

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 December 2010 (09.12.2010)

PCT

(10) International Publication Number
WO 2010/141362 A1

(51) International Patent Classification:

C12Q 1/68 (2006.01) *C12P 19/34* (2006.01)
C07H 21/00 (2006.01)

(21) International Application Number:

PCT/US2010/036606

(22) International Filing Date:

28 May 2010 (28.05.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/182,735 31 May 2009 (31.05.2009) US

(71) Applicants (for all designated States except US): **THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA** [US/US]; 3160 Chestnut Street Suite 200, Philadelphia, PA 19104 (US). **FRED HUTCHINSON CANCER RESEARCH CENTER** [US/US]; 1100 Fairview Avenue North, Seattle, WA 98109 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **NATHANSON, Katherine, L.** [US/US]; 421 Curie Blvd, 351 BRB II/III, Philadelphia, PA 19104 (US). **KANETSKY, Peter, A.** [US/US]; 1134-2 Waverly Street, Philadelphia, PA 19147 (US). **SCHWARTZ, Stephen** [US/US]; 1100 Fairview Ave N., Seattle, WA 98109 (US).

(74) Agents: **BAK, Mary, E.** et al.; Howson & Howson LLP, 501 Office Center Drive, Suite 210, Fort Washington, PA 19034 (US).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every

kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: COMPOSITIONS AND METHODS FOR DIAGNOSING THE OCCURRENCE OR LIKELIHOOD OF OCCURRENCE OF TESTICULAR GERM CELL CANCER

(57) Abstract: A diagnostic composition useful for diagnosing testicular germ cell cancer or tumors or for providing an assessment of the susceptibility of a subject to this cancer includes at least one reagent capable of identifying a genetic variant of the *KITLG* gene or in the *KITLG* genomic region or of the *SPRY4* gene or of the *SPRY4* genomic region. The variant is associated with susceptibility of a human subject to testicular germ cell cancer or tumors. Methods of diagnosis include identifying such genetic variants in a biological sample of a male subject. These variants include one or more single nucleotide polymorphisms (SNP) identified within certain genomic marker sequences of previously unknown function.



WO 2010/141362 A1

COMPOSITIONS AND METHODS FOR DIAGNOSING
THE OCCURRENCE OR LIKELIHOOD OF OCCURRENCE OF
TESTICULAR GERM CELL CANCER

5 STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with government support under Grant Nos. R01CA114478 and R01CA085914 awarded by the National Institutes of Health. The government has certain rights in the invention.

10 BACKGROUND

Testicular germ cell tumors (TGCT) are the most common cancers in young men in the United States, with a peak incidence among those aged 25 to 34 years. The age-adjusted incidence in white men has doubled since 1975 and is now 6.6 per 100,000. The incidence in white non-Hispanic men is nearly fivefold higher than among black men.

15 See, e.g., Ries, L. et al. *SEER Cancer Statistics Review, 1975-2005* (National Cancer Institute, Bethesda, Maryland, 2008). The reasons for the increasing incidence and racial disparity in TGCT rates are unknown.

Although environmental exposures have been postulated to have a role in the increasing incidence of TGCT, there also is evidence for a substantial genetic
20 contribution to TGCT susceptibility. Brothers of individuals with TGCT have an 8- to 12-fold increased risk of disease, with the risk to monozygotic and dizygotic twins 75- and 35-fold increased, respectively, and fathers of affected individuals have a fourfold increased risk (Swerdlow, A.J., et al, 1997 *Lancet* 350, 1723-1728; Hemminki, K. & Li, X., 2004 *Br. J. Cancer* 90, 1765-1770). Consistent with the high familial risks compared
25 to most other cancer types and the ancestry-related differences in TGCT risk, the proportion of TGCT susceptibility accounted for by genetic effects is estimated at 25%, and TGCT has the third highest heritability among all cancers (Czene, K., et al, 2002 *Int. J. Cancer* 99, 260-266).

Results from linkage studies and candidate gene approaches have produced
30 limited insight into TGCT susceptibility factors. For example, an initial report of linkage on Xq27 was not replicated, nor have other loci been identified with significant effects.

This suggests that multiple loci, potentially of weak to moderate effect, contribute to disease susceptibility (Rapley, E.A. et al. 2000 *Nat. Genet.* 24, 197-200; Crockford, G.P. et al. 2006 *Hum. Mol. Genet.* 15, 443-451). The gr/gr deletion on the Y chromosome, studied as a candidate region, increases TGCT risk two- to threefold, but carriage
5 frequency of this variant is low (2-3%), suggesting that it likely accounts for only a small component of risk (Nathanson, K.L. et al. 2005 *Am. J. Hum. Genet.* 77, 1034-1043).

Thus, no genetic risk factor has been identified that can explain an appreciable proportion of TGCT cases. There remains a need in the art for compositions and methods that can identify persons at risk for TGCT, and thereby permit early therapeutic
10 intervention.

SUMMARY

The need in the art as stated above has been met by the compositions and methods described herein. These compositions and methods are based upon the
15 identification of common genetic variants of the *KITLG* and *SPRY4* genes that affect the risk of occurrence and susceptibility to TGCT.

In one aspect, a diagnostic composition includes a reagent that is capable of identifying a single nucleotide polymorphism (SNP) associated with susceptibility of a human subject to TGCT. In one embodiment, such a reagent is a nucleotide sequence or
20 primer capable of hybridizing to, and identifying, one or a combination of SNPs in a sample of the subject's genome. In another embodiment, the SNP is one or more of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952 which map to the *KITLG* (c-KIT ligand) gene region on chromosome 12q22. In another embodiment, the SNP is one or more of rs12521013, rs4324715, rs6897876 which map to
25 2.4 kb downstream of the *SPRY4* (sprouty homolog 4) coding region on chromosome 5q31.3. The use of such a composition in an appropriate diagnostic assay allows for the identification of one or more of the SNPs disclosed herein that are characteristic of increased risk of the occurrence, or the occurrence itself, of TGCT.

In one aspect, a diagnostic composition includes a reagent that is capable of
30 identifying in a biological sample of a human subject a SNP associated with susceptibility of a human subject to TGCT. In one embodiment, the SNP is one or more

of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952 which map to the *KITLG* (c-KIT ligand) gene region on chromosome 12q22. In another embodiment, the SNP is one or more of rs12521013, rs4324715, rs6897876 which map to 2.4 kb downstream of the *SPRY4* (sprouty homolog 4) coding region on chromosome
5 5q31.3. In one embodiment, such a reagent is a nucleotide sequence, such as a genomic probe that hybridizes to the SNP cDNA or mRNA. In another embodiment, the reagent is a nucleotide primer, a nucleotide probe or a set of such primers or probes, capable of amplifying a polynucleotide sequence or mRNA containing a SNP and identifying it. These diagnostic reagents are capable of identifying, one or a combination of SNPs in a
10 biological sample containing the subject's genome. Such reagents are optionally associated with detectable labels that enable the ready identification of the SNP in a biological sample from the subject. In other embodiments, such diagnostic reagents are optionally immobilized on a substrate. The use of such a composition or reagent in an appropriate diagnostic assay allows for the identification of one or more of the SNPs
15 disclosed herein that are characteristic of increased risk of the occurrence, or the occurrence itself, of TGCT.

In another aspect, a diagnostic composition includes multiple reagents identified above in a microarray. In another aspect, a diagnostic composition includes multiple reagents identified above immobilized on a microfluidics card. In another aspect, a
20 diagnostic composition includes multiple reagents identified above on a computer-readable chip or chamber.

In another aspect, a diagnostic kit is provided that includes one or more reagents that are capable of identifying a SNP or combination of SNPs associated with susceptibility of a human subject to TGCT. In one embodiment, such a kit contains for
25 each SNP at least one reagent, e.g., a nucleotide probe sequence or set of primers capable of hybridizing to, and identifying, one or a combination of SNPs in the subject's biological sample. The use of such a composition in an appropriate diagnostic assay allows for the identification of one or more of the SNPs disclosed herein that are characteristic of increased risk of the occurrence, or the occurrence itself, of TGCT.

30 In yet a further aspect, a diagnostic composition provides a combination of two or more reagents that form a hybridization complex or other physical association with one

or more of the specifically identified SNPs when such SNPs are present in a biological sample. Such a combination may be immobilized on a substrate for subsequent evaluation by an instrument suitable for detecting the formation of such associations and revealing the presence of a genetic profile characteristic of the occurrence or high risk of occurrence of TGCT.

In another aspect, use of one or more of the compositions described above for the identification of one or more SNPs in a biological sample for identification of risk of TGCT is provided.

In still another aspect, a diagnostic method is provided that permits the identification of one or more SNPs in a biological fluid or tissue sample of a subject, which SNP(s) are characteristic of increased risk of the occurrence, or the occurrence itself, of TGCT. In one embodiment, the SNP is one or more of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952 which map to the *KITLG* (c-KIT ligand) gene region on chromosome 12q22. In another embodiment, the SNP is one or more of rs12521013, rs4324715, rs6897876 which map to 2.4 kb downstream of the *SPRY4* (sprouty homolog 4) coding region on chromosome 5q31.3. Such a method may include a variety of known assay formats capable of identifying the presence of the characteristic SNPs. In certain embodiments, such a method may employ use of certain machines or computer-programmed instruments that can transform the detectable signals generated from the diagnostic reagents complexed with the SNPs present in the biological sample into numerical or graphical data useful in performing the diagnosis. In one embodiment, the diagnosis based on the identification of the selected SNPs is associated with the presentation of certain clinical symptoms in a subject. In another embodiment, the diagnosis provides a quantitative assessment of the likelihood of TGCT occurrence in a subject that has not yet developed clinical symptoms of TGCT. In other embodiments, such methods employ the reagents described herein.

