

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-529617

(P2004-529617A)

(43) 公表日 平成16年9月30日(2004.9.30)

(51) Int. Cl. ⁷	F I	テーマコード (参考)	
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A	2 G O 4 5
C O 7 K 14/47	C O 7 K 14/47		4 B O 2 4
C O 7 K 16/18	C O 7 K 16/18		4 B O 6 3
C 1 2 N 5/06	C 1 2 Q 1/68	A	4 B O 6 5
C 1 2 Q 1/68	G O 1 N 33/48	M	4 H O 4 5
	審査請求 未請求 予備審査請求 有	(全 131 頁)	最終頁に続く

(21) 出願番号	特願2002-558380 (P2002-558380)	(71) 出願人	500514856 ケンブリッジ ユニバーシティ テクニカル サービスズ リミティド イギリス国, ケンブリッジ シービー2 1 ティーエス, トリニティー レーン, ジ オールド スクールズ
(86) (22) 出願日	平成14年1月18日 (2002.1.18)	(74) 代理人	100064908 弁理士 志賀 正武
(85) 翻訳文提出日	平成15年7月18日 (2003.7.18)	(74) 代理人	100108578 弁理士 高橋 詔男
(86) 国際出願番号	PCT/GB2002/000215	(74) 代理人	100089037 弁理士 渡邊 隆
(87) 国際公開番号	W02002/057307	(74) 代理人	100101465 弁理士 青山 正和
(87) 国際公開日	平成14年7月25日 (2002.7.25)		
(31) 優先権主張番号	0101300.2		
(32) 優先日	平成13年1月18日 (2001.1.18)		
(33) 優先権主張国	英国 (GB)		

最終頁に続く

(54) 【発明の名称】 遺伝子

(57) 【要約】

本発明は、生殖細胞で特異的に発現される、始原生殖細胞のマーカーであり、細胞集団内でこのような細胞を同定するのに使用することができる2つの遺伝子GCR1(Fragilis)およびGCR2(Stella)を提供する。本発明は、GCR1ポリペプチド、あるいはその断片、相同体、変異体または誘導体であって、特に配列番号2に示す配列と少なくとも50%、60%、70%、80%、90%または95%の相同性を有する請求項1に記載のポリペプチドを提供する。またGCR2ポリペプチド、あるいはその断片、相同体、変異体または誘導体であって、特に、配列番号4に示す配列と少なくとも50%、60%、70%、80%、90%または95%の相同性を有する請求項3に記載のポリペプチドを提供する。またこれらのポリペプチドをコードする核酸も提供する。

。

【特許請求の範囲】

【請求項 1】

GCR1ポリペプチド、あるいはその断片、相同体、変異体または誘導体。

【請求項 2】

配列番号2に示す配列と少なくとも50%、60%、70%、80%、90%または95%の相同性を有する請求項1に記載のポリペプチド。

【請求項 3】

GCR2ポリペプチド、あるいはその断片、相同体、変異体または誘導体。

【請求項 4】

配列番号4に示す配列と少なくとも50%、60%、70%、80%、90%または95%の相同性を有する請求項3に記載のポリペプチド。 10

【請求項 5】

請求項 1 から 4 のいずれかに記載のポリペプチドをコードする核酸。

【請求項 6】

配列番号1に記載の配列と少なくとも90%の相同性を有する核酸、あるいはその断片、変異体または誘導体。

【請求項 7】

配列番号3、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9に示す配列と少なくとも75%の相同性を有する核酸、あるいはその断片、変異体、または誘導体。

【請求項 8】

請求項5、6または7に記載の核酸の25個の連続ヌクレオチド配列を含む核酸。 20

【請求項 9】

請求項5から8のいずれかに記載の核酸の15個の連続ヌクレオチド配列を含む核酸。

【請求項 10】

請求項5から9のいずれかに記載の核酸配列の相補体。

【請求項 11】

1つまたは複数のヌクレオチドの置換を含むが、遺伝コードの縮重のためにこのような置換によって前記核酸のコード特異性が改変されない、請求項5から10のいずれかに記載の核酸。

【請求項 12】

請求項 1 から 11 のいずれかに記載の核酸によってコードされるポリペプチド。 30

【請求項 13】

ポリペプチドが配列番号2または配列番号4に示す配列を含む請求項12に記載のポリペプチド。

【請求項 14】

請求項1から4、12、または13のいずれかに記載のポリペプチドまたはその相同体の存在を検出すること、あるいは請求項5から11のいずれかに記載の核酸またはその相同体の発現を検出することを含む、多能性細胞を同定する方法。

【請求項 15】

GCR1および/またはGCR2に特異的な5'および3'プライマーを使用して推定上の多能性細胞から核酸を増幅するステップと、その結果生成された増幅した核酸を検出するステップとを含む、請求項14に記載の方法。 40

【請求項 16】

核酸配列の発現を *in situ*ハイブリダイゼーションによって検出する請求項14に記載の方法。

【請求項 17】

核酸配列の発現が、それにコードされるタンパク質産物を検出することによって決定される請求項8に記載の方法。

【請求項 18】

タンパク質産物を免疫染色によって検出する請求項14または請求項18に記載の方法。 50

【請求項 19】

請求項1から4、12、13のいずれかに記載のポリペプチドに特異的な抗体。

【請求項 20】

GCR1の細胞外ドメインに特異的に結合することができる、請求項19に記載の抗体。

【請求項 21】

多能性細胞の同定および/または単離のための請求項19または請求項20に記載の抗体の使用。

【請求項 22】

請求項14から18、21のいずれかに記載の方法によって同定される多能性細胞。

【請求項 23】

(a)多能性細胞を含む細胞群を提供するステップと、
(b)そこから1つまたは複数の多能性細胞を単離して、単一細胞多能性細胞の単離体を提供するステップと、

(c)単一多能性細胞内に存在する転写された核酸を増幅するステップと、

(d)多能性細胞中に存在するが体細胞中には存在しない転写物を同定するためにサブトラクティブハイブリダイゼーションスクリーンを実施するステップと、

(e)ステップ(d)で同定された1つまたは複数の転写物を用いて核酸ライブラリをプローブし、多能性細胞中で特異的に発現される1つまたは複数の遺伝子をクローニングするステップと

を含む、多能性細胞中で特異的に発現される遺伝子を単離する方法。

【請求項 24】

多能性細胞が、始原生殖細胞(PGC)、胚性幹細胞(ES)、胚性生殖細胞(EG)からなる群から選択される、請求項14から18、23のいずれかに記載の方法、請求項21に記載の使用、請求項23に記載の多能性細胞。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、発生学、分子生物学、および遺伝学の分野に関する。より詳細には、本発明は、始原生殖細胞(PGC)の初期集団内でのみで発現される遺伝子、ならびにこのような遺伝子およびその産物の、PGC、多能性胚性幹細胞(ES)、多能性胚性生殖細胞(EG)など細胞集団内の多能性細胞(pluripotent cell)および多分化能性細胞(multipotent cell)同定における使用に関する。これらはまた、細胞状態が多能性でない状態から多能性である状態になる、変化のマーカー、およびこの状態を多能性でない細胞に与えることができるマーカーである。

【背景技術】

【0002】

受精後、初期の哺乳動物胚は4回の卵割を経て16細胞の桑実胚を形成する。これらの細胞は、さらなる分裂の後、細胞を2つの異なる領域、すなわち、胚を形成する内部細胞塊と胎盤など胚外組織を形成する栄養外胚葉とに分けることができる胚盤胞に発達する。

【0003】

胚盤胞の形成までは、胚の一部を形成する細胞は全能性(totipotent)である。すなわち、各細胞が、完全な個別の胚およびその発達に必要なすべての外胚組織を生じさせる能力を有している。胚盤胞の形成後は、内部細胞塊の細胞はもはや全能性ではなく多能性となり、この場合これらは様々な異なる組織を生じさせることができる。このような細胞の既知のマーカーは、酵素アルカリホスファターゼおよびOct4の発現である。

【0004】

始原生殖細胞(PGC)は、3つの一次胚葉すべてに分化する能力を有する多能性細胞である。哺乳動物では、PGCが尿膜基部から後腸上皮および背側腸間膜を通過して遊走し、生殖腺原基のコロニーを形成する。PGC由来の細胞は、通常は顕著な核小体を有する、特徴的に低い細胞質/核比を有している。胚の生殖隆腺を取り除き、PGCを生殖腺原基から解離し、PG

10

20

30

40

50

Cを回収することによって、PGCを胚から単離することができる。初期PGC集団は、受精後7.25日で新生尿膜基部に見つかる約45(四十五)個のアルカリホスファターゼ陽性細胞のクラスターからなると報告されている(Ginsburg他、1990、Development、110:521-528)。

【0005】

PGCには、最新のバイオテクノロジーおよび分子生物学において多くの用途がある。これらは、遺伝子導入動物の生成に有用であり、この場合PGC由来の胚性生殖(EG)細胞を胚性幹(ES)細胞とほぼ同様に使用することができる(Labosky他、1994、Development、120:3197-3204)。これらはさらに、胎児発生の研究、ならびに変性疾患治療での組織再生および外傷後の破傷組織の再増殖における、多能性幹細胞の供給に有用である。中でも、PGCは、いくつかの分化した特質を有しているが、体細胞の運命を獲得するPGCの創始集団を取り囲む近隣細胞からは失われている、潜在的な多能性を保持している。PGCおよび周囲の体細胞は共通の祖先を共有する。しかし、創始PGCは数が少なく、胚組織および周囲の体細胞から単離するのが困難であるので、これらの研究およびそれを利用する技術の開発が難しくなる。

10

【0006】

PGCの創始集団内での遺伝子の発現、およびPGCに特異的な遺伝子発現とこれら細胞の多能性の保持との関連性については、当分野で少ししか知られていない。特定のPGCマーカーが知られており、たとえば、組織に非特異的なアルカリホスファターゼ(TNAP)の発現が初期PGCのマーカーとして使用されている(Ginsburg他、1990、Development、110:521-528)。Oct4は、PGC内では発現されるが体細胞内では発現されないことで知られている(Yoem他、1996、Development、122:881-894)。BMP4など他のマーカーは、主に体組織で発現されることで知られている(Lawson他、1999、Genes & Dev.、13:424-436)。しかし、これらの遺伝子は、他の種類の組織内でも発現されるので、どれもPGCに特異的でない。

20

【非特許文献1】

Ginsburg他、1990、Development、110:521-528

【非特許文献2】

Labosky他、1994、Development、120:3197-3204

【非特許文献3】

Yoem他、1996、Development、122:881-894

【非特許文献4】

Lawson他、1999、Genes & Dev.、13:424-436

30

【非特許文献5】

Saitou他、1998、J Cell Biol、141、397-408

【非特許文献6】

Deblandre、1995

【非特許文献7】

J.Sambrook、E.F.Fritsch、およびT.Manias、1989、「Molecular Cloning:A Laboratory Manual」、第2版、第1-3巻、Cold Spring Harbor Laboratory Press

【非特許文献8】

Ausubel、F.M.他、(1995および定期的補遺;「Current Protocols in Molecular Biology」第9、13、16章、John Wiley & Sons、ニューヨーク州ニューヨーク)

40

【非特許文献9】

B.Roe、J.Crabtree、およびA.Kahn、1996、「DNA Isolation and Sequencing:Essential Techniques」、John Wiley & Sons

【非特許文献10】

J.M.PolakおよびJames O'D.McGee、1990、「In Situ Hybridization:Principles and Practice」、Oxford University Press

【非特許文献11】

M.J.Gait(編集者)、1984、「Oligonucleotide Synthesis:A Practical Approach」、Irl Press

50

- 【非特許文献 1 2】
D.M.J.LilleyおよびJ.E.Dahlberg、1992、「Methods of Enzymology:DNA Structure Part A:Synthesis and Physical Analysis of DNA Methods in Enzymology」Academic Press
- 【非特許文献 1 3】
Devereux他、1984、Nucleic Acids Research、12:387
- 【非特許文献 1 4】
BergerおよびKimmel、1987、「Guide to Molecular Cloning Techniques,Methods in Enzymology」、第152巻、Academic Press、カリフォルニア州サンディエゴ
- 【特許文献 1】
Winter、ヨーロッパ特許出願第0239400号 10
- 【非特許文献 1 5】
KohlerおよびMilstein、1975、Nature、256:495-497
- 【特許文献 2】
米国特許第4,376,110号
- 【非特許文献 1 6】
HarlowおよびLane、「Antibodies:a Laboratory Manual」、1988、Cold Spring Harbor
- 【特許文献 3】
ヨーロッパ特許第0623679号
- 【特許文献 4】
ヨーロッパ特許第0368684号 20
- 【特許文献 5】
ヨーロッパ特許第0436597号
- 【特許文献 6】
国際公開公報W097/28261号
- 【非特許文献 1 7】
Rabbany他、1994、Crit Rev Biomed Eng、22:307-346
- 【非特許文献 1 8】
Morgan他、1996、Clin Chem、42:193-209
- 【非特許文献 1 9】
Robinson、1991、Biosens Bioelectron、6:183-191 30
- 【非特許文献 2 0】
Ghindilis他、1998、Biosens Bioelectron、1:113-131
- 【非特許文献 2 1】
Attridge他、1991、Biosens Bioelectron、6:201-214
- 【非特許文献 2 2】
BradyおよびIscove、1993
- 【非特許文献 2 3】
Dulac,CおよびAxel、1995
- 【非特許文献 2 4】
Ginsburg,M.、Snow,M.H.L.、およびMcLaren,A.、1990 40
- 【非特許文献 2 5】
Lawson,K.A.、Dunn,N.R.、Roelen,B.A.J.、Zeinstra,L.M.、Davis,A.M.、Wright,C.V.E.、Korving,J.P.W.F.M.、およびHogan,B.L.M.、1999
- 【非特許文献 2 6】
Yoem,Y.II.、Fuhrmann,G.、Ovitt,C.E.、Brehm,A.、Ohbo,K.、Gross,M.、Hubner,K.、およびScholer,H.R.、1996
- 【非特許文献 2 7】
Lawson他、1999
- 【非特許文献 2 8】
Nichols、199、Pesce、1998 50

【非特許文献 29】

Yeom, 1996

【発明の開示】

【発明が解決しようとする課題】

【0007】

したがって、PGCのマーカーとして使用でき、生殖細胞発生の生物学および多能性状態の性質の見識を提供することができる遺伝子の同定が、当分野で必要とされている。

【課題を解決するための手段】

【0008】

本発明者らは、PGCおよび他の多能性細胞内で特異的に発現される2つの遺伝子の配列を開示する。マウス由来のこの遺伝子の配列を、配列番号1(GCR1すなわちFragilis)および配列番号3(GCR2すなわちStella)に示す。マウスのGCR1およびGCR2に対応するアミノ酸配列を、それぞれ配列番号2および配列番号4に示す。ラットGCR2相同体の核酸配列を、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9に示す。

【0009】

本発明の第1の態様によると、本発明者らは、GCR1ポリペプチドまたはその断片、相同体、変異体、誘導体を提供する。好ましくは、このポリペプチドは、配列番号2に示す配列と少なくとも50%、60%、70%、80%、90%または95%の相同性を有する。

【0010】

本発明の第2の態様によると、GCR2ポリペプチドまたはその断片、相同体、変異体、誘導体を提供する。好ましくは、このポリペプチドは、配列番号4に示す配列と少なくとも50%、60%、70%、80%、90%または95%の相同性を有する。

【0011】

本発明の第3の態様によると、本発明者らは、前記請求項のいずれか一項に記載のポリペプチドをコードする核酸を提供する。

【0012】

本発明の第4の態様として、配列番号1に示す配列と少なくとも90%の相同性を有する核酸、またはその断片、変異体、誘導体を提供する。

【0013】

本発明の第5の態様によると、本発明者らは、配列番号3、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9に示す配列と少なくとも75%の相同性を有する核酸、またはその断片、変異体、誘導体を提供する。

【0014】

第6の態様では、本発明は、本発明の第3、第4、または第5の態様に記載の、25個の連続したヌクレオチドの核酸配列を含む核酸を提供する。

【0015】

本発明の第7の態様では、本発明の第3、第4、第5または第6の態様に記載の、15個の連続したヌクレオチドの核酸配列を含む核酸を提供する。

【0016】

本発明第8の態様によると、本発明者らは、本発明の第3から第7の態様のいずれかに記載の核酸配列の相補体(complement)を提供する。

【0017】

好ましくは、このような核酸は、遺伝暗号の縮重のためヌクレオチドの置換で前記核酸のコード特異性が改変されない、1つまたは複数のヌクレオチドの置換を含む。

【0018】

本発明の第9の態様によると、本発明者らは、本発明の前記態様のいずれかに記載の核酸にコードされるポリペプチドを提供する。

【0019】

好ましくは、このポリペプチドは配列番号2または配列番号4に示す配列を含む。

【0020】

本発明の第10の態様によると、本発明の第1、第2、第9または第10の態様に記載のポリペプチドの存在を検出すること、あるいは本発明の第3から第8の態様のいずれかに記載の核酸またはその相同体の発現を検出することを含む、多能性細胞を同定する方法を提供する。

【0021】

好ましくは、この方法は、GCR1(Fragilis)および/またはGCR2(Stella)に特異的な5'および3'プライマーを用いて推定上の多能性細胞由来の核酸を増幅するステップと、増幅して生成した核酸を検出するステップとを含む。好ましくは、核酸配列の発現はin situハイブリダイゼーションによって検出される。

【0022】

核酸配列の発現は、それにコードされるタンパク質産物を検出することで判定することができる。代わりに、またはそれにくわえて、免疫染色によってタンパク質産物を検出することができる。

【0023】

本発明の第11の態様として、本発明者らは、本発明の第1、第2、第9、または第10の態様に記載のポリペプチドに特異的な抗体を提供する。好ましくは、この抗体はGCR1の細胞外ドメインに特異的に結合することができる。

【0024】

本発明の第12の態様によると、本発明者らは、多能性細胞を同定および/または単離するためのこのような抗体の使用を提供する。

【0025】

本発明の第13の態様によると、本発明者らはさらに、前述の方法によって同定した多能性細胞を提供する。

【0026】

本発明の第14の態様によると、(a)多能性細胞を含む細胞集団を提供するステップと、(b)そこから1つまたは複数の多能性細胞を単離して単一細胞の多能性細胞単離体を提供するステップと、(c)単一の多能性細胞内に存在する転写された核酸を増幅するステップと、(d)多能性細胞内に存在するが体細胞内には存在しない転写物を同定するためにサブトラクティブハイブリダイゼーションスクリーンを実施するステップと、(e)ステップ(d)で同定した1つまたは複数の転写物を用いて核酸ライブラリをプローブし、多能性細胞内で特異的に発現される1つまたは複数の遺伝子をクローニングするステップとを含む、多能性細胞内で特異的に発現される遺伝子を単離する方法を提供する。

【0027】

非常に好ましい実施形態では、多能性細胞は、始原生殖細胞(PGC)、胚性幹細胞(ES)、胚性生殖細胞(EG)からなる群から選択される。好ましくは、多能性細胞は始原生殖細胞を含む。

【発明を実施するための最良の形態】

【0028】

GCR1(FRAGILIS)およびGCR2(STELLA)

この開示は、一般的に、GCR1(Fragilis)およびGCR2(Stella)の核酸、ポリペプチドのみならず、その断片、相同体、変異体、誘導体を提供する。

【0029】

「GCR1」と「Fragilis」という名称は互いに同義であり、「GCR2」と「Stella」も同様に同義と見なされるべきである。GCR1/Fragilisの核酸配列およびアミノ酸配列を配列番号1および2に示す一方で、GCR2/Stellaの核酸配列を配列番号3、5、6、7、8および9に示し、GCR2/Stellaのアミノ酸配列を配列番号4に示す。

【0030】

しかし、好ましい実施形態では、GCR1/Fragilisは文脈に応じて配列番号1に示す核酸配列または配列番号2に示すアミノ酸配列を参照するとされるべきである。さらに、好ましい実施形態では、GCR2/Stellaは文脈に応じて配列番号3に示す核酸配列または配列番号4に

10

20

30

40

50

示すアミノ酸配列を参照するとされるべきである。

【0031】

GCR1およびGCR2はPGCに特異的な転写物である。GCR1はPGCの細胞系譜の決定プロセス中に上方制御され、GCR2はGCR1の後に上方制御されてPGCの運命に決定される。第1の遺伝子GCR1(生殖細胞制限-1、Fragilis)は、予測分子量15.0kDの137アミノ酸のタンパク質をコードする。EMBLプログラムPredictProteinで一番合うモデルでは、N末端とC末端のいずれもが外に位置する2つの膜貫通型ドメインが予測される。BLASTP検索により、Fragilisがインターフェロン誘導可能タンパク質ファミリーの新規なメンバーであることが分かった。典型的なメンバーの1つであるヒト9-27(Leu-13抗原と同一)は、白血球および内皮細胞内でインターフェロンによって誘導可能であり、抗増殖性およびホモタイプ付着シグナルの伝達に關与する多量体の複合体の構成成分として細胞表面に位置する(Deblandre、1995)。BLASTN検索により、胚と成体のいずれもからの様々な組織由来のEST内でFragilis配列が見つかることが分かり、これは、様々な発生学および細胞生物学の状況においてFragilisが共通の役割を果たすことを示唆する。データベースの検索により、未知機能のラットインターフェロン誘導可能タンパク質(sp:INIB RAT、pir:JC1241)との配列一致が明らかになった。本発明者らのスクリーン上にGCR1配列が6回現れ、これはPGC内の高発現レベルを示唆する。

10

【0032】

第2の遺伝子GCR2(Stella)は、18kDの150アミノ酸のタンパク質をコードする。これは、既知のどのタンパク質とも配列相同性を有さず、いくつかの核移行共通配列を含み、強度に塩基性のpI(pI=9.67、塩基性残基の含有率=23.3%)であり、潜在的なDNA親和性が示唆される。さらに、潜在的な核外シグナルが同定され、Stellaが核と細胞質の間を往復する可能性が示唆された。BLASTN分析により、Stella配列は移植前の胚および生殖系(新生卵巣、メス12.5中腎や生殖腺など)ESTのみで見つかり、全能性細胞および多能性細胞内で主に発現されることが示唆される。興味深いことに、StellaはそのN末端内に、SAPモチーフとある程度の配列類似性を有する変調ドメインを含む。このモチーフは、染色体の組織化に關与する推定DNA結合ドメインである。さらに、SMARTプログラムにより、スプライシング因子のモチーフ様構造がC末端内に存在することがわかった。これらの発見は、Stellaが染色体の組織化およびRNAプロセッシングに關与している可能性を示唆する。

20

【0033】

GCR1および/またはGCR2ポリペプチドに対して抗体を産生させることができる。具体的には、膜貫通型ポリペプチドであるGCR1の細胞外ドメインに対する抗体を産生させることができる。

30

【0034】

本明細書に開示する抗体および核酸は、細胞集団内でPGCを同定するのに有用である。したがって、本明細書に記載する方法および組成物は、たとえば生殖組織の発生の研究および遺伝子導入動物の作成に有用なPGCを単離する手段、および本明細書に記載の方法によって単離したPGCを提供する。

【0035】

また、GCR1およびGCR2の相同体を使用して、PGCおよびESやEG細胞など他の多能性細胞を同定することができる。

40

【0036】

別段に指示しない限りは、本発明の実施には、当分野の技術者の技量範囲内にある化学、分子生物学、微生物学、組換えDNA学および免疫学の従来技術を利用する。このような技術は文献に記載されている。たとえば、J.Sambrook、E.F.Fritsch、およびT.Maniatis、1989、「Molecular Cloning:A Laboratory Manual」、第2版、第1-3巻、Cold Spring Harbor Laboratory Press;Ausubel、F.M.他、(1995および定期的補遺;「Current Protocols in Molecular Biology」第9、13、16章、John Wiley & Sons、ニューヨーク州ニューヨーク);B.Roe、J.Crabtree、およびA.Kahn、1996、「DNA Isolation and Sequencing:Essential Techniques」、John Wiley & Sons;J.M.PolakおよびJames O'D.McGee、1990、「In Sit

50

u Hybridization:Principles and Practice」、Oxford University Press;M.J.Gait(編集者)、1984、「Oligonucleotide Synthesis:A Practical Approach」、Irl Press;D.M.J.L illeyおよびJ.E.Dahlberg、1992、「Methods of Enzymology:DNA Structure Part A:Synthesis and Physical Analysis of DNA Methods in Enzymology」Academic Press参照。これら一般教科書のそれぞれを、参照により本明細書に組み込む。

【0037】

ポリペプチド

本明細書に開示するポリペプチド配列は、配列番号2および配列番号4に示す具体的な配列またはその断片、あるいはGCR1やGCR2タンパク質から得た配列だけには限定されず、任意の源たとえば関連する細胞性相同体、他の種由来の相同体、それらの変異体または誘導体から得た相同配列も含むことも理解されよう。

10

【0038】

したがって、本開示は、配列番号2および配列番号4に示したアミノ酸配列の変異体、相同体、誘導体のみならず、本明細書に開示するヌクレオチド配列にコードされるアミノ酸配列の変異体、相同体、誘導体も包含する。

【0039】

相同体

開示するポリペプチドには、任意の源、たとえば関連するウイルス/細菌のタンパク質、細胞の相同体、合成ペプチドから得た相同配列のみならず、その変異体や誘導体が含まれる。したがって、ポリペプチドには、哺乳動物(たとえばマウス、ラット、ウサギ)、特に霊長類、とりわけヒトなどの動物を含む他の種由来のGCR1および/またはGCR2の相同体をコードするものも含まれる。より詳細には、相同体にはヒトの相同体が含まれる。

20

【0040】

本文書の文脈では、相同配列または相同体とは、少なくとも30個、好ましくは50、70、90、または100個のアミノ酸にわたるアミノ酸レベルで、GCR1またはGCR2と少なくとも60、70、80、または90%同一、好ましくは少なくとも95または98%同一であるアミノ酸配列、たとえば本明細書中の配列表に示すものを含めることとする。本文書の文脈では、相同配列は、好ましくは少なくとも50または100個、好ましくは200、300、400または500個のアミノ酸にわたるアミノ酸レベルで、GCR1またはGCR2の配列と少なくとも15、20、25、30、40、50、60、70、80または90%同一、好ましくは少なくとも95または98%同一であるアミノ酸配列、たとえばGCR1(配列番号2)およびGCR2(配列番号4)を含めることとする。相同性は類似度の点から考慮することもできるが(すなわち類似の化学特性/機能を有するアミノ酸残基)、本文書の文脈では、配列の同一性の点から相同性を表すことが好ましい。

30

【0041】

相同性の比較は視認で行うことができ、またはより一般的には容易に入手可能な配列比較プログラムを用いて行うことができる。これら市販のコンピュータプログラムで、2つ以上の配列間の%相同性を計算することができる。

【0042】

%相同性は、連続した配列にわたって計算することができる。すなわち一方の配列が他方の配列とアラインしており、一方の配列内の個々のアミノ酸を他方の配列の対応するアミノ酸と、1残基ずつ直接比較する。これは、「ギャップのない(ungapped)」アラインメントと呼ばれる。通常、このようなギャップのないアラインメントは比較的少数の残基でのみ実施される(たとえば50個未満の連続したアミノ酸)。

40

【0043】

この方法は非常に単純で一貫した方法であるが、たとえば、普通なら同一である配列の対で、1つの挿入または欠失がその後のアミノ酸残基にアラインメントのずれを生じさせ、その結果、全体のアラインメントを行った場合に%相同性が大きく減少する可能性があることを考慮に入れていない。したがって、ほとんどの配列比較方法は、全体の相同性スコアに過度のペナルティを与えることなしに挿入または欠失の可能性を考慮に入れる最適なアラインメントを生成するよう設計されている。これは、配列アラインメントに「ギャッ

50

ブ」を挿入して局部相同性を最大にしようとすることによってなされる。

【0044】

しかし、より複雑なこれらの方法では、同数の同一アミノ酸では、できるだけ少ないギャップの配列アラインメント(これは比較している2つの配列間の関連性がより高いことを反映する)がギャップの多いものより高いスコアを得るように、アラインメント内で起こる各ギャップに「ギャップペナルティ」を課す。「擬似ギャップコスト」は通常、ギャップの存在に比較的高いコストを課し、ギャップ内に続く各残基により小さなペナルティを課すように用いられる。これは、最も一般的に使用されているギャップスコア決定システムである。もちろん、高いギャップペナルティは、より少ないギャップの最適化されたアラインメントを生成する。ほとんどのアラインメントプログラムは、ギャップペナルティを変更することができる。しかし、このようなソフトウェアを配列比較に使用する場合は、デフォルト値を使用することが好ましい。たとえば、GCG Wisconsin Bestfitパッケージ(下記参照)を使用する場合は、アミノ酸配列のデフォルトギャップペナルティは、ギャップは-12、各伸長は-4である。

10

【0045】

したがって、最大%相同性の計算には、まず、ギャップペナルティを考慮した最適アラインメントの生成が必要とされる。このようなアラインメントを実施するのに適切なコンピュータプログラムは、GCG Wisconsin Bestfitパッケージ(米国ウィスコンシン大学;Devereux他、1984、Nucleic Acids Research、12:387)である。配列比較を実施することができる他のソフトウェアの例には、それだけには限定されないが、BLASTパッケージ(Ausubel他、1999、ibid、第18章参照)、FASTA(Atschul他、1990、J.Mol.Biol.、403-410)およびGENEWORKSの比較ツール一式が含まれる。BLASTとFASTAのいずれもで、オフラインおよびオンライン検索が利用可能である(Ausubel他、1999、ibid、ページ7-58、7-60)。しかし、GCG Bestfitプログラムを使用することが好ましい。

20

【0046】

最終%相同性を同一性の点から測定することができるが、アラインメントプロセス自体は通常、絶対的な対比較に基づいていない。代わりに、各対の比較に化学的類似度または進化的距離に基づいたスコアを割り当てるスケールド(scaled)類似性スコア行列が一般的に使用される。一般的に使用されるこのような行列の例は、BLASTプログラム一式のデフォルト行列であるBLOSUM62行列である。GCG Wisconsinプログラムは概して、一般的なデフォルト値、または供給されていればカスタムシンボル比較表のいずれかを使用する(詳細な説明には取扱説明書参照)。GCGパッケージには一般的なデフォルト値、他のソフトウェアの場合にはBLOSUM62などデフォルト行列を使用することが好ましい。

30

【0047】

ソフトウェアが最適アラインメントを生成した後は、%相同性、好ましくは%配列同一性を計算することができる。このソフトウェアは通常これを配列比較の一環として行って、数値の結果を作成する。

【0048】

変異体および誘導体

本明細書に記載する、アミノ酸配列に関連する用語「変異体」または「誘導体」には、置換したもの、変異させたもの、改変したもの、交換したもの、1個(または複数の)アミノ酸を配列から欠失させたまたは配列に追加したもののすべてが含まれる。好ましくは、生じたアミノ酸配列は、改変していない配列と実質的に同じ活性を保持し、好ましくは少なくとも、配列表に示すGCR1および/またはGCR2ポリペプチドと同じ活性を有する。したがって、好ましくは、この配列の重要な特徴、すなわちこれらがPGCおよびES細胞またはEG細胞など他の多能性細胞に特異的であり、細胞集団内でこれらの細胞のマーカーとして働くことは、保持されている。

40

【0049】

本明細書に記載の方法および組成物で使用するために、実施例中に示すアミノ酸配列を有するポリペプチドあるいはその断片または相同体を改変することができる。通常、配列の

50

生理活性を維持する改変が行われる。改変した配列が改変していない配列の生理活性を維持するという条件として、アミノ酸の置換、たとえば1、2または3個から10、20または30個の置換を行うことができる。アミノ酸置換には、たとえば治療で投与したポリペプチドの血漿半減期を増加させるための、自然に存在しない類似体の使用が含まれる。

【0050】

GCR1およびGCR2の自然変異体には、おそらく保存的なアミノ酸置換体が含まれる。保存的置換は、たとえば下記の表に従って定義することができる。第2段で同一ブロック内にあり、好ましくは第3段で同一行内にあるアミノ酸は、互いに置換することができる。

【0051】

【表1】

脂肪族	非極性	GAP
		ILV
	極性-非荷電	CSTM
		NQ
	極性-荷電	DE
		KR
芳香族		HFYW

10

20

【0052】

断片

本明細書に開示し、マーカーとして有用であるポリペプチドには、配列番号2および配列番号4に示した配列の断片を含めた上述の完全長ポリペプチドおよびその変異体の断片が含まれる。

【0053】

ポリペプチドには、任意のGCR1および/またはGCR2ポリペプチドの完全長配列の断片も含まれる。好ましくは、断片は少なくとも1つのエピトープを含む。エピトープを同定する方法は当分野で周知である。断片は通常少なくとも6個のアミノ酸、より好ましくは少なくとも10、20、30、50、または100個のアミノ酸を含む。

30

【0054】

GCR1および/またはGCR2のアミノ酸配列からの5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145または150個以上の残基を含む、好ましくはそれからなる断片が含まれる。

40

【0055】

GCRタンパク質のポリペプチド断片ならびにその対立形質および種変異体 (species variant) は、保存的置換を含めて1つまたは複数(たとえば5、10、15、または20個)の置換、欠失、挿入を含む可能性がある。置換、欠失、および/または挿入が起こる場合、たとえば異なる種の場合は、好ましくは配列表に示すアミノ酸残基の50%、40%、または20%未満のアミノ酸残基が改変される。

【0056】

GCR1および/GCR2、ならびにその断片、相同体、変異体、誘導体を、組換え手段によって作成することができる。また一方、これらを固相合成など当業者に周知の技術を用いた合成手段によって作成することもできる。このタンパク質を、たとえば抽出および精製を補

50

助するために、融合タンパク質として生成することもできる。融合タンパク質のパートナーの例には、グルタチオン-S-トランスフェラーゼ(GST)、6×His、GAL4(DNA結合および/または転写活性化ドメイン)、 β -ガラクトシダーゼが含まれる。また、融合タンパク質配列の除去を可能にするために、融合タンパク質のパートナーと目的のタンパク質配列との間にタンパク質分解性の切断部位を含めると便利な場合がある。好ましくは、この融合タンパク質は目的のタンパク質配列の機能を妨げない。動物細胞の細胞抽出物を精製してタンパク質を得ることもできる。

【0057】

本明細書に開示するGCR1および/またはGCR2ポリペプチド、変異体、相同体、断片、および誘導体は、実質的に単離された形にできる。このようなポリペプチドを、タンパク質の意図する目的を妨げないキャリアまたは希釈剤と混合しても、実質的に単離されているとみなされることを理解されたい。GCR1/GCR2の変異体、相同体、断片、誘導体も実質的に精製した形であることができ、この場合、一般的に調製物内にこのタンパク質を含み、調製物内のタンパク質の90%超、たとえば95%、98%、または99%がタンパク質である。

10

【0058】

本明細書に開示するGCR1/GCR2ポリペプチド、変異体、相同体、断片、および誘導体を明示標識(revealing label)で標識することができる。明示標識は、ポリペプチドなどの検出を可能にする任意の適切な標識にすることができる。適切な標識には、放射性同位元素、たとえば ^{125}I 、酵素、抗体、ポリヌクレオチド、およびビオチンなどリンカーが含まれる。試料内のポリペプチドの量を決定するために、標識したポリペプチドを免疫アッセイなど診断上の手順で使用することができる。標準のプロトコルを使用して動物およびヒト内の前記ポリペプチドの免疫反応性を検出するために、ポリペプチドまたは標識したポリペプチドを血清免疫アッセイまたは細胞免疫アッセイで使用することができる。

20

【0059】

本明細書に開示する、任意選択で標識したGCR1/GCR2ポリペプチド、変異体、相同体、断片、および誘導体を固相、たとえば免疫アッセイウェルの表面またはディップスティックに固定化することもできる。このような標識したおよび/または固定化したポリペプチドを、適切な試薬、対照、指示書などと共に、適切な容器のキットにパッケージすることができる。ポリペプチドまたはその対立形質や種変異体に対する抗体の、免疫アッセイによる検出方法に、このようなポリペプチドおよびキットを使用することができる。

30

【0060】

免疫アッセイ方法は当分野で周知であり、一般的に(a)前記タンパク質に対する抗体によって結合可能なエピトープを含むポリペプチドを提供すること、(b)生物試料を前記ポリペプチドと共に、抗体-抗原複合体の形成を可能にする条件下でインキュベートすること、および(c)前記ポリペプチドを含む抗体-抗原複合体が形成されているかどうかを決定することを含む。

