathway.
- Contribuir para a atualização do banco de dados UniProt realizando a anotação SwissProt de sequências agrupadas pelo SeedServer.
- Iniciar outros trabalhos de mineração de texto para gerar vias regulatórias correspondentes a outros processos biológicos que não estão representados no KEGG e outros bancos de dados (vários tipos de câncer, por exemplo).
- Realizar experimentos em células de mamífero para confirmar a interação de Tead4 com as duas proteínas isoladas na triagem de duplo-híbrido e verificar os efeitos destas interações nas células.

9. REFERÊNCIAS

ADJAYE, J. *et al.* Conserved molecular portraits of bovine and human blastocysts as a consequence of the transition from maternal to embryonic control of gene expression. *Physiol Genomics* [S.I.], v. 31, n. 2, p. 315-27, Oct 22 2007.

_____. Primary differentiation in the human blastocyst: comparative molecular portraits of inner cell mass and trophectoderm cells. *Stem Cells* [S.I.], v. 23, n. 10, p. 1514-25, Jan 1 2005.

ALTENHOFF, A. M.; DESSIMOZ, C. Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Comput Biol* [S.I.], v. 5, n. 1, p. e1000262, Jan 2009.

AMBROSETTI, D. C. *et al.* Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem* [S.I.], v. 275, n. 30, p. 23387-97, Jul 28 2000.

ANANIADOU, S. *et al.* Event extraction for systems biology by text mining the literature. *Trends Biotechnol* [S.I.], May 28 2010.

ANBANANDAM, A. *et al.* Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. *Proc Natl Acad Sci U S A* [S.I.], v. 103, n. 46, p. 17225-30, Nov 14 2006.

ANDRIANOPOULOS, A.; TIMBERLAKE, W. E. ATTS, a new and conserved DNA binding domain. *Plant Cell* [S.I.], v. 3, n. 8, p. 747-8, Aug 1991.

BABAIE, Y. *et al.* Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells* [S.I.], v. 25, n. 2, p. 500-10, Feb 1 2007.

BARBOSA-SILVA, A. *et al.* Clustering of cognate proteins among distinct proteomes derived from multiple links to a single seed sequence. *BMC Bioinformatics* [S.I.], v. 9, p. 141, 2008.

_____. LAITOR--Literature Assistant for Identification of Terms co-Occurrences and Relationships. *BMC Bioinformatics* [S.I.], v. 11, p. 70, 2010.

BOUTET, E. *et al.* UniProtKB/Swiss-Prot. *Methods Mol Biol* [S.I.], v. 406, p. 89-112, 2007.

BRAY, S. Drosophila development: Scalloped and Vestigial take wing. *Curr Biol* [S.I.], v. 9, n. 7, p. R245-7, Apr 8 1999.

BROWN, S.; CASTELLI-GAIR HOMBRIA, J. Drosophila grain encodes a GATA transcription factor required for cell rearrangement during morphogenesis. *Development* [S.I.], v. 127, n. 22, p. 4867-76, Nov 2000.

BURGLIN, T. R. The TEA domain: a novel, highly conserved DNA-binding motif. *Cell* [S.I.], v. 66, n. 1, p. 11-2, Jul 12 1991.

CAO, X. *et al.* YAP regulates neural progenitor cell number via the TEA domain transcription factor. *Genes Dev* [S.I.], v. 22, n. 23, p. 3320-34, Dec 1 2008.

CAUFFMAN, G. *et al.* Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. *Hum Reprod* [S.I.], v. 24, n. 1, p. 63-70, Jan 1 2009.

CHAKRAVARTHY, H. *et al.* Identification of DPPA4 and other genes as putative Sox2:Oct-3/4 target genes using a combination of in silico analysis and transcription-based assays. *J Cell Physiol* [S.I.], v. 216, n. 3, p. 651-62, Sep 2008.

CHEN, L. *et al.* Suppression of ES cell differentiation by retinol (vitamin A) via the overexpression of Nanog. *Differentiation* [S.I.], v. 75, n. 8, p. 682-93, Oct 1 2007.