Other aspects and advantages of these methods and compositions are described further in the following detailed description of the preferred embodiments thereof.

30

DETAILED DESCRIPTION

The compositions and methods described herein provide means for diagnosing or identifying the occurrence of, or the likelihood of an increased susceptibility to the occurrence of, testicular germ cell cancer or tumors in a subject, based upon the presence
5 of certain single nucleotide polymorphisms in the genome of the subject.

A. The Relevant SNPs For Identification by the Methods and Compositions

Single nucleotide polymorphisms ("SNPs") are one of the major forms of sequence variation or mutation in the human genome. A SNP is a nucleotide position in
10 a coding or non-coding region of the genome at which at least two alternative bases can occur. SNPs make up about 90% of all human genetic variation and are widespread throughout the genome, *i.e.*, SNPs occur every 100 to 300 bases along the 3-billion-base human genome. Each alternative base occurs at an appreciable frequency (*i.e.*, >1%) in the human population. Two of every three SNPs involve the replacement of cytosine (C)
15 with thymine (T). An allelic SNP occurs when, due to the existence of the polymorphism, some members of a species have the unmutated sequence (*i.e.*, the ancestral, or major, "allele"), while other members of the same species have a mutated sequence (*i.e.*, the variant mutant, or minor, allele).

Each SNP is identified by a specific number (designated "rs#####", hereinafter
20 "rsnumber"). This rsnumber is specific to only one SNP. As used herein, the term SNP may refer to one or more of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, rs12521013, rs4324715, rs6897876. The sequence of each SNP reference marker sequence, which includes the nucleotides flanking the SNP, is provided in the attached sequence listing. These sequences are based on the rsnumbers found at
25 the publicly available NCBI dbSNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). In another embodiment, the SNP reference marker sequences are identified by the publicly available Affymetrix Genome-Wide Human SNP Array 6.0 (www.affymetrix.com/products_services/arrays/specific/genome_wide_snp6/genome_wi
30 [de_snp_6.affx#1_1](http://www.affymetrix.com/products_services/arrays/specific/genome_wide_snp6/genome_wide_snp_6.affx#1_1)), incorporated herein by reference. In the SNP reference marker sequences, the polymorphism is indicated in a given dbSNP reference marker sequence

number by two bases on either side of a slash mark, *e.g.*, the major/minor alleles reported at the SNP are shown in column 2 of Table 1. The major/minor allele designation may differ based on the population being examined. For example, in Table 1, the qualification of allele frequency as “major” or “minor” pertains to a Caucasian European population.

- 5 The designation of major or minor does not affect the scope of the methods and compositions described herein. Rather it is the homozygous or heterozygous genotype of the risk allele that provides a diagnosis of the occurrence of, or an increased risk of the occurrence of, TGCT. The risk allele is pertinent to all populations.

Table 1. Summary results for top 22 markers reaching a statistical significance of $P < 5.0 \times 10^{-6}$ in the discovery phase

Marker ^a	Allele		Chromosome	MAF		Cases	P-value ^b	Per allele	OR (95% CI)	
	Major/Minor (Risk)	Controls		Controls	Homozygote ^d				Heterozygote ^e	
rs4474514	A/G (A)	12	0.198	0.089	3.54 × 10 ⁻¹⁰	0.41 (0.30-0.56)	0.43 (0.30-0.61)	0.12 (0.03-0.50)		
rs3782181	T/G (T)	12	0.202	0.094	8.38 × 10 ⁻¹⁰	0.42 (0.31-0.57)	0.42 (0.30-0.60)	0.18 (0.05-0.58)		
rs3782179	A/G (A)	12	0.199	0.092	1.35 × 10 ⁻⁹	0.42 (0.31-0.58)	0.43 (0.30-0.60)	0.18 (0.06-0.58)		
rs11104952	C/A (C)	12	0.198	0.092	1.90 × 10 ⁻⁹	0.43 (0.31-0.58)	0.43 (0.30-0.61)	0.18 (0.06-0.58)		
rs1472899	T/C (T)	12	0.199	0.095	3.09 × 10 ⁻⁹	0.43 (0.32-0.58)	0.43 (0.30-0.60)	0.19 (0.06-0.63)		
rs1352947	A/G (A)	12	0.183	0.083	4.24 × 10 ⁻⁹	0.42 (0.30-0.58)	0.41 (0.28-0.59)	0.22 (0.07-0.70)		
rs995030	C/T (C)	12	0.172	0.076	7.47 × 10 ⁻⁹	0.42 (0.30-0.58)	0.43 (0.29-0.62)	0.15 (0.04-0.61)		
rs3770112	C/T (T)	2	0.314	0.45	4.93 × 10 ⁻⁸	1.77 (1.44-2.18)	2.70 (1.94-3.76)	2.62 (1.66-4.14)		
rs7486184	C/T (C)	12	0.192	0.103	4.56 × 10 ⁻⁷	0.50 (0.37-0.67)	0.52 (0.37-0.72)	0.20 (0.06-0.65)		
rs2524594	G/A (G)	23	0.17	0.055	8.91 × 10 ⁻⁷	0.53 (0.40-0.71)	0.28e (0.16-0.50)			
rs1549383	G/A (A)	2	0.303	0.418	1.01 × 10 ⁻⁶	1.65 (1.35-2.02)	1.77 (1.32-2.39)	2.61 (1.71-4.00)		
rs6534637	G/T (T)	4	0.449	0.572	1.06 × 10 ⁻⁶	1.73 (1.40-2.14)	4.52 (2.85-7.16)	3.95 (2.36-6.62)		
rs7236484	T/A (A)	18	0.089	0.16	1.61 × 10 ⁻⁶	1.96 (1.47-2.62)	1.63 (1.17-2.27)	13.3 (3.61-48.7)		
rs6897876	C/T (C)	5	0.459	0.345	1.77 × 10 ⁻⁶	0.63 (0.52-0.77)	0.78 (0.58-1.04)	0.34 (0.21-0.53)		
rs3755353	G/A (A)	2	0.364	0.483	1.77 × 10 ⁻⁶	1.79 (1.44-2.23)	3.38 (2.35-4.86)	2.64 (1.59-4.37)		
rs7774545	A/G (G)	6	0.284	0.397	1.83 × 10 ⁻⁶	1.66 (1.35-2.04)	1.92 (1.42-2.60)	2.46 (1.56-3.89)		
rs6961928	C/T (T)	7	0.417	0.539	1.83 × 10 ⁻⁶	1.74 (1.40-2.16)	3.30 (2.20-4.95)	3.23 (1.99-5.24)		
rs12521013	G/T (G)	5	0.478	0.363	1.99 × 10 ⁻⁶	0.63 (0.52-0.76)	0.82 (0.61-1.10)	0.33 (0.21-0.51)		
rs26939101	A/C (A)	12	0.16	0.083	2.16 × 10 ⁻⁶	0.48 (0.34-0.66)	0.48 (0.33-0.69)	0.22 (0.05-0.95)		
rs4324715	T/C (T)	5	0.514	0.398	2.72 × 10 ⁻⁶	0.63 (0.52-0.77)	0.89 (0.65-1.21)	0.34 (0.22-0.53)		
rs2965606	T/A (A)	7	0.265	0.373	4.78 × 10 ⁻⁶	1.61 (1.31-1.98)	1.64 (1.22-2.22)	2.55 (1.63-4.00)		
rs17031166	C/G (G)	2	0.3	0.405	4.98 × 10 ⁻⁶	1.60 (1.31-1.96)	1.67 (1.24-2.24)	2.49 (1.62-3.83)		

^adbSNP rsnumber. ^bDifference in allele frequency determined by Fisher's Exact test. ^cOR for heterozygous carriage of minor allele compared to homozygous carriage of major allele. ^dOR for homozygous carriage of minor allele compared to homozygous carriage of major allele. ^eOR for hemizygous carriage of the minor allele compared to carriage of the major allele.

In one embodiment, the SNP is identified by the sequence of the forward, or positive, DNA strand. In another embodiment, the SNP is identified by the sequence of the reverse, or minus, DNA strand. For example, if the major/minor allele is identified as C/T for the forward strand, the major/minor allele for the reverse strand is G/A, for the same SNP. In one embodiment, as used in this specification, the terms “reference marker sequence” or “marker sequence” or “marker” refer to the NCBI sequences. The NCBI sequences are the short, about 52 nucleotide sequences containing the SNPs (*i.e.*, SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, and 25), but may also be used to refer to the longer FASTA sequences containing additional flanking sequence indicated by SEQ ID NOs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, and 26. In another embodiment, the terms “reference marker sequence” or “marker sequence” or “marker” refer to the sequences complementary to the NCBI sequences. The reverse complementary sequences of SEQ ID NOs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, and 25 respectively, are listed as SEQ ID NOs 27-39. See Table 2.

In one embodiment, a previously unrecognized correlation between certain polymorphisms in a mammalian subject’s genome and the occurrence or increased susceptibility of the occurrence of TGCT has been determined to provide a means for the diagnosis or identification of a risk of TGCT in a subject. Compositions for performing such diagnostic methods are also disclosed. The compositions and methods described herein permit the identification of a human subject having an elevated risk of TGCT by detecting the occurrence of at least one copy of the risk allele in a single nucleotide polymorphism (SNP) in a genomic region containing *KITLG* in a biological sample from the human subject. *KITLG*, also known as stem cell factor, encodes the ligand for the receptor tyrosine kinase, c-KIT, on chromosome 12q22. In one embodiment, the SNP is one or more of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952.