【0061】

本明細書に開示するGCR1/GCR2ポリペプチド、変異体、相同体、断片、および誘導体は、それに対応する遺伝子およびその相同体の、疾病における機能を含めた細胞機能における役割を研究するために、*in vitro*または*in vivo*細胞培養システムで使用することができる。たとえば、切断したまたは改変したポリペプチドを細胞に導入して、細胞内で起こる通常の機能を乱すことができる。組換え発現ベクターからのポリペプチドの*in situ*発現によって(下記参照)ポリペプチドを細胞に導入することができる。発現ベクターは、任意選択でポリペプチドの発現を調節する誘導性プロモーターを有する。

40

【0062】

昆虫細胞や哺乳動物細胞など適切な宿主細胞の使用により、組換え発現産物に最適な生理活性を与えるのに必要な可能性がある翻訳後修飾(たとえば、ミリストル化、グリコシル化、切断、ラビデーション、およびチロシン、セリン、スレオニンリン酸化)が行われることが期待される。本明細書に開示するGCR1/GCR2ポリペプチド、変異体、相同体、断片、および誘導体が発現されるこのような細胞培養システムを、細胞内のポリペプチドの機

50

能を妨害するまたは促進する候補物質を同定するアッセイシステムで使用することができる。

【0063】

GCR1/GCR2核酸

本明細書に記載する方法および組成物は、概して、いくつかのGCR1およびGCR2核酸と共に、その断片、相同体、変異体、誘導体を提供する。これらの核酸配列は好ましくは本明細書中、特に配列表中に開示のポリペプチド配列をコードする。好ましくは、ポリヌクレオチドは、配列番号1、3、5、6、7、8、9またはその断片、相同体、変異体、誘導体からなる群から選択されるStellaおよび/またはFragilis核酸を含む。

【0064】

具体的には、本発明者らは、本明細書に開示する任意のGCR1および/またはGCR2ポリペプチドをコードする核酸を提供する。したがって、用語「GCR核酸」、「GCR1核酸」、および「GCR2核酸」はそれに沿って解釈されるべきである。しかし、好ましくは、このような核酸は、配列番号1、3、5、6、7、8または9に示す任意の配列、あるいは配列番号2および4のポリペプチドの任意のものをコードする配列、ならびにこのような核酸の断片、相同体、変異体、または誘導体を含む。したがって、好ましくは、上記の用語はこれらの配列を指すと解釈すべきである。

【0065】

本文書で使用する用語「ポリヌクレオチド」、「ヌクレオチド」および核酸は、互いに同義であることを意図する。「ポリヌクレオチド」は一般的に、改変していないRNAまたはDNA、あるいは改変したRNAまたはDNAであってよい任意のポリリボヌクレオチドまたはポリデオキシリボヌクレオチドをいう。「ポリヌクレオチド」には、それだけに限定されることはないが、一本鎖および二本鎖DNA、一本鎖と二本鎖領域の混合DNA、一本鎖および二本鎖RNA、一本鎖と二本鎖領域の混合RNA、一本鎖またはより典型的には二本鎖あるいは一本鎖と二本鎖領域の混合であるDNAとRNAを含むハイブリッド分子が含まれる。さらに、「ポリヌクレオチド」とは、RNA、DNA、またはRNAとDNAをいずれも含む三本鎖領域をいう。用語ポリヌクレオチドはまた、1つまたは複数の改変した塩基を含むDNAまたはRNA、および安定性または他の理由で主鎖を改変したDNAまたはRNAを含む。「改変した」塩基には、たとえばトリチル化した塩基、イノシンなど珍しい塩基が含まれる。DNAおよびRNAに様々な改変が行われ、したがって、「ポリヌクレオチド」は自然に存在する化学的、酵素的、代謝的に改変した形のポリヌクレオチドのみならず、ウイルスおよび細胞に特徴的なDNAおよびRNAの化学的形態も包含する。「ポリヌクレオチド」はまた、しばしばオリゴヌクレオチドと呼ばれる比較的短いポリヌクレオチドも包含する。

【0066】

遺伝コードの縮重の結果、多数の様々なポリヌクレオチドおよび核酸が同じポリペプチドをコードすることができることは、当業者には理解されよう。さらに、当業者は日常的な技術を使用して、ポリペプチドを発現させる任意の特定の宿主生物のコドン利用を反映させるために、本明細書に記載のポリヌクレオチドにコードされるポリペプチド配列に影響しないヌクレオチド置換を行うことができることも、理解されよう。

【0067】

変異体、誘導体、および相同体

本明細書に記載のポリヌクレオチドは、DNAまたはRNAを含むことができる。これらは、一本鎖または二本鎖であり得る。これらはまた、内部に合成または改変したヌクレオチドを含むポリヌクレオチドであり得る。いくつかの異なるタイプのオリゴヌクレオチドの改変が当分野で知られている。これらには、メチルホスホネートおよびホスホロチオエート主鎖、分子の3'末端および/または5'末端へのアクリジンまたはポリリジン鎖の付加が含まれる。本文書の目的のために、本明細書に記載されるポリヌクレオチドは、当分野で利用可能な任意の方法で改変できることを理解されたい。このような改変を、ポリヌクレオチドの*in vivo*活性または寿命を増強させるために実施することができる。

【0068】

ポリヌクレオチドが二本鎖の場合は、二重鎖の両鎖が、個別でも組み合わせてでも、本明細書に記載の方法および組成物に包含されている。ポリヌクレオチドが一本鎖の場合は、そのポリヌクレオチドの相補配列も含まれることを理解されたい。

【0069】

ヌクレオチド配列に関連して、用語「変異体」、「相同体」、または「誘導体」には、生じるヌクレオチド配列が多能性細胞に特異的、好ましくはPGC、ES細胞、またはEG細胞に特異的であることを条件として、置換したもの、変異させたもの、改変したもの、交換したもの、1個(または複数の)ヌクレオチドを配列から欠失させたまたは配列に追加したもののすべてが含まれる。最も好ましくは、生じるヌクレオチド配列はPGCに特異的である。

【0070】

上に示したように、配列同一性に関して、「相同体」とは配列表に示す関連する配列と、好ましくは、少なくとも5%の同一性、少なくとも10%の同一性、少なくとも15%の同一性、少なくとも20%の同一性、少なくとも25%の同一性、少なくとも30%の同一性、少なくとも35%の同一性、少なくとも40%の同一性、少なくとも45%の同一性、少なくとも50%の同一性、少なくとも55%の同一性、少なくとも60%の同一性、少なくとも65%の同一性、少なくとも70%の同一性、少なくとも75%の同一性、少なくとも80%の同一性、少なくとも85%の同一性、少なくとも90%の同一性、または少なくとも95%の同一性を有する。

【0071】

より好ましくは少なくとも95%の同一性、より好ましくは少なくとも96%の同一性、より好ましくは少なくとも97%の同一性、より好ましくは少なくとも98%の同一性、より好ましくは少なくとも99%の同一性を有する。ヌクレオチド相同性比較を上記のように行うことができる。好ましい配列比較プログラムは、上記のGCG Wisconsin Bestfitプログラムである。デフォルトスコア決定(scoring)行列は、同一ヌクレオチドそれぞれに対して10、それぞれの不一致に対して-9の一致値(match value)を有する。デフォルトギャップ作成ペナルティは-50であり、デフォルトギャップ伸長ペナルティは各ヌクレオチドに対して-3である。

【0072】

ハイブリダイゼーション

本発明者らはさらに、本明細書に提示する任意の配列、またはその任意の変異体、断片、誘導体、あるいは上記のいずれかの相補体を選択的にハイブリダイズすることができるヌクレオチド配列を記載する。ヌクレオチド配列は、好ましくは少なくとも15ヌクレオチド長、より好ましくは少なくとも20、30、40または50ヌクレオチド長である。

【0073】

本明細書中で使用する用語「ハイブリダイゼーション」には、「核酸の鎖が塩基対形成で相補鎖と結合する方法」のみならず、ポリメラーゼ連鎖反応法技術で行われるような増幅方法が含まれるものとする。

【0074】

本明細書に提示するヌクレオチド配列またはその相補体を選択的にハイブリダイズすることができるポリヌクレオチドは一般的に、少なくとも20個、好ましくは少なくとも25または30個、たとえば少なくとも40、60または100個以上の連続したヌクレオチド領域にわたって、本明細書に提示する対応するヌクレオチド配列と少なくとも70%、好ましくは少なくとも80または90%、より好ましくは少なくとも95%または98%相同である。

【0075】

用語「選択的にハイブリダイズ可能」とは、標的ポリヌクレオチドがバックグラウンドより有意に高いレベルでプローブとハイブリダイズすることが分かっている条件下で、プローブとして用いるポリヌクレオチドを使用することをいう。バックグラウンドハイブリダイゼーションは、たとえばcDNAまたはスクリーニングするゲノムDNAライブラリ内に存在する他のポリヌクレオチドが原因で起こる可能性がある。この場合、バックグラウンドは、プローブとライブラリ内の非特異的DNAメンバーとの間の相互作用によって生じたシグナルレベルを示し、標的DNAで観察される特異的相互作用の1/10未満、好ましくは1/100未

10

20

30

40

50

満の強度である。相互作用の強度は、たとえばプローブを³²Pなどで放射標識することで測定することができる。

【0076】

ハイブリダイゼーション条件は、BergerおよびKimmel(1987、「Guide to Molecular Cloning Techniques, Methods in Enzymology」、第152巻、Academic Press、カリフォルニア州サンディエゴ)に教示されるように核酸結合複合体の融解温度(T_m)に基づいており、「ストリンジェンシー」の定義は以下で説明するように与える。

【0077】

最高ストリンジェンシーは通常、約T_m-5 (プローブのT_mの5 下)で起こり、T_mの約5 から10 下で高ストリンジェンシー、T_mの約10 から20 下で中程度ストリンジェンシー、T_mの約20 から25 下で低ストリンジェンシーである。当業者には理解されるように、最高ストリンジェンシーのハイブリダイゼーションを使用して同一ポリヌクレオチド配列を同定または検出することができ、中程度(または低)ストリンジェンシーのハイブリダイゼーションを使用して類似のまたは関連するポリヌクレオチド配列を同定または検出することができる。

10

【0078】

好ましい態様では、本発明者らは、ストリンジェントな条件下(たとえば、65 および0.1 × SSC{1 × SSC=0.15M NaCl, 0.015M クエン酸Na₃pH7.0})で、GCR1/GCR2核酸、またはその断片、相同体、変異体、誘導体でハイブリダイズすることができるヌクレオチド配列を開示する。

20

【0079】

ポリヌクレオチドが二本鎖の場合は、二重鎖の両鎖が、個別でも組み合わせてでも、本発明の開示に包含されている。ポリヌクレオチドが一本鎖の場合は、そのポリヌクレオチドの相補配列も開示され包含されていることを理解されたい。

【0080】

本明細書に開示する配列と100%相同ではないが開示の範囲内であるポリヌクレオチドは、いくつかの方法で得ることができる。本明細書に記載の配列の他の変異体は、たとえば、様々な個体、たとえば異なる集団からの個体から作成したDNAライブラリをプローブすることによって得ることができる。さらに、他のウイルス性/細菌性、または細胞性の相同体、特に哺乳動物細胞にある細胞性相同体(たとえば、ラット、マウス、ウシおよび霊長類の細胞、ヒト細胞を含む)を得ることができ、このような相同体およびその断片は一般的に、本明細書中の配列表に示す配列に選択的にハイブリダイズすることができる。このような配列は、他の動物種から作成したcDNAライブラリまたは他の動物種からのゲノムDNAライブラリをプローブし、配列番号1または3のすべてまたは一部を含むプローブで、中程度から高いストリンジェンシーの条件下でこのようなライブラリをプローブすることによって得ることができる。GCR1およびGCR2の種相同体ならびに対立形質変異体得るために、同様の考慮事項が当てはまる。

30

【0081】

本明細書に記載のポリヌクレオチドを使用して、プライマー、たとえばPCRプライマーや代替増幅反作用のプライマー、プローブ、たとえば放射性または非放射性標識を使用して従来方法によって明示標識で標識したプローブを生成するか、あるいはポリヌクレオチドをベクター内にクローニングすることができる。このようなプライマー、プローブ、および他の断片は、少なくとも15、好ましくは少なくとも20、たとえば少なくとも25、30、または40ヌクレオチド長であり、また本明細書中で使用される用語ポリヌクレオチドに包含される。好ましい断片は、500、200、100、50、20ヌクレオチド長より短い。

40

【0082】

DNAポリヌクレオチドなどのポリヌクレオチドおよびプローブは、組換え、合成、または当業者に利用可能な任意の方法によって生成することができる。また、これらを標準の技術でクローニングすることもできる。

【0083】

50

一般的に、プライマーは、1ヌクレオチドずつ所望の核酸配列を段階的に製造することを要する合成方法によって生成する。自動化技術を使用してこれを行う技術は、当分野で容易に利用可能である。

【0084】

より長いポリヌクレオチドは一般的に組換え方法を使用して、たとえばPCR(ポリメラーゼ連鎖反応)クローニング技術を使用して生成する。これには、クローニングを所望する配列領域に隣接する1対のプライマー(たとえば約15から30ヌクレオチド)を作成すること、プライマーを動物またはヒトの細胞から得たmRNAまたはcDNAと接触させること、所望の領域の増幅をもたらす条件下でポリメラーゼ連鎖反応を実施すること、増幅した断片を(たとえば反応混合物をアガロースゲル上で精製することによって)単離すること、および増幅したDNAを回収することを要する。プライマーは、増幅したDNAを適切なクローニングベクター内にクローニングできるよう、適切な制限酵素認識部位を含むように設計することができる。

10

【0085】**ヌクレオチドベクター**

ポリヌクレオチドを、組換え複製可能なベクターに組み入れることができる。このベクターは、適合宿主内の核酸を複製するのに使用することができる。したがって、さらなる実施形態では、本発明者らは、ポリヌクレオチドを複製可能なベクターに導入すること、このベクターを適合宿主細胞に導入すること、およびベクターの複製をもたらす条件下で宿主細胞を増殖させることによって、ポリヌクレオチドを作成する方法を提供する。適切な宿主細胞には、E.coliなどの細菌、酵母菌、哺乳動物細胞系、および他の真核生物細胞系、たとえば昆虫Sf9細胞が含まれる。

20

【0086】

好ましくは、ベクター内のポリヌクレオチドは、宿主細胞でコード配列の発現をもたらすことができる調節配列に作動可能に連結している。すなわち、このベクターは発現ベクターである。用語「作動可能に連結」とは、記載した構成成分が、その意図された様式で機能できる関係にあることを意味する。コード配列に「作動可能に連結」している制御配列は、調節配列に適合した条件下でコード配列の発現が成されるようにライゲーションされている。

【0087】

たとえば調節配列に指示される転写レベルの転写モジュレーターに対する応答性をより高めさせる、さらなる転写制御要素を付加することによって、調節配列を改変することができる。

30

【0088】

タンパク質の発現をもたらすために、ベクターを適切な宿主に、下記のように形質転換または形質移入させることができる。この方法に、ベクターにタンパク質をコードするコード配列の発現をもたらす条件下で、上記の発現ベクターを用いて形質転換させた宿主細胞を培養すること、および任意選択で発現させたタンパク質を回収することを含めることができる。

【0089】

ベクターは、たとえば複製起点、任意選択で前記ポリヌクレオチドの発現用のプロモーター、および任意選択でプロモーターのレギュレーターを与えられたプラスミドまたはウイルスベクターにすることができる。このベクターに、1つまたは複数の選択可能マーカータンパク質、たとえば細菌プラスミドの場合はアンピシリン耐性遺伝子または哺乳動物ベクターの場合はネオマイシン耐性遺伝子を含めることができる。たとえば宿主細胞を形質移入またはそれを形質転換させるのにベクターを使用することができる。

40

【0090】

タンパク質をコードする配列に作動可能に連結している調節配列には、プロモーター/エンハンサーおよび他の発現制御シグナルが含まれる。これらの調節配列を、発現ベクターを使用する予定の宿主細胞に適合するように選択することができる。用語「プロモーター

50

」は当分野で周知であり、大きさおよび複雑さが最小プロモーターから上流要素およびエンハンサーを含むプロモーターの範囲である核酸領域を包含する。

【0091】

プロモーターは通常、哺乳動物細胞内で機能的なプロモーターから選択されるが、原核細胞のプロモーターおよび他の真核細胞で機能的なプロモーターを使用することもできる。このプロモーターは通常、ウイルスまたは真核細胞遺伝子のプロモーター配列に由来する。たとえば、発現を引き起こさせる細胞のゲノム由来のプロモーターにすることができる。真核生物のプロモーターに関しては、偏在的に機能するプロモーター(アクチン、アクチン、チューブリンのプロモーターなど)、あるいは組織特異的に機能するプロモーター(ピルビン酸キナーゼの遺伝子のプロモーターなど)にし得る。また、特定の刺激に
10
 応答するプロモーター、たとえばステロイドホルモン受容体を結合するプロモーターにすることもできる。ウイルスプロモーター、たとえばモロニー Maus 白血病ウイルス末端反復配列(MMLV LTR)プロモーター、ラウス肉腫ウイルス(RSV)LTRプロモーター、ヒトサイトメガロウイルス(CMV)IEプロモーターを使用することもできる。

【0092】

異種性遺伝子の発現レベルを細胞の生存期間中制御できるように、プロモーターを誘導性
 にすることが有利な場合がある。誘導性とは、プロモーターを用いて得た発現レベルが制御可能であることを意味する。

【0093】

さらに、さらなる制御配列、たとえばエンハンサー配列を付加することによって、これら
20
 のプロモーターすべてを改変することができる。上に説明したプロモーターの2つ以上の異なるものからの配列要素を含むキメラプロモーターを使用することができる。

【0094】

宿主細胞

本明細書に開示するベクターおよびポリヌクレオチドを、ベクター/ポリヌクレオチドを複製するおよび/またはタンパク質を発現させるために、宿主細胞内に導入することができる。原核細胞を宿主細胞としてタンパク質を生成することもできるが、真核細胞、たとえば酵母菌、昆虫、または哺乳動物の細胞、特に哺乳動物細胞を使用することが好ましい
 。

【0095】

形質移入、形質転換、電気穿孔など当分野で周知の様々な技術を使用してベクター/ポリ
30
 ヌクレオチドを適切な宿主細胞に導入することができる。本明細書に開示するベクター/ポリヌクレオチドを動物に投与する場合は、当分野で既知のいくつかの技術、たとえばレトロウイルス、単純ヘルペスウイルス、アデノウイルスなどの組換えウイルスベクターを用いた感染、核酸の直接注入、微粒子銃による形質転換が知られている。

【0096】

タンパク質の発現および精製

本明細書に開示するポリヌクレオチドを含む宿主細胞を使用してタンパク質を発現させる
40
 ことができる。タンパク質の発現を可能にする適切な条件下で宿主細胞を培養することができる。本明細書に記載するタンパク質の発現は、これらが連続的に生成される構成的なものでも、または発現の開始に刺激を必要とする誘導的なものでもよい。誘導的発現の場合、タンパク質の生産を、たとえば培養基に誘導物質、たとえばデキサメタゾンやIPTGを添加することによって、必要なときに開始させることができる。

【0097】

酵素溶解、化学溶解、および/または浸透性溶解および物理的破壊を含む、当分野で周知
 の様々な技術によって宿主細胞からタンパク質を抽出することができる。

【0098】

組換えStellaおよびFragilisタンパク質

StellaおよびFragilisのヌクレオチド配列をTRIシステムベクター(Qiagen)内にクローニ
50
 ングする。適切な制限酵素部位を用いて、製造者の指示に従って、前方の第2コドン(すな

わち、第1のコドンATGなしのStellaのN末端断片)を含むStella配列をpQEベクター内にクローニングする。QIAexpress pQEベクターは、6×Hisタグタンパク質のE.coli内での高レベル発現を可能にする。

【0099】

Stella遺伝子のN末端部分にHisタグを配置する。Ni-NTAカラム上のアフィニティークロマトグラフィーによって、製造者の指示に従って組換えタンパク質を精製する。適切な制限酵素を使用してHisタグを切断する。

【0100】

組換えで発現したStellaおよびFragilisタンパク質は、生物学的に活性であることが分かった。

【0101】

抗体

本明細書中で使用する抗体とは、選択した標的を結合することができる完全な抗体または抗体断片をいい、Fv、ScFv、Fab'およびF(ab')₂、モノクローナルおよびポリクローナル抗体、キメラ、CDR移植およびヒト化抗体を含む操作した抗体、ならびにファージディスプレイや代替技術を用いて生成した人工的に選択した抗体が含まれる。FvやScFvなど小さな断片は、その小さなサイズとその結果の優れた組織内分布のために、診断用途および治療用途に有利な特性を有する。

【0102】

本明細書の記載に従った抗体は特に、PGCおよびES細胞やEG細胞など他の多能性細胞の検出用に示す。したがって、これらは、標識など効果タンパク質を含む改変した抗体にすることができる。抗体のin vivoまたはin vitro分布のイメージングを可能にする標識が特に好ましい。このような標識は、胚または細胞の塊内で容易に可視化できる放射性標識または金属粒子など放射線不透過性標識であってよい。さらに、これらは蛍光標識または組織試料上で可視化できる他の標識であってよい。

【0103】

組換えDNA技術を使用して、本明細書に記載のように抗体を改良することができる。このように、診断用途または治療用途で、抗体の免疫原性を低下させるためにキメラ抗体を構築することができる。さらに、CDR移植(ヨーロッパ特許出願第0239400号(Winter)参照)および任意選択でフレームワーク改変(ヨーロッパ特許出願第0239400)で抗体をヒト化することによって、免疫原性を最小限にすることができる。

【0104】

抗体は、動物血清から得るか、あるいは、モノクローナル抗体またはその断片の場合は細胞培養物内で産生させることができる。組換えDNA技術を使用して、細菌または好ましくは哺乳動物細胞培養物内で、確立した手順に従って抗体を産生させることができる。選択した細胞培養システムは、好ましくは抗体産物を分泌する。

【0105】

したがって、本発明者らは、前記抗体タンパク質をコードする第2 DNA配列に正しい読み枠で連結されたシグナルペプチドをコードする第1 DNA配列に作動可能に連結されたプロモーターを含む発現カセットを含むハイブリッドベクターを用いて形質転換させた宿主、たとえばE.coliや哺乳動物細胞を培養すること、および前記タンパク質を単離することを含む、抗体を産生させる方法を開示する。

【0106】

ハイブリドーマ細胞または哺乳動物宿主細胞のin vitro増殖は、たとえば、ウシ胎児血清など哺乳動物血清、または微量元素や、正常マウス腹腔浸出細胞など支持細胞、脾臓細胞、骨髄マクロファージ、2-アミノエタノール、インスリン、トランスフェリン、低密度リポタンパク質、オレイン酸など増殖維持サプリメントを任意選択で補充した、ダルベッコ改変イーグル培地(DMEM)やRPMI 1640培地など通常の標準培養基である適切な培養基中で実施する。細菌細胞や酵母菌細胞である宿主細胞の増殖も同様に、当分野で周知の適切な培養基、たとえば、細菌は培地LB、NZCYM、NZYM、NZM、Terrific Broth、SOB、SOC、2×Y

10

20

30

40

50

T、M9最少培地、酵母菌は培地YPD、YEPD、最少培地、完全最少ドロップアウト培地で行う。

【0107】

In vitroの産生は、比較的高純度な抗体調製物を与え、スケールアップにより大量の所望の抗体を与える。細菌細胞、酵母菌細胞、哺乳動物細胞の培養の技術は当分野で周知であり、均一懸濁培養、たとえばエアリフト反応器中または連続攪拌反応器中で、あるいは固定化または包括細胞培養、たとえば中空繊維中、マイクロカプセル中、アガロースマイクロビーズ上、またはセラミックカートリッジ中を含める。

【0108】

大量の所望の抗体を、哺乳動物細胞をin vivoで増殖することで得ることもできる。このために、所望の抗体を産生しているハイブリドーマ細胞を組織適合性哺乳動物に注入して、抗体産生腫瘍の成長を引き起こさせる。任意選択で、注入前に動物を炭化水素、特にプリスタン(テトラメチル-ペンタデカン)など鉱物油で準備刺激する。1から3週間後、これら哺乳動物の体液から抗体を単離する。たとえば、適切なミエローマ細胞とBalb/cマウス由来の抗体産生脾臓細胞との融合によって得られるハイブリドーマ細胞、または所望の抗体を産生するハイブリドーマ細胞系Sp2/0由来の形質移入細胞を、任意選択でプリスタンで前処理したBalb/cマウスに腹腔内注入し、1から2週間後、動物から腹水液を採取する。

10

【0109】

前述のおよび他の技術は、たとえばKohlerおよびMilstein、1975、Nature、256:495-497; 米国特許第4,376,110号; HarlowおよびLane、「Antibodies:a Laboratory Manual」、1988、Cold Spring Harborに記載されており、本明細書に参照により組み込む。組換え抗体分子の調製技術も上記の参照文献に記載されており、また、たとえばヨーロッパ特許第0623679号、ヨーロッパ特許第0368684号、およびヨーロッパ特許第0436597号にも記載されており、これらを参照により本明細書に組み込む。

20

【0110】

細胞培養上清から所望の抗体を、優先的にはPGCまたはES細胞やEG細胞など他の多能性細胞の免疫蛍光染色によって、免疫プロットによって、サンドイッチアッセイやドットプロットアッセイなど酵素免疫アッセイによって、または放射性免疫アッセイによってスクリーニングする。

【0111】

抗体の単離には、培養物上清または腹水液中の免疫グロブリンを、たとえば硫酸アンモニウムを用いた沈殿、ポリエチレングリコールなど吸湿性物質に対する透析、選択的膜での濾過などによって、濃縮することができる。必要かつ/または所望する場合は、抗体を従来のクロマトグラフィー方法、たとえばゲル濾過、イオン交換クロマトグラフィー、DEAE-セルロースおよび/または(免疫)アフィニティークロマトグラフィー、たとえばGCR1やGCR2またはその断片あるいはタンパク質-Aを用いたアフィニティークロマトグラフィーによって精製することができる。

30

【0112】

モノクローナル抗体を分泌しているハイブリドーマ細胞も提供する。好ましいハイブリドーマ細胞は遺伝的に安定であり、所望の特異性のモノクローナル抗体を分泌し、超低温凍結培養物から解凍および再クローニングすることによって活性化させることができる。

40

【0113】

適切な哺乳動物たとえばBalb/cマウスを、1つまたは複数のGCR1またはGCR2ポリペプチドあるいはその抗原性断片で免疫化することを特徴とする、GCR1および/またはGCR2に対するモノクローナル抗体を分泌しているハイブリドーマ細胞系の調製方法が、さらに含まれている。免疫化した哺乳動物の抗体産生細胞を適切なミエローマ細胞系の細胞と融合させ、融合で得られたハイブリッド細胞をクローニングし、所望の抗体を分泌している細胞クローンを選択する。たとえば、GCR1および/またはGCR2を用いて免疫化したBalb/cマウスの脾臓細胞をミエローマ細胞系PA1またはミエローマ細胞系Sp2/0-Ag14の細胞と融合させ、ハイブリッド細胞から所望の抗体の分泌をスクリーニングし、陽性ハイブリドーマ細胞

50

をクローニングする。

【0114】

GCR1および/またはGCR2を発現している10個および 10^7 個と 10^8 個の間の細胞と適切なアジュバントとを数回、たとえば4から6回、数カ月、たとえば2から4カ月の間に皮下注入または腹腔内注入することによってBalb/cマウスを免疫化すること、および、最終注入から2から4日後に免疫化したマウスから脾臓細胞を採取して、融合促進剤、好ましくはポリエチレングリコールの存在下でミエローマ細胞系PA1と融合させることを特徴とする、ハイブリドーマ細胞系の調製方法が好ましい。好ましくは、ミエローマ細胞を、3から20倍過剰の免疫化したマウス由来の脾臓細胞と、約30%から約50%の分子量約4000のポリエチレングリコールを含む溶液中で融合させる。融合後、本明細書に既に記載した適切な培養基中で細胞を増殖させ、正常なミエローマ細胞が所望のハイブリドーマ細胞を過増殖するのを防ぐために選択培地、たとえばHAT培地を定期的な間隔で補充する。

10

【0115】

本明細書に既に記載したGCR1および/またはGCR2に対する抗体の重鎖可変ドメインおよび/または軽鎖可変ドメインをコードする挿入物を含む組換えDNAも、開示している。定義により、このようなDNAには、一本鎖コードDNA(coding DNA)、前記コードDNAおよびそれに相補的なDNAからなる二本鎖DNA、またはこれらの相補的な(一本鎖)DNA自体が含まれる。

【0116】

さらに、GCR1および/またはGCR2に対する抗体の重鎖可変ドメインおよび/または軽鎖可変ドメインをコードするDNAは、重鎖可変ドメインおよび/または軽鎖可変ドメイン、あるいはその変異体をコードする標準DNA配列を有する、酵素的または化学的に合成したDNAであってよい。標準DNAの変異体とは、上で言及した抗体の重鎖可変ドメインおよび/または軽鎖可変ドメインであって、アミノ酸の1つまたは複数欠失している、あるいは1つまたは複数の他のアミノ酸と交換されているドメインをコードするDNAである。好ましくは、前記改変は、抗体の重鎖可変ドメインおよび/または軽鎖可変ドメインのCDRの外側にある。このような変異体DNAはまた、1つまたは複数のヌクレオチドが他のヌクレオチドで置換されているが新しいコドンも同じアミノ酸をコードしている、沈黙変異体であることを意図している。このような変異配列は、縮重配列でもある。縮重配列は、無制限数のヌクレオチドが他のヌクレオチドで置換されても元来コードされていたアミノ酸配列に変化がもたらされないように、遺伝コードの意味の範囲内で縮重している。このような縮重配列は、重鎖マウス可変ドメインおよび/または軽鎖マウス可変ドメインの最適な発現を得るために特定の宿主、特にE.coliが好む、異なった制限部位および/または特定のコドンの頻度のために、有用な可能性がある。

20

30

【0117】

変異体という用語は、当分野で周知の方法に従って標準DNAをin vitroで突然変異させることで得るDNA変異体を含むことを意図する。

【0118】

完全な四量体免疫グロブリン分子の構築およびキメラ抗体の発現のために、重鎖および軽鎖可変ドメインをコードする組換えDNA挿入物を、重鎖および軽鎖定常ドメインをコードする対応したDNAと融合させ、その後、たとえばハイブリッドベクターに組み込ませた後、適切な宿主細胞内に移す。

40

【0119】

ヒト定常ドメインg、たとえば 1、 2、 3、 4、好ましくは 1または 4と融合された、GCR1および/またはGCR2に対する抗体の重鎖マウス可変ドメインをコードする挿入物を含む組換えDNAも開示している。同様に、本発明は、ヒト定常ドメイン または 、好ましくは と融合された、GCR1および/またはGCR2に対する抗体の軽鎖マウス可変ドメインをコードする挿入物を含む組換えDNAに関する。

【0120】

別の実施形態では、本発明者らは、重鎖可変ドメインおよび軽鎖可変ドメインがスパーサー基によって結合されており、任意選択で、宿主細胞内での抗体のプロセッシングを促進

50

するシグナル配列、および/または抗体の精製を容易にするペプチドをコードするDNA、および/または切断部位、および/またはペプチドスペーサー、および/または効果分子を含む組換えポリペプチドをコードする組換えDNAを開示する。

【0121】

効果分子をコードするDNAは、診断用途または治療用途に有用な効果分子をコードするDNAであることが意図される。したがって、毒素または酵素、特にプロドラッグの活性化を触媒することができる酵素であるエフェクタ分子が、特に示される。このようなエフェクタ分子をコードするDNAは、酵素あるいは毒素またはその変異体をコードする自然に存在する配列を有し、当分野で周知の方法によって調製することができる。

【0122】

抗ペプチドStellaおよびFragilis抗体

StellaおよびFragilisペプチド配列に対する抗ペプチド抗体を産生させる。選択した配列は以下のとおりである。

GCR1(Fragilis):ASGGQPPNYERIKEEYEおよびRDRKMVGDTVGAQAYA

GCR2(Stella):MEEPSEKVDPMKDPETおよびCHYQRWDPSENAKIGKN

【0123】

抗体は、たとえばHarlowおよびLane(上記)に記載のように、ウサギに注入すること、および他の従来方法によって産生させることができる。

【0124】

抗体は、Elisaアッセイおよびウエスタンブロットによって確認し、実施例に記載のように免疫染色で使用する。

【0125】

細胞集団内の多能性細胞の検出

本明細書に記載のポリヌクレオチドプローブまたは抗体を、細胞集団内の始原生殖細胞(PGC)など多能性細胞、胚性幹(ES)細胞や胚性生殖(EG)細胞など幹細胞の検出に使用することができる。本明細書で使用する「細胞集団」とは、1つまたは複数のPGC、ES細胞、またはEG細胞を含む、細胞の任意の集まりである。好ましくは、細胞の集まりはPGCのみで構成されず、少なくとも1種の別のタイプの細胞を含む。

【0126】

細胞集団は胚および胚組織を含むが、前述の任意のもの由来の成体細胞ならびに培養物および細胞調製物内で増殖させた組織も含む。

【0127】

本明細書に記載のポリヌクレオチドを、核酸ハイブリダイゼーション技術によって、PGCまたはES細胞やEG細胞など他の多能性細胞内のGCR1およびGCR2の転写物を検出するのに使用することができる。このような技術には、GCR1および/またはGCR2転写物にハイブリダイズさせて転写物を増幅するのにプライマーを使用し、検出可能なシグナルをもたらすPCR、ならびに、GCR1および/またはGCR2転写物内の独特な配列に特異的であるプローブを標的細胞内の転写物を検出するのに使用する標識プローブのハイブリダイゼーションが含まれる。

【0128】

本明細書に既に言及したように、プローブを放射性、放射能不透過性、蛍光、または他の標識で、当分野で周知のように標識することができる。

【0129】

抗体を、GCR1および/またはGCR2を検出するのに使用することもできる。特にGCR1は、抗GCR1抗体を標的にして細胞表面で検出することができる細胞外ドメインを有する。あるいは、細胞内のGCR1および/またはGCR2を検出するために細胞内scFvを使用することができる。

【0130】

免疫染色およびFACS技術を特に示す。適切なフルオロフォアは当分野で周知であり、化学的フルオロフォアおよびGFPやその変異体(国際公開公報97/28261号参照)など蛍光ポリペ

10

20

30

40

50

プチドが含まれる。化学的フルオロフォアは、免疫グロブリン分子の合成中に、これに対する結合部位を免疫グロブリン分子内に組み入れることによって、付着させることができる。

【0131】

好ましくは、フルオロフォアは、有利にGFPまたはその変異体である蛍光タンパク質である。GFPおよびその変異体は、当分野で周知の方法に従って、免疫グロブリンまたは標的分子と共に、融合ポリペプチドとして一緒に発現させることによって合成することができる。たとえば、所望のGFPと免疫グロブリンまたは標的との枠内融合として転写ユニットを構築し、PCRクローニングおよびライゲーション技術を使用して上記のようにベクター内に挿入することができる

10

【0132】

シグナルを生成することができる任意の標識で抗体を標識することができる。シグナルは、検出可能な遺伝子産物の発現誘導など任意の検出可能なシグナルにし得る。検出可能な遺伝子産物の例には、ルシフェラーゼやGFPなど生物発光ポリペプチド、 β -ガラクトシダーゼやCATなど特定のアッセイで検出可能なポリペプチド、HIS3など代謝に必要な酵素やG418など抗生物質耐性遺伝子などの宿主細胞の増殖の特徴を変調させるポリペプチドが含まれる。好ましい態様では、このシグナルは細胞表面で検出可能である。たとえば、このシグナルは、細胞外から検出可能であり、FACSまたは他の光学選別技術による細胞の選別を可能にする、発光または蛍光シグナルにし得る。

【0133】

蛍光標識した抗体の光学検出に基づいた光学免疫センサー技術の使用が好ましい。免疫センサーとは、相補種の結合を検出するシグナルトランスデューサーに結合された抗原または抗体種を含む生化学的検出器である(Rabbany他、1994、Crit Rev Biomed Eng、22:307-346;Morgan他、1996、Clin Chem、42:193-209)。このような相補種には、抗原Zif268および抗Zif268抗体が含まれる。免疫センサーは、血清や全血液など複合試料中に存在する抗体、抗原、またはハプテンの量の定量測定を行う(Robinson、1991、Biosens Bioelectron、6:183-191)。免疫センサーの感度は、スピードと正確さが要される状況に理想的である(Rabbany他、1994、Crit Rev Biomed Eng、22:307-346)。