CHIEN, C. T. *et al.* The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* [S.I.], v. 88, n. 21, p. 9578-82, Nov 1 1991.

CHUNG, Y. G. *et al.* Abnormal regulation of DNA methyltransferase expression in cloned mouse embryos. *Biol Reprod* [S.I.], v. 69, n. 1, p. 146-53, Jul 2003.

COLLINS, J. E. *et al.* Regulation of desmocollin transcription in mouse preimplantation embryos. *Development* [S.I.], v. 121, n. 3, p. 743-53, Mar 1 1995.

CORMIER, S. *et al.* The murine ortholog of notchless, a direct regulator of the notch pathway in *Drosophila melanogaster*, is essential for survival of inner cell mass cells. *Mol Cell Biol* [S.I.], v. 26, n. 9, p. 3541-9, May 2006.

COUCOUVANIS, E.; MARTIN, G. R. BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. *Development* [S.I.], v. 126, n. 3, p. 535-46, Feb 1 1999.

DAVIDSON, E. H. *et al.* A genomic regulatory network for development. *Science* [S.I.], v. 295, n. 5560, p. 1669-78, Mar 1 2002.

DE FELICI, M. *et al.* In or out stemness: comparing growth factor signalling in mouse embryonic stem cells and primordial germ cells. *Curr Stem Cell Res Ther* [S.I.], v. 4, n. 2, p. 87-97, May 2009.

DESHPANDE, N. *et al.* The human transcription enhancer factor-1, TEF-1, can substitute for *Drosophila* scalloped during wingblade development. *J Biol Chem* [S.I.], v. 272, n. 16, p. 10664-8, Apr 18 1997.

DIAZ-CAMINO, C. *et al.* A high-throughput system for two-hybrid screening based on growth curve analysis in microtiter plates. *Anal Biochem* [S.I.], v. 316, n. 2, p. 171-4, May 15 2003.

DUNCAN, S. A. *et al.* Expression of transcription factor HNF-4 in the extraembryonic endoderm, gut, and nephrogenic tissue of the developing mouse embryo: HNF-4 is a marker for primary endoderm in the implanting blastocyst. *Proc Natl Acad Sci USA* [S.I.], v. 91, n. 16, p. 7598-602, Aug 2 1994.

FERNANDES, G. R. *et al.* A procedure to recruit members to enlarge protein family databases--the building of UECOG (UniRef-Enriched COG Database) as a model. *Genet Mol Res* [S.I.], v. 7, n. 3, p. 910-24, 2008.

FIELDS, S.; SONG, O. A novel genetic system to detect protein-protein interactions. *Nature* [S.I.], v. 340, n. 6230, p. 245-6, Jul 20 1989.

FIELDS, S.; STERNGLANZ, R. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet* [S.I.], v. 10, n. 8, p. 286-92, Aug 1994.

FITCH, W. M. Distinguishing homologous from analogous proteins. *Syst Zool* [S.I.], v. 19, n. 2, p. 99-113, Jun 1970.

FONTAINE, J. F. *et al.* MedlineRanker: flexible ranking of biomedical literature. *Nucleic Acids Res* [S.I.], v. 37, n. Web Server issue, p. W141-6, Jul 1 2009.

FUTAKI, S. *et al.* Sox7 plays crucial roles in parietal endoderm differentiation in F9 embryonal carcinoma cells through regulating Gata-4 and Gata-6 expression. *Mol Cell Biol* [S.I.], v. 24, n. 23, p. 10492-503, Dec 1 2004.

GIETZ, R. D.; SCHIESTL, R. H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* [S.I.], v. 2, n. 1, p. 38-41, 2007a.

_____. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* [S.I.], v. 2, n. 1, p. 35-7, 2007b.

GIETZ, R. D. *et al.* Identification of proteins that interact with a protein of interest: applications of the yeast two-hybrid system. *Mol Cell Biochem* [S.I.], v. 172, n. 1-2, p. 67-79, Jul 1997.

GUO, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* [S.I.], v. 18, n. 4, p. 675-85, Apr 20 2010.