In another embodiment, the compositions and methods described herein permit the identification of a human subject having an elevated risk of TGCT by detecting the occurrence of at least one copy of the risk allele in a single nucleotide polymorphism (SNP) in a genomic region containing *SPRY4* (specifically certain markers downstream of *SPRY4*) in a biological sample from the human subject. *SPRY4* (sprouty homolog 4) is

a coding region on chromosome 5q31.3. In one embodiment, the SNP is one or more of rs12521013, rs4324715, rs6897876. Genetic variants at the 12p22 and 5q31 loci are associated with TGCT and strongly implicate *KITLG* as a susceptibility gene in the pathogenesis of TGCT.

5 As described in more detail in the examples below, eight reference marker sequences were identified as relevant to identification of TGCT risk in human subjects (see, e.g., Table 2). Seven of these markers (rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952), including the most significant association at rs4474514, occurred within the *KITLG* (c-KIT ligand) gene region on
10 chromosome 12q22. Sixteen additional markers reached statistical significance (see, e.g., Table 2). Of these, three markers (rs12521013, rs4324715, rs6897876) mapped 2.4 kb downstream of the *SPRY4* (sprouty homolog 4) coding region on chromosome 5q31.3, and two markers (rs17031166, rs1549383) mapped to a gene-free region on chromosome 2p14 that is 500 kb centromeric of *SPRED2* (sprouty-related, EVH1 domain containing
15 2).

Reproducible associations with TGCT were observed with markers rs3782179 and rs4474514 in *KITLG* and with markers rs4324715 and rs6897876 proximal to *SPRY4*, but not with rs17031166 or rs1549383 near *SPRED2*. TGCT risk was increased threefold per copy of the major allele in *KITLG* rs3782179 (Table 2, 5th column) and
20 rs4474514 (Table 2, 5th column). Homozygous carriage of the major alleles at these loci was associated with an over fourfold increased risk of TGCT compared with homozygous carriage of the minor allele. For the two markers close to *SPRY4*, TGCT risk was increased nearly 40% per copy of the major allele in rs4324715 (Table 2, 5th column) and major allele in rs6897876 (Table 2, 5th column). Risk was increased 65-80% with
25 homozygous carriage of the major alleles compared with homozygous carriage of their corresponding minor alleles. A case-parent triad analysis showed that carriage of these risk alleles for the markers in *KITLG* and proximal to *SPRY4* is associated with TGCT. The per-allele relative risks (RR) for rs3782179 and rs4474514 (*KITLG*) were 2.5 and 2.6, respectively. The per-allele relative risks (RR) for rs4324715 and rs6897876
30 (proximal to *SPRY4*) were 1.5 and 1.5, respectively. These family-based estimates provide additional evidence that population stratification did not bias results in the

replication phase. The data showed no interaction between *KITLG* and *SPRY4* marker genotypes. For most cases, *KITLG* and *SPRY4* do not exert their effect solely through mechanisms involving the known TGCT risk factors, *i.e.*, positive family history or personal history of cryptorchidism.

5 SNPs are identified by the Affymetrix or NCBI database reference marker sequences containing the SNPs, e.g., rs4474514. As shown by the NCBI sequences, the reference marker sequences are nucleotide sequences of about 52 nucleotides, with the single nucleotide polymorphism (SNP) occurring at nucleotide position 27, of the forward DNA. For certain SNPs relevant to the present compositions and methods, the following table lists: the NCBI shorter sequences associated with the reference numbers; the larger flanking sequences in which the reference marker sequence is embedded; and the reverse complement to the shorter sequence. Such larger sequences are identified as the SEQ ID NO: following the shorter sequence in the following table of relevant SNPs, with the reverse complement sequence listed thereafter.

15

TABLE 2

SNP Ref No.	SEQ ID NO:	Ancestral Allele/nt position	Variante allele/ nt position	Risk allele	Gene	Chromosome	TGCT Risk
rs995030	1	G/27	A/27	G	within <i>KITLG</i>	12q22	
	2	G/401	A/401	G			
	27	C/26	T/26	C			
rs1352947	3	A/27	G/27	A	within <i>KITLG</i>	12q22	
	4	A/563	G/563	A			
	28	T/24	C/24	T			
rs1472899	5	G/27	A/27	A	within <i>KITLG</i>	12q22	
	6	G/301	A301	A			
	29	C/26	T/26	T			
rs3782179	11	T/27	C/27	T	Within <i>KITLG</i>	12q22	Risk increases 3-fold/copy of A/26; and 4-fold when homozygous for A/26
	12	T/301	C/301	T			
	32	A/26	G/26	A			
rs3782181	13	A/27	C/27	A	Within	12q22	

	14	A/251	C/251	A	<i>KITLG</i>		
	33	T/26	G/26	T			
rs4324715	15	T/27	C/27	T	Downstream of <i>SPRY4</i>	5q31.3	Risk increases 40% per copy of T/27; Risk increases 65-80% when homozygous for T/27
	16	T/341	C/341	T			
	34	A/26	G/26	A			
rs4474514	17	A/27	G/27	A	Within <i>KITLG</i>	12q22	Risk increases 3-fold/copy of A/27; and 4-fold when homozygous for A/27
	18	A/301	G/301	A			
	35	T/26	C/26	T			
rs6897876	19	C/27	T/27	C	Downstream of <i>SPRY4</i>	5q31.3	Risk increases 40% per copy of C/27; Risk increases 65-80% when homozygous for C/27
	20	C/301	T/301	C			
	36	G/26	A/26	G			
rs11104952	21	G/27	T/28	G	Within <i>KITLG</i>	12q22	
	22	G/301	T/301	G			
	37	C/26	A/26	C			
rs12521013	23	C/27	A/27	C	Downstream of <i>SPRY4</i>	5q31.3	
	24	C/501	A/501	C			
	38	G/26	T/26	G			

In one embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "G" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:1 (rs995030) or "C" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:27, which is the corresponding minus strand of SEQ ID NO:1. This same polymorphism may be described as a polymorphism "G" at the nucleotide corresponding to nucleotide 401 of SEQ ID NO:2 (rs995030) or

"C" at the nucleotide corresponding to nucleotide 401 of the corresponding minus strand of SEQ ID NO:2. See Tables 1 and 2.

In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "A" at the nucleotide
5 corresponding to nucleotide 27 of SEQ ID NO:3 (rs1352947) or "T" at the nucleotide corresponding to nucleotide 24 of SEQ ID NO:28, which is the corresponding minus strand of SEQ ID NO:3. This same polymorphism may be described as a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 563 of SEQ ID NO:4 (rs1352947) or "T" at the nucleotide corresponding to nucleotide 563 of the
10 corresponding minus strand of SEQ ID NO:4. See Tables 1 and 2.

In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:5 (rs1472899) or "T" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:29, which is the corresponding minus
15 strand of SEQ ID NO:5. This same polymorphism may be described as a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 301 of SEQ ID NO:6 (rs1472899) or "T" at the nucleotide corresponding to nucleotide 301 of the corresponding minus strand of SEQ ID NO:6. See Tables 1 and 2.

In another embodiment of the compositions and methods described herein, at least
20 one SNP diagnostic of TGCT is a polymorphism comprising "T" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:11 (rs3782179) or "A" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:32, which is the corresponding minus strand of SEQ ID NO:11. This same polymorphism may be described as a polymorphism comprising "T" at the nucleotide corresponding to nucleotide 301 of SEQ ID NO:12
25 (rs3782179) or "A" at the nucleotide corresponding to nucleotide 301 of the corresponding minus strand of SEQ ID NO:12. See Tables 1 and 2.

In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:13 (rs3782181) or "T" at the nucleotide
30 corresponding to nucleotide 26 of SEQ ID NO:33, which is the corresponding minus strand of SEQ ID NO:13. This same polymorphism may be described as a polymorphism

comprising "A" at the nucleotide corresponding to nucleotide 251 of SEQ ID NO:14 (rs3782181) or "T" at the nucleotide corresponding to nucleotide 251 of the corresponding minus strand of SEQ ID NO:14. See Tables 1 and 2.

In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "T" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:15 (rs4324715) or "A" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:34, which is the corresponding minus strand of SEQ ID NO:15. This same polymorphism may be described as a polymorphism comprising "T" at the nucleotide corresponding to nucleotide 341 of SEQ ID NO:16 (rs4324715) or "A" at the nucleotide corresponding to nucleotide 341 of the corresponding minus strand of SEQ ID NO:16. See Tables 1 and 2.

In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:17 (rs4474514) or "T" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:35, which is the corresponding minus strand of SEQ ID NO:17. This same polymorphism may be described as a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 301 of SEQ ID NO:18 (rs4474514) or "T" at the nucleotide corresponding to nucleotide 301 of the corresponding minus strand of SEQ ID NO:18. See Tables 1 and 2.

In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "C" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:19 (rs6897876) or "G" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:36, which is the corresponding minus strand of SEQ ID NO:19. This same polymorphism may be described as a polymorphism comprising "C" at the nucleotide corresponding to nucleotide 301 of SEQ ID NO:20 (rs6897876) or "G" at the nucleotide corresponding to nucleotide 301 of the corresponding minus strand of SEQ ID NO:20.

In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "G" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:21 (rs11104952) or "C" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:37, which is the corresponding minus

strand of SEQ ID NO:21. This same polymorphism may be described as a polymorphism comprising "G" at the nucleotide corresponding to nucleotide 301 of SEQ ID NO:22 (rs11104952) or "C" at the nucleotide corresponding to nucleotide 301 of the corresponding minus strand of SEQ ID NO:22.

5 In another embodiment of the compositions and methods described herein, at least one SNP diagnostic of TGCT is a polymorphism comprising "C" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:23 (rs12521013) or "G" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:38, which is the corresponding minus strand of SEQ ID NO:23. This same polymorphism may be described as or "C" at the
10 nucleotide corresponding to nucleotide 501 of SEQ ID NO:24 (rs12521013) or "G" at the nucleotide corresponding to nucleotide 501 of the corresponding minus strand of SEQ ID NO:24.