20

【0134】

免疫センサーで利用する検出技術には、免疫相互作用の電気化学的、圧電氣的、光学的な検出が含まれる(Ghindilis他、1998、Biosens Bioelectron、1:113-131)。間接的な免疫センサーでは、結合の後にたとえば蛍光または発光によって検出する、個別の標識した種を使用する(Morgan他、1996、Clin Chem、42:193-209)。直接的な免疫センサーは、電位差、電流、抵抗、質量、熱、または光学特性の変化によって結合を検出する(Morgan他、1996、Clin Chem、42:193-209)。間接的な免疫センサーは、非特異的結合のために直面する問題がより少ない(Attridge他、1991、Biosens Bioelectron、6:201-214;Morgan他、1996、Clin Chem、42:193-209)。

30

【0135】

発明のさらなる態様

本発明者らは、配列番号1と少なくとも90%相同な核酸分子および配列番号3と少なくとも75%相同な核酸分子を提供する。

40

【0136】

本発明者らは、配列番号1または配列番号3、配列番号5から9のいずれか、あるいはそれと少なくとも90%相同である配列からのヌクレオチドの連続ストレッチを含むポリヌクレオチドを開示する。有利には、この連続ヌクレオチドのストレッチは50ヌクレオチド長、好ましくは40、35、30、25、20、15または10ヌクレオチド長である。

【0137】

遺伝子GCR1およびGCR2は新規なポリペプチドをコードしており、その配列を配列番号2および配列番号4に示す。したがって、本発明者らは、ここに記載の核酸にコードされるポリペプチドを開示する。好ましくは、ポリペプチドは配列番号2または配列番号4に示す配

50

列を有する。

【0138】

さらに、本発明者らは、(a)PGCまたはES細胞やEG細胞など他の多能性細胞を含む細胞集団を提供するステップと、(b)そこから1つまたは複数のPGCまたはES細胞やEG細胞など他の多能性細胞を単離して単一細胞の単離体を提供するステップと、(c)単一細胞内に存在する転写された核酸を増幅するステップと、(d)PGCまたはES細胞やEG細胞など他の多能性細胞内に存在するが体細胞内には存在しない転写物を同定するためにサブトラクティブハイブリダイゼーションスクリーンを実施するステップと、(e)ステップ(d)で同定した1つまたは複数の転写物を用いて核酸ライブラリをプローブし、特異的に発現される1つまたは複数の遺伝子をクローニングするステップとを含む、PGCまたはES細胞やEG細胞など他の多能性細胞内で特異的に発現される遺伝子を単離することができる方法を提供する。

10

【0139】

ここで、本発明のさらなる態様を、番号付けした段落で示す。本発明が以下の態様を包含することを理解されたい。

【0140】

段落1。配列番号1に示した配列と少なくとも90%の相同性を有する核酸。

【0141】

段落2。配列番号3に示した配列と少なくとも75%の相同性を有する核酸。

【0142】

段落3。段落1または段落2の核酸の25個の連続ヌクレオチド配列を含む核酸。

20

【0143】

段落4。段落1または段落2の核酸の15個の連続ヌクレオチド配列を含む核酸。

【0144】

段落5。前記段落のいずれかに記載の核酸配列の相補体。

【0145】

段落6。1つまたは複数のヌクレオチドの置換を含むが、遺伝コードの縮重のためにこのような置換によって前記核酸のコード特異性が改変されない、段落1から5のいずれかに記載の核酸。

【0146】

段落7。前記段落のいずれかに記載の核酸にコードされるポリペプチド。

30

【0147】

段落8。段落1または段落2に記載の核酸配列またはその相同体の発現を検出することを含む、細胞集団内の始生殖細胞を同定する方法。

【0148】

段落9。GCR1および/またはGCR2に特異的な5'および3'プライマーを使用して推定上のPGCから核酸を増幅するステップと、その結果生成された増幅した核酸を検出するステップとを含む、段落8に記載の方法。

【0149】

段落10。核酸配列の発現がin situハイブリダイゼーションによって検出される、段落8に記載の方法。

40

【0150】

段落11。核酸配列の発現が、それにコードされるタンパク質産物の検出によって決定される、段落8に記載の方法。

【0151】

段落12。タンパク質産物を免疫染色によって検出する段落11に記載の方法。

【0152】

段落13。段落7に記載のポリペプチドに特異的な抗体。

【0153】

段落14。GCR1の細胞外ドメインに特異的な、段落13に記載の抗体。

【0154】

50

段落15. 細胞集団内のPGCを同定するための、段落13または段落14に記載の抗体の使用。

【0155】

段落16. 段落8から12のいずれかに記載の方法によって同定したPGC。

【0156】

段落17. (a)PGCを含む細胞集団を提供するステップと、(b)そこから1つまたは複数のPGCを単離して単一細胞のPGC単離体を提供するステップと、(c)単一のPGC内に存在する転写された核酸を増幅するステップと、(d)PGC内に存在するが体細胞内には存在しない転写物を同定するためにサブトラクティブハイブリダイゼーションスクリーンを実施するステップと、(e)ステップ(d)で同定した1つまたは複数の転写物を用いて核酸ライブラリをプローブし、PGC内で特異的に発現される1つまたは複数の遺伝子をクローニングするステップとを含む、PGC内で特異的に発現される遺伝子を単離する方法。

10

【0157】

実施例

実施例1

単一細胞cDNAディファレンシャルスクリーニングによる、始原生殖細胞(PGC)の初期集団に特異的な遺伝子の同定

生殖細胞系の特異化に關与する遺伝子を同定する単一細胞の分析方法が開発されており、これは、始原生殖細胞(PGC)の創始集団の確立をもたらす。PGCの系譜特異化は、体細胞内では発現されない独特な遺伝子の組の発現に伴う。

【0158】

遺伝子を同定する方法は主に、PGCを含む7.25日齡のマウス胚性断片由来の単一細胞からのライブラリのディファレンシャルスクリーニングに基づいている。単一細胞cDNAのディファレンシャルスクリーンはBradyおよびIscoe、1993によって最初に記載されており、その後Cathaline DulacおよびRichard Axelによって改変され、ラット由来のフェロモン受容体遺伝子の同定に成功した(Dulac,CおよびAxel、1995)。記載した軽微な改変を用いて、Axelのグループの方法を利用する。

20

【0159】

PGCの最も初期集団を保有する胚性断片由来の単一細胞cDNAの構築

マウス内で、PGCの最も初期集団は、妊娠7.25日目に新生尿膜基部に、約40個の細胞のアルカリホスファターゼ陽性クラスターからなると報告されている(Ginsburg,M., Snow,M.H .L., およびMcLaren,A., 1990)。同系交配した129SvおよびC57BL/6系内のPGCクラスターの正確な位置は、全載アルカリホスファターゼ染色およびメチレンブルーのどちらかで染色した準薄切片を使用した顕微鏡観察によって決定する。PGCクラスターを検出できる最も初期段階は後期ストリーク(Streak)段階であり(Downs,K.M.およびDavies,T., 1993)、このとき尿膜芽(Bud)が現れる蓋上皮(epithelial lining)のすぐ下にはっきりと染色された細胞集団が見つかる。この領域は、胚外組織と胚性組織の境界、始原ストリークに最も近位な部分のすぐ後方で上方にある。このクラスターは、少なくとも初期/中期芽段階までこの場所に持続される。同系交配した129Sv系では、PGCクラスターが含む細胞数は僅かに多く、C57BL/6系に比べてより密接に詰まっていることが分かった。得られる初期PGCの回収率がより高いので、129Sv系をその後の実験に使用する。

30

40

【0160】

E7.5に129Sv胚を、DMEMおよび25mMのHEPESで緩衝した10%のFCS中、室温で単離し、各胚の発生段階を解剖顕微鏡の下で決定した。正確な発生段階は、同一腹内の胚でも相当異なる可能性がある。芽なしの段階または初期芽(尿膜)段階にある胚を選択してさらに解体し、これは、解剖顕微鏡の下で見られるPGCを含む領域を同定することの容易さによってある程度規定される。PGCクラスターを含むと予想される断片を、固体ガラス針によって非常に正確に切り出す。37 で10分間、0.25%トリプシン-1mM EGTA/PBSでこの領域を処理し、続いてマウスピペットで穏やかにピペット操作することで、この領域を単一細胞に解離する。解体した断片は通常250から300個の細胞を含んでいた。この穏やかな手順を用いた細胞分散手順により、内臓中胚葉層が無傷の細胞シートのまま残された。

50

【0161】

本発明者らは、マウスピペットで単一細胞を細胞懸濁液から無作為に選び、個々の単一細胞を(空気泡を生じさせるのを避けて)4 μ lの氷冷細胞溶解緩衝液(50mMのトリス-HCl pH8.3、75mMのKCl、3mMのMgCl₂、0.5%のNP-40、80ng/mlのpd(T)24、5 μ g/mlのprime RNase阻害剤、324U/mlのRNA guard、それぞれ10mMのdATP、dCTP、dGTP、dTTPを含む)を含む薄壁PCRチューブに入れた。単一細胞が保有する培地の体積は0.5 μ l未満である。チューブを手短に遠心分離して、細胞が確かに溶解緩衝液中にあることを確認する。それぞれの個別の実験で、本発明者らは合計19個の単一細胞を選び、1つのチューブをPCR増幅手順の陰性対照として細胞なしのままにした。チューブに収集した細胞すべてを、その後の手順を開始する前に氷上に置く。

10

【0162】

チューブを65 $^{\circ}$ Cで1分間インキュベートすることによって細胞を溶解し、オリゴdTをRNAにアニーリングさせるために室温で1から2分間置いた。第1鎖cDNAの合成は、50Uのモロニー Maus白血球ウイルス(MMLV)および0.5Uのトリ骨髄芽球症ウイルス(AMV)逆転写酵素を加えることによって開始させ、次いで37 $^{\circ}$ Cで15分間インキュベートした。65 $^{\circ}$ C、10分間で逆転写酵素を不活性化させる。この逆転写反応は15分間に限定されており、これにより、C末端から500塩基~1000塩基長の比較的均一な大きさのcDNAの合成が可能になる。これは、その後のPCR増幅がかなり表現的(representative)になることを可能にする。

【0163】

次に、合成した第1鎖cDNAの5'(5 prime)末端にポリAテールを付加するために、4.5 μ lの2 \times ターリング緩衝液(200mMのカコジル酸カリウムpH7.2、4mMのCoCl₂、0.4mMのDTT、200mMのdATP、10Uの末端転移酵素を含む)を反応物に加え、次いで37 $^{\circ}$ Cで15分間インキュベートする。試料を65 $^{\circ}$ Cで10分間、熱で不活性化させる。この時点で、反応物には、C末端にポリTテールを有しN末端にポリAストレッチを有する合成したcDNAが含まれ、特異的プライマーを使用したPCRによる増幅の準備ができた。

20

【0164】

10mMのトリス-HCl pH8.3、50mMのKCl、2.5mMのMgCl₂、100 μ g/mlのウシ血清アルブミン、0.05%のTriton-X100、1mMのdATP、dCTP、dGTP、dTTP、10UのTaqポリメラーゼ、5 μ gのAL1プライマーからなる溶液を用いて、各チューブの内容量を100 μ lにする。AL1の配列は、ATTGGA TCC AGG CCG CTC TGG ACA AAA TAT GAA TCC (T)₂₄である。PCR増幅は以下のスケジュールに従って実施する。すなわち、94 $^{\circ}$ Cで1分間、42 $^{\circ}$ Cで2分間、72 $^{\circ}$ Cで6分間で各サイクル10秒間ずつ延長して、25サイクル行う。5ユニットの追加のTaqポリメラーゼを加えて、さらに25サイクル、延長時間なしで同じプログラムを実施する。この時点で、各チューブが単一細胞由来の増幅したcDNA産物を含む。溶液のタンパク質含量をフェノール/クロロホルム処理によって抽出し、増幅したcDNAをエタノールで沈殿させ、最終的に100 μ lのTE pH8.0に懸濁させる。5 μ lのcDNA溶液を1.5%アガロースゲル上に流して増幅が成功したかどうかを確認する。ほとんどの試料で、主に500bpから1200bpの範囲の非常に強い「スマア」バンドが現れ、単一細胞cDNAは効率的に増幅されたことが示された。増幅が成功した試料のみを、その後の「細胞種類分け」分析に使用する。

30

【0165】

実施例 2

マーカー遺伝子の発現の検査によるPGCの同定

理論的に切除される胚の破片には、尿膜中胚葉、PGC、PGCを囲む胚外中胚葉の、3つの主な成分が含まれる。これらの試料からPGC起源の単一細胞cDNAを同定するため、様々な細胞種内のこの領域で発現または抑制されることで知られている4つのマーカー遺伝子(BMP4、TNAP、Hoxb1、Oct4)の発現を検査することによって、構築したcDNAの陽性および陰性選択を実施する。

40

【0166】

芽なし/初期芽段階では、BMP4は新生尿膜中および新生羊膜、漿膜、内臓卵黄囊の中胚葉成分中で発現されると報告されている(Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinst

50

ra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., および Hogan, B.L.M., 1999)。BMP4発現の境界は非常にはっきりしており、羊膜中胚葉につながっている蓋上皮の下の推定PGCが決定される中胚葉領域で、発現は完全に排除されている。したがって、BMP4は選択の陰性マーカーとして使用する。BMP4のC末端部分を増幅するためにプライマーの対を設計する(5':GCC ATA CCT TGA CCC GCA GAA G、3':AAA TGG CAC TCA GTT CAG TGG G)。PCR増幅は、0.5 μlのcDNA溶液を鋳型として使用し、以下のスケジュールに従って実施する。すなわち、95 で1分間、55 で1分間、72 で1分間を20サイクルで行う。試験した83個の試料のうち、57個の試料が予想した大きさのバンドを示し、これら単一細胞内でのBMP4の発現が示された。これらの試料は尿膜中胚葉起源のものと考えられ、したがって、PGC起源の細胞を表現する候補から除外する。

10

【0167】

続いて、長い間PGCの初期マーカーとして使用されてきた組織非特異的アルカリホスファターゼ(TNAP)の発現(Ginsburg, M., Snow, M.H.L., および McLaren, A., 1990)を検査する。プライマーの対を設計し(5':CCC AAA GCA CCT TAT TTT TCT ACC、3':TTG GCG AGT CTC TG C AAT TGG)、上と同じPCR反応を実施する。26個の試料のうち、22個の試料がTNAPに陽性であると判断した。切片にした胚のアルカリホスファターゼ染色から、PGCを囲む体細胞もある程度の量のTNAPを発現するが、その発現レベルはPGCより少し低い。したがって、これら22個の陽性試料のうちには、PGCになる運命の細胞のみならず、依然として体細胞になる運命の細胞もあるはずである。

【0168】

全能性PGCで発現されるが体細胞では発現されないことで知られている遺伝子の1つはOct4である(Yoem, Y.H., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., および Scholer, H.R., 1996)。この段階でPGCを体細胞から区別するためのマーカーとしてOct4が使用できる可能性を調査するために、Oct4発現を22個の試料でPCRによって確認した(5':CAC TCT ACT CAG TCC CTT TTC、3':TGT GTC CCA GTC TTT ATT TAA G)。22個の試料すべてが比較可能なレベルでOct4を発現し、この段階では体細胞も依然として活発にOct4のRNAを転写していることが示された。

20

【0169】

TNAPの発現量を、22個の試料でサザンプロット分析によって定量した(逆ノーザンプロット分析)。単一ES細胞cDNAを増幅することで確認される、単一細胞法のかなり表現的な増幅が与えられれば、サザンプロットにより最初の単一細胞で発現される遺伝子の量の半定量的測定が可能になるが、細胞のアイデンティティの完全な指標としては役立たない。しかし、このTNAP分析の結果、22個の試料中10個が、同じレベルの比較的強いバンドを示したが、残りの12個の試料はより弱いシグナルを示した。これらの結果は、これら22個の試料を少なくとも2つのグループ、すなわちより強いTNAP発現(したがって推定上のPGC由来)のグループと、より弱いTNAPのグループに分けることができることを示す。

30

【0170】

PGCを囲む体細胞がHoxb1を発現し始め、PGCが発現しない可能性(Kirstie Lawson博士からの個人的な情報)も調査した。プライマーの対を設計し(5':AAC TCA TCA GAG GTC GAA GGA、3':CGG TGC TAT TGT AAG GTC TGC)、上と同じPCR反応を実施した。試験した22個の試料のうち、12個が陽性であり、より重要なことに、これら12個の試料は、サザンプロット分析でより弱いTNAPシグナルを示すものと完全に一致した。

40

【0171】

これらの結果すべてを考慮して、83個の試料のうち、Oct4(+), TNAP(++), BMP4(-)、およびHoxb1(-)である10個の試料がPGC起源であると結論付けられた。PGCの創始集団の数が40個であり、破片中の細胞の数が250から300個であったことを考慮すると、この比(10/83)は妥当である。

【0172】

実施例 3

単一細胞cDNAライブラリのディファレンシャルスクリーニング

50

cDNAの増幅効率はチューブ毎に異なるので、ライブラリの構築のためには最も効率的に増幅されたcDNAの試料を選択することが非常に重要である。10個のPGC候補試料で、6つの異なる遺伝子(リボソームタンパク質S12、中間径フィラメントタンパク質ビメンチン、 チューブリン-5、 アクチン、 Oct4、 E-カドヘリン)の増幅をサザンブロット分析によって調査した。これら6つの遺伝子すべての全体的な増幅プロファイルから判断して、3つのcDNA調整物をライブラリの構築のために選択する。

【0173】

最大量の二本鎖cDNAを得るために、上に記載のPCR緩衝液100 μ l(1 μ lのAmplitaqを含む)中、5 μ lの細胞cDNAを用いて以下のスケジュールに従って伸長ステップを実施する。すなわち、94 $^{\circ}$ Cで5分間、42 $^{\circ}$ Cで5分間、72 $^{\circ}$ Cで30分間で行う。この溶液をフェノール/クロロホルム処理によって抽出し、増幅したcDNAをエタノールで沈殿させ、TEに懸濁させ、EcoRIを用いて完全に消化した。PCRプライマーおよび過剰量のdNTPをQIAGEN PCR Purification Kitによって取り除き、精製した全cDNAを2%の低融点アガロースゲル上に流した。500bpより大きなcDNAを切り出し、QIAGEN Gel Purification Kitで精製した。精製したcDNAsをエタノールで沈殿させ、TEに懸濁させ、ZAP IIベクターのアーム内にライゲーションさせた。ライゲーションさせたベクターをパッケージし、力価を測定し(titered)、20個の溶菌斑からのT3およびT7プライマーで挿入物を増幅することによってライゲーションが成功したクローンの比をモニターした。95%を超えるファージが挿入物を含んでいることが分かった。

10

【0174】

5000個の溶菌斑をスクリーニングすることで3つの遺伝子、リボソームタンパク質S12、チューブリン-5、Oct4の表現を定量し、3つ(S12 0.62%、チューブリン0.4%、Oct4 0.5%)のうち最高品質のものライブラリをディファレンシャルスクリーニングに使用した。PGCプローブの比較パートナーとして、最も効率的に増幅される周囲体細胞cDNAの1つ(Oct4(+)、TNAP(+/-)、BMP(-)、Hoxb1(+))を同様のサザンブロット分析によって選ぶ。

20

【0175】

このライブラリを、15cmの皿に1000個の溶菌斑の密度で植えて大きな溶菌斑(直径2mm)を得、AmershamのHybond N+フィルターを用いて2つの複製写し(duplicate lift)を取った。1%ウシ血清アルブミンおよび4%SDSを含む0.5Mリン酸ナトリウム緩衝液(pH7.3)中、65 $^{\circ}$ Cでフィルターをプレハイブリダイズした。本発明者らは、冷dCTPを存在させずに100 μ Ciの新しく入手した³²PdCTPを用いて、1 μ lの最初の細胞cDNAをAL1プライマーと共に50 μ lの全反応物に入れて10サイクル増幅することによって細胞cDNAプローブを調製し、その後Amersham Nick(商標)Spin Columnを使用して精製した。フィルターを少なくとも16時間 1.0×10^7 cpm/mlでハイブリダイズさせた(第1のフィルターを体細胞プローブとハイブリダイズさせ、第2のフィルターをPGCプローブとハイブリダイズさせた)。ハイブリダイゼーション後、フィルターを65 $^{\circ}$ C、0.5 \times SSC、0.5%SDSで3回洗浄し、適切なシグナルが得られるまでX線フィルムに暴露させた(通常1から2日間)。

30

【0176】

2つの複製フィルターの陽性溶菌斑を非常に注意深く比較する。スクリーニングした5000個の溶菌斑のうち、示差的に発現された遺伝子を表現する候補として280個を選ぶ。280個の溶菌斑すべての挿入物を、T3およびT7プライマーで増幅して1.5%アガロースゲル上に流し、ダブルサンドイッチサザンブロットを行う。スクリーニングと同じ条件を使用して、各膜をそれぞれPGCプローブと体細胞プローブとハイブリダイズさせる。280個のうち38個のクローンが遺伝子の示差的発現として選択される。次に、これらのクローンを第2のPGCプローブおよび体細胞cDNAプローブとハイブリダイズさせ、その結果、38個のうち20個のクローンがいずれのPGC cDNAでも共通であったが、いずれの体細胞cDNAでも含まれないまたはより少量であった。20個のクローンすべての配列を決定した。

40

【0177】

PGCの初期集団に高度に特異的な遺伝子

20個のクローンは11個の異なる遺伝子を表現する(2つのクローンが2回現れ、1つのクロー

50

ンが3回現れ、1つのクローンが6回現れる)。さらにストリンジェントに発現の特異性を確認するために、これら11個のクローンのためにプライマーの対を設計し、PCRによってその発現を10個の異なる単一PGC候補cDNAおよび10個の異なる単一体細胞cDNA中で確認する。このうち2個がPGC cDNAに高度に特異的な発現を示す。

【0178】

第1の遺伝子GCR1(生殖細胞制限-1、Fragilis)は、予測分子量15.0kDの137アミノ酸のタンパク質をコードする。マウスFragilisのヌクレオチドおよびアミノ酸配列を図1に示す。

【0179】

EMBLプログラムPredictProteinで一番合うモデルでは、N末端とC末端のいずれもが外に位置する2つの膜貫通型ドメインが予測される。BLASP検索により、Fragilisがインターフェロン誘導可能タンパク質ファミリーの新規なメンバーであることが分かった。典型的なメンバーの1つであるヒト9-27(Leu-13抗原と同一)は、白血球および内皮細胞内でインターフェロンによって誘導可能であり、抗増殖性およびホモタイプ付着シグナルに關与する多量体の複合体の構成成分として細胞表面に位置する(Deblandre、1995)。BLASTN検索により、胚と成体のいずれもからの様々な組織由来のEST内でFragilis配列が見つかることが分かり、これは、様々な発生学および細胞生物学の状況においてFragilisが共通の役割を果たすことを示唆する。データベースの検索により、未知機能のラットインターフェロン誘導可能タンパク質(sp:INIB RAT、pir:JC1241)との配列一致が明らかになった。本発明者らのスクリーン上にGCR1配列が6回現れ、これはPGC内の高発現レベルを示唆する。

【0180】

第2の遺伝子GCR2(Stella)は、18kDの150アミノ酸のタンパク質をコードする。マウスFragilisのヌクレオチドおよびアミノ酸配列を図2に示す。

【0181】

これは、既知のどのタンパク質とも配列相同性を有さず、いくつかの核移行共通配列を含み、強度に塩基性のpI(pI=9.67、塩基性残基の含有率=23.3%)であり、潜在的なDNA親和性が示唆される。さらに、潜在的な核外シグナルが同定され、Stellaが核と細胞質の間を往復する可能性が示唆された。BLASTN分析により、Stella配列は移植前の胚および生殖系(新生卵巢、メス12.5中腎や生殖腺など)ESTのみで見つかり、全能性細胞および多能性細胞内で主に発現されることが示唆される。興味深いことに、StellaはそのN末端内に、SAPモチーフとある程度の配列類似性を有する変調ドメインを含む。このモチーフは、染色体の組織化に關与する推定DNA結合ドメインである。さらに、SMARTプログラムにより、スプライシング因子のモチーフ様構造がC末端内に存在することがわかった。これらの発見は、Stellaが染色体の組織化およびRNAプロセッシングに關与している可能性を示唆する。

【0182】

実施例 4

GCR1およびGCR2の発現をスクリーニングすることによるPGCの同定

BMP4、TNAP、Hoxb1、Oct4の分析によって実施例2でPGCを同定したが、いずれも、これらの遺伝子単一ではPGC状態のマーカーとして捉えることができなかった。しかし、GCR1およびGCR2はどちらも、このように使用することができる。

【0183】

GCR1の発現を検査した。プライマーの対を設計し(5':CTACTCCGTGAAGTCTAGG、3':AATGAGTTTACACCTGCGTG)、上と同じPCR反応を行った。GCR1の発現を生殖細胞の、反応能を有する細胞(competent cell)内で検出した。GCR1の発現を示しているこの細胞集団から、最終的なPGCを収集した。

【0184】

GCR2の発現の境界は特によく定義されており、この発現は実質的にPGCに限定されている。したがって、GCR2はPGCの選択のための陽性マーカーとして使用される。GCR2のC末端部分を増幅するためにプライマーの対を設計する(5':GCCATTCAGATGTCTCTGCAC、3':CTCACAGCTTGAGGCTTCTAA)。実施例1のPGCから得たcDNA溶液0.5μlを鋳型として用いて、以下のスケジュールに従ってPCR増幅を行った。すなわち、95℃で1分間、55℃で1分間、72℃で1分間

10

20

30

40

50

を20サイクルで行った。試験した83個の試料のうち、PGCから取ったものだけがGCR2の発現を示した。したがって、GCR2はPGC運命の陽性マーカーである。

【0185】

多能性細胞を検出するために、GCR1およびGCR2に対する抗体を同様に使用することができる。好ましくは、GCR1に対する抗体は生殖細胞の、反応能を有する細胞を検出するのに使用し、GCR2に対する抗体はPGCを検出するのに使用する。

【0186】

したがって、GCR1およびGCR2のいずれもが、PGCを肯定的に同定するのに使用できるPGC運命の陽性マーカーである。

【0187】

ISHによるPGCの同定

2つの遺伝子の *in vivo* 発現を *in situ* ハイブリダイゼーションで検査する。GCR1の発現は、E6.0からE6.5(プレストリーク段階)に原外胚葉全体で非常に弱く開始され、原外胚葉近位縁の数細胞層で強くなる。胚外の外胚葉で発現されるBMP4は、生殖細胞反応能の誘導およびGCR1の発現に重要なシグナル分子の1つである。インターフェロンなど他のシグナルがGCR1の誘導に関与している可能性が高い。初期/中期ストリーク段階で、発達中の原始ストリークの近位-後部末端で発現がより強くなり、後期ストリーク段階以降、この位置で非常に強くなる。発現は初期頭褶段階まで持続し、最終的には徐々に消える。E8.5の遊走PGCで発現は検出されない。

【0188】

GCR2の発現は、中期/後期ストリーク段階で、発達中の原始ストリークの近位-後部末端で開始され、それ以降の段階、同じ位置で徐々に強くなる。この発現は特異的であり、ドットプロット様式で染色した個々の単一細胞は、PGCが細胞のクラスターとして分化し始めると考えられる領域で見られる。後期芽/初期頭褶段階では、最初のクラスターから遊走していると考えられる一部の細胞も、クラスター内の細胞と同様によく染まる。E8.5およびE9.5では、遊走PGCと考えられる細胞群が非常に特異的に染まる。

【0189】

これらの結果から、GCR2がGCR1の後に開始されてPGC運命を固定するとき、GCR1は細胞系譜特異化および生殖細胞反応能のプロセス中、およびその後のPGCのプロセス中に上方制御される遺伝子であることが結論付けられる。

【0190】

したがって、GCR1の発現は、細胞系譜特異化および/または生殖細胞反応能など多能性を検出する方法で検出することができる。同様に、GCR2の発現は、細胞運命の決定、たとえば始原生殖細胞としての運命の決定を検出するために検出することができる。

【0191】

実施例5

生殖系の発達中におけるFragilisおよびStellaの発現

StellaおよびFragilisに対する抗体を使用して、初期胚中のこれら遺伝子の発現を検出する。これらの遺伝子それぞれが始原生殖細胞中で発現されることが分かった。具体的には、本発明者らは、生殖細胞の割当の時点でPGCの、反応能を有する細胞を標識する最初の遺伝子がFragilisであることを見出した。Stellaは系譜限定創始PGC内でのみ発現され、その後は生殖細胞系内で発現される。

【0192】

図3は、胚性幹(ES)細胞中のFragilisの発現を示す。

【0193】

Fragilisは、多能性ES細胞およびEG細胞内で発現される。PGCからEG細胞が誘導される間に、EG細胞でFragilisの発現が再出現することが分かった。これら細胞の特異化が完了したあとは、後期PGCはFragilis陰性である。

【0194】

図5は、E7.2マウス胚の全載 *in situ* ハイブリダイゼーションによって検出されるFragilis

10

20

30

40

50

の発現を示す。

【0195】

創始PGC集団がE7.25の胚内で分化する初発性尿膜の基部で、強いFragilis発現がある。Fragilisの発現はE7.5まで持続したが、E8.5の遊走PGC中では検出されなかった。Fragilisはまず、生殖細胞の、反応能を有する近位原外胚葉中で検出される。Fragilisの発現は、BMP4源である組織外胚の外胚葉組織と混合した場合に原外胚葉細胞中で誘導される。BMP4変異マウスではFragilisの発現はなく、これは、これらの胚中でPGCが存在しないことと矛盾していない(Lawson他、1999)。

【0196】

図4は、PGC中のStellaの発現を示す。

10

【0197】

PGC内の強いStellaの発現は、EG細胞内で下方制御される。ES細胞内でもStellaの発現レベルが低い。StellaおよびFragilisは、ES細胞およびEG細胞内でノーザンブロット分析によって検出することができる。Stellaはまず、E7.0の細胞系譜限定PGC特有クラスター内の単一細胞内で検出され、その後は遊走PGC内、次いでこれらが生殖腺に入るときに検出される。図7は、E9.0胚の、生殖腺への遊走プロセス中のPGCでのStella発現を示す。Stellaは、PGCの創始集団の決定的なマーカーとして現在知られている唯一の遺伝子である。

【0198】

図6は、E7.2マウス胚の全載in situハイブリダイゼーションによって検出したStellaの発現を示す。

20

【0199】

図8。単一細胞cDNAのPCR分析によって検出した、単一細胞のFragilisおよびStellaの発現を示す。Stellaの発現を示している単一細胞に比べて、Fragilisの発現を示している単一細胞の方が多くに注目せよ。Fragilis発現が最も高いレベルである細胞のみがStellaを発現し、生殖細胞の運命を獲得することが分かった。Stellaを発現する細胞は、Hoxb1の発現を示さないことが分かった。より低いレベルのFragilisを発現してStellaを発現しない細胞は、体細胞となってHoxb1の発現を示す。PGCの創始集団も高レベルのTnapを示す。創始PGCおよび体細胞のどちらもOct4、T(短尾)、およびFgf8の発現を示す。

【0200】

実施例 6

30

個別細胞中のFragilisおよびStellaの発現

StellaおよびFragilisの細胞内局在化も決定した。Fragilisは、ゴルジ体の単一細胞質スポット内のみならず原形質膜内にも局在化する。Stellaは推定核移行シグナルおよび核外移行シグナルを含み、細胞質と核のどちらもに局在化している。

【0201】

Fragilisは、ゴルジ体のみならずPGCの原形質膜内で観察される。Fragilisの細胞表面局在化は、インターフェロン誘導性遺伝子ファミリーのメンバーとして予測されている(Deblandre, 1995)。原外胚葉近位縁でのFragilisの発現は、生殖細胞反応能の開始を示す。Fragilisは、そのエキソン1の上流にIFN応答要素を有しているため、近位の原外胚葉細胞のBMP4による開始初回刺激の後にIFNによって誘導される可能性が高い。これらのIFN誘導性タンパク質は、抗増殖性シグナルの伝達が可能なTAPA1など他のタンパクと共に多量体の複合体を形成することができる。これは、体細胞は急速に分割し続けるのに対して、創始PGCの細胞サイクル時間が6から16時間増加する理由である可能性がある。

40

【0202】

推定核移行シグナルおよび核外移行シグナルを有するStellaは、細胞質および核のどちらもで観察された。Stellaの開始に続き、E8.5ではFragilisの発現が損失される。したがって、Fragilisの発現は生殖細胞反応能の開始を示し、Stellaの発現は特異化プロセスの終了を示す。創始PGC中のStellaの発現は、体細胞運命の回避を示し、その多能性状態と矛盾しない。これらの研究は、特定の遺伝子セットが、普通なら体細胞になる可能性がある細胞に生殖系運命を課さなければならないことを示す。多くの生物が、生殖細胞が体細胞

50

になるのを防ぐ精巧な転写機構を有するので、Stellaは、その核と細胞質間を往復する潜在性により、転写および翻訳の制御の一役を担っている可能性がある。卵母細胞および移植前の胚中でのStellaの発現は、これが全能性および多分化能性においてより幅広い役割を担っていることを示す。

【0203】

実施例7

FragilisとStellaの関連性

Fragilisを発現する細胞の一部のみが最後にStellaの発現を示した。Fragilisの発現レベルが最も高い細胞のみがPGCとなり、Stellaを発現し始めた。さらに、Stella陽性PGCは決してHoxb1の発現を示さない。より重要なことに、Fragilisの発現レベルがより低い体細胞のみがHoxb1発現を示す。さらに、体細胞のみがホメオボックスを含む2つの他の遺伝子Lim1およびEvx-1の発現を示す。したがって、Hoxb1、Evx-1およびLim1の発現を損失することが、生殖細胞運命の特異化に重要と考えられる。

10

【0204】

図8aおよび8bは、PCR分析による、単一細胞PGCおよび体細胞中の様々な遺伝子の発現を示す。

【0205】

本実験はまた、Oct4はPGCの決定的なマーカーではないことを示す。これまでに、全能性細胞および多能性細胞でOct4の発現が実証されている(Nichols、199、Pesce、1998;Yeom、1996)。しかし、本発明者らは、Oct4はすべてのPGCおよび体細胞で同程度発現されていることを見出した。しかし、本発明者らは、PGC中にT(短尾)およびFgf8の発現を見出し、これは、PGCが最初中胚葉細胞になる運命であった胚性細胞の中から収集されることを示す。

20

【0206】

実施例8

PGCの特異化

創始PGCおよびその近隣の体細胞は、近位原外胚葉細胞由来の共通起源を共有する。創始PGCおよびその近隣の体細胞を分析することにより、生殖細胞運命の特異化に重大な遺伝子の系統的なスクリーンが確立された。Fragilisは、推定生殖細胞をその近隣の体細胞から区別するために生殖細胞反応能および同型の会合を促進することができるインターフェロン(IFN)誘導性遺伝子であり、このような例は発生中の他の状況に適用される場合もある。Stellaの発現は、Fragilisの発現が高い細胞で起こる。ひとたび生殖細胞の特異化が完了するとFragilisはもはや必要ないが、生殖細胞系でStellaの発現は続く。Stellaは卵母細胞内および初期の移植前の発生胚内でも発現されるので、これは全能性/多能性細胞全体で重要である可能性もある。

30

【0207】

実施例9

生殖細胞系および多能性幹細胞

多能性胚性(EG)細胞を誘導するのにPGCを使用することができる。しかし、EG細胞と異なると、未分化胚芽細胞に導入された場合は、PGCは発生に関与しない。これらは、シグナル分子に応答できないか、あるいは転写的に抑制される。ひとたび特異化されると、PGCはその細胞表面にFragilisを発現しない。しかし、EG細胞はその細胞表面上にはっきりとFragilisの発現を示し、ES細胞もそうである。EG細胞とES細胞のどちらもがノーザン分析で判断するとStellaを発現するが、StellaはPGCに比べてより低いレベルで、ES細胞やEG細胞で発現される。したがって、FragilisおよびStellaは多能性幹細胞中で、ある役割を担っている。したがって、これらの遺伝子はこれら多能性幹細胞のマーカーであり、これら幹細胞に多能性を与えるのに役割を担っている可能性もある。

40

【0208】

[参考文献]

Brady, G. and Iscove, N.N. (1993). Construction of cDNA libraries from single cells. *Methods Enzymol.* 225, 611-623.