HAYASHI, Y. *et al.* BMP4 induction of trophoblast from mouse embryonic stem cells in defined culture conditions on laminin. *In Vitro Cell Dev Biol Anim* [S.I.], v. 46, n. 5, p. 416-30, May 1 2010.

HECKER, M. *et al.* Gene regulatory network inference: data integration in dynamic models-a review. *Biosystems* [S.I.], v. 96, n. 1, p. 86-103, Apr 2009.

HEMMING, R. *et al.* Human growth factor receptor bound 14 binds the activated insulin receptor and alters the insulin-stimulated tyrosine phosphorylation levels of multiple proteins. *Biochem Cell Biol* [S.I.], v. 79, n. 1, p. 21-32, 2001.

HIROTA, K. *et al.* AP-1 transcriptional activity is regulated by a direct association between thioredoxin and Ref-1. *Proc Natl Acad Sci U S A* [S.I.], v. 94, n. 8, p. 3633-8, Apr 15 1997.

HOFFMANN, R.; VALENCIA, A. A gene network for navigating the literature. *Nat Genet* [S.I.], v. 36, n. 7, p. 664, Jul 2004.

HOLMGREN, A. Thioredoxin. *Annu Rev Biochem* [S.I.], v. 54, p. 237-71, 1985.

HUANG, J. *et al.* The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila Homolog of YAP. *Cell* [S.I.], v. 122, n. 3, p. 421-34, Aug 12 2005.

HWANG, C. Y. *et al.* Thioredoxin modulates activator protein 1 (AP-1) activity and p27Kip1 degradation through direct interaction with Jab1. *Oncogene* [S.I.], v. 23, n. 55, p. 8868-75, Nov 25 2004.

HWANG, M. S. *et al.* The caudal homeodomain protein activates Drosophila E2F gene expression. *Nucleic Acids Res* [S.I.], v. 30, n. 23, p. 5029-35, Dec 1 2002.

- HYSLOP, L. *et al.* Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells* [S.I.], v. 23, n. 8, p. 1035-43, Sep 1 2005.
- IMMINK, R. G.; ANGENENT, G. C. Transcription factors do it together: the hows and whys of studying protein-protein interactions. *Trends Plant Sci* [S.I.], v. 7, n. 12, p. 531-4, Dec 2002.
- IWABUCHI, K. *et al.* Stimulation of p53-mediated transcriptional activation by the p53-binding proteins, 53BP1 and 53BP2. *J Biol Chem* [S.I.], v. 273, n. 40, p. 26061-8, Oct 2 1998.
- JENSEN, L. J. *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* [S.I.], v. 37, n. Database issue, p. D412-6, Jan 2009.
- JIANG, J. *et al.* A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* [S.I.], v. 10, n. 3, p. 353-60, Mar 1 2008.
- JOHNSON, M. H.; MCCONNELL, J. M. Lineage allocation and cell polarity during mouse embryogenesis. *Semin Cell Dev Biol* [S.I.], v. 15, n. 5, p. 583-97, Oct 2004.
- KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* [S.I.], v. 28, n. 1, p. 27-30, Jan 1 2000.
- KANEKO, K. J. *et al.* Transcription factor mTEAD-2 is selectively expressed at the beginning of zygotic gene expression in the mouse. *Development* [S.I.], v. 124, n. 10, p. 1963-73, May 1997.
- KANEKO, K. J.; DEPAMPHILIS, M. L. Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. *Dev Genet* [S.I.], v. 22, n. 1, p. 43-55, 1998.
- KARP, P. D. *et al.* Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* [S.I.], v. 33, n. 19, p. 6083-9, 2005.
- KEMP, C. *et al.* Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Dev Dyn* [S.I.], v. 233, n. 3, p. 1064-75, Jul 1 2005.
- KERNER, P. *et al.* Orthologs of key vertebrate neural genes are expressed during neurogenesis in the annelid *Platynereis dumerilii*. *Evol Dev* [S.I.], v. 11, n. 5, p. 513-24, Sep-Oct 2009.
- KONDO, M. *et al.* A role for Id in the regulation of TGF-beta-induced epithelial-mesenchymal transdifferentiation. *Cell Death Differ* [S.I.], v. 11, n. 10, p. 1092-101, Oct 1 2004.
- KRAWETZ, R.; KELLY, G. M. Wnt6 induces the specification and epithelialization of F9 embryonal carcinoma cells to primitive endoderm. *Cell Signal* [S.I.], v. 20, n. 3, p. 506-17, Mar 1 2008.
- KURIMOTO, K. *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res* [S.I.], v. 34, n. 5, p. e42, Jan 1 2006.
- LALONDE, S. *et al.* Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations. *Plant J* [S.I.], v. 53, n. 4, p. 610-35, Feb 2008.