In certain embodiments, the occurrence of at least one copy of at least one of these SNPs is indicative of an elevated risk of TGCT. In certain embodiments, the
15 occurrence of homozygous copies of at least one of these SNPs is indicative of an elevated risk of TGCT. In certain embodiments, the occurrence of two or more of these SNPs (homozygous or heterozygous) is indicative of an elevated risk of TGCT.

B. Compositions for Diagnosis/Identification of Risk of TGCT

20 A diagnostic composition, including either individual reagents or kits containing multiple reagents for diagnosing the risk, occurrence, stage or progression of TGCT in a mammalian subject is provided as described herein. In one embodiment, a diagnostic composition is a reagent that is capable of identifying a genetic variation or mutant, e.g., a SNP, in a genomic region containing the gene *KITLG* on chromosome 12 at locus
25 12q22. In one embodiment, the SNP is one or more of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952. In another embodiment, a diagnostic composition is a reagent that is capable of identifying a genetic variation or mutant, e.g., a SNP, in a genomic region containing the gene *SPRY4* gene on chromosome 5 at locus
30 5q31.3. In another embodiment, the SNP is one or more of rs12521013, rs4324715, rs6897876 which map to 2.4 kb downstream of the *SPRY4* (sprouty homolog 4) coding region on chromosome 5q31.3. These genomic regions have been found by the inventors

to be associated with susceptibility of a human subject to TGCT or tumors. These reagents are capable of detecting such genetic variations in a biological sample from the human subject.

By the term "biological sample" is meant a cell-containing fluid or tissue obtained
5 from the subject containing genomic material. This sample includes, without limitation, whole blood, serum or plasma, saliva, semen, urine, cheek cells, and cellular exudates from a mammalian subject, as well as tissue samples, including biopsied tissue. Such samples may further be diluted with saline, buffer or a physiologically acceptable diluent. Alternatively, such samples are concentrated by conventional means. In the methods
10 described herein or in concert with the compositions, the sample is in one embodiment, examined *ex vivo*, i.e., outside of the subject's body. In one embodiment, each sample is obtained from the same subject to provide a diagnosis.

By the terms "patient" or "subject" as used herein means preferably a human male, but can also include a human female, a human of ambiguous sex, and non-human
15 mammals, including a veterinary or farm animal, a domestic animal or pet, and animals normally used for clinical research. More specifically, the subject of these methods and compositions is a human. In one aspect of the methods described herein, the subject undergoing the diagnostic method is asymptomatic for TGCT. In another aspect, the subject undergoing the diagnostic methods described herein shows clinical signs of
20 TGCT. In another aspect, the subject undergoing the diagnostic methods described herein has a familial history of TGCT. In another aspect, the subject undergoing the diagnostic methods described herein has no familial history of TGCT. In another aspect, the subject undergoing the diagnostic methods described herein has a personal history of undescended testes, an uncommon but major risk factor for TGCT. In another aspect, the
25 subject undergoing the diagnostic methods described herein has no personal history of undescended testes.

The genetic variant detected by the reagent in the region of the *KITLG* gene is, in one embodiment, a nucleotide sequence containing a SNP selected from the group consisting of: rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514,
30 rs11104952, and a combination thereof. The genetic variant detected by the reagent in the region of the *SPRY4* gene is, in one embodiment, a nucleotide sequence containing a

single nucleotide polymorphism (SNP) selected from the group consisting of rs4324715, rs6897876, rs12521013, and a combination thereof. In still other embodiments, the reagents may be designed to detect genetic variants of any combination of the genes and genomic loci described herein.

5 In certain embodiments, reagents useful herein are capable of forming a physical association with a selected SNP in the subject's biological sample containing one or combination of the SNPs described herein. One such reagent is a nucleic acid sequence capable of hybridizing to a SNP-containing marker sequence in the sample. For example, when the reagent is a genomic probe, the physical association formed by contact of the
10 reagent with the sample is the hybridization of the probe to the cDNA or mRNA of a sequence containing the SNP or SNPs. Where the reagent is a PCR primer or primer pair, the physical association is the hybridization of the primer sequences to different strands or different portions of the nucleic acid (e.g., mRNA) of a marker sequence containing the SNP or SNPs.

15 The polynucleotide sequences for genomic probes or primer sets useful to identify or amplify a nucleotide sequence in the sample containing the SNPs, their length and labels used in the composition are designed based upon the SNP reference numbers and sequences associated with the SEQ ID NOs described herein. Preferably the nucleic acid probes or primers are from about 8 or more nucleotides in length, wherein the
20 nucleotides are complementary to portions of the "non-coding" or "coding" strands of the gene sequences or non-gene sequences flanking or encompassing the selected SNP. Such probes are, for example, oligo or polynucleotide sequences corresponding to the region surrounding (and/or comprising) any of the SNP marker sequences identified on the human chromosomes 12 or 5. Such a fragment usually has a length comprised between 8
25 and 50 nucleotides, preferably 12 to 35 nucleotide or 15 to 25 nucleotides. It may be a fragment of naturally occurring or synthetic DNA or RNA. In certain embodiments, each primer or probe is at least 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or at least 20
nucleotides in length. In other embodiments, the primers and /or probes may be longer than 20 nucleotides in length. Given the information provided herein, one of skill in the
30 art may design any number of suitable primer/probe sequences useful for identifying the

SNPs and genomic regions associated with the diagnosis of TGCT or susceptibility thereto.

In certain embodiments of diagnostic compositions defined herein, the reagent is associated with a directly-detectable, or indirectly-detectable, label. Detectable labels for attachment to nucleic acid sequences useful in diagnostic assays of this invention may be easily selected from among numerous compositions known and readily available to one skilled in the art of diagnostic assays. Among common labels for such use are radioactive, enzymatic, luminescent and fluorescent markers. Non-exclusive examples of such labels include radioactive compounds, radioisotopes, such as ^{32}P , ^{125}I , technetium; fluorescent or chemiluminescent compounds, such as FITC, rhodamine or luciferin; and proteins such as biotin or enzymes and enzyme co-factors, such as alkaline phosphatase, β -galactosidase or horseradish peroxidase; and/or molecular labels such as FLAG, etc. Commonly used fluorochromes include fluorescein isothiocyanate (FITC), phycoerythrin (PE), allophycocyanin (APC), and also include the tandem dyes, PE-cyanin-5 (PC5), PE-cyanin-7 (PC7), PE-cyanin-5.5, PE-Texas Red (ECD), rhodamine, PerCP, fluorescein isothiocyanate (FITC) and Alexa dyes. Combinations of such labels, such as Texas Red and rhodamine, FITC + PE, FITC + PECy5 and PE + PECy7, among others may be used. Association of a nucleic acid primer or probe sequence with a suitable label is conventional in the art. Other elements of the label systems include substrates useful for generating the signals upon interaction with the other components of the label system employed.

In certain embodiments, these above-described diagnostic reagents are immobilized on a suitable substrate. Certain substrates include solid support, plates, sticks, or beads, a computer chip or computer-readable chamber, or microfluidics card.

Still another diagnostic composition is a kit comprising one or more reagents that are capable of identifying a SNP or combination of SNPs associated with susceptibility of a human subject to TGCT. Such a kit employing multiple diagnostic reagents for diagnosing the occurrence or susceptibility of TGCT in a biological sample of a mammalian subject can identify one or more than one of the selected genomic variations identified herein. Such a kit or other multi-reagent composition includes one or more genomic probes or PCR primer-probe sets that amplifies a nucleic acid sequence

containing one or more of the selected SNPs. Thus, in one embodiment, the composition contains one or a plurality of polynucleotides immobilized on a substrate, wherein at least one polynucleotide is a genomic probe that hybridizes to a marker nucleotide sequence (RNA, mRNA, DNA, cDNA) containing at least one of the above-identified SNPs. In
5 another aspect, the composition contains one or a plurality of PCR primer-probe sets, wherein at least one primer-probe set amplifies a polynucleotide (mRNA) sequence of a SNP or a marker sequence containing a SNP as identified above.

Thus, in one embodiment, a diagnostic kit contains oligonucleotides specific for identifying one or more SNP located in or near KITLG or SPRY4 in a biological sample
10 of a subject, e.g., a nucleic acid sample. The kit also contains additional reagents for carrying out a procedure that uses the oligonucleotides to identify the nucleotide at position 27 of the reference marker sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 and 25 in the nucleic acid sample or the corresponding nucleotide in the complementary sequence thereto.

15 The diagnostic compositions of the invention can be presented in the format of a microfluidics card, a microarray, a chip or chamber that employs PCR, RT-PCR or Q-PCR techniques described below. In one aspect, the diagnostic composition is a TAQMAN® Quantitative PCR low density array containing multiple probes and primer sequences. When a biological sample from a selected subject is contacted with the
20 primers and probes in the diagnostic composition, PCR amplification of gene sequences or marker sequences containing the SNP from the subject permits detection of the genetic variations that are indicative of a susceptibility or diagnosis of TGCT.

Such diagnostic reagents and kits containing them are useful for the detection of homozygous or heterozygous genetic variations or polymorphisms related to TGCT
25 identified herein, and enable a diagnosis of TGCT or an increased susceptibility thereto. Such diagnostic kits optionally also contain miscellaneous reagents and apparatus for reading labels, e.g., certain substrates that interact with an enzymatic label to produce a color signal, etc., apparatus for taking biological samples, as well as appropriate vials and other diagnostic assay components.

30 The diagnostic kits may optionally contain a positive or negative control. By positive control is meant genetic material reflecting a predisposition to TGCT, for

example a DNA sample from a person affected by TGCT. By negative control is meant genetic material reflecting the absence of a predisposition to TGCT. By definition, the means for detecting the SNP alleles of the markers present in the kit lead to a negative result when applied to the negative control, whereas they lead to a positive result when
5 applied to the positive control.