Dulac, C. and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* 83, 195-206.

Ginsburg, M., Snow, M.H.L., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* 110, 521-528.

10

Downs, K.M., and Davies, T. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* 118, 1255-1266.

Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., and Hogan, B.L.M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes&Dev.* 13, 424-436.

Yoem, Y.II., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 1996, 881-894. 10

1. Weismann, A. Das Keimplasma. Eine theorie der Vereburg. *Jenna. Gustav Fischer* (1892).

2. Eddy, E. M. Germ plasm and the differentiation of the germ cell line. *Int Rev Cytol* 43, 229-80 (1975).

3. Seydoux, G. & Strome, S. Launching the germline in *Caenorhabditis elegans*: regulation of gene expression in early germ cells. *Development* 126, 3275-83. (1999). 20

4. Wylie, C. Germ cells. *Cell* 96, 165-74. (1999).

5. Lawson, K. A. et al. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* 13, 424-36. (1999). 30

6. Lawson, K. A. & Hage, W. J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* 182, 68-84 (1994).

7. Tam, P. P. & Zhou, S. X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* 178, 124-32. (1996). 40

8. Yoshimizu, T., Obinata, M. & Matsui, Y. Stage-specific tissue and cell interactions play key roles in mouse germ cell specification. *Development* **128**, 481-90. (2001).

9. McLaren, A. Signaling for germ cells. *Genes Dev* **13**, 373-6. (1999).

10. Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol Endocrinol* **14**, 1053-63. (2000). 10

11. Ying, Y., Qi, X. & Zhao, G. Q. Induction of primordial germ cells from murine epiblasts by synergistic action of BMP4 and BMP8B signaling pathways. *Proc Natl Acad Sci U S A* **98**, 7858-7862. (2001). 20

12. Ying, Y. & Zhao, G. Q. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev Biol* **232**, 484-92. (2001).

13. Chiquoine, A. D. The identification, origin and migration of the primordial germ cells in the mouse embryo. *Anat Rec* **118**, 135-146 (1954). 30

14. Ginsburg, M., Snow, M. H. & McLaren, A. Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521-8. (1990).

15. MacGregor, G. R., Zambrowicz, B. P. & Soriano, P. Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. *Development* **121**, 1487-96. (1995). 40

16. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-91. (1998).

17. Pesce, M., Gross, M. K. & Scholer, H. R. In line with our ancestors: Oct-4 and the mammalian germ. *Bioessays* **20**, 722-32. (1998).
18. Yeom, Y. I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-94. (1996).
19. Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255-66. (1993). 10
20. Brady, G. & Iscove, N. N. Construction of cDNA libraries from single cells. *Methods Enzymol* **225**, 611-23 (1993).
21. Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195-206. (1995). 20
22. Frohman, M. A., Boyle, M. & Martin, G. R. Isolation of the mouse Hox-2.9 gene; analysis of embryonic expression suggests that positional information along the anterior-posterior axis is specified by mesoderm. *Development* **110**, 589-607. (1990).
23. Deblandre, G. A. et al. Expression cloning of an interferon-inducible 17-kDa membrane protein implicated in the control of cell growth. *J Biol Chem* **270**, 23860-6. (1995). 30
24. Friedman, R. L., Manly, S. P., McMahon, M., Kerr, I. M. & Stark, G. R. Transcriptional and posttranscriptional regulation of interferon- induced gene expression in human cells. *Cell* **38**, 745-55. (1984). 40
25. Evans, S. S., Collea, R. P., Leasure, J. A. & Lee, D. B. IFN-alpha induces homotypic adhesion and Leu-13 expression in human B lymphoid cells. *J Immunol* **150**, 736-47. (1993).

26. Evans, S. S., Lee, D. B., Han, T., Tomasi, T. B. & Evans, R. L. Monoclonal antibody to the interferon-inducible protein Leu-13 triggers aggregation and inhibits proliferation of leukemic B cells. *Blood* **76**, 2583-93. (1990).

27. Aravind, L. & Koonin, E. V. SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**, 112-4. (2000).

10

28. Gurdon, J. B., Lemaire, P. & Kato, K. Community effects and related phenomena in development. *Cell* **75**, 831-4. (1993).

29. Reid, L. E. et al. A single DNA response element can confer inducibility by both alpha- and gamma-interferons. *Proc Natl Acad Sci USA* **86**, 840-4. (1989).

30. Kita, M. et al. [Expression of cytokines and interferon-related genes in the mouse embryo]. *C R Seances Soc Biol Fil* **188**, 593-600 (1994).

20

31. Gomperts, M., Garcia-Castro, M., Wylie, C. & Heasman, J. Interactions between primordial germ cells play a role in their migration in mouse embryos. *Development* **120**, 135-41. (1994).

32. Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-22. (1990).

30

33. Herrmann, B. G. Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* **113**, 913-17

34. Crossley, P. H. & Martin, G. R. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439-51. (1995).

40

35. Barnes, J. D., Crosby, J. L., Jones, C. M., Wright, C. V. & Hogan, B. L. Embryonic expression of Lim-1, the mouse homolog of *Xenopus* Xlim-1, suggests a role in lateral mesoderm differentiation and neurogenesis. *Dev Biol* **161**, 168-78. (1994).
36. Fujii, T. et al. Expression patterns of the murine LIM class homeobox gene *lim1* in the developing brain and excretory system. *Dev Dyn* **199**, 73-83. (1994). 10
37. Bastian, H. & Gruss, P. A murine even-skipped homologue, *Evx 1*, is expressed during early embryogenesis and neurogenesis in a biphasic manner. *Embo J* **9**, 1839-52. (1990).
38. Rogers, M. B., Hosler, B. A. & Gudas, L. J. Specific expression of a retinoic acid-regulated, zinc-finger gene, *Rex- 1*, in preimplantation embryos, trophoblast and spermatocytes. *Development* **113**, 815-24. (1991). 20
39. Sutton, J. et al. Genesis, a winged helix transcriptional repressor with expression restricted to embryonic stem cells. *J Biol Chem* **271**, 23126-33. (1996).
40. Cox, D. N. et al. A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev* **12**, 3715-27. (1998). 30
41. Fujiwara, Y. et al. Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila* vasa and its specific expression in germ cell lineage. *Proc Natl Acad Sci USA* **91**, 12258-62. (1994).
42. Dixon, K. E. Evolutionary aspects of primordial germ cell formation. *Ciba Found Symp* **182**, 92-110 (1994). 40
43. Mahowald, A. P. Assembly of the *Drosophila* germ plasm. *Int Rev Cytol* **203**, 187-213 (2001).

44. Nieuwkoop, P. D. & Satasurya, L. A. Primordial germ cells in the chordates. *Cambridge University Press, Cambridge, UK* (1979).
45. Johnson, A. D., Bachvarova, R. F., Drum, M. & Masi, T. Expression of axolotl *dazl* rna, a marker of germ plasm: widespread maternal rna and onset of expression in germ cells approaching the gonad. *Dev Biol* **234**, 402-15. (2001). 10
46. Johnson, A. D., Bachvarova, R. F., Masi, T. & Drum, M. Expression of *Vasa* and *Daz-like* genes demonstrate that Axolotl primordial germ cells (PGCs) are not predetermined. *Germ cells Cold Spring harbor laboratory*, 61 (2000).
47. Toyooka, Y. et al. Expression and intracellular localization of mouse *Vasa*-homologue protein during germ cell development. *Mech Dev* **93**, 139-49. (2000). 20
48. Saitou, M. et al. Occludin-deficient embryonic stem cells can differentiate into polarized epithelial cells bearing tight junctions. *J Cell Biol* **141**, 397-408. (1998).
49. Henrique, D. et al. Expression of a Delta homologue in prospective neurons in the chick. *Nature* **375**, 787-90. (1995).
50. Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol* **225**, 361-73 (1993). 30
51. Winnier, G., Blessing, M., Labosky, P. A. & Hogan, B. L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev* **9**, 2105-16. (1995).

【 0 2 0 9 】

本文書中で言及する各出願や特許、各出願や特許の実施中を含めて上記の各出願や特許中で引用または参照された各文書(「出願に引用された文書」)、各出願や特許および出願に引用された文書すべてで引用または言及された製造者の指示書やカタログすべてが、本明細書に参照により組み込まれる。さらに、本文中で引用したすべての文書、本文中で引用した文書中で引用または参照された文書すべて、本文中で引用または言及した製品すべての製造者の指示書やカタログすべてが、本明細書に参照により組み込まれる。 40

【 0 2 1 0 】

本発明の範囲および精神から逸脱することなしに、本発明に記載した方法およびシステムの様々な改変や変化が当業者には明らかであろう。本発明を具体的な好ましい実施形態と関連させて記載したが、請求する本発明はこのような具体的な実施形態だけには過度に限定さるべきでないことを理解されたい。実際、本発明を実施するために記載した様式の、 50

分子生物学または関連分野の当業者に明らかな改変は、特許請求の範囲内であることを意図する。

【図面の簡単な説明】

【0211】

【図1】Fragilisのヌクレオチド配列および推定アミノ酸配列を示す図である。2つの膜貫通型ドメイン(TM IおよびTM II)の予測位置に下線を引き、太字で表示した。ポリ(A)シグナルに下線を引いた。

【図2】Stellaのヌクレオチド配列および推定アミノ酸配列を示す図である。3つの核移行シグナルに下線を引いた。潜在的な核外移行シグナルに二重下線を引き、疎水性残基を太字で示した。SAPドメインに類似性のあるモチーフ内のらせん構造(アミノ酸28からアミノ酸63)に赤の下線を引き、保存された残基を青の下線で示した。スプライシング因子様のモチーフに下線を引き、保存された残基を緑で示した。ポリ(A)シグナルにも下線を引いた。

【図3】胚性幹(ES)細胞内でのFragilisの発現を示す図である。ES細胞は、PBS中4%のパラホルムアルデヒドで10分間、室温で固定し、Saitou他、1998、J Cell Biol、141、397-408(1998)に記載のように免疫組織化学用に処理した。生殖細胞の、反応能を有する細胞であるE6.5の近位胚盤葉上層細胞内、および新しく特異化された生殖細胞内で、Fragilisの発現を同様に検出した。この発現は、生殖細胞運命の特異化が完了した後のE8.5で減少する。

【図4】PGC内のStellaの発現を示す図である。E12.5の生殖腺をPBS中4%のパラホルムアルデヒドで10分間、室温で固定し、Saitou他、1998、J Cell Biol、141、397-408(1998)に記載のように免疫組織化学用に処理した。StellaはE7.25からE13.5のPGC、多能性ES細胞およびEG細胞内で検出された。Stellaはまた、全能性卵母細胞、接合子、ならびに、移植前発生および発生中の配偶子内の発生中の全能性および多能性割球でも検出される。EG細胞がPGC由来の場合(Labosky他、1994、Development、120:3197-3204)、Fragilisの発現はここでもまた、ES細胞と同様に多能性EG細胞中で検出される。したがって、FragilisおよびStellaは多能性幹細胞のマーカーでもある。

【図5】E7.2マウス胚の全載in situハイブリダイゼーションによるFragilisの発現を示す図である。

【図6】E7.2マウス胚の全載in situハイブリダイゼーションによるStellaの発現を示す図である。

【図7】E9.0胚における、生殖腺内に遊走中のPGC中のStellaの発現を示す図である。

【図8A】単一細胞cDNAのPCR分析によって検出された単一細胞中のFragilisおよびStellaの発現を示す図である。図8B中に記号*で示した数値はPGCである。Stellaの発現を示す単一細胞よりFragilisの発現を示している単一細胞の方が多くに注目されたい。最も高いレベルでFragilisを発現する細胞のみがStellaを発現して生殖細胞の運命を獲得することが分かった。Stellaを発現する細胞はHoxb1の発現を示さないことが分かった。より低いレベルでFragilisを発現しStellaを発現しない細胞は体細胞となり、Hoxb1の発現を示した。PGCの創始集団も高レベルのTnapを示す。創始PGCと体細胞のいずれもがOct4、T(短尾)、およびFgf8の発現を示す。

【図8B】単一細胞cDNAのPCR分析によって検出された単一細胞中のFragilisおよびStellaの発現を示す図である。図8B中に記号*で示した数値はPGCである。Stellaの発現を示す単一細胞よりFragilisの発現を示している単一細胞の方が多くに注目されたい。最も高いレベルでFragilisを発現する細胞のみがStellaを発現して生殖細胞の運命を獲得することが分かった。Stellaを発現する細胞はHoxb1の発現を示さないことが分かった。より低いレベルでFragilisを発現しStellaを発現しない細胞は体細胞となり、Hoxb1の発現を示した。PGCの創始集団も高レベルのTnapを示す。創始PGCと体細胞のいずれもがOct4、T(短尾)、およびFgf8の発現を示す。

10

20

30

40

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
25 July 2002 (25.07.2002)

PCT

(10) International Publication Number
WO 02/057307 A2

- (51) International Patent Classification: C07K 14/705
- (21) International Application Number: PCT/GB02/00215
- (22) International Filing Date: 18 January 2002 (18.01.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 0101300.2 18 January 2001 (18.01.2001) GB
- (71) Applicant (for all designated States except US): CAMBRIDGE UNIVERSITY TECHNICAL SERVICES LIMITED [GB/GB]; The Old Schools, Trinity Lane, Cambridge CB2 1TS (GB).
- (81) Designated States (national): AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SI, TR), OAPI patent (BE, BI, CE, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
of inventorship (Rule 4.17(iv)) for US only

Published:
without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

- (72) Inventors: and
- (75) Inventors/Applicants (for US only): SAITOU, Mitsuori [JP/JP]; Wellcome CRC Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR (GB); SURANI, Azim [GB/GB]; Wellcome CRC Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR (GB).

- (74) Agents: MASCHIO, Antonio et al.; D Young & Co, 21 New Fetter Lane, London EC4A 1DA (GB).

(54) Title: GENES

WO 02/057307 A2



- (57) Abstract: The invention provides two primordial germ cell-specifically expressed genes, GCR1 (Fragilis) and GCR2 (Stella), which are markers for primordial germ cells and may be used to identify such cells in cell populations.

WO 02/057307

PCT/GB02/00215

1

GENES**FIELD**

The present invention relates to the fields of development, molecular biology and genetics. More particularly, the invention relates to genes which are expressed exclusively
5 in the earliest populations of primordial germ cells (PGCs) and the use of such genes and the products thereof in identification of pluripotent and multipotent cells such as PGCs, pluripotent embryonic stem cells (ES) and pluripotent embryonic germ cells (EG), in cell populations. They are also markers for a change in the state of cells from being non pluripotent to becoming pluripotent, and in being able to confer this state on a non
10 pluripotent cell.

INTRODUCTION

Post fertilisation, the early mammalian embryo undergoes four rounds of cleavage to form a morula of 16 cells. These cells, following further rounds of division, develop into a blastocyst in which the cells can be divided into two distinct regions; the inner cell
15 mass, which will form the embryo, and the trophectoderm, which will form extra-embryonic tissue, such as the placenta.

The cells that form part of the embryo up until the formation of the blastocyst are totipotent; in other words, each of the cells has the ability to give rise to a complete individual embryo, and to all the extra-embryonic tissues required for its development.
20 After blastocyst formation, the cells of the inner cell mass are no longer totipotent, but are pluripotent, in that they can give rise to a range of different tissues. A known marker for such cells is the expression of the enzyme alkaline phosphatase and *Oct4*.

Primordial germ cells (PGCs) are pluripotent cells that have the ability to differentiate into all three primary germ layers. In mammals, the PGCs migrate from the base of the allantois, through the hindgut epithelium and dorsal mesentery, to colonise the
25 gonadal anlage. The PGC-derived cells have a characteristically low cytoplasm/nucleus

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

2

ratio, usually with prominent nucleoli. PGCs may be isolated from the embryos by removing the genital ridge of the embryo, dissociating the PGCs from the gonadal anlage, and collecting the PGCs. The earliest PGC population is reported to consist of a cluster of some 45 (forty-five) alkaline phosphatase positive cells, found at the base of the emerging allantois, 7.25 days post-fertilisation (Ginsburg *et al.*, (1990) Development 110:521-528).

PGCs have many applications in modern biotechnology and molecular biology. They are useful in the production of transgenic animals, where embryonic germ (EG) cells derived from PGCs may be used in much the same manner as embryonic stem (ES) cells (Labosky *et al.*, (1994) Development 120:3197-3204). Moreover, they are useful in the study of foetal development and the provision of pluripotent stem cells for tissue regeneration in the therapy of degenerative diseases and repopulation of damaged tissue following trauma. Above all, PGCs while having some specialised properties, retain an underlying pluripotency, which is lost from the neighbouring cells that surround the founder population of PGCs that acquire a somatic cell fate. PGCs and the surrounding somatic cells share a common ancestry. However, the founder PGCs are few in number and difficult to isolate from embryonic tissue and the surrounding somatic cells, which complicates their study and the development of techniques which make use thereof.

Little is known in the art about the expression of genes in the founder population of PGCs and the relationship between PGC-specific gene expression and the retention of pluripotency in these cells. Certain markers for PGCs are known – for example, the expression of tissue non-specific alkaline phosphatase (TNAP) has been used as a marker for early PGCs (Ginsburg *et al.*, (1990) Development 110:521-528). Oct4 is known to be expressed in PGCs, but not somatic cells (Yoem *et al.*, (1996) Development 122:881-894). Other markers, such as BMP4, are known to be expressed primarily in somatic tissues (Lawson *et al.*, (1999) Genes & Dev. 13:424-436). However, none of these genes is specific for PGCs, since they are also expressed in other tissue types. There is therefore a need in the art for the identification of genes which may be used as markers for PGCs and which may provide an insight into the biology of germ cell development and the nature of the pluripotent state.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

3

SUMMARY

We disclose the sequences of two genes which are expressed specifically in PGCs and other pluripotent cells. The sequence of the genes from mouse is set forth in **SEQ ID NO: 1** (GCR1 or Fragilis) and **SEQ ID NO: 3** (GCR2, or Stella). Corresponding amino acid sequences for mouse GCR1 and GCR2 are set out in **SEQ ID NO: 2** and **SEQ ID NO: 4** respectively. Nucleic acid sequences of rat GCR2 homologues are set out in **SEQ ID NO: 5**, **SEQ ID NO: 6**, **SEQ ID NO: 7**, **SEQ ID NO: 8**, and **SEQ ID NO: 9**.

According to a first aspect of the present invention, we provide a GCR1 polypeptide, or a fragment, homologue, variant or derivative thereof. Preferably, the polypeptide has at least 50%, 60%, 70%, 80%, 90% or 95% homology to a sequence shown in **SEQ ID NO: 2**.

There is provided, according to a second aspect of the present invention, GCR2 polypeptide, or a fragment, homologue, variant or derivative thereof. Preferably, the polypeptide has at least 50%, 60%, 70%, 80%, 90% or 95% homology to a sequence shown in **SEQ ID NO: 4**.

We provide, according to a third aspect of the present invention, a nucleic acid encoding a polypeptide according to any preceding claim.

As a fourth aspect of the present invention, there is provided a nucleic acid having at least 90% homology with the sequence set forth in **SEQ ID NO: 1**, or a fragment, variant or derivative thereof.

We provide, according to a fifth aspect of the present invention, a nucleic acid having at least 75% homology with the sequence set forth in **SEQ ID NO: 3**, **SEQ ID NO: 5**, **SEQ ID NO: 6**, **SEQ ID NO: 7**, **SEQ ID NO: 8** or **SEQ ID NO: 9**, or a fragment, variant or derivative thereof

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

4

The present invention, in a sixth aspect, provides a nucleic acid comprising a sequence of 25 contiguous nucleotides of a nucleic acid according to the third, fourth or fifth aspect of the invention.

In a seventh aspect of the present invention, there is provided a nucleic acid
5 comprising a sequence of 15 contiguous nucleotides of a nucleic acid according to the third, fourth, fifth or sixth aspect of the invention.

According to an eighth aspect of the present invention, we provide a complement of a nucleic acid sequence according to any of the third to seventh aspect of the invention.

Preferably, such a nucleic acid comprises one or more nucleotide substitutions,
10 wherein such substitutions do not alter the coding specificity of said nucleic acid as a result of the degeneracy of the genetic code.

We provide, according to a ninth aspect of the invention, a polypeptide encoded by a nucleic acid according to any preceding aspect of the invention.

Preferably, the polypeptide comprises a sequence shown in SEQ ID NO: 2 or SEQ
15 ID NO: 4.

There is provided, in accordance with a tenth aspect of the present invention, a method for identifying a pluripotent cell, comprising detecting the presence of a polypeptide according to the first, second, ninth or tenth aspect of the invention or the expression of a nucleic acid according to any of the third to eighth aspect of the invention,
20 or a homologue thereof.

Preferably, the method comprises the steps of amplifying nucleic acids from a putative pluripotent cell using 5' and 3' primers specific for GCR1 (Fragilis) and/or GCR2 (Stella), and detecting amplified nucleic acid thus produced. Preferably, the expression of the nucleic acid sequence is detected by *in situ* hybridisation.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

5

The expression of the nucleic acid sequence may be determined by detecting the protein product encoded thereby. Alternatively or in addition, the protein product may be detected by immunostaining.

As an eleventh aspect of the invention, we provide an antibody specific for a polypeptide according to the first, second, ninth or tenth aspect of the invention. preferably, the antibody is capable of specifically binding to an extracellular domain of GCR1.

We provide, according to a twelfth aspect of the invention, there is provided use of such an antibody for the identification and/ or isolation of a pluripotent cell.

We further provide, according to a thirteenth aspect of the invention, a pluripotent cell identified by a method as set out previously.

There is provided, according to a fourteenth aspect of the present invention, a method for isolating a gene specifically expressed in a pluripotent cell, comprising the steps of: (a) providing a population of cells containing a pluripotent cell; (b) isolating one or more pluripotent cells therefrom and providing single-cell pluripotent cell isolates; (c) amplifying the transcribed nucleic acid present in a single pluripotent cell; (d) conducting a subtractive hybridisation screen to identify transcripts present in pluripotent cells but not in somatic cells; and (e) probing a nucleic acid library with one or more transcripts identified in (d) to clone one or more genes which are specifically expressed in pluripotent cells.

In a highly preferred embodiment, the pluripotent cell is selected from the group consisting of: a primordial germ cell (PGC), an embryonic stem cell (ES) and an embryonic germ cell (EG). Preferably, the pluripotent cell comprises a primordial germ cell.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

6

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Nucleotide and deduced amino acid sequence of *Fragilis*. Predicted positions of the two transmembrane domains (TM I and TM II) are underlined and indicated by bold letters. The poly(A) signal is underlined.

5 Figure 2: Nucleotide and deduced amino acid sequence of *Stella*. Three nuclear localization signals are underlined. A potential nuclear export signal is underlined twice, and the hydrophobic residues are indicated in bold. Helical structures in a motif with similarity to SAP domain (a.a.28 to a.a.63) are underlined in red, and the conserved residues are indicated by blue. A splicing factor-like motif is underlined and the conserved
10 residues are indicated in green. Poly(A) signals are also underlined.

Figure 3: Expression of *Fragilis* in embryonic stem (ES) cells. ES cells are fixed in 4% paraformaldehyde in PBS for 10min. at room temperature and processed for immunohistochemistry as described by Saitou et al., (1998). *J Cell Biol* 141, 397-408. (1998). *Fragilis* expression is similarly detected in E6.5 proximal epiblast cells, which are
15 germ cell competent cells, and in newly specified germ cells. The expression declines after E8.5 following completion of the specification of germ cells fate.

Figure 4: Expression of *Stella* in PGCs. PGCs from E12.5 genital ridges are fixed in 4% paraformaldehyde in PBS for 10min. at room temperature and processed for immunohistochemistry as described by Saitou et al., (1998). *J Cell Biol* 141, 397-408.
20 (1998). *Stella* is detected in PGCs from E 7.25-13.5, as well as in pluripotent ES cells and in EG cells. *Stella* is also detected in the totipotent oocyte, zygote and in the totipotent and pluripotent blastomeres during preimplantation development and in developing gametes. When EG cells are derived from PGCs (Labosky et al., (1994) *Development* 120:3197-3204). *Fragilis* expression is again detected in the pluripotent EG cells as it is in ES cells.
25 Therefore, *Fragilis* and *Stella* are also markers for the pluripotent stem cells.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

7

Figure 5. Fragilis expression by whole-mount *in situ* hybridization in E7.2 mouse embryos.

Figure 6. Stella expression by whole mount *in situ* hybridisation in E 7.2 mouse embryos.

5 Figure 7. Stella expression in PGCs in the process of migration into the gonads in E9.0 embryos.

Figure 8a and 8b. Expression of Fragilis and Stella in single cells detected by PCR analysis of single cell cDNAs. Numbers marked by symbol* in 8b are the PGCs. Note that there are more single cells showing expression of Fragilis compared to those showing expression of Stella. Only cells with the highest levels of Fragilis expression were found to express Stella and acquire the germ cell fate. Cells that express Stella were found not to show expression of Hoxb1. Cells that express lower levels of Fragilis and no Stella become somatic cells and showed expression of Hoxb1. The founder population of PGCs also show high levels of Tnap. Both the founder PGCs and the somatic cells show
10 expression of Oct4, T(Brachyury), and Fgf8.
15

DETAILED DESCRIPTION

GCR1 (FRAGILIS) AND GCR2 (STELLA)

The disclosure provides generally for GCR1 (Fragilis) and GCR2 (Stella) nucleic acids, polypeptides, as well as fragments, homologues, variants and derivatives thereof.

20 The names "GCR1" and "Fragilis" should be understood as synonymous with each other, and likewise, "GCR2" and "Stella" should be considered synonyms. Nucleic acid and amino acid sequences of GCR1/Fragilis are set out in SEQ ID NO: 1 and 2, while nucleic acid sequences of GCR2/Stella are set out in SEQ ID NO: 3, 5, 6, 7, 8 and 9, with an amino acid sequence of GCR2/Stella shown in SEQ ID NO: 4.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

8

In preferred embodiments, however, GCR1/ Fragilis should be taken to refer to the nucleic acid sequence shown in SEQ ID NO: 1, or the amino acid sequence shown in SEQ ID NO: 2, as the context requires. Furthermore, in preferred embodiments, GCR2/ Stella should be taken to refer to the nucleic acid sequence shown in SEQ ID NO: 3, or the amino acid sequence shown in SEQ ID NO: 4, as the context requires.

GCR1 and GCR2 are PGC-specific transcripts. GCR1 is upregulated during the process of lineage commitment of PGCs, while GCR2 is upregulated after GCR1, and marks commitment to the PGC fate. The first gene, GCR1 (Germ cell restricted-1, Fragilis), encodes a 137 amino acid protein with a predicted molecular weight of 15.0kD. The best fit model of the EMBL program PredictProtein predicts two transmembrane domains, both N and C terminus ends being located outside. The BLASTP search revealed that Fragilis is a novel member of the interferon-inducible protein family. One prototype member, human 9-27 (identical to Leu-13 antigen), is inducible by interferon in leukocytes and endothelial cells, and is located at the cell surface as a component of a multimeric complex involved in the transduction of antiproliferative and homotypic adhesion signals (Deblandre, 1995). The BLASTN search revealed that the Fragilis sequence was found in ESTs derived from many different tissues both from embryos and adults, indicating that Fragilis may play a common role in different developmental and cell biological contexts. Database searches reveal a sequence match with the rat interferon-inducible protein (sp:INIB RAT, pir:JC1241) with unknown function. The GCR1 sequence appears six times in our screen, indicating high level expression in PGCs.

The second gene, GCR2, (Stella) encodes a 150 amino acid protein, of 18kD. It has no sequence homology with any known protein, contains several nuclear localisation consensus sequences and is highly basic pI (pI=9.67, the content of basic residues=23.3%), indicating a possible affinity to DNA. Furthermore a potential nuclear export signal was identified, indicating that Stella may shuttle between the nucleus and the cytoplasm. BLASTN analysis revealed that the Stella sequence was found only in the preimplantation embryo and germ line (newborn ovary, female 12.5 mesonephros and gonad etc.) ESTs indicating its predominant expression in totipotent and pluripotent cells. Interestingly, we found that Stella contains in its N terminus a modular domain which has

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

9

some sequence similarity with the SAP motif. This motif is a putative DNA-binding domain involved in chromosomal organisation. Furthermore, the SMART program revealed the presence of a splicing factor motif-like structure in its C-terminus. These findings indicate a possible involvement of Stella in chromosomal organisation and RNA processing.

Antibodies may be raised against the GCR1 and/or GCR2 polypeptides. In particular, antibodies may be raised against the extracellular domain of GCR1, which is a transmembrane polypeptide.

Antibodies and nucleic acids disclosed here are useful for the identification of PGCs in cell populations. The methods and compositions described here therefore provide a means to isolate PGCs, useful for example for the study of germ tissue development and the generation of transgenic animals, and PGCs when isolated by a method described here.

Homologues of GCR1 and GCR2 may also be used to identify PGCs and other pluripotent cells, such as ES or EG cells.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al. (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, and A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; J. M. Polak and James O'D. McGee, 1990, *In Situ Hybridization: Principles and Practice*; Oxford University Press; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; and, D. M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A:*

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

10

Synthesis and Physical Analysis of DNA Methods in Enzymology, Academic Press. Each of these general texts is herein incorporated by reference.

POLYPEPTIDES

5 It will be understood that polypeptide sequences disclosed here are not limited to the particular sequences set forth in SEQ ID NO: 2 and SEQ ID NO: 4, or fragments thereof, or sequences obtained from GCR1 or GCR2 protein, but also include homologous sequences obtained from any source, for example related cellular homologues, homologues from other species and variants or derivatives thereof.

10 This disclosure therefore encompasses variants, homologues or derivatives of the amino acid sequences set forth in SEQ ID NO: 2 and SEQ ID NO: 4, as well as variants, homologues or derivatives of the amino acid sequences encoded by the nucleotide sequences disclosed here.

Homologues

15 The polypeptides disclosed include homologous sequences obtained from any source, for example related viral/bacterial proteins, cellular homologues and synthetic peptides, as well as variants or derivatives thereof. Thus polypeptides also include those encoding homologues of GCR1 and/or GCR2 from other species including animals such as mammals (e.g. mice, rats or rabbits), especially primates, more especially humans. More specifically, homologues include human homologues.

20 In the context of the present document, a homologous sequence or homologue is taken to include an amino acid sequence which is at least 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical at the amino acid level over at least 30, preferably 50, 70, 90 or 100 amino acids with GCR1 or GCR2, for example as shown in the sequence listing herein. In the context of this document, a homologous sequence is taken to include
25 an amino acid sequence which is at least 15, 20, 25, 30, 40, 50, 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical at the amino acid level, preferably over at

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

11

least 50 or 100, preferably 200, 300, 400 or 500 amino acids with the sequence of GCR1 or GCR2, for example GCR1 (SEQ ID NO: 2) and GCR2 (SEQ ID NO: 4). Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present document it is
5 preferred to express homology in terms of sequence identity.

Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

% homology may be calculated over contiguous sequences, i.e. one sequence is
10 aligned with the other sequence and each amino acid in one sequence directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues (for example less than 50 contiguous amino acids).

Although this is a very simple and consistent method, it fails to take into
15 consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without
20 penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared
25 sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

12

system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package (see below) the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux *et al.*, 1984, *Nucleic Acids Research* 12:387). Examples of other software than can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel *et al.*, 1999 *ibid* - Chapter 18), FASTA (Atschul *et al.*, 1990, *J. Mol. Biol.*, 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel *et al.*, 1999 *ibid*, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program.

Although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

13

Variants and Derivatives

The terms "variant" or "derivative" in relation to the amino acid sequences as described here includes any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) amino acids from or to the sequence. Preferably, the resultant amino acid sequence retains substantially the same activity as the unmodified sequence, preferably having at least the same activity as the GCR1 and/or GCR2 polypeptides shown in the sequence listings. Thus, the key feature of the sequences – namely that they are specific for PGCs and other pluripotent cells, such as ES or EG cells, and can serve as a marker for these cells in a cell population – is preferably retained.

Polypeptides having the amino acid sequence shown in the Examples, or fragments or homologues thereof may be modified for use in the methods and compositions described here. Typically, modifications are made that maintain the biological activity of the sequence. Amino acid substitutions may be made, for example from 1, 2 or 3 to 10, 20 or 30 substitutions provided that the modified sequence retains the biological activity of the unmodified sequence. Amino acid substitutions may include the use of non-naturally occurring analogues, for example to increase blood plasma half-life of a therapeutically administered polypeptide.

Natural variants of GCR1 and GCR2 are likely to comprise conservative amino acid substitutions. Conservative substitutions may be defined, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

14

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
Polar - charged	D E	
	K R	
AROMATIC		H F W Y

Fragments

Polypeptides disclosed here and useful as markers also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ ID NO:2 and SEQ ID NO: 4.

Polypeptides also include fragments of the full length sequence of any of the GCR1 and/or GCR2 polypeptides. Preferably fragments comprise at least one epitope. Methods of identifying epitopes are well known in the art. Fragments will typically comprise at least 6 amino acids, more preferably at least 10, 20, 30, 50 or 100 amino acids.

Included are fragments comprising, preferably consisting of, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145 or 150, or more residues from a GCR1 and/or GCR2 amino acid sequence.

Polypeptide fragments of the GCR proteins and allelic and species variants thereof may contain one or more (e.g. 5, 10, 15, or 20) substitutions, deletions or insertions, including conserved substitutions. Where substitutions, deletion and/or insertions occur, for example in

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

15

different species, preferably less than 50%, 40% or 20% of the amino acid residues depicted in the sequence listings are altered.

GCR1 and/ GCR2, and their fragments, homologues, variants and derivatives, may be made by recombinant means. However, they may also be made by synthetic means using techniques well known to skilled persons such as solid phase synthesis. The proteins may also be produced as fusion proteins, for example to aid in extraction and purification. Examples of fusion protein partners include glutathione-S-transferase (GST), 6xHis, GAL4 (DNA binding and/or transcriptional activation domains) and β -galactosidase. It may also be convenient to include a proteolytic cleavage site between the fusion protein partner and the protein sequence of interest to allow removal of fusion protein sequences. Preferably the fusion protein will not hinder the function of the protein of interest sequence. Proteins may also be obtained by purification of cell extracts from animal cells.

The GCR1 and/or GCR2 polypeptides, variants, homologues, fragments and derivatives disclosed here may be in a substantially isolated form. It will be understood that such polypeptides may be mixed with carriers or diluents which will not interfere with the intended purpose of the protein and still be regarded as substantially isolated. A GCR1/GCR2 variant, homologue, fragment or derivative may also be in a substantially purified form, in which case it will generally comprise the protein in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the protein in the preparation is a protein.

The GCR1/GCR2 polypeptides, variants, homologues, fragments and derivatives disclosed here may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide, etc to be detected. Suitable labels include radioisotopes, e.g. ^{125}I , enzymes, antibodies, polynucleotides and linkers such as biotin. Labelled polypeptides may be used in diagnostic procedures such as immunoassays to determine the amount of a polypeptide in a sample. Polypeptides or labelled polypeptides may also be used in serological or cell-mediated immune assays for the detection of immune reactivity to said polypeptides in animals and humans using standard protocols.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

16

GCR1/GCR2 polypeptides, variants, homologues, fragments and derivatives disclosed here, optionally labelled, may also be fixed to a solid phase, for example the surface of an immunoassay well or dipstick. Such labelled and/or immobilised polypeptides may be packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like. Such polypeptides and kits may be used in methods of detection of antibodies to the polypeptides or their allelic or species variants by immunoassay.

Immunoassay methods are well known in the art and will generally comprise: (a) providing a polypeptide comprising an epitope bindable by an antibody against said protein; (b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and (c) determining whether antibody-antigen complex comprising said polypeptide is formed.