LENGYEL, J. A.; IWAKI, D. D. It takes guts: the *Drosophila* hindgut as a model system for organogenesis. *Dev Biol* [S.I.], v. 243, n. 1, p. 1-19, Mar 1 2002.

LETUNIC, I. *et al.* iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci* [S.I.], v. 33, n. 3, p. 101-3, Mar 2008.

LI, E.; DAVIDSON, E. H. Building developmental gene regulatory networks. *Birth Defects Res C Embryo Today* [S.I.], v. 87, n. 2, p. 123-30, Jun 2009.

LI, L. *et al.* OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* [S.I.], v. 13, n. 9, p. 2178-89, Sep 2003.

MAKINO, Y. *et al.* Direct association with thioredoxin allows redox regulation of glucocorticoid receptor function. *J Biol Chem* [S.I.], v. 274, n. 5, p. 3182-8, Jan 29 1999.

MARIKAWA, Y.; ALARCÓN, V. B. Establishment of trophoctoderm and inner cell mass lineages in the mouse embryo. *Mol Reprod Dev* [S.I.], v. 76, n. 11, p. 1019-32, Nov 1 2009.

MATIN, M. M. *et al.* Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. *Stem Cells* [S.I.], v. 22, n. 5, p. 659-68, Jan 1 2004.

MEILHAC, S. M. *et al.* Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. *Dev Biol* [S.I.], v. 331, n. 2, p. 210-21, Jul 15 2009.

MIZUNO, S. *et al.* Expression and subcellular localization of GSE protein in germ cells and preimplantation embryos. *J Reprod Dev* [S.I.], v. 52, n. 3, p. 429-38, Jun 1 2006.

MONK, M. *et al.* Differential expression of the embryo/cancer gene ECSA(DPPA2), the cancer/testis gene BORIS and the pluripotency structural gene OCT4, in human preimplantation development. *Molecular Human Reproduction* [S.I.], v. 14, n. 6, p. 347-55, Jun 1 2008.

MORRIS, S. M. *et al.* Dual roles for the Dab2 adaptor protein in embryonic development and kidney transport. *EMBO J* [S.I.], v. 21, n. 7, p. 1555-64, Apr 2 2002.

MULLER, J. *et al.* eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res* [S.I.], v. 38, n. Database issue, p. D190-5, Jan 2010.

MURAKAMI, A. *et al.* SOX7 and GATA-4 are competitive activators of Fgf-3 transcription. *J Biol Chem* [S.I.], v. 279, n. 27, p. 28564-73, Jul 2 2004.

_____. Retinoic acid-regulated expression of fibroblast growth factor 3 requires the interaction between a novel transcription factor and GATA-4. *J Biol Chem* [S.I.], v. 274, n. 24, p. 17242-8, Jun 11 1999.

NIESSING, D. *et al.* Bicoid associates with the 5'-cap-bound complex of caudal mRNA and represses translation. *Genes Dev* [S.I.], v. 16, n. 19, p. 2576-82, Oct 1 2002.

NISHIOKA, N. *et al.* The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophoctoderm from inner cell mass. *Dev Cell* [S.I.], v. 16, n. 3, p. 398-410, Mar 1 2009.