In one embodiment, the diagnostic composition includes a microarray of two or more reagents capable of identifying the presence of two or more SNPs in a biological sample. In certain embodiments, the two or more SNPs are selected from SNPs identified in Tables 1 and 2. In certain embodiments, the two or more SNPs are selected
10 from rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, and a combination thereof. In another embodiment, the two or more SNPs are selected from those within the markers rs4324715, rs6897876, rs12521013, and a combination thereof. In still another embodiment, the two or more SNPs include SNPs selected from the marker sequence rs3782179, rs4324715, rs4474514 and rs6897876. In still other
15 embodiments, the two or more SNPs include polymorphisms on the coding strand of the marker sequence. In another embodiment, the two or more SNPs include polymorphisms on the non-coding “minus” strand of the marker sequence. Any kit or composition containing multiple reagents can include reagents in addition to one or more of the reagents specifically identified herein.

20

C. Methods for Diagnosis/Identification of Risk of TGCT

In another aspect, a method is provided for diagnosing or identifying the occurrence or the susceptibility or risk of occurrence of TGCT in a subject. Such a diagnostic method includes obtaining a biological sample from a subject. The biological
25 sample is then contacted with a diagnostic reagent that is capable of identifying one or more genetic variants of the *KITLG* gene or the *SPRY4* gene, as discussed above. In certain embodiments, one or more SNPs are selected from rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, and a combination thereof. In another embodiment, one or more SNPs are selected from those within the markers
30 rs4324715, rs6897876, rs12521013, and a combination thereof. The occurrence of, or an increased susceptibility to, TGCT is diagnosed or identified when at least one of these

variants are identified in said sample. In another embodiment, the occurrence of, or an increased susceptibility to, TGCT is diagnosed or identified when two or more of these variants are identified in said sample. In another embodiment, the occurrence of, or an increased susceptibility to, TGCT is diagnosed or identified when four or more of these
5 variants are identified in said sample. In another embodiment, the occurrence of, or an increased susceptibility to, TGCT is diagnosed or identified when six or more of these variants are identified in said sample. As used herein, the terms risk and susceptibility are used interchangeably and refer to the likelihood that a subject has or will develop TGCT as compared to a control population.

10 In one embodiment of these methods, the control population is formed by a population of male subjects having environmental, phenotypic and/or genotypic similarities and characteristics to those of the subject, e.g., age, race, general physical health. In other embodiments of these methods, the control population is formed by a population of male subjects having a broader selection of environmental, phenotypic
15 and/or genotypic characteristics. In one embodiment, the other subjects are human males who are not affected by TGCT. In another embodiment, the other subjects are human males who are affected by TGCT. In still another embodiment, the other subjects are human males having a blood relationship with the subject to be diagnosed. Still other control populations may be selected by the diagnostician.

20 In certain embodiments, the variant of the KITLG gene region includes a nucleic acid sequence comprising a single nucleotide polymorphism (SNP) within that region. The SNP may be one or a combination of two or more of the SNPs found within the marker sequences identified by the NCBI or Affymetrix reference numbers rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, rs4324715,
25 rs6897876, rs12521013. The SNPs may also be those discussed in the marker sequences in any of the tables and included in this specification (see, e.g., Tables 1 and 2). In other embodiments of the method, any combination of the specifically identified SNPs with one or more other SNPs found subsequently to be relevant to TGCT may be detected to provide a diagnosis or identification of a risk of TGCT in a subject or population.

30 Useful methods or assays for performing such diagnoses include methods based on hybridization analysis of polynucleotide genomic probes or primer/probe sets useful

for amplification of the SNP and sequences flanking it. Such methods include sequencing of polynucleotides, proteomics-based methods or immunochemistry techniques. The most commonly used methods known in the art for the detection and/or quantification of mRNA expression in a sample include northern blotting and *in situ* hybridization; RNase protection assays; and PCR-based methods, such as reverse transcription polymerase chain reaction, real-time PCR (RT-PCR), or qPCR.

Detection of the nucleotides hereinbefore described can be performed by any method which is suitable for genotyping. Methods for detecting nucleic acid polymorphisms are well-known (allelotyping or genotyping) and use as diagnostic reagents chip microarrays on which oligonucleotides are immobilized, as described above. Conventional genotyping procedures are indicated in the following references: Tang K, et al. (1999) "Chip-based genotyping by mass spectrometry", Proc. Natl. Acad. Sci. USA 96: 10016-10020; Bansal et al. (2002) "Association testing by DNA pooling-- An effective initial screen", Proc. Natl. Acad. Sci. USA, December 24; 99 (26): 16871-16784; Werner, M. et al. "Large scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry", Hum. Mutat. 2002 July; 20 (1): 57-64; Stoerker J, Mayo et al. "Rapid genotyping by MALDI-monitored nuclease selection from probe libraries", Nat. Biotechnol. 2000 November; 18 (11): 1213-1216.

Thus, in one embodiment, a method for performing the diagnosis involves detecting the allele of the selected SNP marker sequence as described herein using sequencing devices which make it possible to determine the sequence of a sample of DNA or RNA. In order to detect the alleles of one or more of the SNPs described herein, a nucleic acid probe may be used which hybridizes with only one of the alleles and not with the other under stringent conditions. Stringent conditions in performing the hybridization can ensure the hybridization of a probe with the specific SNP allele in the sample only in the case of strict complementarity. The stringency of the conditions for strict complementarity can be determined by the specialist skilled in the art. Such conditions depend in particular on the length of the probe. The stringency increases when the concentrations of salts (NaCl for example), detergents (SDS, for example), non-specific material (salmon sperm, for example) and the temperature increase.

In other embodiments the SNPs are identified in the biological sample by PCR (Polymerase Chain Reaction) amplification procedure. In this situation, a technique developed from the MALDI-TOF mass spectrometry technology includes use of a microarray chip which enables several tens of samples (384) to be examined at once.

5 Other methods may include mini-sequencing of the DNA in the vicinity of the polymorphic site, as a result of an elongation behind the primers in the neighborhood of the polymorphism. Identification of the alleles of a selected SNP present in a sample may also be obtained by performing PCR in real time.

Still other methods useful in performing the diagnostic steps described herein are
10 known and well summarized in, e.g., US Patent No. 7,081,340. The methods described herein are not limited by the particular techniques selected to perform them. Exemplary commercial products for generation of reagents or performance of assays include Affymetrix Genome-Wide Human SNP Array 6.0, the protocol for which is hereby incorporated by reference, TRI-REAGENT, Qiagen RNeasy mini-columns,
15 MASTERPURE Complete DNA and RNA Purification Kit (EPICENTRE®, Madison, Wis.), Paraffin Block RNA Isolation Kit (Ambion, Inc.) and RNA Stat-60 (Tel-Test), the MassARRAY-based method (Sequenom, Inc., San Diego, CA), differential display, amplified fragment length polymorphism (AFLP), and BeadArray™ technology (Illumina, San Diego, CA) using the commercially available Luminex 100 LabMAP
20 system and multiple color-coded microspheres (Luminex Corp., Austin, Tex.) and high coverage expression profiling (HiCEP) analysis.

Thus, in performing the diagnosis on a subject's biological sample, a method involves forming a physical association between the diagnostic reagent and the variant in
the sample. The method of diagnosis can involve contacting the biological sample with
25 one or more of the diagnostic reagents described above.

When a reagent used in the diagnostic method is associated with a directly or indirectly detectable label, the method involves transforming the detectable signals generated from the diagnostic reagent in association with a SNP present in the biological sample into numerical or graphical data. Desirably, the transforming is performed by a
30 suitably-programmed machine or instrument that can detect the detectable signals generated from the diagnostic reagents associated with the SNPs present in the biological

sample. Transformation by the instrument of the detection of the SNP in a biological sample into into numerical or graphical data useful for comparison with similar results in a selected "control" population assists in performing the diagnosis.

In certain embodiments of the diagnostic method, the identification of the selected
5 SNPs is coupled with the presentation of clinical symptoms of TGCT in a subject to confirm a diagnosis of the cancer and/or to confirm the level of risk of susceptibility to the cancer. In other embodiments, the diagnosing includes coupling the identification of the selected SNPs with evidence of a familial history of testicular cancer and/or a personal history of undescended testes to confirm a diagnosis assessing level of risk of
10 susceptibility to the cancer in a particular subject. In certain embodiments, the method provides a quantitative assessment of the likelihood or risk of TGCT occurrence in a subject that has not yet developed clinical symptoms of TGCT, based upon the results of the SNP identification.

In one exemplary method, a subject's biological sample is contacted with PCR
15 primers and/or probes that are designed for amplification and/or detection of the selected SNP in the selected marker sequence. According to the first step of this process the samples are amplified by the PCR, the target being the nucleic acid sequence in the sample that contains or may contain the selected SNP. An elongation reaction (starting from a primer close to the SNP) is carried out. It is the difference in size (tiny, usually a
20 difference of between 1 and 4 nucleotides) between the product obtained by elongation for the allele by default (A for example) and that of the other allele (G for example) detected by MALDI-TOF, which is recorded and makes it possible to type the genotype AA or AG or GG, for example. The treatment of the results obtained can be performed by means of the method "MassARRAY".

25 Thus one method for predicting a risk of TGCT includes detecting, in at least one nucleic acid sample, one or more polymorphisms (SNPs) within the genomic region of the KITGL or SPRY4 gene, wherein said one or more polymorphisms is associated with said risk of TGCT. In one embodiment, the polymorphism in said KITLG or SPRY4 gene modulates the level of transcription of the gene. In certain embodiments, the one or
30 more SNPs are selected from rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, and a combination thereof. In another embodiment, the one or

more SNPs are selected from those within the markers rs4324715, rs6897876, rs12521013, and a combination thereof. In still another embodiment, the two or more SNPs include SNPs selected from the marker sequence rs3782179, rs4324715, rs4474514 and rs6897876.