The GCR1/GCR2 polypeptides, variants, homologues, fragments and derivatives disclosed here may be used in *in vitro* or *in vivo* cell culture systems to study the role of their corresponding genes and homologues thereof in cell function, including their function in disease. For example, truncated or modified polypeptides may be introduced into a cell to disrupt the normal functions which occur in the cell. The polypeptides may be introduced into the cell by *in situ* expression of the polypeptide from a recombinant expression vector (see below). The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

The use of appropriate host cells, such as insect cells or mammalian cells, is expected to provide for such post-translational modifications (e.g. myristylation, glycosylation, truncation, lipidation and tyrosine, serine or threonine phosphorylation) as may be needed to confer optimal biological activity on recombinant expression products. Such cell culture systems in which the GCR1/GCR2 polypeptides, variants, homologues, fragments and derivatives disclosed here are expressed may be used in assay systems to identify candidate substances which interfere with or enhance the functions of the polypeptides in the cell.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

17

GCR1/GCR2 NUCLEIC ACIDS

The methods and compositions described here provide generally for a number of GCR1 and GCR2 nucleic acids, together with fragments, homologues, variants and derivatives thereof. These nucleic acid sequences preferably encode the polypeptide sequences disclosed here, and particularly in the sequence listings. Preferably, the polynucleotides comprise *Stella* and/or *Fragilis* nucleic acids, preferably selected from the group consisting of: SEQ ID NO: 1, 3, 5, 6, 7, 8 or 9, fragments, homologues, variants and derivatives thereof.

In particular, we provide for nucleic acids which encode any of the GCR1 and/or GCR2 polypeptides disclosed here. Thus, the terms "GCR nucleic acid", "GCR1 nucleic acid" and "GCR2 nucleic acid" should be construed accordingly. Preferably, however, such nucleic acids comprise any of the sequences set out as SEQ ID NO: 1, 3, 5, 6, 7, 8 or 9 or a sequence encoding any of the polypeptides SEQ ID NO: 2 and 4, and a fragment, homologue, variant or derivative of such a nucleic acid. The above terms therefore preferably should be taken to refer to these sequences.

As used here in this document, the terms "polynucleotide", "nucleotide", and nucleic acid are intended to be synonymous with each other. "Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications has been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

18

modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

5 It will be understood by a skilled person that numerous different polynucleotides and nucleic acids can encode the same polypeptide as a result of the degeneracy of the genetic code. In addition, it is to be understood that skilled persons may, using routine techniques, make nucleotide substitutions that do not affect the polypeptide sequence encoded by the polynucleotides described here to reflect the codon usage of any particular host organism in which the polypeptides are to be expressed.

10 *Variants, Derivatives and Homologues*

The polynucleotides described here may comprise DNA or RNA. They may be single-stranded or double-stranded. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present document, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or life span of polynucleotides.

20 Where the polynucleotide is double-stranded, both strands of the duplex, either individually or in combination, are encompassed by the methods and compositions described here. Where the polynucleotide is single-stranded, it is to be understood that the complementary sequence of that polynucleotide is also included.

The terms "variant", "homologue" or "derivative" in relation to a nucleotide sequence 25 include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleotides from or to the sequence providing the resultant

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

19

nucleotide sequence is specific for pluripotent cells, preferably specific for PGCs, ES cells or EG cells. Most preferably, the resultant nucleotide sequence is specific for PGCs.

As indicated above, with respect to sequence identity, a "homologue" has preferably at least 5% identity, at least 10% identity, at least 15% identity, at least 20%
5 identity, at least 25% identity, at least 30% identity, at least 35% identity, at least 40% identity, at least 45% identity, at least 50% identity, at least 55% identity, at least 60% identity, at least 65% identity, at least 70% identity, at least 75% identity, at least 80% identity, at least 85% identity, at least 90% identity, or at least 95% identity to the relevant sequence shown in the sequence listings.

10 More preferably there is at least 95% identity, more preferably at least 96% identity, more preferably at least 97% identity, more preferably at least 98% identity, more preferably at least 99% identity. Nucleotide homology comparisons may be conducted as described above. A preferred sequence comparison program is the GCG Wisconsin Bestfit
15 program described above. The default scoring matrix has a match value of 10 for each identical nucleotide and -9 for each mismatch. The default gap creation penalty is -50 and the default gap extension penalty is -3 for each nucleotide.

Hybridisation

We further describe nucleotide sequences that are capable of hybridising selectively to any of the sequences presented herein, or any variant, fragment or derivative
20 thereof, or to the complement of any of the above. Nucleotide sequences are preferably at least 15 nucleotides in length, more preferably at least 20, 30, 40 or 50 nucleotides in length.

The term "hybridisation" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" as well as
25 the process of amplification as carried out in polymerase chain reaction technologies.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

20

Polynucleotides capable of selectively hybridising to the nucleotide sequences presented herein, or to their complement, will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% or 98% homologous to the corresponding nucleotide sequences presented herein over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

The term "selectively hybridisable" means that the polynucleotide used as a probe is used under conditions where a target polynucleotide is found to hybridize to the probe at a level significantly above background. The background hybridization may occur because of other polynucleotides present, for example, in the cDNA or genomic DNA library being screened. In this event, background implies a level of signal generated by interaction between the probe and a non-specific DNA member of the library which is less than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target DNA. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ³²P.

Hybridisation conditions are based on the melting temperature (T_m) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about T_m-5°C (5°C below the T_m of the probe); high stringency at about 5°C to 10°C below T_m; intermediate stringency at about 10°C to 20°C below T_m; and low stringency at about 20°C to 25°C below T_m. As will be understood by those of skill in the art, a maximum stringency hybridisation can be used to identify or detect identical polynucleotide sequences while an intermediate (or low) stringency hybridisation can be used to identify or detect similar or related polynucleotide sequences.

In a preferred aspect, we disclose nucleotide sequences that can hybridise to a GCR1/GCR2 nucleic acid, or a fragment, homologue, variant or derivative thereof, under

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

21

stringent conditions (e.g. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na₃ Citrate pH 7.0}).

Where a polynucleotide is double-stranded, both strands of the duplex, either individually or in combination, are encompassed by the present disclosure. Where the polynucleotide is single-stranded, it is to be understood that the complementary sequence of that polynucleotide is also disclosed and encompassed.

Polynucleotides which are not 100% homologous to the sequences disclosed here but fall within the disclosure can be obtained in a number of ways. Other variants of the sequences described herein may be obtained for example by probing DNA libraries made from a range of individuals, for example individuals from different populations. In addition, other viral/bacterial, or cellular homologues particularly cellular homologues found in mammalian cells (e.g. rat, mouse, bovine and primate cells, including human cells), may be obtained and such homologues and fragments thereof in general will be capable of selectively hybridising to the sequences shown in the sequence listing herein. Such sequences may be obtained by probing cDNA libraries made from or genomic DNA libraries from other animal species, and probing such libraries with probes comprising all or part of SEQ ID NOs: 1 or 3 under conditions of medium to high stringency. Similar considerations apply to obtaining species homologues and allelic variants of GCR1 and GCR2.

The polynucleotides described here may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides as used herein. Preferred fragments are less than 500, 200, 100, 50 or 20 nucleotides in length.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

22

Polynucleotides such as a DNA polynucleotides and probes may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

5 In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

10 Longer polynucleotides will generally be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15 to 30 nucleotides) flanking a region of the sequence which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme
15 recognition sites so that the amplified DNA can be cloned into a suitable cloning vector

NUCLEOTIDE VECTORS

The polynucleotides can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, we provide a method of making polynucleotides by introducing a
20 polynucleotide into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells include bacteria such as *E. coli*, yeast, mammalian cell lines and other eukaryotic cell lines, for example insect Sf9 cells.

25 Preferably, a polynucleotide in a vector is operably linked to a control sequence that is capable of providing for the expression of the coding sequence by the host cell, i.e.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

23

the vector is an expression vector. The term "operably linked" means that the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control
5 sequences.

The control sequences may be modified, for example by the addition of further transcriptional regulatory elements to make the level of transcription directed by the control sequences more responsive to transcriptional modulators.

Vectors may be transformed or transfected into a suitable host cell as described
10 below to provide for expression of a protein. This process may comprise culturing a host cell transformed with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the protein, and optionally recovering the expressed protein.

The vectors may be for example, plasmid or virus vectors provided with an origin
15 of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used, for example, to transfect or transform a host cell.

20 Control sequences operably linked to sequences encoding the protein include promoters/enhancers and other expression regulation signals. These control sequences may be selected to be compatible with the host cell for which the expression vector is designed to be used in. The term "promoter" is well-known in the art and encompasses nucleic acid regions ranging in size and complexity from minimal promoters to promoters
25 including upstream elements and enhancers.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

24

The promoter is typically selected from promoters which are functional in mammalian cells, although prokaryotic promoters and promoters functional in other eukaryotic cells may be used. The promoter is typically derived from promoter sequences of viral or eukaryotic genes. For example, it may be a promoter derived from the genome of a cell in which expression is to occur. With respect to eukaryotic promoters, they may be promoters that function in a ubiquitous manner (such as promoters of α -actin, β -actin, tubulin) or, alternatively, a tissue-specific manner (such as promoters of the genes for pyruvate kinase). They may also be promoters that respond to specific stimuli, for example promoters that bind steroid hormone receptors. Viral promoters may also be used, for example the Moloney murine leukaemia virus long terminal repeat (MMLV LTR) promoter, the Rous sarcoma virus (RSV) LTR promoter or the human cytomegalovirus (CMV) IE promoter.

It may also be advantageous for the promoters to be inducible so that the levels of expression of the heterologous gene can be regulated during the life-time of the cell. Inducible means that the levels of expression obtained using the promoter can be regulated.

In addition, any of these promoters may be modified by the addition of further regulatory sequences, for example enhancer sequences. Chimeric promoters may also be used comprising sequence elements from two or more different promoters described above.

HOST CELLS

Vectors and polynucleotides disclosed here may be introduced into host cells for the purpose of replicating the vectors/polynucleotides and/or expressing the proteins. Although the proteins may be produced using prokaryotic cells as host cells, it is preferred to use eukaryotic cells, for example yeast, insect or mammalian cells, in particular mammalian cells.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

25

Vectors/polynucleotides may introduced into suitable host cells using a variety of techniques known in the art, such as transfection, transformation and electroporation. Where vectors/polynucleotides as disclosed here are to be administered to animals, several techniques are known in the art, for example infection with recombinant viral vectors such as retroviruses, herpes simplex viruses and adenoviruses, direct injection of nucleic acids and biolistic transformation.

PROTEIN EXPRESSION AND PURIFICATION

Host cells comprising polynucleotides disclosed here may be used to express proteins. Host cells may be cultured under suitable conditions which allow expression of the proteins. Expression of the proteins described here may be constitutive such that they are continually produced, or inducible, requiring a stimulus to initiate expression. In the case of inducible expression, protein production can be initiated when required by, for example, addition of an inducer substance to the culture medium, for example dexamethasone or IPTG.

Proteins can be extracted from host cells by a variety of techniques known in the art, including enzymatic, chemical and/or osmotic lysis and physical disruption.

RECOMBINANT STELLA AND FRAGILIS PROTEINS

Nucleotide sequences of Stella and Fragilis are cloned into a TRI-system vector (Qiagen). Stella sequence comprising the second codon onwards (i.e., an N terminal fragment of Stella without the first ATG codon) is cloned into a pQE vector using appropriate restriction enzyme sites, and according to the manufacturers instructions. QIAexpress pQE vectors enable high-level expression of 6xHis-tagged proteins in *E. coli*.

A His tag is placed in the N terminal portion of the Stella gene. Recombinant protein is purified by affinity chromatography on a Ni-NTA column, according to manufacturer's instructions. The His tag is cleaved using a suitable protease.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

26

Recombinantly expressed Stella and Fragilis protein are found to be biologically active.

ANTIBODIES

Antibodies, as used herein, refers to complete antibodies or antibody fragments
5 capable of binding to a selected target, and including Fv, ScFv, Fab' and F(ab')₂,
monoclonal and polyclonal antibodies, engineered antibodies including chimeric, CDR-
grafted and humanised antibodies, and artificially selected antibodies produced using
phage display or alternative techniques. Small fragments, such as Fv and ScFv, possess
10 advantageous properties for diagnostic and therapeutic applications on account of their
small size and consequent superior tissue distribution.

The antibodies according described here are especially indicated for the detection
of PGCs and other pluripotent cells, such as ES or EG cells. Accordingly, they may be
altered antibodies comprising an effector protein such as a label. Especially preferred are
15 labels which allow the imaging of the distribution of the antibody *in vivo* or *in vitro*. Such
labels may be radioactive labels or radioopaque labels, such as metal particles, which are
readily visualisable within an embryo or a cell mass. Moreover, they may be fluorescent
labels or other labels which are visualisable on tissue samples.

Recombinant DNA technology may be used to improve the antibodies as described
here. Thus, chimeric antibodies may be constructed in order to decrease the
20 immunogenicity thereof in diagnostic or therapeutic applications. Moreover,
immunogenicity may be minimised by humanising the antibodies by CDR grafting [see
European Patent Application 0 239 400 (Winter)] and, optionally, framework modification
[EP 0 239 400].

Antibodies may be obtained from animal serum, or, in the case of monoclonal
25 antibodies or fragments thereof, produced in cell culture. Recombinant DNA technology
may be used to produce the antibodies according to established procedure, in bacterial or

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

27

preferably mammalian cell culture. The selected cell culture system preferably secretes the antibody product.

Therefore, we disclose a process for the production of an antibody comprising culturing a host, e.g. *E. coli* or a mammalian cell, which has been transformed with a
5 hybrid vector comprising an expression cassette comprising a promoter operably linked to a first DNA sequence encoding a signal peptide linked in the proper reading frame to a second DNA sequence encoding said antibody protein, and isolating said protein.

Multiplication of hybridoma cells or mammalian host cells *in vitro* is carried out in suitable culture media, which are the customary standard culture media, for example
10 Dulbecco's Modified Eagle Medium (DMEM) or RPMI 1640 medium, optionally replenished by a mammalian serum, e.g. foetal calf serum, or trace elements and growth sustaining supplements, e.g. feeder cells such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages, 2-aminoethanol, insulin, transferrin, low density lipoprotein, oleic acid, or the like. Multiplication of host cells which are bacterial cells or
15 yeast cells is likewise carried out in suitable culture media known in the art, for example for bacteria in medium LB, NZCYM, NZYM, NZM, Terrific Broth, SOB, SOC, 2 x YT, or M9 Minimal Medium, and for yeast in medium YPD, YEED, Minimal Medium, or Complete Minimal Dropout Medium.

In vitro production provides relatively pure antibody preparations and allows scale-
20 up to give large amounts of the desired antibodies. Techniques for bacterial cell, yeast or mammalian cell cultivation are known in the art and include homogeneous suspension culture, e.g. in an airlift reactor or in a continuous stirrer reactor, or immobilised or entrapped cell culture, e.g. in hollow fibres, microcapsules, on agarose microbeads or ceramic cartridges.

25 Large quantities of the desired antibodies can also be obtained by multiplying mammalian cells *in vivo*. For this purpose, hybridoma cells producing the desired antibodies are injected into histocompatible mammals to cause growth of antibody-

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

28

producing tumours. Optionally, the animals are primed with a hydrocarbon, especially mineral oils such as pristane (tetramethyl-pentadecane), prior to the injection. After one to three weeks, the antibodies are isolated from the body fluids of those mammals. For example, hybridoma cells obtained by fusion of suitable myeloma cells with antibody-producing spleen cells from Balb/c mice, or transfected cells derived from hybridoma cell line Sp2/0 that produce the desired antibodies are injected intraperitoneally into Balb/c mice optionally pre-treated with pristane, and, after one to two weeks, ascitic fluid is taken from the animals.

The foregoing, and other, techniques are discussed in, for example, Kohler and Milstein, (1975) Nature 256:495-497; US 4,376,110; Harlow and Lane, Antibodies: a Laboratory Manual, (1988) Cold Spring Harbor, incorporated herein by reference. Techniques for the preparation of recombinant antibody molecules is described in the above references and also in, for example, EP 0623679; EP 0368684 and EP 0436597, which are incorporated herein by reference.

The cell culture supernatants are screened for the desired antibodies, preferentially by immunofluorescent staining of PGCs or other pluripotent cells, such as ES or EG cells, by immunoblotting, by an enzyme immunoassay, e.g. a sandwich assay or a dot-assay, or a radioimmunoassay.

For isolation of the antibodies, the immunoglobulins in the culture supernatants or in the ascitic fluid may be concentrated, e.g. by precipitation with ammonium sulphate, dialysis against hygroscopic material such as polyethylene glycol, filtration through selective membranes, or the like. If necessary and/or desired, the antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose and/or (immuno-) affinity chromatography, e.g. affinity chromatography with GCR1 or GCR2, or fragments thereof, or with Protein-A.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

29

Hybridoma cells secreting the monoclonal antibodies are also provided. Preferred hybridoma cells are genetically stable, secrete monoclonal antibodies of the desired specificity and can be activated from deep-frozen cultures by thawing and recloning.

Also included is a process for the preparation of a hybridoma cell line secreting monoclonal antibodies directed to GCR1 and/or GCR2, characterised in that a suitable mammal, for example a Balb/c mouse, is immunised with a one or more GCR1 or GCR2 polypeptides, or antigenic fragments thereof; antibody-producing cells of the immunised mammal are fused with cells of a suitable myeloma cell line, the hybrid cells obtained in the fusion are cloned, and cell clones secreting the desired antibodies are selected. For example spleen cells of Balb/c mice immunised with GCR1 and/or GCR2 are fused with cells of the myeloma cell line PAI or the myeloma cell line Sp2/0-Ag14, the obtained hybrid cells are screened for secretion of the desired antibodies, and positive hybridoma cells are cloned.

Preferred is a process for the preparation of a hybridoma cell line, characterised in that Balb/c mice are immunised by injecting subcutaneously and/or intraperitoneally between 10^6 and 10^7 and 10^8 cells expressing GCR1 and/or GCR2 and a suitable adjuvant several times, e.g. four to six times, over several months, e.g. between two and four months, and spleen cells from the immunised mice are taken two to four days after the last injection and fused with cells of the myeloma cell line PAI in the presence of a fusion promoter, preferably polyethylene glycol. Preferably the myeloma cells are fused with a three- to twentyfold excess of spleen cells from the immunised mice in a solution containing about 30 % to about 50 % polyethylene glycol of a molecular weight around 4000. After the fusion the cells are expanded in suitable culture media as described hereinbefore, supplemented with a selection medium, for example HAT medium, at regular intervals in order to prevent normal myeloma cells from overgrowing the desired hybridoma cells.

Recombinant DNAs comprising an insert coding for a heavy chain variable domain and/or for a light chain variable domain of antibodies directed to GCR1 and/or GCR2 as described hereinbefore are also disclosed. By definition such DNAs comprise coding

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

30

single stranded DNAs, double stranded DNAs consisting of said coding DNAs and of complementary DNAs thereto, or these complementary (single stranded) DNAs themselves.

Furthermore, DNA encoding a heavy chain variable domain and/or for a light chain variable domain of antibodies directed to GCR1 and/or GCR2 can be enzymatically or chemically synthesised DNA having the authentic DNA sequence coding for a heavy chain variable domain and/or for the light chain variable domain, or a mutant thereof. A mutant of the authentic DNA is a DNA encoding a heavy chain variable domain and/or a light chain variable domain of the above-mentioned antibodies in which one or more amino acids are deleted or exchanged with one or more other amino acids. Preferably said modification(s) are outside the CDRs of the heavy chain variable domain and/or of the light chain variable domain of the antibody. Such a mutant DNA is also intended to be a silent mutant wherein one or more nucleotides are replaced by other nucleotides with the new codons coding for the same amino acid(s). Such a mutant sequence is also a degenerated sequence. Degenerated sequences are degenerated within the meaning of the genetic code in that an unlimited number of nucleotides are replaced by other nucleotides without resulting in a change of the amino acid sequence originally encoded. Such degenerated sequences may be useful due to their different restriction sites and/or frequency of particular codons which are preferred by the specific host, particularly *E. coli*, to obtain an optimal expression of the heavy chain murine variable domain and/or a light chain murine variable domain.

The term mutant is intended to include a DNA mutant obtained by in vitro mutagenesis of the authentic DNA according to methods known in the art.

For the assembly of complete tetrameric immunoglobulin molecules and the expression of chimeric antibodies, the recombinant DNA inserts coding for heavy and light chain variable domains are fused with the corresponding DNAs coding for heavy and light chain constant domains, then transferred into appropriate host cells, for example after incorporation into hybrid vectors.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

31

Also disclosed are recombinant DNAs comprising an insert coding for a heavy chain murine variable domain of an antibody directed to GCR1 and/or GCR2 fused to a human constant domain μ , for example $\gamma 1$, $\gamma 2$, $\gamma 3$ or $\gamma 4$, preferably $\gamma 1$ or $\gamma 4$. Likewise the invention concerns recombinant DNAs comprising an insert coding for a light chain
5 murine variable domain of an antibody directed to GCR1 and/or GCR2 fused to a human constant domain κ or λ , preferably κ .

In another embodiment, we disclose recombinant DNAs coding for a recombinant polypeptide wherein the heavy chain variable domain and the light chain variable domain are linked by way of a spacer group, optionally comprising a signal sequence facilitating
10 the processing of the antibody in the host cell and/or a DNA coding for a peptide facilitating the purification of the antibody and/or a cleavage site and/or a peptide spacer and/or an effector molecule.

The DNA coding for an effector molecule is intended to be a DNA coding for the effector molecules useful in diagnostic or therapeutic applications. Thus, effector
15 molecules which are toxins or enzymes, especially enzymes capable of catalysing the activation of prodrugs, are particularly indicated. The DNA encoding such an effector molecule has the sequence of a naturally occurring enzyme or toxin encoding DNA, or a mutant thereof, and can be prepared by methods well known in the art.

ANTI-PEPTIDE STELLA AND FRAGILIS ANTIBODIES

20 Anti-peptide antibodies are produced against Stella and Fragilis peptide sequences. The sequences chosen are as follow:

GCR1 (Fragilis): ASGGQPPNYERIKEEYE and RDRKMVGDVTGAQAYA

GCR2 (Stella): MEEPSEKVDPMKDPET and CHYQRWDPSNAKIGKN

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

32

Antibodies are produced by injection into rabbits, and other conventional means, as described in for example, Harlow and Lane (supra).

Antibodies are checked by Elisa assay and by Western blotting, and used for immunostaining as described in the Examples.

5 DETECTION OF PLURIPOTENT CELLS IN CELL POPULATIONS

Polynucleotide probes or antibodies as described here may be used for the detection of pluripotent cells such as primordial germ cells (PGCs), stem cells such as embryonic stem (ES) and embryonic germ (EG) cells in cell populations. As used herein, a "cell population" is any collection of cells which may contain one or more PGCs, ES or EG cells. Preferably, the collection of cells does not consist solely of PGCs, but comprises at least one other cell type.

Cell populations comprise embryos and embryo tissue, but also adult tissues and tissues grown in culture and cell preparations derived from any of the foregoing.

Polynucleotides as described here may be used for detection of GCR1 and GCR2 transcripts in PGCs or other pluripotent cells, such as ES or EG cells, by nucleic acid hybridisation techniques. Such techniques include PCR, in which primers are hybridised to GCR1 and/or GCR2 transcripts and used to amplify the transcripts, to provide a detectable signal; and hybridisation of labelled probes, in which probes specific for a unique sequence in the GCR1 and/or GCR2 transcript are used to detect the transcript in the target cells.

As noted hereinbefore, probes may be labelled with radioactive, radioopaque, fluorescent or other labels, as is known in the art.

The antibodies may also be used to detect GCR1 and/or GCR2. GRC1, in particular, possesses an extracellular domain which may be targeted by an anti-GCR1

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

33

antibody and detected at the cell surface. Alternatively, intracellular scFv may be used to detect GCR1 and/or GCR2 within the cell.

Particularly indicated are immunostaining and FACS techniques. Suitable fluorophores are known in the art, and include chemical fluorophores and fluorescent polypeptides, such as GFP and mutants thereof (see WO 97/28261). Chemical fluorophores may be attached to immunoglobulin molecules by incorporating binding sites therefor into the immunoglobulin molecule during the synthesis thereof.

Preferably, the fluorophore is a fluorescent protein, which is advantageously GFP or a mutant thereof. GFP and its mutants may be synthesised together with the immunoglobulin or target molecule by expression therewith as a fusion polypeptide, according to methods well known in the art. For example, a transcription unit may be constructed as an in-frame fusion of the desired GFP and the immunoglobulin or target, and inserted into a vector as described above, using conventional PCR cloning and ligation techniques.

Antibodies may be labelled with any label capable of generating a signal. The signal may be any detectable signal, such as the induction of the expression of a detectable gene product. Examples of detectable gene products include bioluminescent polypeptides, such as luciferase and GFP, polypeptides detectable by specific assays, such as β -galactosidase and CAT, and polypeptides which modulate the growth characteristics of the host cell, such as enzymes required for metabolism such as HIS3, or antibiotic resistance genes such as G418. In a preferred aspect, the signal is detectable at the cell surface. For example, the signal may be a luminescent or fluorescent signal, which is detectable from outside the cell and allows cell sorting by FACS or other optical sorting techniques.

Preferred is the use of optical immunosensor technology, based on optical detection of fluorescently-labelled antibodies. Immunosensors are biochemical detectors comprising an antigen or antibody species coupled to a signal transducer which detects the binding of the complementary species (Rabbany *et al.*, 1994 *Crit Rev Biomed Eng* 22:307-

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

34

- 346; Morgan *et al.*, 1996 *Clin Chem* 42:193-209). Examples of such complementary species include the antigen Zif 268 and the anti-Zif 268 antibody. Immunosensors produce a quantitative measure of the amount of antibody, antigen or hapten present in a complex sample such as serum or whole blood (Robinson 1991 *Biosens Bioelectron* 6:183-191).
- 5 The sensitivity of immunosensors makes them ideal for situations requiring speed and accuracy (Rabbany *et al.*, 1994 *Crit Rev Biomed Eng* 22:307-346).

- Detection techniques employed by immunosensors include electrochemical, piezoelectric or optical detection of the immunointeraction (Ghindilis *et al.*, 1998 *Biosens Bioelectron* 1:113-131). An indirect immunosensor uses a separate labelled species that is
- 10 detected after binding by, for example, fluorescence or luminescence (Morgan *et al.*, 1996 *Clin Chem* 42:193-209). Direct immunosensors detect the binding by a change in potential difference, current, resistance, mass, heat or optical properties (Morgan *et al.*, 1996 *Clin Chem* 42:193-209). Indirect immunosensors may encounter fewer problems due to non-specific binding (Attridge *et al.*, 1991 *Biosens Bioelectron* 6:201-214; Morgan *et al.*, 1996
- 15 *Clin Chem* 42:193-209).

FURTHER ASPECTS OF THE INVENTION

- We provide a nucleic acid molecule which is at least 90% homologous to SEQ ID NO: 1 and a nucleic acid molecule which is at least 75% homologous to SEQ ID NO: No. 3.

- 20 We disclose polynucleotides which comprise a contiguous stretch of nucleotides from SEQ ID NO: 1 or SEQ ID NO: 3, or any of SEQ ID NOs: 5 to 9, or of a sequence at least 90% homologous thereto. Advantageously, this stretch of contiguous nucleotides is 50 nucleotides in length, preferably 40, 35, 30, 25, 20, 15 or 10 nucleotides in length.

- The genes GCR1 and GCR2 encode novel polypeptides, the sequences of which
- 25 are set forth in SEQ ID NO: 2 and SEQ ID NO: 4. We therefore disclose polypeptides encoded by the nucleic acids described here. Preferably, the polypeptides have the sequences set forth in SEQ ID NO: 2 and SEQ ID NO: 4.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

35

Moreover, we provide a method by which genes specifically expressed in PGCs or other pluripotent cells, such as ES or EG cells, may be isolated, comprising the steps of: (a) providing a population of cells containing PGCs or other pluripotent cells, such as ES or EG cells; (b) isolating one or more PGCs or other pluripotent cells, such as ES or EG cells, therefrom and providing single-cell isolates; (c) amplifying the transcribed nucleic acid present in a single cell; (d) conducting a subtractive hybridisation screen to identify transcripts present in the PGCs or other pluripotent cells, such as ES or EG cells, but not in somatic cells; and (e) probing a nucleic acid library with one or more transcripts identified in d) to clone one or more genes which are specifically expressed.

Further aspects of the invention are now set out in the following numbered paragraphs; it is to be understood that the invention encompasses these aspects:

Paragraph 1. A nucleic acid having at least 90% homology with the sequence set forth in SEQ. ID. No. 1.

Paragraph 2. A nucleic acid having at least 75% homology with the sequence set forth in SEQ. ID. No. 3.

Paragraph 3. A nucleic acid comprising a sequence of 25 contiguous nucleotides of the nucleic acid of Paragraph 1 or Paragraph 2.

Paragraph 4. A nucleic acid comprising a sequence of 15 contiguous nucleotides of the nucleic acid of Paragraph 1 or Paragraph 2.

Paragraph 5. The complement of a nucleic acid sequence according to any preceding Paragraph.

Paragraph 6. A nucleic acid according to any one of Paragraphs 1 to 5, comprising one or more nucleotide substitutions, wherein such substitutions do not alter the coding specificity of said nucleic acid as a result of the degeneracy of the genetic code.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

36

Paragraph 7. A polypeptide encoded by a nucleic acid according to any preceding Paragraph.

Paragraph 8. A method for identifying a primordial germ cell in a population of cells, comprising detecting the expression of a nucleic acid sequence according to
5 Paragraph 1 or Paragraph 2, or a homologue thereof.

Paragraph 9. A method according to Paragraph 8, comprising the steps of amplifying nucleic acids from putative PGCs using 5' and 3' primers specific for GCR1 and/or GCR2, and detecting amplified nucleic acid thus produced.

Paragraph 10. A method according to Paragraph 8, wherein the expression of the
10 nucleic acid sequence is detected by *in situ* hybridisation.

Paragraph 11. A method according to Paragraph 8, wherein the expression of the nucleic acid sequence is determined by detecting the protein product encoded thereby.

Paragraph 12. A method according to Paragraph 11, wherein the protein product is detected by immunostaining.

15 Paragraph 13. An antibody specific for a polypeptide according to Paragraph 7.

Paragraph 14. An antibody according to Paragraph 13, specific for the extracellular domain of GCR1.

Paragraph 15. Use of an antibody according to Paragraph 13 or Paragraph 14 for the identification of a PGC in a population of cells.

20 Paragraph 16. A PGC when identified by a method according to any one of Paragraphs 8 to 12.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

37

Paragraph 17. A method for isolating a gene specifically expressed in PGCs, comprising the steps of: a) providing a population of cells containing PGCs; b) isolating one or more PGCs therefrom and providing single-cell PGC isolates; c) amplifying the transcribed nucleic acid present in a single PGC; d) conducting a subtractive hybridisation screen to identify transcripts present in PGCs but not in somatic cells; and e) probing a
5 nucleic acid library with one or more transcripts identified in d) to clone one or more genes which are specifically expressed in PGCs.

EXAMPLES

10 **Example 1. Identification of Genes Specific to the Earliest Population of Primordial Germ Cells (PGCs) by Single Cell cDNA Differential Screening**

A method for single cell analysis is developed to identify genes that are involved in the specification of the germ cell lineage, which results in the establishment of a founder population of Primordial Germ Cells (PGCs). It is determined that the lineage specification of PGCs accompanies the expression of a unique set of genes, which are not
15 expressed in somatic cells.

The method for the identification of the genes is mainly based on the differential screening of the libraries made from single cells from day 7.25 mouse embryonic fragments that contain PGCs. The single cell cDNA differential screen was originally described by Brady and Iscove (1993), and subsequently modified by Cathaline Dulac and
20 Richard Axel which resulted in the successful identification of the pheromone receptor genes from rat (Dulac, C. and Axel, 1995). The method of Axel's group is employed, with slight modifications as described.

Construction of single cell cDNAs from embryonic fragment bearing the earliest population of PGCs

25 In the mouse, the earliest population of the PGCs is reported to consist of alkaline phosphatase positive cluster of some 40 cells, at the base of the emerging allantois at day 7.25 of gestation (Ginsburg, M., Snow, M.H.L., and McLaren, A. (1990)). The precise

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

38

location of the PGC cluster in the inbred 129Sv and C57BL/6 strain is determined by microscopy using both whole-mount alkaline phosphatase staining and semi-thin sections stained by methylene blue. The earliest stage at which a cluster of PGCs can be detected is at the Late Streak stage (Downs, K.M., and Davies, T. (1993)), when a distinctively
5 stained population of cells is found just beneath an epithelial lining from which the allantoic bud appears. This region is at the border between the extraembryonic and embryonic tissues just posterior to and above the most proximal part of the primitive streak. The cluster persists at this position at least until Early/Mid Bud stage. In the inbred 129Sv strain, the PGC cluster is found to contain a slightly larger number of the cells,
10 which are more tightly packaged than in the C57BL/6 strain. The 129Sv strain is used for subsequent experiments, as a better recovery of the earliest PGCs is obtained.

129Sv embryos are isolated at E7.5 in DMEM plus 10% FCS buffered with 25mM HEPES at room temperature and the developmental stage of each embryo is determined under a dissection microscope. The precise developmental stage can differ substantially
15 even amongst embryos within the same litter. Embryos that are at the no bud or early bud (allantoic) stage are chosen for further dissection, which in part is dictated by the ease of identification of the region containing PGCs as seen under the dissection microscope. The fragment that is expected to contain the PGC cluster is cut out very precisely by means of solid glass needles. This region is dissociated it into single cells using 0.25% trypsin-1mM
20 EGTA/PBS treatment at 37°C for 10 min, followed by gentle pipetting with a mouth pipette. The dissected fragment usually contained between 250-300 cells. The procedure for cell dispersal with this gentle procedure left the visceral endoderm layer remained as an intact cellular sheet.

We picked single cells randomly from the cell suspension by a mouth pipette and
25 put individual single cells (but avoiding generating air bubbles), into a thin-walled PCR tube containing 4µl of ice-cold cell lysis buffer (50mM Tris-HCl pH8.3, 75mM KCl, 3mM MgCl₂, 0.5% NP-40, containing 80ng/ml pd(T)24, 5µg/ml prime RNase inhibitor, 324U/ml RNA guard, and 10mM each of dATP, dCTP, dGTP, and dTTP). The volume of medium carried with the single cell is less than 0.5µl. The tube is briefly centrifuged to

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

39

ensure that the cell is indeed in the lysis buffer. During each separate experiment, we picked a total of 19 single cells, and left one tube without a cell, to serve as a negative control for the PCR amplification procedure. All the cells that are collected in tubes are kept on ice before starting the subsequent procedure.

- 5 The cells are lysed by incubating the tubes at 65°C for 1min, and then kept at room temperature for 1-2 min to allow the oligo dT to anneal to the RNA. First-strand cDNA synthesis is initiated by adding 50U of Moloney murine leukaemia virus (MMLV) and 0.5U of avian myeloblastosis virus (AMV) reverse transcriptase followed by incubation for 15min at 37°C. The reverse transcriptases are inactivated for 10min at 65°C. This
- 10 reverse transcription reaction is restricted to 15 min, which allows the synthesis of relatively uniform size cDNAs of between 500 base -1000 bases in length from the C termini. This enables the subsequent PCR amplification to be fairly representative.

- Next, in order to add the poly A tail to the 5 prime end of the synthesised first-strand cDNA, 4.5µl of 2X tailing buffer (200mM potassium cacodylate pH7.2, 4mM
- 15 CoCl₂, 0.4mM DTT, 200mM dATP containing 10U of terminal transferase) is added to the reaction followed by incubation for 15min at 37 °C. The samples are heat inactivated for 10 min at 65°C. The reaction now contained synthesised cDNAs bearing poly T tail at their C termini and poly A stretch at their N termini, ready for the amplification by the PCR using the specific primer.

- 20 The contents of each tube is brought to 100µl with a solution made of 10mM Tris-HCl pH8.3, 50mM KCl, 2.5mM MgCl₂, 100µg/ml bovine serum albumin, 0.05% Triton-X 100, 1mM of dATP, dCTP, dGTP, dTTP, 10U of Taq polymerase, and 5µg of the AL1 primer. The AL1 sequence is ATT GGA TCC AGG CCG CTC TGG ACA AAA TAT GAA TCC (T)₂₄. The PCR amplification is performed according to the following
- 25 schedule: 94°C for 1 min, 42°C for 2 min, and 72°C for 6 min with 10 s extension per cycle for 25 cycles. Five additional units of Taq polymerase are added before performing 25 more cycles with the same programme but without the extension time. Each tube at this point contains amplified cDNA products derived from a single cell. The protein contents

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

40

of the solution are extracted by phenol/chloroform treatment, and the amplified cDNAs are precipitated by ethanol and eventually suspended in 100µl of TE pH8.0. 5µl of the cDNA solution is run on a 1.5% agarose gel to check the success of the amplification. Most of the samples show a very intense 'smear' band ranging mainly between 500bp to 1200bp, indicating the efficient amplification of the single cell cDNA. Only the successfully amplified samples are used for the subsequent 'cell typing' analysis.