- NOVERE, N. L. *et al.* The Systems Biology Graphical Notation. *Nat Biotech.* v. 27. n. 82009. p. 735-741.
- OKUMURA-NAKANISHI, S. *et al.* Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J Biol Chem* [S.I.], v. 280, n. 7, p. 5307-17, Feb 18 2005.
- OOI, H. S. *et al.* Databases of protein-protein interactions and complexes. *Methods Mol Biol* [S.I.], v. 609, p. 145-59, 2010.
- PAN, G. *et al.* A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *FASEB J* [S.I.], v. 20, n. 10, p. 1730-2, Aug 1 2006.
- PEREIRA, L. *et al.* Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* [S.I.], v. 26, n. 20, p. 7479-91, Oct 1 2006.
- PETER, I. S.; DAVIDSON, E. H. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev Biol* [S.I.], v. 340, n. 2, p. 188-99, Apr 15 2010.
- PLUSA, B. *et al.* Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *J Cell Sci* [S.I.], v. 118, n. Pt 3, p. 505-15, Feb 1 2005.
- QIN, J. *et al.* Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NF kappa B. *Structure* [S.I.], v. 3, n. 3, p. 289-97, Mar 15 1995.
- RALSTON, A. *et al.* Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* [S.I.], v. 137, n. 3, p. 395-403, Feb 1 2010.
- RALSTON, A.; ROSSANT, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev Biol* [S.I.], v. 313, n. 2, p. 614-29, Jan 15 2008.
- RILEY, P. *et al.* The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nat Genet* [S.I.], v. 18, n. 3, p. 271-5, Mar 1 1998.
- ROBSON, P. *et al.* Inner cell mass-specific expression of a cell adhesion molecule (PECAM-1/CD31) in the mouse blastocyst. *Dev Biol* [S.I.], v. 234, n. 2, p. 317-29, Jun 15 2001.
- SAITO, S. *et al.* Animal embryonic stem (ES) cells: self-renewal, pluripotency, transgenesis and nuclear transfer. *Hum Cell* [S.I.], v. 17, n. 3, p. 107-15, Sep 2004.
- SCAFFIDI, P.; BIANCHI, M. E. Spatially precise DNA bending is an essential activity of the sox2 transcription factor. *J Biol Chem* [S.I.], v. 276, n. 50, p. 47296-302, Dec 14 2001.
- SHIMODA, M. *et al.* Sox17 plays a substantial role in late-stage differentiation of the extraembryonic endoderm in vitro. *J Cell Sci* [S.I.], v. 120, n. Pt 21, p. 3859-69, Nov 1 2007.
- SHIN, M. R. *et al.* Identification of mouse blastocyst genes that are downregulated by double-stranded RNA-mediated knockdown of Oct-4 expression. *Mol Reprod Dev* [S.I.], v. 70, n. 4, p. 390-6, Apr 1 2005.
- SPIELBERGER, J. C. *et al.* Oxidation and nuclear localization of thioredoxin-1 in sparse cell cultures. *J Cell Biochem* [S.I.], v. 104, n. 5, p. 1879-89, Aug 1 2008.

SRIVASTAVA, R.; LAL, S. K. A liquid synchronized-growth culture assay for the identification of true positive and negative yeast three-hybrid transformants. *Lett Appl Microbiol* [S.I.], v. 34, n. 4, p. 300-3, 2002.

STRUMPF, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophoderm in the mouse blastocyst. *Development* [S.I.], v. 132, n. 9, p. 2093-102, May 1 2005.

SUN, C. *et al.* Dax1 binds to Oct3/4 and inhibits its transcriptional activity in embryonic stem cells. *Mol Cell Biol* [S.I.], v. 29, n. 16, p. 4574-83, Aug 1 2009.

SUZEK, B. E. *et al.* UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* [S.I.], v. 23, n. 10, p. 1282-8, May 15 2007.

SUZUKI, A. *et al.* Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. *Proc Natl Acad Sci USA* [S.I.], v. 103, n. 27, p. 10294-9, Jul 5 2006.

SUZUKI, T. *et al.* Expression of c-MYC in nuclear speckles during mouse oocyte growth and preimplantation development. *J Reprod Dev* [S.I.], v. 55, n. 5, p. 491-5, Oct 2009.

TAM, W. L. *et al.* T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells* [S.I.], v. 26, n. 8, p. 2019-31, Aug 1 2008.