5 In another embodiment, a method of determining genetic predisposition for TGCT in a subject uses single nucleotide polymorphism (SNP) analysis. A biological sample is taken from a subject and a SNP genotyping assay is performed to identify one or more genetic variations in the indicated KITGL or SPRY4 genomic regions. A SNP panel comprising predetermined identifier SNPs that define a genetic predisposition for
10 TGCT is used for comparison with the experimental results. The SNP analysis from the sample is then compared with the predetermined identifiers. The presence of a genetic predisposition or susceptibility for TGCT is reported if the subject's SNP panel meets the predetermined criterion. In such a method, the predetermined identifier SNPs include one or more of the SNP mutations identified in Table 1 or 2 herein or in the other tables
15 forming this specification.

 In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "G" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:1 (rs995030) or "C" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO: 27, which is the corresponding minus
20 strand of SEQ ID NO:1.

 In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "G" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:3 (rs1352947) or "C" at the nucleotide corresponding to nucleotide 24 of SEQ ID NO:28, which is the corresponding minus
25 strand of SEQ ID NO:3.

 In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:5 (rs1472899) or "T" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:29, which is the corresponding minus
30 strand of SEQ ID NO:5.

In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "C" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:11 (rs3782179) or "G" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:32, which is the corresponding minus strand of SEQ ID NO:11.

In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:13 (rs3782181) or "T" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:33, which is the corresponding minus strand of SEQ ID NO:13.

In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "T" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:15 (rs4324715) or "A" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:34, which is the corresponding minus strand of SEQ ID NO:15.

In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "A" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:17 (rs4474514) or "T" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:35, which is the corresponding minus strand of SEQ ID NO:17.

In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "C" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:19 (rs6897876) or "G" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:36, which is the corresponding minus strand of SEQ ID NO:19.

In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "G" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:21 (rs11104952) or "C" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:37, which is the corresponding minus strand of SEQ ID NO:21.

In one embodiment, a method for diagnosing a predisposition to TGCT in a subject includes detecting a polymorphism comprising "C" at the nucleotide corresponding to nucleotide 27 of SEQ ID NO:23 (rs12521013) or "G" at the nucleotide corresponding to nucleotide 26 of SEQ ID NO:38, which is the corresponding minus strand of SEQ ID NO:23.

In certain embodiments, the occurrence of at least one copy of at least one of these SNPs is indicative of an elevated risk of TGCT. In certain embodiments, the occurrence of homozygous copies of at least one of these SNPs is indicative of an elevated risk of TGCT. In certain embodiments, the occurrence of two or more of these SNPs (homozygous or heterozygous) is indicative of an elevated risk of TGCT.

In performing the diagnosis the results of the SNP identification can also involve comparing the allelic form of the SNP marker found in the subject's sample with that of other subjects or controls or populations as described above.

The examples that follow do not limit the scope of the embodiments described herein. A genome-wide scan was conducted among 277 TGCT cases and 919 controls and found that seven markers at 12p22 within *KITLG* (c-KIT ligand) reached genome-wide significance ($P < 5.0 \times 10^{-8}$ in discovery). In independent replication, TGCT risk was increased threefold per copy of the major allele at rs3782179 and rs4474514 (OR = 3.08, 95% CI = 2.29-4.13; OR = 3.07, 95% CI = 2.29-4.13, respectively) (see Table 2). Associations were found with rs4324715 and rs6897876 at 5q31.3 near *SPRY4* (sprouty4; $P < 5.0 \times 10^{-6}$ in discovery). In independent replication, risk of TGCT was increased nearly 40% per copy of the major allele (OR = 1.37, 95% CI = 1.14-1.64; OR = 1.39, 95% CI = 1.16-1.66, respectively). All of the genotypes were associated with both seminoma and nonseminoma TGCT subtypes. These results demonstrated that common genetic variants affect TGCT risk and implicate *KITLG* and *SPRY4* as genes involved in TGCT susceptibility.

One skilled in the art will appreciate that modifications can be made in the following examples which are intended to be encompassed by the spirit and scope of the invention.

30

EXAMPLE 1: Genes Associated with TGCT Development

To identify genes associated with TGCT development, a genome-wide association study was conducted. Cases were 277 white, non-Hispanic men with pathologically defined TGCT seen at the University of Pennsylvania Health System (UPHS) or Fox Chase Cancer Center (FCCC) in Philadelphia, Pennsylvania. DNA extracted from venous blood was genotyped using the Affymetrix Genome Wide Human SNP Array 6.0™. The frequency of observed genotypes among TGCT cases was compared to those available from 919 white, non-Hispanic males from the Philadelphia region genotyped on the same Affymetrix™ platform (Table 3).

Eight markers reached statistical significance at a genome-wide threshold of $P < 5.0 \times 10^{-8}$ (Table 1). Seven of these (rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952), including the most significant association ($P = 3.54 \times 10^{-10}$) at rs4474514, occurred within the *KITLG* (c-KIT ligand) gene region on 12q22. These markers were in strong linkage disequilibrium (LD) with each other; pair wise D and r^2 measures were all > 0.99 (data not shown).

The eighth marker (rs3770112, $P = 4.93 \times 10^{-8}$) mapped to the integrin alpha 4 (*ITGA4*) gene on 2q31.3. No other markers in this genomic region (± 10 Mb) reached statistical significance at $P < 1.0 \times 10^{-3}$. To investigate the possibility that this association arose by chance, genotypes near rs3770112 were imputed on the basis of publicly available HapMap genotypic data (Frazer, K.A. et al. 2007 *Nature* 449, 851-861). After imputation, the test of association at rs3770112 no longer surpassed the genome-wide threshold ($P = 0.05$); as well, all other markers in the region remained below the threshold for advancing to replication. The correlation between observed and imputed P values for the 23 markers that were in the same LD block with rs3770112 was very high ($r = 0.96$). Information content and maximum posterior call probability for rs3770112 were both > 0.998 . Taken together, these results strongly suggested that the association observed in the discovery phase was a false positive (data not shown).

Two markers in *KITLG* (rs3782179, rs4474514) were selected for replication. Sixteen additional markers reached statistical significance at the $P < 5.0 \times 10^{-6}$ level (Table 1). Of these, three (rs12521013, rs4324715, rs6897876) mapped 2.4 kb downstream of the *SPRY4* (sprouty homolog 4) coding region on 5q31.3, and two

(rs17031166, rs1549383) mapped to a gene-free region on 2p14 that is 500 kb centromeric of *SPRED2* (sprouty-related, EVH1 domain containing 2) (Table 4).

Table 3 Age, Family history of TGCT and tumor type in the discovery and replication samples

Status	Discovery			Replication		
	Case	Control	Total	Case	Control	Total
	N=277	N=919	N=1196	N=371	N=860	N=1231
Total	No.	No.	No.	No.	No.	No.
	%	%	%	%	%	%
Age (median, interquartile range)	31 ^a (24, 39)	57 (52, 62)	57 (52, 62)	34 (28, 38)	35 (30, 39)	35 (30, 39)
Family history of TGCT						
No	239	-	239	314	769	1083
Yes	30 ^b	-	30	11 ^c	10 ^c	21
Unknown	8	-	8	46	81	89
Personal history of cryptorchidism						
No	245	-	245	330	844	1074
Yes	26	-	26	38	16	54
Unknown	6	-	6	3	0	3
Tumor type						
Seminoma	85	-	85	230	-	230
Nonseminoma	180	-	180	141	-	141
Unknown	12	-	12	0	-	0

^aAge of diagnosis missing for six TGCT cases. ^bSixteen cases were selected on the basis of family history of TGCT; among nonselected (n=261) cases, the proportion reporting any family history of TGCT was 5.4%. ^cDenotes reported family history of TGCT among first-degree relatives only.

Table 4 Associations of TGCT with replicated SNP markers

Gene	Marker ^a	Risk allele	Genotype count ^b				OR (95% CI)				P trend ^e
			Controls	Cases	Phase	Per allele	Heterozygote ^c	Homozygote ^d			
KITLG	rs3782179A/G	A	597/276/44	229/45/3	Discovery	2.36 (1.73-3.21)	2.39 (0.71-8.03)	5.63 (1.73-18.3)	1.95 x 10 ⁻⁸		
			515/285/38	309/49/5	Replication	3.08 (2.29-4.13)	1.31 (0.49-3.48)	4.56 (1.78-11.7)	5.88 x 10 ⁻¹⁵		
SPRY4	rs4474514A/G	A	599/276/44	229/45/2	Discovery	2.45 (1.79-3.35)	3.59 (0.84-15.3)	8.41 (2.02-35.0)	7.34 x 10 ⁻⁹		
			517/285/38	310/49/5	Replication	3.07 (2.29-4.13)	1.31 (0.49-3.48)	4.56 (1.77-11.7)	5.88 x 10 ⁻¹⁵		
SPRY4	rs4324715T/C	T	230/433/255	87/145/33	Discovery	1.59 (1.30-1.93)	2.59 (1.72-3.89)	2.92 (1.89-4.53)	3.57 x 10 ⁻⁶		
			191/437/197	119/171/68	Replication	1.37 (1.14-1.64)	1.13 (0.82-1.57)	1.81 (1.26-2.58)	6.77 x 10 ⁻⁴		
SPRY4	rs6897876C/T	C	282/429/207	114/135/28	Discovery	1.59 (1.31-1.94)	2.33 (1.50-3.61)	2.99 (1.90-4.69)	2.96 x 10 ⁻⁶		
			251/428/154	156/149/57	Replication	1.39 (1.16-1.66)	0.94 (0.66-1.34)	1.68 (1.17-2.42)	3.67 x 10 ⁻⁴		

^arsSNP rsnumber and major/minor alleles. ^bNumber of individuals genotyped as homozygous for the risk allele/heterozygous for the risk allele/homozygous for the nonrisk allele. Nomenclature for major/minor alleles based on calls from Affymetrix Genome-wide Human SNP Array 6.0. MAF for discovery phase markers given in Supplementary Table 1. ^cOR for heterozygous carriage of risk allele compared to homozygous carriage of nonrisk allele. ^dOR for homozygous carriage of risk allele compared to homozygous carriage of nonrisk allele. ^eCochran-Armitage test for trend.