Example 2. Identification of PGCs by Examination of the Expression of Marker Genes

The embryonic fragment which is excised theoretically contains three major components: the allantoic mesoderm, PGCs, and extraembryonic mesoderm surrounding PGCs. In order to identify the single cell cDNA of PGC origin amongst these samples, positive and negative selection of the constructed cDNAs is performed, by examining the expression of four marker genes (BMP4, TNAP, Hoxb1, and Oct4), which are known to be either expressed or repressed in various cell types in this region.

At the No/Early Bud stage, BMP4 is reported to be expressed in the emerging allantois and mesodermal components of the developing amnion, chorion, and visceral yolk sac (Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., and Hogan, B.L.M. (1999)). The boundary of BMP4 expression is very sharp, and the expression is completely excluded in the mesodermal region beneath the epithelial lining continuous from the amnionic mesoderm where the putative PGCs are determined. Therefore, BMP4 is used as a negative marker for the selection. Primer pairs are designed for amplifying the C terminal portion of BMP4 (5': GCC ATA CCT TGA CCC GCA GAA G, 3': AAA TGG CAC TCA GTT CAG TGG G). The PCR amplification is performed using 0.5µl of the cDNA solution as a template according to the following schedule: 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min for 20 cycles. Among 83 samples tested, 57 samples show the expected size of bands, indicating expression of BMP4 these single cells. These samples are considered to be of

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

41

allantoic mesodermal origin, and therefore excluded from amongst the candidates representing cells of PGC origin.

The expression of tissue non-specific alkaline phosphatase (TNAP), which has long been used as an early marker for PGCs (Ginsburg, M., Snow, M.H.L., and McLaren, A. (1990)), is then examined. Primer pairs are designed (5': CCC AAA GCA CCT TAT TTT TCT ACC, 3': TTG GCG AGT CTC TGC AAT TGG) and the same PCR reaction as above is performed. Amongst the 26 samples, 22 samples are judged to be positive for TNAP. From the alkaline phosphatase staining of the sectioned embryos, it is known that the somatic cells surrounding PGCs also express some amount of TNAP, although the level of expression is slightly lower than that in PGCs. Therefore, amongst these 22 positive samples there should be still be cells destined to become somatic cells as well as PGCs.

One of the genes known to be expressed in the totipotent PGCs but not in somatic cells is Oct4 (Yoem, Y.II., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996)). To examine the possibility that Oct4 can be used as a marker to distinguish PGCs from somatic cells at this stage, Oct4 expression is checked in the 22 samples by PCR (5': CAC TCT ACT CAG TCC CTT TTC, 3': TGT GTC CCA GTC TTT ATT TAA G). All the 22 samples express Oct4 at comparable levels, indicating that the somatic cells at this stage are still actively transcribing Oct4 RNA.

The amount of expression of TNAP is quantitated in 22 samples by Southern blot analysis (reverse northern blot analysis). Given the fairly representative amplification of the single cell method, confirmed by amplifying single ES cell cDNA, Southern blot analysis allows semi-quantitative measurement of the amount of the genes expressed in the original single cells, although it does not serve as a perfect indicator of cell identity. However, as a result of this TNAP analysis, 10 samples out of 22 show relatively stronger bands at an equivalent level, while the remaining 12 samples exhibit weaker signals. These results indicate that these 22 samples can be divided at least into two groups, one with

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

42

stronger TNAP expression (therefore from putative PGCs) and the other with weaker TNAP.

The possibility that somatic cells surrounding PGCs start to express Hoxb1, while PGCs do not (personal communication from Dr. Kirstie Lawson) is also examined. Primer pairs are designed (5': AAC TCA TCA GAG GTC GAA GGA, 3': CGG TGC TAT TGT AAG GTC TGC) and the same PCR reaction as above is performed. Among the 22 samples tested, 12 are positive, and more importantly, these 12 samples perfectly match the ones which show weaker TNAP signals, by Southern blot analysis.

Taking all these results into consideration, it is concluded that 10 samples out of 83, which are Oct4 (+), TNAP (++) , BMP4 (-), and Hoxb1(-), are of PGC origin. This ratio (10/83) is reasonable, considering the number of the founding population of PGCs as 40 and the number of cells in the fragment as 250-300.

Example 3. Differential Screening of Single Cell cDNA Libraries

As the efficiency of the amplification of cDNA differs in each tube, it is very important to select the samples with the most efficiently amplified cDNA for the construction of libraries. The amplification of six different genes (ribosomal protein S12, intermediate filament protein vimentin, β tubulin-5, α actin, Oct4, E-cadherin) is examined in the 10 PGC candidate samples, by Southern blot analysis. Judging from the overall profile of the amplification of all these six genes, three cDNA preparations are selected for the construction of libraries.

To obtain the maximum amount of double strand cDNA, an extension step is performed with 5 μ l of cell cDNA in 100 μ l of the PCR buffer described as above (including 1 μ l of Amplitaq) according to the following schedule: 94°C for 5min, 42°C for 5min, 72°C for 30min. The solution is extracted by phenol/chloroform treatment, and the amplified cDNAs are precipitated by ethanol, suspended in TE, and completely digested with EcoRI. The PCR primer and excess amount of dNTPs are removed by QIAGEN PCR

WO 02/057307

PCT/GB02/00215

43

Purification Kit, and all the purified cDNAs are run on a 2% low melting agarose gel. cDNAs above 500bp are cut and purified by QIAGEN Gel Purification Kit. The purified cDNAs are precipitated by ethanol and suspended in TE and ligated into λ .ZAP II vector arms. The ligated vector is packaged, titered and the ratio of the successfully ligated clones is monitored by amplifying the inserts with T3 and T7 primers from 20 plaques. More than 95% of the phage are found to contain inserts.

The representation of the three genes, ribosomal protein S12, β tubulin-5, Oct4, is quantitated by screening 5000 plaques, and the library of the best quality among the three (S12 0.62%, β tubulin 0.4%, Oct4 0.5%) is used for the differential screening. As a comparison partner with the PGC probe, one of the most efficiently amplified surrounding somatic cell cDNA (Oct4 (+), TNAP(+/-), BMP(-), and Hoxb1(+)) is selected by the similar Southern blot analysis.

The library is plated at a density of 1000 plaques per 15cm dish to obtain large plaques (2mm diameter) and two duplicate lifts are taken using Hybond N+ filters from Amersham. The filters are prehybridized at 65°C in 0.5M sodium phosphate buffer (pH7.3) containing 1% bovine serum albumin and 4% SDS. We prepared the cell cDNA probes by reamplifying for 10 cycles 1 μ l of the original cell cDNA into 50 μ l of total reaction with the AL1 primer, in the absence of cold dCTP and with 100 μ Ci of newly received ³²PdCTP, followed by the purification using Amersham NickTM Spin Column. The filters are hybridised for at least 16 hrs with 1.0X10⁷cpm/ml (The first filter is hybridised with somatic cell probe and the second filter is hybridised with the PGC probe). After the hybridisation, the filters are washed three times at 65°C in 0.5X SSC, 0.5% SDS and exposed to X ray films until the appropriate signal is obtained (usually one to two days).

The positive plaques in the two duplicate filters are compared very carefully. Among 5000 plaques screened, 280 are picked as candidates representing the differentially expressed genes. The inserts of all the 280 plaques are amplified with T3 and T7 primers, run on 1.5% gels, and double sandwich Southern blotted. Each membrane is hybridised

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

44

with the PGC and somatic cell probe, respectively, using the same conditions as the screening. 38 clones amongst the 280 are selected as differentially expressed genes. These clones are next hybridised with the second PGC and somatic cell cDNA probes, which resulted in 20 clones out of 38 to be common in both PGC cDNAs but they are either not included or less abundant in both somatic cell cDNAs. The sequences of all the 20 clones are determined.

Genes highly specific to the earliest population of PGCs

The 20 clones represent 11 different genes (two clones appear two times, one clone appears three times, and one clone appears 6 times). To further stringently check the specificity of expression, primer pairs are designed for these 11 clones and their expression checked in 10 different single PGC-candidate cDNAs and 10 different single somatic cell cDNAs by PCR. Two of them show highly specific expression to PGC cDNAs.

The first gene, GCR1 (Germ cell restricted-1, Fragilis), encodes a 137 amino acid protein with a predicted molecular weight of 15.0kD. Nucleotide and amino acid sequences of mouse Fragilis are shown in **Figure 1**.

The best fit model of the EMBL program PredictProtein predicts two transmembrane domains, both N and C terminus ends being located outside. The BLASP search revealed that Fragilis is a novel member of the interferon-inducible protein family. One prototype member, human 9-27 (identical to Leu-13 antigen), is inducible by interferon in leukocytes and endothelial cells, and is located at the cell surface as a component of a multimeric complex involved in the transduction of antiproliferative and homotypic adhesion signals (Deblandre, 1995). The BLASTN search revealed that the Fragilis sequence was found in ESTs derived from many different tissues both from embryos and adults, indicating that Fragilis may play a common role in different developmental and cell biological contexts. Database searches reveal a sequence match with the rat interferon-inducible protein (sp:INIB RAT, pir:JC1241) with unknown

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

45

function. The GCR1 sequence appears six times in our screen, indicating high level expression in PGCs.

The second gene, GCR2, (Stella) encodes a 150 amino acid protein, of 18kD. Nucleotide and amino acid sequences of mouse *Fragilis* are shown in **Figure 2**.

5 It has no sequence homology with any known protein, contains several nuclear localisation consensus sequences and is highly basic pI (pI=9.67, the content of basic residues=23.3%), indicating a possible affinity to DNA. Furthermore a potential nuclear export signal was identified, indicating that Stella may shuttle between the nucleus and the cytoplasm. BLASTN analysis revealed that the Stella sequence was found only in the
10 preimplantation embryo and germ line (newborn ovary, female 12.5 mesonephros and gonad etc.) ESTs indicating its predominant expression in totipotent and pluripotent cells. Interestingly, we found that Stella contains in its N terminus a modular domain which has some sequence similarity with the SAP motif. This motif is a putative DNA-binding domain involved in chromosomal organisation. Furthermore, the SMART program
15 revealed the presence of a splicing factor motif-like structure in its C-terminus. These findings indicate a possible involvement of Stella in chromosomal organisation and RNA processing.

Example 4. Identification of PGCs by Screening for GCR1 and GCR2 Expression

Although PGCs are identified in Example 2 by analysis of BMP4, TNAP, Hoxb1,
20 and Oct4, no single one of these genes can be taken as a marker for the PGC state. However, both GCR1 and GCR2 may be used as such.

The expression of GCR1 is examined. Primer pairs are designed (5':
CTACTCCGTGAAGTCTAGG, 3': AATGAGTGTTACACCTGCGTG) and the same
PCR reaction as above is performed. GCR1 expression was detected in germ cell
25 competent cells. The definitive PGCs were recruited from amongst this group of cells showing expression of GCR1.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

46

The boundary of GCR2 expression in particular is well-defined, and the expression is substantially limited to PGCs. Therefore, GCR2 is used as a positive marker for the selection of PGCs. Primer pairs are designed for amplifying the C terminal portion of GCR2 (5': GCCATTCAGATGTCCTCTGCAC, 3': CTCACAGCTTGAGGCTTCTAA).

5 The PCR amplification is performed using 0.5µl of the cDNA solution obtained from PGCs in Example 1 as a template according to the following schedule: 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min for 20 cycles. Among 83 samples tested, only those taken from PGCs show expression of GCR2. Hence, GCR2 is a positive marker for the PGC fate.

10 Antibodies against GCR1 and GCR2 can be similarly used to detect pluripotent cells. Preferably, antibodies against GCR1 are used to detect germ cell competent cells, and antibodies against GCR2 are used to detect PGCs.

Accordingly, both GCR1 and GCR2 are positive markers for the PGC fate which can be used to positively identify PGC.

15 *Identification of PGC by ISH*

The *in vivo* expression of the two genes is examined by *in situ* hybridisation. The expression of GCR1 starts very weakly in the entire epiblast at E6.0-E6.5 (PreStreak stage) and becomes strong in the few cell layers of the proximal rim of the epiblast. BMP4 that is expressed in the extraembryonic ectoderm is one signalling molecule that is important for the induction of germ cell competence and expression of GCR1. Other signals, such as interferons are likely to be involved in the induction of GCR1. The expression becomes more intense at the proximo-posterior end of the developing primitive streak at the Early/Mid Streak stage and becomes very strong at this position from Late Streak stage onward. The expression persists until Early Head Fold stage and eventually disappears gradually. No expression is detected in the migrating PGCs at E8.5.

20
25

The expression of GCR2 starts at the proximo-posterior end of the developing primitive streak at Mid/Late Streak stage and becomes gradually strong at the same

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

47

position from the later stage onward. The expression is specific and individual single cells stained in a dotted manner can be seen in the region where PGCs are considered to start differentiating as a cluster of cells. At Late Bud/Early Head Fold stage, some cells considered to be migrating from the initial cluster are stained as well as cells in the cluster.

- 5 At E8.5 and E9.5, a group of cells considered to be the migrating PGCs are very specifically stained.

From these results, it is concluded that GCR1 is a gene which is upregulated during the process of lineage specification and germ cell competence, and subsequently of PGCs, when GCR2 is turned on after GCR1 to fix the PGC fate.

- 10 Accordingly, expression of GCR1 may be detected in a method of detecting lineage specification, and/or pluripotency, such as germ cell competence. Similarly, expression of GCR2 may be detected to detect commitment to cell fate, for example, commitment to fate as a primordial germ cell.

Example 5. Expression of Fragilis and Stella During Germ Line Development

- 15 Antibodies against Stella and Fragilis are used to detect expression of these genes in early embryos. It is found that each of these genes is expressed in primordial germ cells. In particular, we find that Fragilis is the first gene to mark PGC competent cells at the time of germ cell allocation. Stella is expressed only in the lineage-restricted founder PGCs and thereafter in the germ cell lineage.

- 20 **Figure 3** shows expression of Fragilis in embryonic stem (ES) cells.

Fragilis is expressed in pluripotent ES and EG cells. During the derivation of EG cells from PGCs, it is found that Fragilis expression re-appears on EG cells. Late PGCs are negative for Fragilis after specification of these cells is completed.

WO 02/057307

PCT/GB02/00215

48

Figure 5 shows expression of *Fragilis* as detected by whole-mount *in situ* hybridization in E7.2 mouse embryos.

There is strong *Fragilis* expression at the base of incipient allantois where the founder PGC population differentiates in the E7.25 embryos. *Fragilis* expression persisted until E7.5, but it was not detected in migrating PGCs at E8.5. *Fragilis* is first detected in germ cell competent proximal epiblast cells. *Fragilis* expression can be induced in the epiblast cells when combined with the tissues extraembryonic ectoderm tissues, which is the source of BMP4. In the BMP4 mutant mice, there is no expression of *Fragilis*, consistent with the absence of PGCs in these embryos (Lawson et al., 1999).

Figure 4 shows expression of *Stella* in PGCs.

Stella expression which is strong in PGCs is downregulated in EG cells. There is also low level expression of *Stella* in ES cells. *Stella* and *Fragilis* are detectable in ES and EG cells by Northern blot analysis. *Stella* is first detected at E7.0 in single cells within the distinctive cluster of lineage-restricted PGCs, and thereafter in migrating PGCs and subsequently when they enter the gonads. Figure 7 shows *Stella* expression in PGCs in the process of migration into the gonads in E9.0 embryos. *Stella* is the only gene so far known to be a definitive marker for the founder population of PGCs.

Figure 6 shows expression of *Stella* as detected by whole-mount *in situ* hybridization in E7.2 mouse embryos.

Figure 8. Expression of *Fragilis* and *Stella* in single cells detected by PCR analysis of single cell cDNAs. Note that there are more single cells showing expression of *Fragilis* compared to those showing expression of *Stella*. Only cells with the highest levels of *Fragilis* expression are found to express *Stella* and acquire the germ cell fate. Cells that express *Stella* were found not to show expression of *Hoxb1*. Cells that express lower levels of *Fragilis* and no *Stella* become somatic cells and show expression of *Hoxb1*. The

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

49

founder population of PGCs also show high levels of Tnap. Both the founder PGCs and the somatic cells show expression of Oct4, T(Brachyury), and Fgf8.

Example 6. Expression of Fragilis and Stella in Individual Cells

Intracellular localisation of Stella and Fragilis is also determined. Fragilis localised
5 to a single cytoplasmic spot at the Golgi apparatus, as well as in the plasma membrane. Stella comprises a putative nuclear localisation signal and nuclear export signal, and is localised in both the cytoplasm and nucleus.

Fragilis is observed in the Golgi apparatus as well as in the plasma membrane of PGCs. The cell surface localization of Fragilis is expected as a member of the interferon inducible gene family [Deblandre, 1995]. Expression of Fragilis in the proximal rim of the epiblast marks the onset of germ cell competence. *Fragilis* has an IFN response element upstream of its exon 1, so it is very likely to be induced by IFN after initial priming by BMP4 of the proximal epiblast cells. These IFN inducible proteins can form a multimeric complex with other proteins such as TAPA1, which is capable of transduction of
15 antiproliferative signals, which may be why the cell cycle time in founder PGCs increases from 6 to 16hr, while the somatic cells continue to divide rapidly.

Stella, which has the putative nuclear localization signal and a nuclear export signal, was observed in both the cytoplasm and the nucleus. The onset of *Stella* is followed by the loss of *Fragilis* expression by E8.5. Therefore, *Fragilis* expression marks the onset
20 of germ cell competence and *Stella* expression marks the end of this specification process. Expression of *Stella* in the founder PGCs marks an escape from the somatic cell fate and consistent with their pluripotent state. These studies indicate that specific set of genes are required to impose a germ line fate on cells that may otherwise become somatic cells. Stella, with its potential to shuttle between the nucleus and cytoplasm, could have a role in
25 transcriptional and translational regulation, since many organisms possess elaborate transcriptional mechanisms to prevent germ cells from becoming somatic cells. Expression

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

50

of Stella in the oocyte and preimplantation embryos indicates that it has a wider role in totipotency and pluripotency.

Example 7. The Link Between Fragilis and Stella

Only some of the cells that express Fragilis, ended up showing expression of
5 Stella. Only those cells with the highest levels of Fragilis expression become PGCs and
began to express Stella. Furthermore, Stella positive PGCs never show expression of
Hoxb1. More importantly, only somatic cells with lower levels of Fragilis expression,
show Hoxb1 expression. Furthermore, only the somatic cells show expression of two other
10 homeobox-containing genes, Lim1 and Evx-1. Therefore lack of expression of Hoxb1,
Evx-1 and Lim1, appears to be important for the specification of germ cell fate.

Fig 8a and 8b show expression of various genes in single cell PGCs and somatic cells by
PCR analysis.

Our experiments also show that Oct4 is not a definitive marker of PGC,
15 Previously, Oct4 expression is demonstrated in totipotent and pluripotent cells [Nichols,
199, Pesce, 1998; Yeom, 1996]. However, we find that Oct4 is expressed to the same
extent in all PGCs and somatic cells. We do however find expression of T (Brachyuri) and
Fgf 8 in PGCs indicating that PGCs are recruited from amongst embryonic cells that are
initially destined to become mesodermal cells.

Example 8 PGC Specification

20 The founder PGCs and their somatic neighbours share common origin from the
proximal epiblast cells. By analysing the founder PGC and the somatic neighbour, a
systematic screen for critical genes for the specification of germ cell fate has been
established. Fragilis is an interferon (IFN) inducible gene that can promote germ cell
competence and homotypic association to demarcate putative germ cells from their
25 somatic neighbours, and such an example may apply to other situation during

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

51

development. Expression of Stella occurs in cells with high expression of Fragilis. Fragilis is no longer required once germ cell specification is complete, but Stella expression continues in the germ cell lineage. Stella may also be important throughout in the totipotent/pluripotent cells since it is also expressed in oocytes and early preimplantation development embryos.

Example 9 Germ Line and Pluripotent Stem Cells

PGCs can be used to derive pluripotent embryonic germ (EG) cells. However, unlike EG cells, PGCs do not participate in development if introduced into blastocysts. They either cannot respond to signalling molecules, or that they are transcriptionally repressed. PGCs once specified do not express Fragilis on their cell surface. However, EG cells clearly show expression of Fragilis on their cell surface as do ES cells. Both EG and ES cells express Stella as judged by Northern analysis, although Stella is expressed at a lower level in ES and EG cells than in PGCs. Fragilis and Stella therefore have a role in pluripotent stem cells. These genes are therefore markers of these pluripotent stem cells, where they may also have a role in conferring pluripotency on these stem cells.

REFERENCES

- Brady, G. and Iscove, N.N. (1993). Construction of cDNA libraries from single cells. *Methods Enzymol.* 225, 611-623.
- Dulac, C. and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* 83, 195-206.
- Ginsburg, M., Snow, M.H.L., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* 110, 521-528.
- Downs, K.M., and Davies, T. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* 118, 1255-1266.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

52

- Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., and Hogan, B.L.M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes&Dev.* 13, 424-436.
- Yoem, Y.II., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 1996, 881-894.
1. Weismann, A. Das Keimplasma. Eine theorie der Vereburg. *Jenna. Gustav Fischer* (1892).
 2. Eddy, E. M. Germ plasm and the differentiation of the germ cell line. *Int Rev Cytol* 43, 229-80 (1975).
 3. Seydoux, G. & Strome, S. Launching the germline in *Caenorhabditis elegans*: regulation of gene expression in early germ cells. *Development* 126, 3275-83. (1999).
 4. Wylie, C. Germ cells. *Cell* 96, 165-74. (1999).
 5. Lawson, K. A. et al. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* 13, 424-36. (1999).
 6. Lawson, K. A. & Hage, W. J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* 182, 68-84 (1994).
 7. Tam, P. P. & Zhou, S. X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* 178, 124-32. (1996).

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

53

8. Yoshimizu, T., Obinata, M. & Matsui, Y. Stage-specific tissue and cell interactions play key roles in mouse germ cell specification. *Development* **128**, 481-90. (2001).
9. McLaren, A. Signaling for germ cells. *Genes Dev* **13**, 373-6. (1999).
- 5 10. Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of *Bmp8b* for the generation of primordial germ cells in the mouse. *Mol Endocrinol* **14**, 1053-63. (2000).
11. Ying, Y., Qi, X. & Zhao, G. Q. Induction of primordial germ cells from murine epiblasts by synergistic action of *BMP4* and *BMP8B* signaling pathways. *Proc Natl Acad Sci U S A* **98**, 7858-7862. (2001).
- 10 12. Ying, Y. & Zhao, G. Q. Cooperation of endoderm-derived *BMP2* and extraembryonic ectoderm-derived *BMP4* in primordial germ cell generation in the mouse. *Dev Biol* **232**, 484-92. (2001).
13. Chiquoine, A. D. The identification, origin and migration of the primordial germ cells in the mouse embryo. *Anat Rec* **118**, 135-146 (1954).
- 15 14. Ginsburg, M., Snow, M. H. & McLaren, A. Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521-8. (1990).
- 15 15. MacGregor, G. R., Zambrowicz, B. P. & Soriano, P. Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. *Development* **121**, 1487-96. (1995).
- 20 16. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor *Oct4*. *Cell* **95**, 379-91. (1998).

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

54

17. Pesce, M., Gross, M. K. & Scholer, H. R. In line with our ancestors: Oct-4 and the mammalian germ. *Bioessays* **20**, 722-32. (1998).
18. Yeom, Y. I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-94. (1996).
- 5 19. Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255-66. (1993).
20. Brady, G. & Iscove, N. N. Construction of cDNA libraries from single cells. *Methods Enzymol* **225**, 611-23 (1993).
- 10 21. Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195-206. (1995).
22. Frohman, M. A., Boyle, M. & Martin, G. R. Isolation of the mouse Hox-2.9 gene; analysis of embryonic expression suggests that positional information along the anterior-posterior axis is specified by mesoderm. *Development* **110**, 589-607. (1990).
- 15 23. Deblandre, G. A. et al. Expression cloning of an interferon-inducible 17-kDa membrane protein implicated in the control of cell growth. *J Biol Chem* **270**, 23860-6. (1995).
24. Friedman, R. L., Manly, S. P., McMahon, M., Kerr, I. M. & Stark, G. R. Transcriptional and posttranscriptional regulation of interferon- induced gene expression in human cells. *Cell* **38**, 745-55. (1984).
- 20 25. Evans, S. S., Collea, R. P., Leasure, J. A. & Lee, D. B. IFN-alpha induces homotypic adhesion and Leu-13 expression in human B lymphoid cells. *J Immunol* **150**, 736-47. (1993).

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

55

26. Evans, S. S., Lee, D. B., Han, T., Tomasi, T. B. & Evans, R. L. Monoclonal antibody to the interferon-inducible protein Leu-13 triggers aggregation and inhibits proliferation of leukemic B cells. *Blood* **76**, 2583-93. (1990).
27. Aravind, L. & Koonin, E. V. SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**, 112-4. (2000).
28. Gurdon, J. B., Lemaire, P. & Kato, K. Community effects and related phenomena in development. *Cell* **75**, 831-4. (1993).
29. Reid, L. E. et al. A single DNA response element can confer inducibility by both alpha- and gamma-interferons. *Proc Natl Acad Sci U S A* **86**, 840-4. (1989).
- 10 30. Kita, M. et al. [Expression of cytokines and interferon-related genes in the mouse embryo]. *C R Seances Soc Biol Fil* **188**, 593-600 (1994).
31. Gomperts, M., Garcia-Castro, M., Wylie, C. & Heasman, J. Interactions between primordial germ cells play a role in their migration in mouse embryos. *Development* **120**, 135-41. (1994).
- 15 32. Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-22. (1990).
33. Herrmann, B. G. Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* **113**, 913-17
- 20 34. Crossley, P. H. & Martin, G. R. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439-51. (1995).

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

56

35. Barnes, J. D., Crosby, J. L., Jones, C. M., Wright, C. V. & Hogan, B. L. Embryonic expression of Lim-1, the mouse homolog of Xenopus Xlim-1, suggests a role in lateral mesoderm differentiation and neurogenesis. *Dev Biol* **161**, 168-78. (1994).
36. Fujii, T. et al. Expression patterns of the murine LIM class homeobox gene
5 lim1 in the developing brain and excretory system. *Dev Dyn* **199**, 73-83. (1994).
37. Bastian, H. & Gruss, P. A murine even-skipped homologue, Evx 1, is expressed during early embryogenesis and neurogenesis in a biphasic manner. *Embo J* **9**, 1839-52. (1990).
38. Rogers, M. B., Hosler, B. A. & Gudas, L. J. Specific expression of a
10 retinoic acid-regulated, zinc-finger gene, Rex- 1, in preimplantation embryos, trophoblast and spermatocytes. *Development* **113**, 815-24. (1991).
39. Sutton, J. et al. Genesis, a winged helix transcriptional repressor with expression restricted to embryonic stem cells. *J Biol Chem* **271**, 23126-33. (1996).
40. Cox, D. N. et al. A novel class of evolutionarily conserved genes defined
15 by piwi are essential for stem cell self-renewal. *Genes Dev* **12**, 3715-27. (1998).
41. Fujiwara, Y. et al. Isolation of a DEAD-family protein gene that encodes a murine homolog of Drosophila vasa and its specific expression in germ cell lineage. *Proc Natl Acad Sci U S A* **91**, 12258-62. (1994).
42. Dixon, K. E. Evolutionary aspects of primordial germ cell formation. *Ciba*
20 *Found Symp* **182**, 92-110 (1994).
43. Mahowald, A. P. Assembly of the Drosophila germ plasm. *Int Rev Cytol* **203**, 187-213 (2001).

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

57

44. Nieuwkoop, P. D. & Satasurya, L. A. Primordial germ cells in the chordates. *Cambridge University Press, Cambridge, UK* (1979).
45. Johnson, A. D., Bachvarova, R. F., Drum, M. & Masi, T. Expression of axolotl *dazl* rna, a marker of germ plasm: widespread maternal rna and onset of expression in germ cells approaching the gonad. *Dev Biol* **234**, 402-15. (2001).
46. Johnson, A. D., Bachvarova, R. F., Masi, T. & Drum, M. Expression of *Vasa* and *Daz-like* genes demonstrate that Axolotl primordial germ cells (PGCs) are not predetermined. *Germ cells Cold Spring harbor laboratory*, 61 (2000).
47. Toyooka, Y. et al. Expression and intracellular localization of mouse *Vasa*-homologue protein during germ cell development. *Mech Dev* **93**, 139-49. (2000).
48. Saitou, M. et al. Occludin-deficient embryonic stem cells can differentiate into polarized epithelial cells bearing tight junctions. *J Cell Biol* **141**, 397-408. (1998).
49. Henrique, D. et al. Expression of a Delta homologue in prospective neurons in the chick. *Nature* **375**, 787-90. (1995).
50. Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol* **225**, 361-73 (1993).
51. Winnier, G., Blessing, M., Labosky, P. A. & Hogan, B. L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev* **9**, 2105-16. (1995).
- Each of the applications and patents mentioned in this document, and each document cited or referenced in each of the above applications and patents, including during the prosecution of each of the applications and patents ("application cited documents") and any manufacturer's instructions or catalogues for any products cited or

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

58

mentioned in each of the applications and patents and in any of the application cited documents, are hereby incorporated herein by reference. Furthermore, all documents cited in this text, and all documents cited or referenced in documents cited in this text, and any manufacturer's instructions or catalogues for any products cited or mentioned in this text, are hereby incorporated herein by reference.

Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the claims.

WO 02/057307

PCT/GB02/00215

SEQUENCE LISTING

SEQ ID NO: 1 (MOUSE GCRI/FRAGILIS NUCLEIC ACID)

Mouse GCRI (Fragilis) full length nucleotide sequence

5 GCGCAGAAAGGGCAGACCCGCGAGCGGCTCCATCCTTTGCCCCCAGTGCTGCCCTTGCCTCCGCA
 CCATGAACCCACACTTCTCAAGCCTTCATCACCCTGCCAGTGGAGGACAGCCCCAACTACGAAA
 GAATCAAGGAAGAAATAGAGGTGGCTGAGATGGGGCACCCGACGGATCGGCTTCTGTAGAACTA
 CTGTGATCAACATGCCAGAGAGGTTCGGTGCCTGACCATGTGGTCTGGTCCCTGTTCAATACAC
 TCTTCATGAACCTTCTGCTGCCCTGGGCTCATAGCCTATGCCTACTCCGTGAAGTCTAGGGATCGGA
 10 AGATGGTGGTGTGACTGGAGCCAGGCCTACGCCTCCACTGCTAAGTGCCTGAACATCAGCA
 CCTGGTCCCTCAGCATCCTGATGGTGTGTATCACCATTGTAGTGCATCATCTGTTCTTAAAG
 CTCAAAACCTTCAACATTAATAGAGGATCCGACTCCGGTCCGAAAGTGTTCACCCCTCCGACG
 TCGTCCCTCCTTGCCTCCCTACACGCGAGGTAAACATCATTATCTATCCACAGTGGATTCA
 ATRAAGTGCCTTGATAACCACC

SEQ ID NO: 2 (MOUSE GCRI/FRAGILIS AMINO ACID)

Mouse GCRI (Fragilis) amino acid sequence

15 MNHTSQAFITTAASGGQPPNYERIKEEYEVAEMGAPHGSASVRTTVINMPREVSVDPHVVWSLFNPL
 FMNFCLGFATAYAVSKSRDRRMVGDVTAQAYASTARCLNISTLVLSILMVVITIVSVIIIVLNA
 QNLHT

SEQ ID NO: 3 (MOUSE GCR2/STELLA NUCLEIC ACID)

Mouse GCR2 (Stella) full length nucleotide sequence

20 GGATCACAGACTGACTGCTAATGGGTCITGGTITTAGGTCTTTTCAAAGACTAAGCAATCTTGTT
 CCGAGCTAGCTTTGAGGCTTCTGCCATCGCATCGCCATGGAGCAACCATCAGAGAAAGTCGACC
 CRAATGAAGGACCCGAAACTCCTCAGAAGAAAGATGAGAGGAGCGCTITGGATGATACAGACCTCC
 TRACACAGAAACACTAGTAAGGTCAATGAAAGGTAAACCTTAACCCCGGTGTCAAGCGGTCCG
 25 CAGCCCGGCGAGTCTACGGAACCGCATTCGAGCGTACCTGTGGAGAACAAAGTGAAGAAATCC
 GGAGGCAACTTCAAAGCGCTTCCCAAGAGAGGTCGCCACTTTGTTGTGGTCTGAAAGACC
 CTATAGCAAGATGAGAAGACTTGTTCGGATTGAGCAGAGACAAAAAGGCTCGAAGGAAATGAGT
 TTGAACGGGACAGTGAAGCATTCAGATGTCTCTGCATTTCTGCCATTATCAAAGATGGGATCCCT
 30 CTGAGAATGCGAAAATCGGGAAGAAATAGGAGCTTACATTTGACGCTGCCCTGGCTGCGACGATG
 CCGCACAGCAGATGTAAAGCTATTTTGGTTAAGATTAACTTTTCTGGTGTGGGAAATCTT
 AACTTGTAACTTTAAATGTAGATAGGATGCACAAAGATCCAGATTTATGTGAAGTTTAGAAGC
 CTCAAGCTGTGAGGCCCGGGCTGAGGAATARAATAAATAGAAATTTGGAGTATGTACGTTCTAAT
 TCCAGAAATTTGATAAAAGCAATTTTGT

WO 02/057307

PCT/GB02/00215

60

SEQ ID NO: 4 (MOUSE GCR2/STELLA AMINO ACID)

Mouse GCR2 (Stella) amino acid sequence

MEEPSEKVDPMKDPETPQKKDEEDALDDTDVLQPETLVKVMKLLTLPNGVKRSARRRSLRNRIAAV
PVENKSEKLRREVQSAFPKRRVRLTLLSVLKDPIAKMRRRLVRIEQRQKRLGNEFERDSBPFRCLCT
5 FCHYQRWDPSENAKIGKN

SEQ ID NO: 5 (RAT GCR2 HOMOLOGUE NUCLEIC ACID)

Rat GCR2 (Stella) homologue genomic sequence; similar intron-exon structure as
mouse-Stella. AC094826 contig No.5 (22671 - 27595: contig of 4925 bp in length)