TANAKA, T. S. *et al.* Esg1, expressed exclusively in preimplantation embryos, germline, and embryonic stem cells, is a putative RNA-binding protein with broad RNA targets. *Dev Growth Differ* [S.I.], v. 48, n. 6, p. 381-90, Aug 1 2006.

TARASOV, K. V. *et al.* B-MYB is essential for normal cell cycle progression and chromosomal stability of embryonic stem cells. *PLoS ONE*. v. 3. n. 62008. p. e2478.

TATUSOV, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* [S.I.], v. 4, p. 41, Sep 11 2003.

_____. A genomic perspective on protein families. *Science* [S.I.], v. 278, n. 5338, p. 631-7, Oct 24 1997.

TIRODE, F. *et al.* A conditionally expressed third partner stabilizes or prevents the formation of a transcriptional activator in a three-hybrid system. *J Biol Chem* [S.I.], v. 272, n. 37, p. 22995-9, Sep 12 1997.

TORRES, J.; WATT, F. M. Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3. *Nat Cell Biol* [S.I.], v. 10, n. 2, p. 194-201, Feb 1 2008.

TORRES-PADILLA, M. E. *et al.* The anterior visceral endoderm of the mouse embryo is established from both preimplantation precursor cells and by de novo gene expression after implantation. *Dev Biol* [S.I.], v. 309, n. 1, p. 97-112, Sep 1 2007.

TUCKER, C. L. *et al.* Towards an understanding of complex protein networks. *Trends Cell Biol* [S.I.], v. 11, n. 3, p. 102-6, Mar 2001.

UENO, M. *et al.* Thioredoxin-dependent redox regulation of p53-mediated p21 activation. *J Biol Chem* [S.I.], v. 274, n. 50, p. 35809-15, Dec 10 1999.

- UHLER, J. *et al.* The Nk-2 box of the Drosophila homeodomain protein, Vnd, contributes to its repression activity in a Groucho-dependent manner. *Mech Dev* [S.I.], v. 124, n. 1, p. 1-10, Jan 2007.
- VAN CRIEKINGE, W.; BEYAERT, R. Yeast Two-Hybrid: State of the Art. *Biol Proced Online* [S.I.], v. 2, p. 1-38, Oct 4 1999.
- VAN DEN BERG, D. L. *et al.* Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Mol Cell Biol* [S.I.], v. 28, n. 19, p. 5986-95, Oct 1 2008.
- VAUTI, F. *et al.* Arp3 is required during preimplantation development of the mouse embryo. *FEBS Lett* [S.I.], v. 581, n. 29, p. 5691-7, Dec 11 2007.
- VIDALAIN, P. O. *et al.* Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* [S.I.], v. 32, n. 4, p. 363-70, Apr 2004.
- VOSS, A. K. *et al.* Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development* [S.I.], v. 127, n. 24, p. 5449-61, Dec 1 2000.
- WANG, H.; DEY, S. K. Roadmap to embryo implantation: clues from mouse models. *Nat Rev Genet* [S.I.], v. 7, n. 3, p. 185-99, Mar 2006.
- WEI, S. J. *et al.* Thioredoxin nuclear translocation and interaction with redox factor-1 activates the activator protein-1 transcription factor in response to ionizing radiation. *Cancer Res* [S.I.], v. 60, n. 23, p. 6688-95, Dec 1 2000.
- WU, S. *et al.* The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. *Dev Cell* [S.I.], v. 14, n. 3, p. 388-98, Mar 2008.
- YAMADA, K. *et al.* Genomic organization, chromosomal localization, and the complete 22 kb DNA sequence of the human GCMa/GCM1, a placenta-specific transcription factor gene. *Biochem Biophys Res Commun* [S.I.], v. 278, n. 1, p. 134-9, Nov 11 2000.
- YAMANAKA, Y. *et al.* Cell and molecular regulation of the mouse blastocyst. *Dev Dyn* [S.I.], v. 235, n. 9, p. 2301-14, Sep 2006.
- YANG, D. H. *et al.* Disabled-2 is essential for endodermal cell positioning and structure formation during mouse embryogenesis. *Dev Biol* [S.I.], v. 251, n. 1, p. 27-44, Nov 1 2002.
- YU, J. N. *et al.* 5-AZA-2'-deoxycytidine (5-AZA-CdR) leads to down-regulation of Dnmt1o and gene expression in preimplantation mouse embryos. *Zygote* [S.I.], v. 17, n. 2, p. 137-45, May 1 2009.
- ZAFFRAN, S. *et al.* The NK-2 homeobox gene scarecrow (scro) is expressed in pharynx, ventral nerve cord and brain of Drosophila embryos. *Mech Dev* [S.I.], v. 94, n. 1-2, p. 237-41, Jun 2000.
- ZHANG, J. *et al.* Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* [S.I.], v. 8, n. 10, p. 1114-23, Oct 1 2006.
- ZHANG, X. *et al.* Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* [S.I.], v. 283, n. 51, p. 35825-33, Dec 19 2008.