Both *SPRY4* and *SPRED2* have been implicated in the KIT-KITLG signaling pathway (Wakioka, T. et al. 2001 *Nature* 412,647-651; Frolov, A. et al. 2003 *Mol. Cancer Ther.* 2, 699-709). These two regions were the only ones that contained more than one marker surpassing threshold significance. Two markers at each of these loci
5 (*SPRY4*: rs4324715, rs6897876; 2p14: rs17031166, rs1549383) were selected to bring forward for replication.

The replication set consisted of a population-based set of 371 TGCT cases and 860 controls, all white non-Hispanic, recruited from residents of the metropolitan Seattle-Puget Sound region, and parents of 204 of the cases. Associations with rs3782179 ($P_{\text{trend}} = 5.88 \times 10^{-15}$) and rs4474514 ($P_{\text{trend}} = 5.88 \times 10^{-15}$) in *KITLG* and with rs4324715 ($P_{\text{trend}} = 6.77 \times 10^{-4}$) and rs6897876 ($P_{\text{trend}} = 3.67 \times 10^{-4}$) proximal to *SPRY4* (Table 5), but not
10 with rs17031166 ($P_{\text{trend}} = 0.90$) or rs1549383 ($P_{\text{trend}} = 0.88$) near *SPRED2* were observed. TGCT risk was increased threefold per copy of the major A allele in *KITLG* rs3782179 and rs4474514 (OR = 3.08, 95% CI = 2.29-4.13; and OR = 3.07, 95% CI = 2.29-4.13,
15 respectively). Homozygous carriage of the major A allele at these loci was associated with an over fourfold increased risk of TGCT (OR = 4.56, 95% CI = 1.78-11.7; and OR = 4.56, 95% CI = 1.77-11.7, respectively) compared with homozygous carriage of the minor G allele. Weaker associations were noted for the two markers close to *SPRY4*. TGCT risk was increased nearly 40% per copy of the major T allele in rs4324715 (OR = 1.37,
20 95% CI = 1.14-1.64) and major C allele in rs6897876 (OR = 1.39, 95% CI = 1.16-1.66). Risk was increased 65-80% with homozygous carriage of the major alleles (OR = 1.81, 95% CI = 1.26-2.58; and OR = 1.68, 95% CI = 1.17-2.42, respectively) compared with homozygous carriage of their corresponding minor alleles.

In addition to the case-control analysis, a case-parent triad analysis was
25 conducted, which also showed that carriage of the risk allele for the markers in *KITLG* and proximal to *SPRY4* is associated with TGCT. The per-allele relative risks (RR) for rs3782179 and rs4474514 (*KITLG*) were 2.5 (95% CI = 1.6-0.9) and 2.6 (95% CI = 1.6-4.0), respectively. The per-allele relative risks (RR) for rs4324715 and rs6897876 (proximal to *SPRY4*) were 1.5 (95% CI 1.2, 2.1) and 1.5 (95% CI 1.1, 2.0), respectively.
30 These family-based estimates provide additional evidence that population stratification did not bias results in the replication phase.

Table 5 Associations of *KITLG* and *SPRY4* SNP markers with seminoma and nonseminoma TGCT

		Seminoma			Nonseminoma		
		OR (95% CI)			OR (95% CI)		
Gene	Marker	Risk allele	Per allele	Heterozygote ^a	Homozygote ^b	Per allele	Heterozygote ^a
<i>KITLG</i>	rs4474514 A/G	A	2.97 (2.08-4.23)	1.42 (0.42-4.87)	4.70 (1.44-15.4)	3.27 (2.05-5.18)	1.13 (0.25-5.1)
	(1.03-18.23)						4.34
<i>KITLG</i>	rs3782179 A/G	A	2.95 (2.07-4.21)	1.42 (0.42-4.87)	4.67 (1.43-15.3)	3.30 (2.08-5.24)	1.13 (0.25-5.1)
	(1.04-18.45)						4.39
<i>SPRY4</i>	rs6897876 C/T	C	1.38 (1.11-1.72)	1.11 (0.72-1.73)	1.78 (1.14-2.79)	1.39 (1.07-1.81)	1.55
(0.93-2.56)							
<i>SPRY4</i>	rs4324715 T/C	T	1.40 (1.13-1.73)	1.39 (0.92-2.09)	1.98 (1.27-3.09)	1.32 (1.01-1.71)	1.60
(0.97-2.62)							

Analyses reflect the replication set only.

^aOR for heterozygous carriage of risk allele compared to homozygous carriage of non-risk allele. ^bOR for homozygous carriage of risk allele compared to homozygous carriage of non-risk allele.

An interaction between *KITLG* and *SPRY4* marker genotypes was not observed. In the replication set, marker genotypes in *KITLG* and *SPRY4* were associated with both seminoma and nonseminoma germ cell tumors without indication that genotype
5 associations differed between the two subtypes (Table 5).

In subgroup analyses among those without a family history of TGCT and among those without cryptorchidism, two strong and well-established risk factors for TGCT, the genotypic ORs associated with *KITLG* and *SPRY4* markers were only negligibly
10 attenuated (data not shown). These findings indicate that for most cases, *KITLG* and *SPRY4* do not exert their effect solely through mechanisms involving these known risk factors. Because of limited numbers, it was not possible to examine the effect of *KITLG* and *SPRY4* among those with positive family history or personal history of cryptorchidism.

Variation at 12q22 was identified as a major risk locus for TGCT susceptibility.
15 For rs3782179 and rs4474514, a threefold increased risk of disease per major allele and a 4.5-fold increased risk of disease for homozygous carriage of the major allele were identified. The identified region contains *KITLG*, also known as stem cell factor, encoding the ligand for the receptor tyrosine kinase, c-KIT. The *KITLG*-KIT signaling pathway has an important role in gametogenesis, hematopoiesis and melanogenesis
20 (Roskoski, R. Jr. 2005 *Biochem. Biophys. Res. Commun.* 337, 1-13). In mouse models, *Kitl* (encoded at the *steel (Sl)* locus) is required for multiple aspects of primordial germ cell (PGC) development, including proliferation, migration and survival (Mahakali Zama, A. *et al*, 2005 *Biol. Reprod.* 73, 639-647; Runyan, C. *et al*. 2006 *Development* 133, 4861-4869).

Kitl has a crucial role in the migration of PGCs from the hindgut and subsequent
25 targeting to the genital ridges, and down regulation of *Kitl* in the midline triggers localized apoptosis of PGCs (Runyan *et al*, 2006 cited above). Given the similarity in cellular ultrastructure, patterns of imprinting, and gene expression, multiple human studies have suggested that TGCT arises from PGCs (Oosterhuis, J.W. & Looijenga, L.H.
30 2005 *Nat. Rev. Cancer* 5, 210-222). Delayed differentiation of PGCs has been associated with development of testicular germ cell carcinoma *in situ* among individuals with

intersex conditions and abnormalities of chromosomal number (Rajpert-de Meyts, E. & Høeie-Hansen, C.E., 2007 *Ann. NY Acad. Sci.* 1120, 168-180).

These data support a role for *KITLG* in TGCT susceptibility. Furthermore, loss of the transmembrane form of *Kitl*, which leads to decreased PGC number, has been
5 identified as a TGCT susceptibility locus in the *129/Sv* mouse (Heaney, J.D. *et al*, 2008 *Cancer Res.* 68, 5193-5197). In humans, activating mutations of *KIT* are the most common somatic point mutations in TGCT, present in 25% of seminomas, although rarely identified in nonseminomas (Forbes, S. *et al.* 2006 *Brit. J. Cancer* 94, 318-322). Thus, both germline variation and somatic mutations in the *KITLG-KIT* signaling
10 pathway are associated with TGCT. In addition, *KITLG-KIT* signaling has an important role in male fertility (Blume-Jensen, P. *et al.* 2000 *Nat. Genet.* 24, 157-162), and mutations in *Kitl* lead to decreased germ cell number. The findings suggest that the reported epidemiological association between TGCT and male infertility (Richiardi, L. & Akre, O., 2005 *Cancer Epidemiol. Biomarkers Prev* 14, 2557-2562) may be due, in part,
15 to a common genetic basis.

As *KITLG* has a role in determining level of pigmentation (Miller, C.T. *et al.* 2007 *Cell* 131, 1179-1189), it was postulated that inherited variation at this locus could provide a genetic explanation for the observed differences in TGCT incidence in whites and blacks. *KITLG* has undergone strong positive selection in the European and East
20 Asian populations, with an extended haplotype of 400 kb (Sulem, P. *et al.* 2007 *Nat. Genet.* 39, 1443-1452). Data from HapMap phase 3 show significant differences ($P = 4.3 \times 10^{-20}$) in the frequency of the risk alleles of *KITLG* (rs3782179 and rs4474514) when comparing the CEU (major allele frequency = 0.80) and ASW (African ancestry in Southwest United States: major allele frequency = 0.25) populations (Frazer *et al.*, 2007,
25 cited above). This finding suggests that inherited variation in *KITLG* may explain, in part, the observed differences in TGCT incidence between whites and blacks. An association between TGCT risk and variation at 5q31 was observed just downstream of *SPRY4*. As with *KITLG*, the major allele was associated with increased risk. *SPRY4* is one of a family of four genes (*SPRY1-4*) that have been implicated as negative regulators
30 of the RAS-ERK-MAPK signaling pathway in response to growth factors (Sasaki, A. *et al.*, 2003 *Nat. Cell Biol.* 5, 427-432). Expression analyses and tumor studies have shown

that *SPRY4* is the most significantly down regulated gene when KIT signaling is inhibited by imatinib mesylate in gastrointestinal stromal tumors, supporting a functional relationship between the two proteins (Frolov 2003, cited above).