10 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCTCCGACGTATGATGGCTCCTAGACGCAA
CAGGAGGGGACTCCCGCATCATTCACGTAGACCCGCCTTCTGCTTCCCTGTCGGGGTTTTGGG
AAGCCCGCGCCCTCTCTTCCACCTTGCTCCACTAGCACGCGGCTGTTCCTACTGAGCCAGCA
CTGGCTAAGTGGAGCACCAGGAGTTTCCAGGCTTCCCTCAGAGGGCAAGGTGTAGTCCATGGTGGG
CTCAGAGAGACCCCTCTCTCCGTGATACAGAGGCGCAACCCARGCCAGACAGGGGTGATGATT
15 AGGAACATACCTTCGTCGGGGAGAAAAATACCGGTTCATATAGGAATAGAGGAACCCAGGAGGTAGT
TAAGGCTGTGGTGTCTGGTTCCGGGGTTTTGACTCTCAACAACCACGTTCCAGAACGTCGCTGAGTT
TTTTATGATGGTGTAGAATTCCTTATCAGCAATGGTCTCCGGGGTGTTCCTTTTCTTTTAAAT
TTTTTAAGTATAATTTGGTGTGTAAGCAACTGTACTTGGACTAGAATCCCTGTGTAATCCAGAA
TGGAAATCCCAATCCTAGGATAAAAGTTTATGTGGGCTGAGTGTGGGTGGGGTTGTTTTGAT
20 TAGCTGTAGCCAGGCTGGGCTCAATCTCAATCCCTCCCTGCTGCTCCCTTCTAAACGCTAGGATTA
AAAGTCTGGCCATGATCCTGCTGTAGCTTTATTTTATTTATTTATTTATTTATTTATTTTGGCTCTT
TTTTTTGGAGCTGGGACCGAACCCGAGGCTTGTGCTTCCTAGGCAAGCGCTCTACCACCTGAGC
TAAATCCCAACCCAGTGTAGCTTTATTTTAAAGAACAGGAGTCTTGTTCCTCAAAACAGTTCCT
CTGTAGCCCTGGTGTCTCCTGGAACFCGGTAAACAGGCTGGTTGGGACTCTGCCCTTAAACACAT
25 GGGACTAAAGGCGGTACCCTCCCTGGGCTACACCGGAATCTTAAAGTTCATTTGAACCGGGG
CTTTCTCTTTTCTCACCACCTTCTGGAAGCGATTTCCCTGCTAAATTCATTCCTGGTAAATG
ACTCTGAGGGGAAATAGGAACCCAGAATAGATTTGAGCCGGGGCTACCCTGGGACCCCGCACTCCCC
ACCCCCAGCCGCTTTGAGGCTCTTTGCCTGAGGGCCCTCCCGGTTGATACCTCCTAGCACTCC
GGGCTGAGGGCGTGGCTCGGGAGGAGCCATTCCTTTGAGAGGAAACAACTGCTGGCTTGAATC
30 TGCCCTAATACCTGACAGTTACATGGGACCTCCTTATTTCCACAGGATTCCTTAGTCTTTGTTTGG
GAGATTTCAAAATCTTGAGACTGCTCAACCCCTTCCCTGGCCTAACACTCACAGGCCAGGCTAGACC
CAAAATCTGTCAACCCCTTCTGTGTCCAAAACGGTGGGTGGCTAGCTGGCTCACCCCTGGTGTAC
TTTTCTTTAACAATTCGGAAAAGTTGTGGTAAAGTTCCCTGATAAAAATAGGACCATCTACTGGGTGT
GGTCCCATGTAAAGCAAGGTTGGTTTCCAAAATACCCCTGTTTACATAGATGTCGGGAGCATGG
35 AGCAGGTCATATAGATTTAGGTGGAACAGCCTGTTTTTGGAAAGCTTCCAGGGCGGAAAATGAA
CCCAGAGGCACTATTGGGCAAGCCCTCCGGCTAAGCAACACAAATGGCTGCAGGGGCTCTGGGAG
AGGTGTAGACAAAGAGAAATATGCAAGTTCCAGGACCTCTGAACCTAGAGTATAGGCTGCTGTAACA
TTGTACATTTGCTGTAGCAGAACAGCCCATGGTAAGAAGCTCAGTGGATCTCAGCAAACTAGG
ATATCTGCTCAGGGTTATGACCAAGCCCTGTGCATATGGTTTGTCTGTTGTCCTGGCCCTCTCTTG
40 AAGGGGGTATATCTGTATACCACCTTCTGTGTTCTCTGGGCTATTACCTTGCAAAATGCAAAA
TGTATATCTCACTATCTCTCCATCTCTGTTTTCAGAAATCCTCAACCAAGAAACACTAGTAAAG
GTCATGAAAAGCTAACCCGAAACCCAGTCCCAAGCCCAACAAATATCATCTGCTCAAGGGTT
CCTCTCCAGGTTAAGAGCCAGCTGTGGAGAACAGAGTGAAGAATCATGAGGAAAGTTCAAAGC

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

62

CCAGGATTGACGAGCTAGGCCATGATGATGGAGACCTTCCCTCAAGCAGAAATAAACAGGGTAG
 CACACATGAACTCTGAACATCAGGAGTGTGCACACACCCACACATGCATCTGTAAAAACGAGTC
 CCCATCTCCAAATGGCTCGTCTAATCTGTCTGTGTATTTATTAAGATAACAAATTTGCCCTCAT
 TACAATTTCTCTGCAACFAGAAAATCTAAAATAAAGATCTATTCCAATTACCTCTAAATCAA
 5 ACTACCGGGCTTTGACCTCAFGCTCAATCTTGGTAAATCTGTGATTCCTCAATGAATCCAAATGCA
 CACATCCCTATAATAATTAAAGGTAGCAAGTAGAGATTTCCCCAGCACCAGAAAAGTTTAACTT
 AAACAAAACAGCTTTCACATCTGCTGCATGGCCCTTACCGCAAGGACAGTGTATGATCA
 AACTGCCCTAATTTCTGTCTGCTGCCCCAATTTAGCATTCTCAGAAGGATCCCACTCTGTATAAT
 GGCAGAAAGTACAGAGACATCTGAATGGCTCAACTCTTCTCTCATTTCCCTCAAGCTGTCTTTGTC
 10 TCTGCTCAATCCGAAACAAATCTTCTCATCTTGTGTATGGGGCTTTTCAGCACCAGCAACAGTGTGC
 GGACCTTCTCTTGGGAAAGGGCTTTGAACTCCCTCATGATTCTTTCACTTCTGTCTCCACAG
 GCTGGCTCTTAATCTGGAGACGAAACCTTTGACGAAGATGATATTTGGCCGATTTGAGATGAAATA
 TCAAAACACATTTAACATTTAAATAACTTAACGATATACACACCTTTTTTTTTCCACCTCCCA
 CACAGACAAAACAAACCTATTTTCTTTTACAACCCCGCTTAAAGCAAGGAGCATAGTAACT
 15 GACCAATCATAGAAAGGAAACACACACAGACACATCAAAATAAAATAAAATCACCGCCAAACCCA
 CCCCATAAAAACCCCGCACACACACATATACTCCCCCCCCCGCACCATCACTACATCAC
 CCTTCCACCCCTTTCCACCTCCCCCCCCAACATTAACCCCAACCCATCAGGAAACCCCAACAC
 CACAAATAAATAGACACATCGCATACATAAATGACACAGACCCACCCAAAGAGGACGAA
 20 AGATTAGAGCCACATCTCTGGCCCAACACATACACTCAACCTGCATAGTATCTATCTCACCCCA
 ACCTAGAACAAAATCTAATCAGCACCAGGCCCCAGTATCAGCAGACCTCAAACACATACCCA
 CCAATTAACACGCCCAACCCACCAACCCACCCGCTGACACACACTTCGAACTACCCCTC
 AACATCACAAAAGCAATGGCAAGTTACGATGACTCAAACCACTCACTCTCTCATTTG

SEQ ID NO: 7 (Rat GCR2 HOMOLOGUE NUCLEIC ACID)

Rat GCR2 (Stella) homologue genomic sequence; different intron-exon structure
 25 from mouse-Stella (fused exons). AC093991 (1 - 7657; contig of 7657 bp in length)

ACTCCAAAGTGTTCATCATTTACACATCAAAGAAAGAGAAATAAAAAACAAAGGTGTCATGATCC
 CTCCAAAAGAGTGGAAACATTTCAACTGCCAGATCCAGATACTGAAATGGGTAGCATGCTGGAGAA
 AGAATTTCAAAGTTAGGTAGAGAAATCTGGTTGAGCAGAGCACCTTGCTTTTCTTCCAGAGGATCTGA
 GTTCAAGTCCCAAGGACCTATATCACAGTTTCTGTAACTCTAGCTCCAGAGGGCTGTGACACTCTG
 30 TTCACTGTGGGCACCTGCATTTACAGACAAACATAAAGTAGTTCAACACCCCTTTTCACAGAAAACC
 CACAGCATGTGAGGAAATCCGGGCTCTGCGCAATGCCCCACAGCAGAAAGGGGGAGCTGGAGAG
 ATGGTTCATCTGTAGCCATTATTTGCTCTTGAAGAGAACCCAGGGTCAATCCATAGCACCCATAG
 CAGCTCACAACCATCTCCAGTCCAGGAGATCCAATGCCCTGTGTGACCTCAGGTACCAGGCATA
 CACAAATGAACCTGCACACATACAAAAGTCCATAGAGCCATAGTTACCATTGTGAGCTCTGAGAACC
 35 AAATCCGTGTTCTGCAAGAGGACATGCACGGTGAAGAACAGGCACCTTTCCCACTGCCTCTG
 AGACAAATCTCACATAGTAGTTACACTGGCTTCCGACTTGCACCCATCCTCTGCTCTGCTT
 TTTTTCACCCCATGAATGATTTTGAATAATGTTGAGCTGTTTACATTAATAAACAAAATCAGA
 TGGAGACTATATGTCATATTTCAATGAATCAATGACTAGTACAAATACAGTATTTTTATAGCT
 40 TTTCTATTTTGTGTTTAAATTTATTTTTCTTTTTTTTTTTTCTTTTGTAGTTTGTCTTGT
 TTGTTTGGAGGGCTCTCACTGTGTAGTCTTGGGTGATCTGGAACCTACTAGGTAACAAAGGATA
 GCCTTAAACTCAGAAATTTGCTTGCCTGTCTCCAGAGTGTGAGTTAAAGTTGTACACCCG
 ATGTTAGGTGTTTTTATAGTGTCTGTATGCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT
 45 GAGGCCATGTAGCGCATGCTTGAACCAAGAACAGAGGAAGTGTGTATACAGTTACCCCTGGGAGCC
 CAGAAAGGGCAGGAGATGCCCTGGAATGGAATTTTGGTAGTGTGTTAACTGCCATAAAGTGTGG
 GACCTAACACTCTTAACCTCTGAGCCATGGCTTAGTCTGGGGTCCCCCTCTCTTTTATGA
 CTATGAGACTATACAAATTTATTTATATATTAAGTCTACGGGAGCAGTTTCCCTGGCAGAGA

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

GTATATATATCTCAATGGTGACATACATATCTCATGGTGACACACATATCTCATGGTGACACACATA
 TCCATGGTGACATACATATCTCATGGTGACATACATATCTCATGGTGACACAAATTGAGCATT
 GAGAGCAGCTACAGACCGATTAGATCAGACTTATTAATCTTGCCAAAGTATGGTGACCGCAGGC
 CTGCAATGCCAGTAACTTTGGAGACTGAGCCAAGCAGATCACCTGAGCCTAGAGACTCAAGGCCAC
 5 CTTGGACAACATAGACATATCTCTGTTCAAAATGAACAAGCTAAGTCTTTGTACATAGCAGCCT
 CTCTATTGACTGGCAGGGCAGCTGACAGTGTCTCACCTAGTCACAGATGTTCTTTAGAGGG
 AACAGACCCGATGAATACAAACATTTTACGCTCAAGTAAAGTCTATACTATGAGGAACACTCTC
 TTCAAACATCATAACATTAATAATGAGAGATTTTACAACCTTTTTTAAAGATTTATTTGTTTAT
 10 GATTAAGTACACIGTCACTGCTCTCAGACACACCAGAATTTGGCCTCAGATCTCATTTACAGATGTT
 GTCGCCACCAATGTTGTTGTTGGAAATTGAACCTCAGSACCCTGAGAGGACAGTCAAGCCTCTTTT
 TTTTTTTTTTTTTTTTCTTCATTTTTGGAGCTGGGACCGAACCCAGGGCTTGTGCTTCTTA
 GGCAGGCTCTACCACTGAGCTAAATCCCAACCCAGCCAGCTGCTCTTAAGTCTGAGCCATC
 TCCAGGCCCAACATCAATTTTGGTCTAGATGTTTACCTGGTGGTCCATGCCATCTCGATG
 15 GCCCTTGTGGCAGGGTCCCGGTAAAGCAGCCCTAGGSCATGAGTTAGGGAGAGCAAAACCTGAC
 CCAGAACCTGACTGCCATGAAGTATGGAGATGCCCTTTGAGTACATGGGGTTTTTGGTGGTGT
 TCTTTTCTTTTGTTTTTTGTGTGACTTGACACATGTCACAGTCACTGAGAGTGAACCTTA
 ATTGAGAAAATGCCCTGATTTTTCTCCGGCCCCCTAAGTGTCTTTGATGAGTGTATTTTATCA
 CAGCAATAGAACTCTAATAGATAGATTTGATAGAGTGAATATGCTGTAAACAGCCCTA
 ACCATGTTCTCTTTGGGAGGATTTGGGAAAGACTTTGAACTTGGAACTTGGAAACAGGAGAAGCA
 20 TGGGTACTTAGACTTAATGGGCTGTTCTGTGGAGCTTGGAAAGTCTGGAGAAATGGGGATGA
 TACTGTAAAGTTGACAGCACCCTAAAGATGTTCAAGACAGTGTGCAATACATTTAGATTAAG
 AATCTGTGGTGTCTGGTCACTGAGCTGAGATTCAGCTGATTAATAAGACCACTAAGTAAA
 ACTTTGCTTTACGTTGACATCAGTCTGGTGTAGCTAAGGTTGACAGATGACAGTACTAATA
 AGAGACTGGCATCAGAACTGATCCAGAGAGGCCAAGCTGCTCAAACTGGCAGCCAAATTT
 25 GATCACATGTAAGAATCTCCCTCAATGGGGTTGGGATTTAGCTCAGTGGTAGAGCGCTTGCCTAG
 GAAGCACAGGTCCTGGGTTCCGTTCCAGCTCCGAAAAAAGAAACAAAAAAGAAAAA
 GAATCTCCCTCATGTTACAGGCTTGGTGGCATGAGAGCTTTAGGTTGAAGGATCATGGAGAGA
 GCCGAGGCTCCGACCATGTTGGCGGGCAGAGGTACAGCCAGTTACCACAGAGACACCAGCATA
 TTGGAGGTTGCCAGGATCATGGATAATTTGCTAAGACAGGAGGCTGCCCTGACTTTTGTAGGACAAGC
 30 TCCATGATCTGTTTGGCAGGACTGGAGAAACAGAGCTGTAAGGGAAATCAGGACACAGCTGTTC
 AAGATATGATTTGAGAGAGAGGTTTCATTGACAGATCTGAGGAGAGGACGCCAGAGAGGCTCTG
 GAAGGTTCCAGATTGAACCTGGTCTAGAGAGGAGAGGGCTAAGAGCACCAAAAGGCTGTGAC
 CAATTTACAGGTTTATAGAGAAAAGATGCTTGGGAAAGAGAGGGGGAGCCCTGAGCTGGAG
 AGATTTAAGTAGGGGAGGATGAGAAGTGGCTGGGACAGGATGAGAAGTGTGAGGAGCCAAAG
 35 GCCTCAGTGAACCTAGAGGCCAAGATACATTTGACATGCTAATAGGCTTTTGTCTATTGTCT
 CTGCATTTCTTTAGGACAGGCCAAGCTGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
 AAAATGTTTTTAAAGAGAGGAGGGGAGGTTGGGAGGTTGCTGAAATTAAGAGACTTTCCAG
 TATTAGACATTTGATATTTTAAAGAAAATTTTGAACCTTTAAGAGACTGACCTTTTAAAGTGT
 TGAATTTTAAAGACCAGGATACATCAGGTTGAGGACACATGACCCCTGCTCCGCCCCCCOCCC
 40 CAAAATATAATTTTAAAAAGACTGTTGGGAGCTGGTGGTGTATAGGCTTTAATCCTAGCA
 CCCAGGAGGACAGGACAGGAGATCTGAGTTTGGAGCCAGCCTGATATAGCATGATTTCCAG
 GACATCAAGGCTACACACTGAAGCTATCTTAGAAAAAAGACTTGTACTTTTGTGTTGGGATG
 TATTTTATATTTAGGTTGCTGACATTAATATGAAATCTTTGTGAGTGGCAGAAAATAAGACTAA
 AGCTGAATACTGATGCCACTTGTGTGCTCAGATGACAAAGGGTTTTGGAATTTTTTATTTTTTA
 45 TTTTTTTTAGGATATATCAACCAATTTGTTTATACACAGCATGAACAACACAAAAATCAAGCC
 TTTTCCAGATCTGCTGACAAGCTTATGTTCAAACTCGGAAACGAGGAGGAGGAGCCAGGACT
 AAAAGACCGAGGCTCATGGAGACTTGTCTCAAGCAGAAAATAACAGGGTGGTAGCACACA
 CGAATCTGAACACACAGGTTGACACTACCCACACTGACCTGTAATAACAAATCCCCATCT
 CCAATGCTCCTTCAATCTGTTCTGTTATTTAAGATTAACAATTTCCCTTTATTAACAAT
 50 TCTCTGCAACTAGAAAATCTGAAAGATCTATCCATTTACCTCTAATCAACTACCGGCTTT
 GACTCATGCTCAATCTTTGGTAAATTTGTCATCGCATGAATCCAAATGTCACACATCTATATA
 ATTTAAAGGTTAAACAGTAGAAGAGATGCCCTAGCACCAGAAAAGTTAATCTTAACAGAAAAC

WO 02/057307

PCT/GB02/00215

64

AGCTTTCACATCTGCTGTGTGGCACCTTTAACGGCAAGGACGGCGTACAATTGGAACTCGCCCTAAT
 TCTTCTGGTTCGCCCCGATTTTAGCATTCTCAGACGGATCCCATCTCTGATAATGGCAGAAAAGTGC
 AGAGACATCTAAATGGCTCATCTCTGTTCATTTCCCTCAAGCTGCTTTGTCTCTGCTCAATCC
 GAACAAATCTTTCATCCTTGCCTACAGGTTCTTTCAGCACCGACGACAACAATGTGTGGACCCCTC
 5 TCTTGGGAAGGCGCTTTGAACTCCCTCATGATTTTCACTTCTGTCTCCACAGGCTGGTCT
 GAACCCGGTACGAAAGGCTGTGATGACGATGATATTTGGCCACTTGGCACTGGGGTTCAGGGTTA
 GCTTTTTCATGACCTTTACTAGTGTCTGGTGTAGGTTTCTGAATCATTTGGGTTGAGTCTCT
 CCACCTTCTTGTGAGATCTATCATCTGAGTTTCTGGATACACAACCTGGGTCACACTTCTGTGATG
 GCCTCCATGGCGTGGGCGAAGCCCTCAAAGGCCAGCTCCGAACAAATTCCTAGCTAATCTTT
 10 GCAAGACCTAGACTTTGGCCCCACTAGCAGACTGAAGTGTGGAATTTTTTTTTTTTTTTTTT
 TTTTFTTGTAAATCAACTTGAACAACAATTGAGAAAATGCTCCATAAGGTTAAATCCTTGTGCC
 ACCATGCTGGACCTAAGCTTTTCATGGCCACTATTCCTCGAGGCTCGGATCAGAAAGCTTGTGTAT
 TTCATTTCCGGATTTGCTGCTCACTCCAGATTAAGTCCAAATGAAAGCAATAGCCATGTAATAAT
 GCCTAGATAAATCTTCCCTGTTTCCAGCAGCAAAATGCAATAGCAATAAGCTTAGCTGGTGGGATC
 15 TTCCAAAGCTACTCTGCTCTTTTCTTCTTGGACATAGGATTCAGCAACATTTACTTCTTGATC
 CCCTTTATCTTGTGAACATACATTTTACTTTTCTTTCGTAGCTTCTTCTTTCATCAAAAAGA
 TCTTTCATAAGATGAAATTTGGGTTAGAGAGATGTTTCCAGTGGTAAATAGCCTGACTGCTTCT
 CCAGAGGTCCTGAATCAATTCCTAGCAGCCACATGGTAGCTCAATACCACTGTAAATAGGATCTG
 ATGCCCTCTTTGGTGTCTGAAGAAGACAGCAACAGTACTCAACATACATAAAATAAAAATAAA
 20 TCAACATACATAAAATAAAATAATTTTAAAAAAAAGGTGAATTTAACCAACACAGAAAT
 TATGCCAGGCTGTTTGGTGTCTGAAGAAGACAGCAACAGTACTCAACATACATAAAATAAAAATA
 TAGCTCTCTGGACATGGCAAAAGCAGCCACATTTCTCATCAAAATATACAAAGCCGCTCT
 CAGCCACATACIAAAATTTCTCTCTGAAACTTCTAGAGCCAGGCTTCCACAGTTCAAACACCTTC
 AGCAACAAAGTCTTCTATATTTCTACGATGATAGCCCTTAAGCCCCACTAAAGCAATTCACCTGA
 25 ATTCCAAATCTAAGTCTCCAAATCTATATTTCCAAATAAAGCAAGGTCAGACCTACCTATCA
 CAGCAATATCCAGTCCCTGGTACCACCTCTGTCTTAGTTAGGGTTCCATGTTGTGAAGAGAC
 ACCATGACCAAGAAACACTTTTTTTTTTTTAAATTTATTTTTATGCTATGAGTACACTGTGTC
 TGTCTTCCAGACACACCAGAAAGGGCATCAGATCTCATTACAAATGGCTGTAGCCACTACGTAGT
 TGCTGGAAATGAACTCAGGACCTCTGCAAGAGCAGCCACTGCTTTAACCCGCGGACCAATTTCT
 30 CCAGTCCCAAAGAACACTTATAAAGGACAAATTTTTTTGGTTTTTTTAAAGGTTTATTTAT
 TTATGTATGAGTACACTGTAGCTGTCTTCAAGTACACCAGAAAGGGCATCAGATCTTACTATA
 GATGGTTGTGAACCACCATGTGGTGTGGGATTAAGTCAAGGACCTCTGGAAGAGCAGTCAAGT
 CTCTAACCCCTTAGCCATCTCCAGTCTTAAGGACAAATGTTAATCGGGGCTGGTCCACAGT
 TCAAGGTTTCACTCCATTTATCATTTGAGACAGGAGGCTGGCAGCATCCAGCCAGGTTGGGGCTGA
 35 GAGCTCAAAGTCTTACTCTTGAATCCAAAGGACAGCCAAAATAAAGTCTGGTACGGGCTTACC
 ATAGCAGCTAAGAGGAAAGTCTCAAAGCCACCTTACAGTGGCATGTTCTCCAAACAGGCCACAT
 CTCCATAATGAGCCACTCCCCGGCCATGCAATTCAGTCCGCCACCCACTGAGCCATCTCTCC
 AACCTGCTCCAGACCTCTCCCTGCTTTTACCTAAGCTCATTAGGCAGCAATATGCTCTTATTTG
 TTTGAGCTCAGATCCTGTTTTTCAAAGGCTGCTGTGATCAGAGTGGTTGTTCCACAACCTCTC
 40 CCAGTTCTTTGTNAAAACACCAATGCTTAGAGAGATGCTCTTCTGATATATCGCATGTGCAGAA
 GAAAGGTCGCCAGATCCTTTTCATGTGACCTGTGATGCTTTACCCACGTAAGTCTGCTCTGA
 CTCTCTCGAGATGCTGANAATGATTTGAGCGTAGGATGCTTGGGATGTCATGGGCAATTT
 G

SEQ ID NO: 8 (Rat GCR2 HOMOLOGUE NUCLEIC ACID)

45 Rat GCR2 (Stella) homologue genomic sequence; different intron-exon structure
 from mouse-Stella (fused exons). AC103122 (11084 - 13244: contig of 2161 bp in length)

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

ACCTCTGGA AAAAGCAGTCAAGGAAGCCAGAGTGGCCGGAACTCCTGAAAATGGAGTAACAACAGGT
 TETTGTGAGGGTAAATTGAACTCAGGTCTTATGCAAGAGCAACAGAGGTCTTAGCCCTTATTATT
 TTTAATATCTAATTATTTTATTTTATTTTATTTTATTTATTTATATATAAAGTACACTGTA
 5 GCTGCTTTCAGATACACCAGAAGAGGGCATCAGATCTCTTTACAGATGGTGTGAGCCACCATGTG
 GTTGGTGGGAATTGAACTCATGACCTCTGGAGAGCAGTCCGGTGTCTTAACCACTGAGCCATCT
 CTCCAGCCCTAATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT
 AGGGCTCTGGGTATCAGTCCACCTTACAGATGGTGTGAGCCACCCTGTTGGTGTGGGAATTA
 ACTCAGGACCTCTGAAGAGCAGTCAAGCATTTTAAACGACTGAGCCATCTCTCCAGCCCAACCCCC
 10 CCICCATTTTTTTTTATACAAAAGGAGCTTCTTGCAGAGACATGGCCATATACATCCACCCC
 TCTTCTTTGAGGTTTGTAGTGTCTGCTCTCTCTCTGGAAAAGAAAATCTCTAGGACTA
 AGCTAAAAGAGCCAGATGGATGGAATGGGGTCCCATGGCAACACCATCTGAGGATCTGAGCCCT
 GCTGTCTCTCCAGTTATGTTGACATTTGGTGGTTCATGCTTGAACACTGAAGTGTCTCTCC
 ACCTATGAAAGAGAGGGCCGTTCCAGAGGTCTTAATTTATCTGCTCCATCAGTAGCATTTGGACTG
 15 CTACATTTATGCTGACAAACCATTGGCCAGGAGGTAGAAGAGGATGGAGGAAGGCCAGCCTG
 GCTGGGTACTATCGGATCTAGTGAAGCTGTATAGAATCTGTGGGGTTTATTACTCCCACTGG
 AGCAGAGCCAGTGTCTCAGGAAGCAGTAAAGAGTCAAGCTTACCACAGAAAATAAAGTACTAC
 TGTGATACCAATTTGACATGATCACCCTGAGCCACTCCACCTCAGAACAAAGTACCATATCG
 TTAAGTGTCTGAGCTCAGGGGAAGGCCCTGTGCTGTGAGTAGAGCCAGGTAACCTTAAACA
 GCCCTACTACACTTCTTAAAGCATCTGTCTACATACAAAGATTTACTCTTTAAGAGCAG
 20 ACTTTAAAANAANTGAGCCACTTACACTTTCGAAAGTTGATCCTGATGTCACATGCTGAGAC
 AGTGGCCAGTCTCAAGGACAGCCCTCCACACTGAAGTAGTCTCAGCAGTATGCTATGTCACC
 TAGCCACCAATAAGAGCTCACCTAAGAAAATTTCCACTTTACTGGTAAAGAGGCTATCTTCCCTC
 CTTTCTCTCAATTAGCATCTCTCCTCCAGACTTCCCTACTACCCGACTTAAAAGATCAAGCC
 AGGCACCATAGCAGAGGCTGAGGTCCGAAGGCAGAAAGCCAGAAAGATCTATGTGATTTCCAGGCTA
 25 CTAGCACCACAGTGTGAGCCCTGTCTAACAATGGAGGTTGGGAGGATGGCAGTAACTTGAAC
 CTACAAAATTTACAAAATTTCAATTAAGAACATTTTGTGTTTTGAGGCGAATCTCACTACG
 TAGAGTGGGCTTACACCCAGTTCCAATTAAGAACATTTTAAAGGCTGAGAGATGGCTCAGCTGTT
 AAGAGCACTGGCCACTCTTCCCAAGCTCCTGAGTACAATTTCCAGCAACACATGATGGCTACAA
 CCATCTGTAATGAGCCCTGATGCCCTCTCTCTGTGTCTGAAGCAGCTACAGTCTCTCAITTA
 30 AATAAAAACATTTTAAATAGAAAATCCAAACAGGGAGGCTGATGAGAAAACGACATAACCTTTGTC
 CAGGAGTGTGGTTAAGGGGAATGGAAACATAGTAGAGTCCATTTCTTTCTCTTTGAGCCAAA
 AAGTETATTTATTCAGGCTTCCATTTGAAGTACTCTTGGTGGCATCTTAAGCCTGAGATCTTT
 TGCCATACGTAGTCTTACCACACTCCCAACTGCAACCAACTGTTTCTGTGGCATCCCTCTGAT
 GACTTTTACACAGGGGTTGGGGATTTAGCTCAGTGTGAGAGCCCTTCCCTAGGAAGCACAGCCCT
 35 TGGGTTGGTCCCAGCTCCGAAAANAANAAGATTTTACCGGGCACACCCACTCCCACTAGTTT
 CTCATGATCAAGTATATCAGATGATCTGGTCTCGGCACAAAGTSCCTCCICAGCTCGACACA
 CACGAGCTCATCAGTGGATTCCAGCCACAGATGGGTTGGCACTTGTCTAAGGCTTCAGGAG
 CTTTGTGTTGCCAACGCTGCTGGGTATCGTGGATGAGGGGGCTTTCAGCACCTCTTTAGAGCA
 40 GTGTTGACATCCACACCTCCAGTGGCAGTGCCTGCTCCGCTCCGGAAGCTGAGGTGGAATAGCA
 AGTCAAGTTCTCTCTCATTTCCAGACACCAATTTGATGCTCAGTGTGAGTGTCTATTTGTC
 ACTTACTTTTACAAATTTGTTTATTATTATTGATAGATTTATGCTCTCTGTCTACTAGCTACCAGGC
 AGGCTCTCAGGACTTATCCAATTTGTTCTGCTCCTCGAGCTAAGCCTGAAGGCATATATGAA
 TCATCTCACCAGCAGCATCAGCTTTAAGAGTTTCTGAACGTCACACGTTAACACTGGGGCCAT
 45 ATATGTACGATGTAATTAATCCTCAGCAACTGGCCACAGCCCTAAAAGAAAANAATCCAG
 ARCCAAACAACAAAACAGGCAGAAATGGTGGCACACACTTCAATCTTTTACTTGGAGGTTG
 GRTCCAGGAGGATGGAATTCAGAGCCGGCTAGAGTACCAGTGTGAAGGCCAGCATCTGCT
 CAAGCCAAACAAGATAAATAAGTACTTGTTCAGCTGGGAGTGGTGGTACATTTGGAGGGGAGA
 GGCAGCCCTGAACACTGGGTTCAAGGCCAGCCTGGTCTAGAGTCAAGTCCCHAAACAGCCAGG
 50 GATAGACAGAGAAGCCCTGCTCAAAAAGTGGGCTGGAGAGATGGCTTACTGGTTAAGAGCCACTG
 ACTGCTCTCTAGAGATCTGAGTTCAATTTCCAGCAGCTATATGGTGGCTCACAAACATCTGTA
 TGGGATCTGATGCCCTCTCTGTGTCTGAAGACAGCTACAGTGTACTTATATACATGAATAAAA
 TCTAAAATAATAAAGCTGCACAATTTCTGCTGCTATATGGCTGCAAGCCATCCCTCCAAC

WO 02/057307

PCT/GB02/00215

67

CCAATAAATAAATATTAATAAAAAAAAAAAGCACAAAACCAACCAAAAGTAAATAAATAAACAA
 CTTTTATTCTACCAAGAGAACACACATTCCTTGAGAACTAAGGACAAACATGTTTATGGTTAGAA
 CACAGAGAAATAAGAGCAGCTCAGCTGGAGAGAACAAAGTGTCTGGGGACAAAGGAGCCTTC
 TTCCCTGCCCCATAACAGTGGCCAGATTGAACCTCTGGTACGACAGTCAAGTTGGTGTGAGTTC
 5 AAGTTGGAAGTCACTTTCTAAATCAGGATCAAAGCAAGCTGGAGGCTCCCTCAGTCACTCAC
 AAGTCTGTGAATCAGGAAAAAATAATCAGTTAGACACTGAGTFCGAGGAGCCCAAAACCAA
 GATTTCCACCACCAAGACAGGATATCTGGATTTCCAGGGAACAGATGAGACTTATATCTC
 TGACTGGCATTTAAATCTCACGCCATCCCTCTCCAGCACATCCCTTTCTCCAGGGAATGGTCCCA
 GCACCCATGTCAAGCACTACCCCAAGTAGTCTCCATCAGAGAGCCAAATAGCAAACTGGGAGGGA
 10 AAGGGAGAAAGGATGGTGGTGGGGCCCAACCCCATTCGAGGCTTCTGTATCATATTCCTGCT
 CATGGACACAGAGCACAGAGCCCAACCAACTGTGGATGGCAAGAGGTCAACAGCGCAGATGGGA
 AAGAGCTTGTCCCAACCTGATGACCTGACCTCCACCCCAAAATCCACAGCAGCATGGGATGACC
 TGAAGGCGGTCTAAATGTCAACTGTGGCGAGTGTATGCCCAACATCCACATAAAATATGTTCT
 ACRAAAGAAACGAGAAACCCACAGCTGTGAGTCTGAATGATGACTTGGATTATTAATCTCA
 15 CTACCCAGGAGGCTAAGGAGGCGAGTCAAGCAAGAGACTCACAAATGTCTTTGCTACACGTG
 TCCCTACAATCTCAAGCGTATCTCATCGTCTGCTGAATACAATGCTCTGGAAAGGAGAGAG
 CAGGCTCATCAAGCAGACTCAGGCTGGTCTCATCCCTCTCACCAACTCCPCCTCATTCGCTCAC
 CTATCCATGCTTGTAAACAGGGGGTTCGAATTTGGATCAAACTCCATCTCTGAAGGATGGA
 20 TGAAAGGAATTTGACACAAGGTAGCATTTCAAAATAGTGCATCAAGGATGAGACTCAGGGGCG
 TGGTTCTCTCTCCCTCGGCTCACCCACACGCCAGACTCAGCTGTCGAGAGTGAAGCAGGACAT
 GGCCCAATTTCTCTGAAAAGTCCAACTAGAAGGAAATGACCGTCTTCAAAAGCTCTCAAGCA
 TCTTTACCTGATTTCTAGGCACATTAATCATGTTTCTTAACAGTTAAATGTAGCATTTGTTTAA
 ATTTCTCTCTGTGATTTCTTCAATTTCTTACATTTTGTCTTCAATTTTAAIGTGAAGAAAT
 25 ATTTGACCTCACATGTGCTGTGACCACTGACCTGCAGTGGCCATGGAAAGCCAGGAGGGGTAT
 TGGGACCTGCAGAAATAGGAGTTACAGATTAATGTGAGCCATGGGCTGGTGTGGGAGTCAAAAC
 CCAGGCTTATAGAACCCAGTAGTGTCTTAAACCACTGAGCTATAGACCCCTTAGCCTTTAAGAAA
 CTTAAATTTCTGAGGCTAGAGATAGCTCAAGTGGTTAAGAGCACTGACTGCTCTCCATGGCTCCT
 GAGTTCAAATCCGAGCAACCAATGGTGGCTCACAAACCATCTGTAATGAGATCTGATGCCCTCTC
 30 TGGTGTCTCTGAAGAGACTACAGAGGAGTGTGATAATAAATAAATCAGGAGGCTAGAGAGATGGC
 TCAGCGGTTAAGAGCACTGATTTGCTTCCCAATGATCATGAGTCAATTCIAGCAATCACATAGT
 GGCTCAATAATCATCTGTAATGGGATCTGATGCCCTCTCTGATGTCTGAAGACACAGTGTACT
 CATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
 35 ACTAAGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
 GAATTTGAACCTCTCACTTAGGAAATACCCGTAACCTTTCTTTTTTTTTTTTTTTTTTTTTTTTT
 TTTTTTGGAGCTCAGGACCGAACCCAGGACTTGCATCTCTAGGCAAGGCTTACCACCTGAGC
 CAATCCCAACCCCAATACCTTTCTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA
 40 CAATCCCAACCTTCCCTTTAAAGGGCCGGAATAAATAAATAAATAAATAAATAAATAAATAAATAA
 AACCTCGCCCGGCTGGATCACACCAGCACCATATTTGCGATGGTAGTACAGCAATCACAAAGC
 CATATCTGCAGAAAGATGAAAAAAGACAGTACTGTATGTGAAGAGCCTTAAAAAGCCACAGC
 AATAGTCTGCGTGTGATGAAACCTCTGCTCGAACAGCTCGATGACCAAGAGAGACAGAACTCAGA
 45 TTAGCACCTGAAATATTAATGCTCTCTCACAATTTGACAGTAAATGCCAAGAGGACACAGATA
 TGCTGACATACACCTATCTCTCAGTACCAGGACTTCCAGTCAATGGTGGAGACAGGCTCTTTCGA
 AAACCAAAATCAGACAGAAAAATTTGACGAAAACTTTAATCCAGCACTCAGTGGCAGGAGTT
 CTCTGAATTAGAGGCGAGCTTGGTCCACATAGTGAAGCCATCTCCAAACCAAAACATTTGCATAA
 50 TAACCGTCTGATGCGCATAGGGAAGAAAAATTTGGTTTAGCAACCTTTAGAAAGCCCAAAATAG
 GCAAAATAGCTGCTCGGATGCTGGAGTGGTGAAGAGTCTCTGAGTGAAGTAAAGCAAGCCCT
 GACTGAAGGAGTGAAGTAGAGTTACAGAGTAGGTTATTTGCTGATTCAGGAGCAGACTG
 TGAATCAGACACTTACTTCCAGTGCAGTCCCTCCCTCCCTCGTGAACAGCTCTGCTTACCCT
 GTTGAACCAAGTTGGACCAAACTAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
 55 AGAAAGGAGATGAAGACTCCGCCAGCCACTGAGAACAGGAGGCGAACCCTGGCCCTCCAGCC
 TCCTCCTGCTCCCTCCCTCAGCCGTAACCCGCTCCAGCTCACATGATAAACAATCTTCTGAGAA
 GCTTGGACCCGAGAGCCAGAACTCCCAAGGAGGACCTCGCCGGAAGCACTAGCAGAACTCCCA