LISTA DE ANEXOS

ANEXO A – Meios de cultura para leveduras

ANEXO B – Meios de cultura para bactérias

ANEXO C – Tabela S1

ANEXO A - Meios de cultura para leveduras

Mix de aminoácidos

Adenina (1,8 g), L-Arginina (1,2 g), L-Histidina (1,2 g), L-Isoleucina (1,8 g), L-Leucina (1,8 g), L-Lisina (1,8 g), L-Metionina (1,2 g), L-Fenilalanina (3 g), L-Serina (1,2 g), L-Triptofano (1,2 g), L-Tirosina (11,8 g), L-Treonina (1,2 g), Uracila (1,2 g), L-Valina (9 g), Mio-Inositol (2 g), PABA (0,2 g). Todos SIGMA. A mistura era utilizada na formulação do meio SD seletivo.

Meio SD seletivo (SDs) com aminoácidos (pH 5,9)

YNB (BD) sem aminoácidos e sulfato de amônio 0,17%, dextrose 2% (Cromoline), sulfato de amônio 0,5% (VETEC) e mix de aminoácidos 0,06% (SIGMA). Esterilizado em autoclave (soc. fabbe ltda.) a uma pressão de 120kgf/cm² durante 15 minutos.

Meio YPAD

Extrato de levedura 1% (BD), peptona 2% (Fisher Scientific), dextrose 2% (Cromoline) e Adenina 0,004% (SIGMA). Esterilizado em autoclave (soc. fabbe ltda.) a uma pressão de 120kgf/cm² durante 15 minutos.

Aos meios SDs e YPAD sólidos, era adicionado ágar a 1,6% (BD).

ANEXO B - Meios de cultura para bactérias

Meio LB

Extrato de levedura 0,5% (BBL), NaCl 0,5% (VETEC), pH 7. Em meio LB sólido, foi adicionado ágar a 1,5% (BD). Esterilizado em autoclave (soc. fabbe ltda.) a uma pressão de 120kgf/cm² durante 15 minutos.

Meio 2xYT

Bactopeptona 1% (BD), extrato de levedura 0,5% (BBL), NaCl 1% (VETEC), pH 7. Esterilizado em autoclave (soc. fabbe ltda.) a uma pressão de 120kgf/cm² durante 15 minutos.

Meio SOB

Peptona 2% (INLAB), extrato de levedura 0,5% (BBL), NaCl 10 mM (VETEC), KCl 2,5 mM (SYNTH), MgCl₂ 10 mM (Grupo Química), MgSO₄ 10 mM (SYNTH). Esterilizado em autoclave (soc. fabbe ltda.) a uma pressão de 120 kgf/cm² durante 15 minutos.

Meio SOC

Em 100 ml de meio SOB, foi adicionado 2 ml de solução de glicose 2 M (VETEC) e 1 ml de solução de MgCl₂ 2M (Grupo Química), esterilizados por filtração a vácuo.

ANEXO C

Tabela S1: Identificação do grupo taxonômico que contém o último ancestral comum para cada gene da via de pré implantação.

GENE	UniProt ID	LCA	Profundidade
AFP	P02771	Eutheria	20
aPKC	P17252	Fungi/Metazoa	3
ARP3	P61158	Fungi/Metazoa	3
ATP6AP2	O75787	Eumetazoa	5
BMP4	P12644	Metazoa	4
BORIS	A2APF3	Coelomata	7
CDH1	P09803	Euteleostomi	14
CDX2	Q99626	Bilateria	6
CER1	O55233	Euteleostomi	14
CGA	P01216	Gnathostomata	12
CGB	P01233	Euteleostomi	14
CtBP	Q99708	Euteleostomi	14
CTNNB1	Q02248	Eumetazoa	5
DAB2	P98082	Metazoa	4
DAX1	Q61066	Chordata	9
DDB1	Q3U1J4	Fungi/Metazoa	3
DNMT1	P26358	Fungi/Metazoa	3
DPP3	Q9NY33	Fungi/Metazoa	3
DPPA4	Q7L190	Euarthontoglires	21
DSC2	P55292	Tetrapoda	16
ECSA	Q7Z7J5	Euarthontoglires	21
EOMES	O54839	Euteleostomi	14
ESG1	A6NC42	Euarthontoglires	21
ESRRB	Q61539	Coelomata	7
ETIF2B	Q14232	Fungi/Metazoa	3
FGF3	P11487	Eumetazoa	5
FGF4	P08620	Eumetazoa	5
FGFR2	P21802	Metazoa	4
FOXD3	Q9UJU5	Bilateria	6
GATA2	P23769	Fungi/Metazoa	3
GATA3	P23771	Fungi/Metazoa	3
GATA4	Q08369	Fungi/Metazoa	3
GATA6	Q92908	Fungi/Metazoa	3
GCM1	P70348	Coelomata	7
GJA3	Q9Y6H8	Euteleostomi	14
GJB3	O75712	Euteleostomi	14
GJC1	P28229	Euteleostomi	14
GK003	Q9NRH1	Euarthontoglires	21
GSE	Q9DAG5	Metazoa	4
HAND1	O96004	Eumetazoa	5
HNF4a	P49698	Metazoa	4
ID2	P41136	Eumetazoa	5
JAK2	Q62120	Coelomata	7

GENE	UniProt ID	LCA	Profundidade
KLF2	Q60843	Chordata	9
KLF4	Q60793	Coelomata	7
KLF5	Q9Z0Z7	Eumetazoa	5
KRT18	P05784	Chordata	9
LAMB3	Q13751	Eutheria	20
LATS2	Q9NRM7	Coelomata	7
LEFTY1/LEFTY2	O75610	Deuterostomia	8
LIF	P15018	Eutheria	20
LIFR	P42702	Euteleostomi	14
MYBL2	P48972	Euteleostomi	14
MYC	P01106	Eumetazoa	5
NANOG	Q5TM83	Metazoa	4
NFKB1	P19838	Coelomata	7
NLE1	Q9NVX2	Fungi/Metazoa	3
OCT4	P20263	Euteleostomi	14
PARD3	Q8TEW0	Eumetazoa	5
PECAM1	P16284	Euteleostomi	14
PKA	P05132	Fungi/Metazoa	3
PSG2	P11465	Homo sapiens	30
RENT1	Q92900	Fungi/Metazoa	3
REX1	P22227	Euarchontoglires	21
RPS14	P62263	Fungi/Metazoa	3
RUNX1	Q01196	Metazoa	4
SALL4	Q9UJQ4	Bilateria	6
SAP18	O55128	Fungi/Metazoa	3
SMAD1	P70340	Metazoa	4
SNAI1	O95863	Eumetazoa	5
SOX17	Q61473	Eumetazoa	5
SOX2	P48431	Euarchontoglires	21
SOX7	Q9BT81	Eumetazoa	5
STAT3	P40763	Eumetazoa	5
TBN	Q7Z7C8	Metazoa	4
TCF3	P15923	Bilateria	6
TDGF1	P13385	Euteleostomi	14
TEAD4	Q15561	Fungi/Metazoa	3
THY1	P04216	Euteleostomi	14
TLE2	Q04725	Metazoa	4
WNT6	P22727	Coelomata	7
WNT9A	O14904	Coelomata	7
YAP1	P46938	Eumetazoa	5