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

CCAAGTCTCCGAGTCGCTCCGAGATTGAGTCTTAACGCCATGGGCGGGGAAACGTGAAGCCC
 CGCCCTCAGGCCTCCCATCAGCGCTCATCAGCACAGCCAGGATTACACAGAAAAACCCGGTCTC
 GAAAAACCTTAAAAAANAAAAAANAAAAAANAAAAAAGGTTAAGAGGTTCTGGCTTGTCCGCCACAT
 GCCTTTAAACCCAGCCGTGGCAGACAGATCTCTAAATCAAGGCTAAGCCACATCTACAAAAGTGA
 5 TTCCAGGATAACCAAGACTGTGTATACAAACCTATAAAAAAATTTGTTTTGGGGTTGGGGATTT
 GGCTCAGAGTACAGGCGCTGGCTAGCAACCGCAAGGCCCTGGGTTCCGTTCCCGAGTCCGAAAA
 GAGAAAAAATTTGTTTTTAAATTTTATTTAGGGGCTGAGGATTAGCTCAGTCTTAAAGGC
 ACTTCCAGCCCCACAGGATAGCTCACRRTCTTCTGTAACTCAGTTCCAGAGAGAACTGACAC
 CCTCTTCCGCTTCAATTCAGCACTGCATGCTAGTCTCACACAGACATAATCCAGGAGAACCCGA
 10 TGCTGTAAAAAABAAAAAABAGATGAGGTAGTTGGGGAGATTGCTCAACAGTTAAAAAACAATGGT
 GCTCTCCGAAGGATCCAGGTTGATTCCTAGAACAACATGGTAACCTCACTAGCTATATTTCAA
 TCTAGGGGATCCAGTGCCATCTGGGGCTCCATGGACACTTCTCCCTTGTGGTGAACAGGCATAG
 ATACAGCCAGAACATTCATACATATAAAAAAATAAAGGTTTTTACACATAAAAAAATAAA
 GCTCTCGAAGGAGCCTGAGTTCAATTACTAACACTGCACCCGAGGTTCAACAATCCAGCTCGAA
 15 GGGATCTGAACCTTTCTCATTGCTCAGGAGTACCAGCACTTGTGGGTTGTACTCACATACAG
 ATACAGACATCATAGTACACCTAATTAAGAAGAGTCACTTGGAGTGTGGCACAGCCCTTAA
 ATCCCAATATTAGGAACAACAGGAGGTTGGGTTCTCAAGTTCAAGGCCAACCTTGTCTACAGCAT
 GAGTCCAGAACAGCCAGGATACATAAAAAAGAGGTTGTGGGGTTGGGGATTTAGCTCAGCGG
 TAGAGCGCTTGGCTAGCAAGTGCAGGCCCTGGGTTCCGTTCCCGAGTCCGAAAAAATAAAG
 20 AGTCCCTGTGTAACAAAACAAAAGACAAAGCAAAAGATTAATGTTGGGACCTGGATGACCTGAGATTGCT
 CTACTTTCTTTCTTTTGGGGACGGTTTTATATGTGACCATGGATGACCTGAGATTGCTGTT
 GTAGATAGCTTGCCTGAACTTTTTTTCCCTGGAGCTGAGGACCTAACCCAGGTTGCTGGGT
 TTAAGGCAGCGCTTACACACTGAGCTAAATCCCAACCCCACTTCACTTTTAGGATACCA
 AGCAGACTCTTGGCTTAGGAACAACCTCAGCTCCGGACTTTTTTTTTTACACTAGGTTCCG
 25 CTCTGTTAGACTAGACTCTTCCACCCCTCAGTACATTAATACTAGGACACTAGGACAAACCAT
 AGCAAACTCTGTACAGCACAGCTGACAGCCCTAAGCCGACTCCATCTTTCTTTTTTTTTTAA
 ATATATATTTATTTATGATATAGTACACTGTCATGTTCTCAGACACACAGAGAGGGGACCTG
 GATCCCATACAGATGCTTGTGAGCCACATGTGGTTGCTGGGAATTGAACCTCAGGACCTCTGGGA
 GAGCAGTCACTGCTCTTAAACCGCTGAGCCATCTCTCCAGCCCCCACTGAAGACTTTGATCTGGTT
 30 ACCAICTGACCCCAATCTCTGCAAAAGCCCTCCCTTCCCTCGAAGAACTCTTACGCTCTTTA
 TGCTTGGCCCATGACTTGTATTAATCAGCAACAATGACAAGACCTGTATGTTCTCCCTAGC
 TCAGAAGCAGATCCTTGTCTCTGTTAATGTTTGTATTTCTGGTCTGCTCCGTTGGGGACAGTCTG
 ATAGTCTTAGACTGATAGCTTGGGGATTCTAAACTCACACAGGGGCTATGTTACCAGTGGGC
 ACATACAGAGCTGGCATTTGCTTTGGAGTGGGACCATATCTTGACAGAAAGATTAACATAAACC
 35 CTAGCTGTGATGCTCCGGGATCCATGCTAATGAAACACTGCCACCGCCCTCAGGAACTTCTC
 ACAGAGTCTGGCTCTGGAACTGCTGTGAACCTCTACTGTCCACCTGCAGCAGCCATACCGA
 AATCAGCTCAATAACCTCTCAACTTCTGCATCTTGTCTTGGTGAACCTTTCCGCTCCAATGT
 CATGACCTTTCAAAGTCACTCATAGCAGTCTGCAGGAGAACAGGTAATTCAGGGCTGGGA
 40 TTAGCTCAGTGTAGAGCGCTTACCAGGAGCGCAAGGCCCTGGGTTCCGTTCCCGAGCTCCGGA
 AAAAAAAGAACAAAAAANAAAAAANAAAAAAGAGGACAGGTAATTCAGCTAAGACTGGTACACA
 AGTGAATTTTAAACTTAGGAGGTTGAGCCGAGCGCATCTGGAGTTGGATTAACTGGACTCCA
 TAGTGAATTTGGGCTAGCTTAGGCTACATAAGCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT
 CTCTATCTCTCTCTCTCTCTCAACCACAAAAGAGAGAACGGAAAAAAGGAGAAATTAAGAGAA
 45 AGAAAAACAAAAGAAATTTCTTAAGCAAGCATAATTTAATTTATTTATTTGTTTTTCAAGACA
 GTGTTGTCTATGTAGCATTGGCTGTCTAGAACAAATCGTTGTAGCCAAAGTGGCTTGAACCTA
 TAGGCCTGCTTTGGCTTCCAAATCTGGAATTAAGCCTTGTGGCAGCACTGCCAGCGACACT
 GGAATTTTAAAAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT
 50 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT
 CTTCACAGAGTGTCTCCAAACCACCCACCCCTCTAATTTTATTTATTTATTTATTTATTTATTT
 CTT
 CTTCCTAGGTAAGCGCTTACCAGTGAAGTCCCGAGCCCTACATTCCTCTTTCTACTTCT
 TTGGCACAGACTTGGAGGTTGCAAAATCAAGAGACAGCTTTCTTTCTTTTGTGATGCCAATCT

WO 02/057307

PCT/GB02/00215

ACTTAAAGTGAACCAAGCCATGCTTTAAATCGTAAGGACTGGGAGGCAGTCAGGCACATATC
 CAGGTTCCAGACCAGCCCTGATGTATGTAATGAGTCCAGACCAATTAGGGCTATATCATCAGACCA
 TGCTCAAACCAAAAACAAAAGAAAAGAAAAGAAAAGAACATCAAGTCAAGCATGATATAAT
 CACATAATCCTATAATCCTAATATGGGAGGCTGAAGCAGAAATGGCCATGGCTTAGCTTAGCC
 5 TGGCAGGACCAACCACTGGGTACACAGGAATACATAATACACTGCCATAGAAAACAAAGCATG
 GCTGACTTCGCTCACTGCTAGTTGGGGCTTGGGTTAGGCTTTTCAAACACTAAGCAATTTGGTTC
 GGAGCTAGTTTTGAGCCCTCTGCCACCGCCATGGAGGAGCCACAGAGAAAGTCAGCCAGTTG
 TAGTCCCAGAAAGCTCCTCAATGAAGATGACGAGGACCGCTCCGCTGATTCAGAAGTCCTACAAC
 CAGAAACACTAGTAAAGGTGATGAAAACGGTAACCTGAACCCAGTGCAGGAGGTCAGCAGCTC
 10 ATCAGAGCCCTCAGTGTCCGGATTCAGGGCAGGCCCTGGGAGAACAGATGTGAAGAACTCTTAGGG
 AAGTTCAAAGCCCTTCCAGAGAGAGGGTCCACACATTTGTTGGTGGTCTGAGAGATGCCGGAGCA
 AGGATGAGAGATTTGTTGGGATTTGACGAGAGACAAAGCCCTGAAGGAAATGAGTAGGAGGG
 AAGAGTGAGCCACTCAGACGCTCTCTGCTTCTGCCATCGTCAGAGATGGAATCCGCTAAGAAA
 GCTAAAATCCGGAGAAATAGGACAGTCCGGTTTATGACACTATCCTTGTCTCATGATGCCATG
 15 CAGCAGACTGAAAACCTGGTTTTGTTTTTAAAGATAAACTTTCCCTGGTGGTGGGAGACAGT
 CTTGTAACTTTCAACTATGTAGGAAGTGTGACGGTTGAATTCATGTGAAGACTTAAATTTACC
 CAAATGATGGAGAAAGTAAAGCAATCTGTGAACCTTAGAAGCCTCAAGCTGGGGCTGAGAAA
 CACTTAACATAGAAATTTGGGGTAGTTTGGCTTAGAAGGTAAATGGAATAGCCCTTTGGATTTCTA
 GTTTCCGAAATGTGTAATAAAGCAATTTGTATCTTAAACAAACACACAGAACAGATAGAAT
 20 GACCATTTGAGATGGGGGTTGTTTTACAGAGACACTGTGCGTCCGCCACTCCTCATGATGCCA
 GATTCACATCTGCTGCTAACCCGTTTTATTTCTGCTCCAGGCGGGTCTCCATGACCTAGCCCA
 GCTCTCAGCTCGTGGTCTGCTCCCTTGTGGCCAGTTTTGAGCCACAGGCTTACAGCAAG
 ATCTAGAAAATGCTGTCTGATTTTGTGTTTGTTCATGCTGTGTAATAAAAAGAACAAATTTGTTG
 25 ATGTATTCCTAAATTTAAAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA
 AAGCTCAGAAAGCCAGAGCGGTTGGATCTCTGAATTCATGGCCAGCCAGGGCTACACAGAAAAC
 CCTGTCTTGAAGAAAGAGACTTGGGGGTTGGGATTTGGCTCAGTGGTAGAGCCCTGCTACCCCT
 GGGTCCGCTCCCAAGCTCCGAAAAGAAATGAGAAAAGAAAAGAAAAGAAAAGAGACTCGTA
 ACCAAGCAAGCTTGGTACTTAAGAAATGAGAAATCCTTAGAGCTACCTTAGAGCTAGAAAAGG
 30 CAGGACATTTAGGAGAGAGCTGGTACGGCAAGCCAAAGGCTCAGCCGCGGGTTATACCATGT
 AAGGTTATCCTCAGGGGCTGGAGAGAAATGCACAGCAACACTAACAGCTCATACTGTCTGGCCAA
 GTATCACTACCATGGCTTATAGATCCTGCTCTTGAAGAAAGGGTAGATCAAGGGGTAATCAAG
 GATAGATACCCCTTTGGCAATAGGACGGAGGGTGGTAGATCCCTCAACAGTGTGAGTAGTCC
 AAGATATGAATCACTATGGCTCCTAATAAACACTGCTTAGGCTAATTTACCATTGAGCTACATCC
 35 CAAATTCAAAAGTTGTTTTGGAGAGGGATGCAATGGGAGACAGGTTCTAATGTGAATCTTACTG
 TCCTGGAACTCCCTCCATAGACCTGCTGCTTGAACCTACAGAGTTCTCACAGGAGACTTAACT
 GCCTTTGTCTCCAAAGTGGTGGATCAAAGGCGTGCACCACCACTCCAGCCCTTATTTAATTAAT
 TATATCAATTTAATTAATTAATTAATTAATTAATTTAATTTAGTTTTGATCATATTTATCGATGAT
 TATGGAAGTGGGCCCTGTCATGTCATCTTGTGGTAAAGGTCAGGAGATAAATACTACTTGGT
 40 AAATAAGAAAACCCAAAGTTAAGAAAGATGGAGAAAAAACAATATATATCTAAAAAAGAAAAA
 ACTTGGCTTTTAAAAAATAAATACAGGGGGCTGGGATTTAGCTCAGTGGTAGAGCCCTTACCTA
 GGAAGCACAAAGCCCTGGTTCGGTCCCAAGCTCTGAAAAGAAAGCAAAAAAAGAAAAGAAAA
 AAGAAAATACAGGGCTGGAGAGATGCTCAGCGGCTAAGAGCACAGACTGCTCTCCAGAGGTCCTG
 AGTTCAAATCCAGCAACACATGTTGGCTCACACCAATTTGTAATGGGATCTGATGCCCTCTCTCT
 45 GGTGTCTCTGAAGACAGTACACTGTACATGAATACATAAATAAATCTTTAAAAAATGAAAAAT
 AAAATACATCTCATATGATTTATCAAAAAAATACTACTTGGACAGGGTGGAGATTTAGCTCA
 GFGGGCAGGACTTGCCTAGCAAGTCAAGACCCCTGGTTCGGTCCCTCAGCTCTGAAAAAATAAAT
 TACTACTTGGAGAGTGGTTCTCCCTTCCACTCAAGTTGTAGAAATCCAACTTAGATCTCAGGA
 50 GGCAGACTCCTTACCACCGAACTTAAGATTTGGTTTTGAAAGTCTTCTAGAGACCGAGGCTATC
 CTGAAATCARGATTTAATTTACCCAGCTCCAAAAAAGAAAAGAAATTAATTTAAAGTAGCTG
 TTCCATGCTTTGATCCAGCACTTGGACAAGAGAGGAGATGCAGGTTGGTGTGAGTTTGG
 ATCAGTCTCAAGCTTGGTCCACATGAAAAGTTCTAGAACAGCCAGGCTTCATGAGATCGTCTCT
 CAAAACAGCAAGCACTGACGATGACGTCATGATGATGAGCAACATAGACTCAAGCCCTCTAGGC

WO 02/057307

PCT/GB02/00215

73

CLAIMS

1. A GCR1 polypeptide, or a fragment, homologue, variant or derivative thereof
2. A polypeptide according to Claim 1, which has at least 50%, 60%, 70%, 80%, 90% or 95% homology to a sequence shown in SEQ ID NO: 2.
- 5 3. A GCR2 polypeptide, or a fragment, homologue, variant or derivative thereof
4. A polypeptide according to Claim 3, which has at least 50%, 60%, 70%, 80%, 90% or 95% homology to a sequence shown in SEQ ID NO: 4.
5. A nucleic acid encoding a polypeptide according to any preceding claim.
6. A nucleic acid having at least 90% homology with the sequence set forth in SEQ
10 ID NO: 1, or a fragment, variant or derivative thereof.
7. A nucleic acid having at least 75% homology with the sequence set forth in SEQ
ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8 or SEQ ID NO:
9, or a fragment, variant or derivative thereof
8. A nucleic acid comprising a sequence of 25 contiguous nucleotides of a nucleic
15 acid according to Claim 5, 6 or 7.
9. A nucleic acid comprising a sequence of 15 contiguous nucleotides of a nucleic
acid according to any of Claims 5 to 8.
10. The complement of a nucleic acid sequence according to any of Claims 5 to 9.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

74

11. A nucleic acid according to any of Claims 5 to 10, comprising one or more nucleotide substitutions, wherein such substitutions do not alter the coding specificity of said nucleic acid as a result of the degeneracy of the genetic code.
12. A polypeptide encoded by a nucleic acid according to any preceding claim.
- 5 13. A polypeptide according to Claim 12, in which the polypeptide comprises a sequence shown in SEQ ID NO: 2 or SEQ ID NO: 4.
14. A method for identifying a pluripotent cell, comprising detecting the presence of a polypeptide according to any of Claims 1 to 4, 12 or 13 or the expression of a nucleic acid according to any of Claims 5 to 11, or a homologue thereof.
- 10 15. A method according to Claim 14, comprising the steps of amplifying nucleic acids from a putative pluripotent cell using 5' and 3' primers specific for GCR1 and/or GCR2, and detecting amplified nucleic acid thus produced.
16. A method according to Claim 14, wherein the expression of the nucleic acid sequence is detected by *in situ* hybridisation.
- 15 17. A method according to Claim 8, wherein the expression of the nucleic acid sequence is determined by detecting the protein product encoded thereby.
18. A method according to Claim 14 or Claim 18, wherein the protein product is detected by immunostaining.
19. An antibody specific for a polypeptide according to any of Claims 1 to 4, 12 or 13.
- 20 20. An antibody according to Claim 19, which is capable of specifically binding to an extracellular domain of GCR1.

SUBSTITUTE SHEET (RULE 26)

WO 02/057307

PCT/GB02/00215

75

21. Use of an antibody according to Claim 19 or Claim 20 for the identification and/ or isolation of a pluripotent cell.
22. A pluripotent cell identified by a method according to any one of Claims 14 to 18 and 21.
- 5 23. A method for isolating a gene specifically expressed in a pluripotent cell, comprising the steps of:
- (a) providing a population of cells containing a pluripotent cell;
 - (b) isolating one or more pluripotent cells therefrom and providing single-cell pluripotent cell isolates;
 - 10 (c) amplifying the transcribed nucleic acid present in a single pluripotent cell;
 - (d) conducting a subtractive hybridisation screen to identify transcripts present in pluripotent cells but not in somatic cells; and
 - (e) probing a nucleic acid library with one or more transcripts identified in (d) to clone one or more genes which are specifically expressed in pluripotent cells.
- 15 24. A method according to any of Claims 14 to 18 or 23, a use according to Claim 21, a pluripotent cell according to Claim 23, in which the pluripotent cell is selected from the group consisting of: a primordial germ cell (PGC), an embryonic stem cell (ES) and an embryonic germ cell (EG).

WO 02/057307

PCT/GB02/00215

1/9

```

GCGGAGAAAGGGCAGACCCGCGCGGCTCCATCCCTTTGCCCTCCAGTGTGCTGCCFTTGC 60
TCCGCACCATGAACCACACTTTCAGCCCTTCATCACCGCTGCCAGTGGAGGACAGCCCC 120
M N H T S Q A F I T A A S G G Q P P

CAACTACGAAAGCAATCAAGGAAGAATAATGAGGTGGCTGAGATGGGGCAACCGCAAGGAT 180
N Y E R I K E E Y E V A E M G A P H G S

CGGCTTCGTGACAACTACTGTGATCAACATGCCAGAGGTTGGTGGCTGACCATG 240
A S V R T T V I N M P R E V S V P D H V

TGGTCTGGTCCCTGGTCAATACACTCTTCATGAACCTTCGCTGCCCTGGGCTTCATAGCCT 300
V W S L F N T L F M N F C C L G F I A Y
TM I

ATGCCACTCCGTTGAAGTCTAGGGATCGGAAGATGGTGGGTGATGTGACTGGAGCCAGG 360
A Y S V K S R D R K M V G D V T G A Q A

CCTACGCCCTCCACTGCTAAGTGCCTGAACATCAGCACCTTGGTCCCTCAGCATCCIGATGG 420
Y A S T A K C L N I S T L V L S I L M V
TM II

TTGTTATCACCAATTGTTAGTGTATCATCATTTGTTCTTAACGCTCAAAACCTTCACACTT 480
V I T I V S V I I I V L N A Q N L H T *

AATAGAGGATTCGACCTCCGCTCCGAGTGGCTTCACCCCTCCGAGCTGGTCCCTCCT 540
TGCCTCCCTACAGCGAGGTGTAACACTCATTTATCTATCCACAGTGGATTCAATAAAG 600
TGCACCTTGATAACCACC

```

FIGURE 1

WO 02/057307

PCT/GB02/00215

3/9

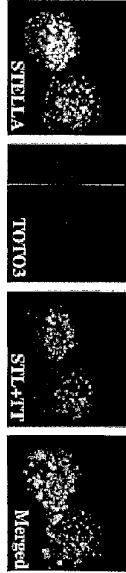


Figure 3

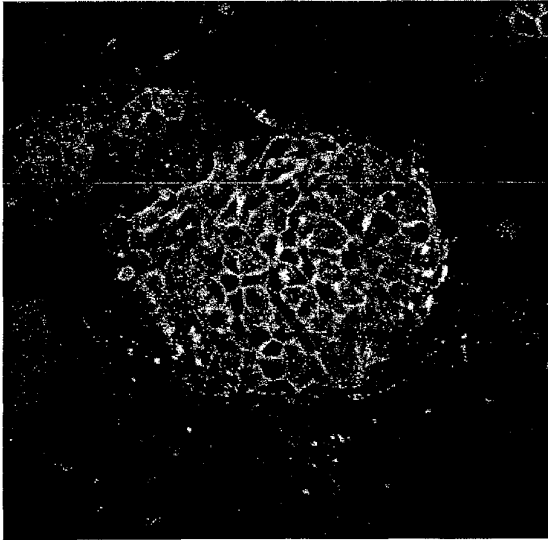


Figure 4

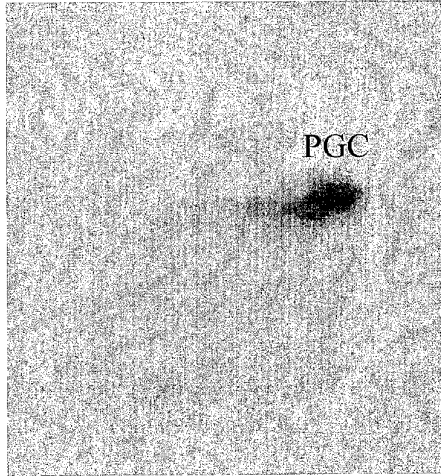


Figure 5

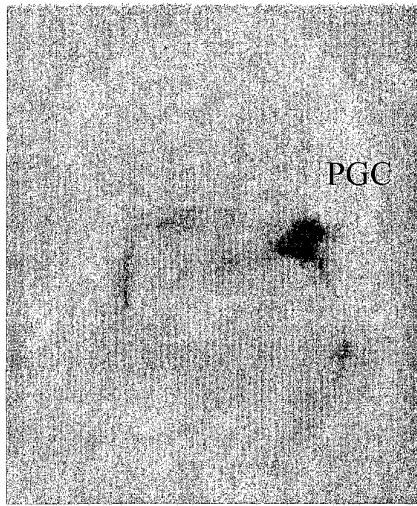


Figure 6

WO 02/057307

PCT/GB02/00215

7/9



FIGURE 7

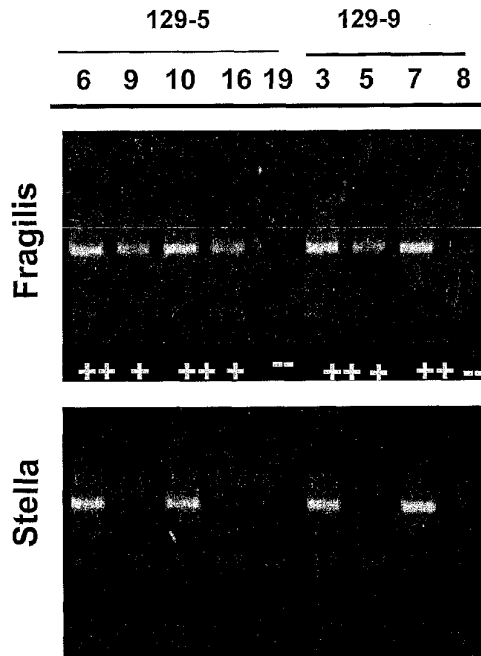


Figure 8A

【国際公開パンフレット(コレクトバージョン)】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
25 July 2002 (25.07.2002)

PCT

(10) International Publication Number
WO 02/057307 A3

- (51) International Patent Classification: C15N 15/12, C12N 5/06, 5/08, C07K 14/47, 14/705, 16/18, 16/28, C12Q 1/68, G01N 33/50
- (21) International Application Number: PCT/GB02/00215
- (22) International Filing Date: 18 January 2002 (18.01.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 0101300.2 18 January 2001 (18.01.2001) GB
- (71) Applicant (for all designated States except US): CAMBRIDGE UNIVERSITY TECHNICAL SERVICES LIMITED [GB/GB]; The Old Schools, Trinity Lane, Cambridge CB2 1TS (GB).
- (72) Inventors: and
(73) Inventors/Applicants (for US only): SAITOU, Mitinori [JP/JP]; Wellcome CRC Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR (GB); SURANI, Azim [GB/GB]; Wellcome CRC Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR (GB).
- (74) Agents: MASCHIO, Antonio et al.; D Young & Co, 21 New Fetter Lane, London EC4A 1DA (GB).
- (81) Designated States (national): AI, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IL, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Declaration under Rule 4.17:
— of inventorship (Rule 4.17(ii)) for US only
- Published:
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- (88) Date of publication of the international search report: 20 February 2003
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/057307 A3

(54) Title: GENES

(57) Abstract: The invention provides two primordial germ cell-specifically expressed genes, GCR1 (Fragilis) and GCR2 (Stella), which are markers for primordial germ cells and may be used to identify such cells in cell populations.

【 國際調查報告 】

INTERNATIONAL SEARCH REPORT		International Application No PCT/GB 02/00215
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C15N15/12 C12N5/06 C07K14/47 C07K14/705 C07K16/18 C07K16/28 C12Q1/68 G01N33/50		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C07K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, MEDLINE, BIOSIS, EMBL		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 38973 A (CORIXA CORP) 5 August 1999 (1999-08-05) SEQ ID NOs:119 and 124, SALT-T8 page 30, line 16 -page 31, line 12	1, 2, 5, 8-12, 19, 20
X	LEWIN A R ET AL: "MOLECULAR ANALYSIS OF A HUMAN INTERFERON-INDUCIBLE GENE FAMILY" EUROPEAN JOURNAL OF BIOCHEMISTRY, BERLIN, DE, vol. 199, no. 2, 1991, pages 417-423, XP000906854 ISSN: 0014-2956 figure 3A page 419, left-hand column, paragraph 1 page 421, right-hand column, paragraph 5 --- -/--	1, 2, 5, 8-12, 19, 20
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *C* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *Z* document member of the same patent family		
Date of the actual completion of the international search 4 December 2002		Date of mailing of the international search report 20/12/2002
Name and mailing address of the ISA European Patent Office, P.B. 5816 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2340, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Barnas, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 02/00215

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SHIMADA HIROYUKI ET AL: "Analysis of genes under the downstream control of the t(8;21) fusion protein AML1-MTG8: Overexpression of the TIS11b (ERF-1, cM61) gene induces myeloid cell proliferation in response to G-CSF." BLOOD, vol. 96, no. 2, 15 July 2000 (2000-07-15), pages 655-663, XP002222767 ISSN: 0006-4971 J66550, Table 1 page 657, left-hand column, paragraph 3 -right-hand column, paragraph 1 -& DATABASE EMBL 'Online! 31 January 2000 (2000-01-31) Database accession no. AB030416 XP002223547 ---	6,8-12, 19,20
X	DATABASE EMBL 'Online! 3 March 2000 (2000-03-03) Database accession no. AA059844 XP002222772 ---	5,6,8-11
X	DATABASE EMBL 'Online! 26 May 1992 (1992-05-26) Database accession no. X61381 XP002222773 ---	1,2,5, 8-12,19, 20
X	SASAKI NOBUYA ET AL: "Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones." GENOMICS, vol. 49, no. 2, 15 April 1998 (1998-04-15), pages 167-179, XP002222768 ISSN: 0888-7543 01B00056N105, 01B00058P605 page 168, right-hand column, paragraph 5 page 178, left-hand column, paragraph 2 -& DATABASE EMBL 'Online! 28 May 1998 (1998-05-28) Database accession no. C89109 XP002223548 abstract -& DATABASE EMBL 'Online! 28 May 1998 (1998-05-28) Database accession no. C89043 XP002223549 ---	7-12,19, 20
X	DATABASE EMBL 'Online! 4 March 2000 (2000-03-04) Database accession no. AI606389 XP002222774 ---	7-12,19, 20
	-/--	

Form PCT/IS/M210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		International Application No PCT/GB 02/00215
C/(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YEOM YOUNG IL ET AL: "Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells." DEVELOPMENT (CAMBRIDGE), vol. 122, no. 3, 1996, pages 881-894, XP002222769 ISSN: 0950-1991 cited in the application page 887, right-hand column, paragraph 2 figure 4	1-13,19, 20
A	NURCOMBE VICTOR ET AL: "MK: A pluripotential embryonic stem-cell-derived neuroregulatory factor." DEVELOPMENT (CAMBRIDGE), vol. 116, no. 4, 1992, pages 1175-1183, XP002222817 ISSN: 0950-1991 page 1177, right-hand column, paragraph 2 -page 1178, left-hand column, paragraph 2 page 1180, right-hand column, paragraph 2 -page 1181, left-hand column, paragraph 1	1-14
P,X	RIKEN GENOME EXPLORATION RESEARCH GROUP PHASE II TEAM AND FANTOM CONSORTIUM: "FUNCTIONAL ANNOTATION OF A FULL-LENGTH MOUSE CDNA COLLECTION" NATURE, MACMILLAN JOURNALS LTD. LONDON, GB, vol. 409, no. 6821, 8 February 2001 (2001-02-08), pages 685-690, XP001009930 ISSN: 0028-0836 page 685, left-hand column, paragraph 4 -& DATABASE NCBI 'Online! 8 February 2001 (2001-02-08) Database accession no. AK010723 XP002223550 -& DATABASE NCBI 'Online! 8 February 2001 (2001-02-08) Database accession no. AK003407 XP002223551 -& DATABASE EMBL 'Online! 8 February 2001 (2001-02-08) Database accession no. AK007919 XP002223552	1,2,5,6, 8-13,19, 20
P,X	DATABASE NCBI 'Online! 15 September 2001 (2001-09-15) Database accession no. AC094826 XP002222775 cited in the application	7-11
	-/--	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 02/00215

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE NCBI 'Online! 20 December 2001 (2001-12-20) Database accession no. AC097234 XP002222776 cited in the application ---	7-11
P,X	DATABASE NCBI 'Online! 20 December 2001 (2001-12-20) Database accession no. AC093991 XP002222777 cited in the application ---	7-11
P,X	DATABASE NCBI 'Online! 21 December 2001 (2001-12-21) Database accession no. AC103122 XP002222778 cited in the application ---	7-11
P,X	DATABASE NCBI 'Online! 12 January 2002 (2002-01-12) Database accession no. AC106604 XP002222779 ---	7-11
T	SAITOU MITINORI ET AL: "A molecular programme for the specification of germ cell fate in mice." NATURE. ENGLAND 18 JUL 2002, vol. 418, no. 6895, 18 July 2002 (2002-07-18), pages 293-300, XP002222771 ISSN: 0028-0836 page 293, right-hand column, paragraph 4 -page 295, right-hand column, paragraph 1 -----	1-24

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT			International Application No. PCT/GB 02/00215	
Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 9938973 A	05-08-1999	AU 2344399 A	16-08-1999	
		BR 9907259 A	22-01-2002	
		CA 2319117 A1	05-08-1999	
		CN 1292029 T	18-04-2001	
		EP 1051489 A2	15-11-2000	
		JP 2002516659 T	11-06-2002	
		NO 20003853 A	27-09-2000	
		WO 9938973 A2	05-08-1999	
		ZA 9900693 A	21-10-1999	

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
G 0 1 N 33/48	G 0 1 N 33/53	D
G 0 1 N 33/53	G 0 1 N 33/53	M
G 0 1 N 33/566	G 0 1 N 33/566	
	C 1 2 N 5/00	E

(81) 指定国 AP(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW

(74) 代理人 100094400
弁理士 鈴木 三義

(74) 代理人 100107836
弁理士 西 和哉

(74) 代理人 100108453
弁理士 村山 靖彦

(74) 代理人 100110364
弁理士 実広 信哉

(72) 発明者 斎藤 通紀
イギリス・ケンブリッジ・シービー２・１キューアール・テニス・コート・ロード・(番地なし)
・ユニバーシティ・オブ・ケンブリッジ・ウエルカム・シーアールシー・インスティテュート

(72) 発明者 アジム・スラニ
イギリス・ケンブリッジ・シービー２・１キューアール・テニス・コート・ロード・(番地なし)
・ユニバーシティ・オブ・ケンブリッジ・ウエルカム・シーアールシー・インスティテュート

F ターム(参考) 2G045 AA24 CB01 DA13 DA36 FB02 FB03 FB08
4B024 AA01 AA11 BA80 CA01 CA09 CA11 DA02 DA05 DA11 EA02
EA04 GA11 HA01 HA03 HA11
4B063 QA01 QA18 QQ08 QQ42 QQ52 QR08 QR42 QR55 QR62 QS25
QS34 QS36 QX02
4B065 AA01X AA57X AA90X AA91Y AB01 BA01 CA24 CA25 CA44 CA46
4H045 AA10 AA11 AA20 AA30 BA10 CA40 EA20 EA50 FA72 FA74

专利名称(译)	基因		
公开(公告)号	JP2004529617A	公开(公告)日	2004-09-30
申请号	JP2002558380	申请日	2002-01-18
[标]申请(专利权)人(译)	剑桥大学技术服务有限公司		
申请(专利权)人(译)	剑桥大学技术服务Rimitido		
[标]发明人	斎藤通紀 アジムスラニ		
发明人	斎藤 通紀 アジム・スラニ		
IPC分类号	G01N33/48 C07K14/47 C07K14/705 C07K16/18 C12N5/06 C12N5/08 C12N15/09 C12Q1/68 G01N33/53 G01N33/566		
CPC分类号	C07K14/705 C07K14/47		
FI分类号	C12N15/00.ZNA.A C07K14/47 C07K16/18 C12Q1/68.A G01N33/48.M G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.E		
F-TERM分类号	2G045/AA24 2G045/CB01 2G045/DA13 2G045/DA36 2G045/FB02 2G045/FB03 2G045/FB08 4B024/AA01 4B024/AA11 4B024/BA80 4B024/CA01 4B024/CA09 4B024/CA11 4B024/DA02 4B024/DA05 4B024/DA11 4B024/EA02 4B024/EA04 4B024/GA11 4B024/HA01 4B024/HA03 4B024/HA11 4B063/QA01 4B063/QA18 4B063/QQ08 4B063/QQ42 4B063/QQ52 4B063/QR08 4B063/QR42 4B063/QR55 4B063/QR62 4B063/QS25 4B063/QS34 4B063/QS36 4B063/QX02 4B065/AA01X 4B065/AA57X 4B065/AA90X 4B065/AA91Y 4B065/AB01 4B065/BA01 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/CA40 4H045/EA20 4H045/EA50 4H045/FA72 4H045/FA74		
代理人(译)	渡边 隆 正和青山 村山彦		
优先权	2001001300 2001-01-18 GB		
其他公开文献	JP4704666B2		
外部链接	Espacenet		

摘要(译)

本发明涉及两种基因，GCR1 (Fragilis) 和GCR2 (Fragilis)，它们是在生殖细胞中特异性表达的原始生殖细胞的标记，可用于在细胞群中鉴定这些细胞。斯特拉) 提供。本发明涉及GCR1多肽，或其片段，同源物，变体或衍生物，其与SEQ ID NO：2中所示的序列具有至少50%，60%，70%，80%，90%或95%的同源性。2.根据权利要求1的多肽，其具有同源性。所述GCR2多肽或其片段，同源物，变体或衍生物，特别是至少50%与SEQ ID NO：2所示序列：95%的同源性4，60%，70%，80%，90%或具有SEQ ID NO：3所示氨基酸序列的多肽。还提供了编码这些多肽的核酸。

脂肪族	非極性	GAP
		ILV
	極性-非荷電	CSTM
		NQ
	極性-荷電	DE
		KR
芳香族		HFWD