In summary, the results demonstrate that common genetic variants at the 12p22
5 and 5q31 loci are associated with TGCT and strongly implicate *KITLG* as a susceptibility gene in the pathogenesis of TGCT. In addition, these observations may explain, in part, two important features of the disease: the increased incidence in whites and the epidemiological association with male infertility.

10 EXAMPLE 2: Genome-wide association study.

For the discovery phase, 353 individuals with TGCT were initially selected who were seen at UPHS ($n = 303$) and FCCC ($n = 50$): all cases were from the Philadelphia region. TGCT cases from UPHS were from an ongoing clinic-based case-control study of genetic susceptibility of TGCT for which study participants were asked to complete a
15 self-administered questionnaire that elicited information on known and presumptive risk factors for TGCT. TGCT cases from FCCC were obtained from the Biosample Repository Facility, which collects and stores blood samples and obtains information on family history of cancer, risk factors and demographics from participating subjects. Each individual with TGCT was classified according to the histological diagnosis of his tumor:
20 seminoma or nonseminoma (including yolk sac, choriocarcinoma, embryonal, teratoma and mixed cell type TGCT) germ cell tumor. Only those with primary disease in the testis were included.

Male controls ($n = 932$) were selected from Penn CATH, a UPHS single center, hospital-based study of angiographic coronary artery disease (CAD) in almost 4,000
25 subjects undergoing cardiac catheterization. This study investigates the association of biochemical and genetic factors for CAD and its risk factors (Lehrke, M. *et al.* 2007 *J. Am. Coll. Cardiol.* 49, 442-449). Information on personal history of cancer was not collected. All controls were from the Philadelphia region, 90% were 46 years or older and had already passed the peak age of TGCT development. On the basis of available
30 age-specific TGCT rates, it was estimated that only four TGCT cases would be expected

to have arisen in this control group (Reis *et al.*, cited above). It is unlikely that this potential small misclassification of phenotype would have biased results appreciably.

Controls had been genotyped previously using the Affymetrix Genome Wide Human SNP Array 6.0™ platform and had passed genotyping quality controls measures analogous to those used for TGCT cases (see below). The Affymetrix Genome-Wide Human SNP Array 6.0™ was used to obtain genotypes for TGCT cases. The Birdseed algorithm was to determine genotypes for the combined TGCT case and CAD control sample set (McCarroll, SA *et al.* 2008 *Nat. Genet.* 40, 1166-1174). Among the 353 case samples, 18 subsequently were excluded for not meeting case eligibility (two Leydig cell tumors, one female germ cell tumor erroneously coded as TGCT, 15 non-TGCT samples), and 11 replicate samples with lower genotyping call rates were excluded. Of the 324 unique samples from TGCT cases, 19 (5.9%) were excluded because of a low (<95%) genotyping call rate, 8 (2.5%) because of lower than expected genotypic heterozygosity across called markers ($F_{ST} \geq 0.06$) and 20 (6.2%) because of Asian or African ancestry as determined by multidimensional scaling (MDS) (The Wellcome Trust Case Control Consortium. 2007 *Nature* 447, 661-678). No cases were excluded for cryptic relatedness (proportion of genotypes identical by descent for all cases was <0.20). Among the 932 CAD controls, 13 were excluded because of female or ambiguous sex. After exclusion of 224,705 (24.7%) markers with a minor allele frequency (in the total sample) <0.05, 1,594 (0.2%) that deviated from Hardy-Weinberg equilibrium (HWE: $P < 1 \times 10^{-7}$), 71,978 (7.9%) with an individual genotype call rate <0.95 and 233 (0.03%) invalid markers, 611,112 markers remained in the discovery phase. To further investigate potential bias that could arise from our choice of control group, a comparison was made of the minor allele frequencies of markers brought into replication between those controls with verified coronary heart disease ($n = 700$) and those without ($n = 219$). No statistically significant differences were noted for the six markers, nor were differences observed when comparing the Penn Cath study subjects to population-based controls used in the replication phase (data not shown).

30

EXAMPLE 3: Replication study.

To replicate findings of the discovery phase, 371 cases, 860 controls and 204 sets of mothers and fathers of cases were used from a population-based case-control study of TGCT in western Washington. Methods for recruitment of TGCT cases and parents in
5 this study previously have been published (Starr, J.R. *et al.* 2005 *Cancer Epidemiol. Biomarkers Prev.* 14, 2183-2190). Briefly, all cases had first, primary TGCT diagnosed between 1999 and 2007 and were residents of three urban counties of western
10 Washington aged 18 to 44 years at diagnosis. Control subjects did not have a personal history of TGCT and were frequency-matched on age and ascertained from the general population of the three counties using random digit telephone dialing. Family history of TGCT among first-degree relatives and personal history of cryptorchidism was
ascertained through self-administered questionnaires. Only cases and controls who self-identified as white, non-Hispanic were included in the replication study.

Genotyping was accomplished using predesigned TaqMan SNP Genotyping
15 Assays™ according to manufacturer's specifications. Genotyping was run in duplicate for 1,034 marker pairs (an average of 172 sample pairs per each of the six markers in replication). In total, six (0.58%) calls were discordant; the Spearman correlation coefficient was > 0.99. Genotyping calls were made without knowledge of case or
20 duplicate status. The majority (94-99%) of TGCT cases from the discovery phase were re-genotyped for markers in Table 5. Concordance between genotype calls obtained from the Affymetrix™ chip and TaqMan™ assays for these four makers was 100%.

For both the genome-wide scan and replication study, all participants provided written informed consent approved by their local institutional review boards.

25 EXAMPLE 4: Statistical analysis.

For the discovery phase, PLINK software was used to adjust for missing genotypes and calculate rates of heterozygosity (Purcell, S. *et al.* 2007 *Am. J. Hum. Genet.* 81, 559-575). Population stratification was assessed using multi-dimensional
30 scaling (MDS) methods and all markers were tested for HWE. PLINK also was used to determine genotypic associations among the 277 TGCT cases and 919 CAD controls. Statistical significance was assessed using Fisher's exact test, and for top hits ORs and

95% CIs were determined for the per-allele, heterozygous and homozygous effects of the minor allele (Fig. 6). Imputation was conducted using a computationally efficient hidden Markov model-based algorithm as implemented in software MACH (Li, Y. & Abecasis, G.R., 2006 *Am. J. Hum. Genet.* S79, 2290). MACH combines our genotyped data with phased chromosomes from the HapMap CEU samples and then infers the unknown genotypes in the study sample probabilistically by searching for similar stretches of flanking haplotype in the HapMap CEU reference sample. Only markers were analyzed that passed the following imputation quality control criteria: $R^2 > 0.3$ and MAF > 0.05 in both cases and controls. To account for uncertainty involved in the imputation, case-control associations were analyzed for imputed SNP markers using software SNPTEST (Marchini, J., *et al*, 2007 *Nat. Genet.* 39,906-913).

For the replication phase, analyses were conducted using SAS v9.1.3. Unconditional logistic regression was used to determine per-allele associations and associations of homozygous and heterozygous carriage of risk alleles with case status (overall and among specified subgroups), and present unadjusted ORs because age was not a confounder in our data. Trend across genotype categories was assessed by the Cochran-Armitage test for trend.

Models containing markers coded on an ordinal scale (additive model) and a cross-product term were made to test for marker-marker interaction. To estimate and compare the associations within TGCT subtypes, multi nomial logit models were used to obtain simultaneously the OR and 95% CI for the association between markers and each level of outcome after adjusting for age.

Unless defined otherwise in this specification, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs and by reference to published texts. While the invention has been described with reference to specific embodiments, it is appreciated that modifications can be made without departing from the spirit of the invention. Such modifications are intended to fall within the scope of the appended claims.

All documents listed in this specification are incorporated herein by reference. The publication P. Kanetsky *et al*, "Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer", *Nature Genetics*, 41(7):811-815 (July, 2009),

as well as its on-line publication of May 31, 2009, identified as <http://dx.doi.org/10.1038/ng393> and all supplemental materials published therewith are incorporated herein by reference.

The priority US provisional patent application No. 61/182735 is expressly
5 incorporated by reference herein in its entirety.

Applicants hereby incorporate by reference the Sequence Listing material filed in electronic form herewith. This file is labeled "UPN-V5217PCT_ST25.txt".

What is claimed is:

1. A method for identifying a risk for testicular germ cell cancer comprising:
contacting a biological sample from a human subject with a reagent that is capable of identifying one or more genetic variants of the *KITLG* gene or the *SPRY4* gene that is associated with susceptibility of a human subject to testicular germ cell tumors (TGCT), wherein the genetic variant of the *KITLG* gene is a nucleotide sequence containing a single nucleotide polymorphism (SNP) selected from the group consisting of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, and a combination thereof or wherein the genetic variant of the *SPRY4* gene is a nucleotide sequence containing a single nucleotide polymorphism (SNP) selected from the group consisting of rs4324715, rs6897876, rs12521013 and a combination thereof; and
identifying the occurrence of, or an increased susceptibility to, testicular germ cell cancer when one or more said variants are identified in said sample.
2. A method for identifying a risk for testicular germ cell cancer comprising:
contacting a biological sample from a human subject with a reagent that is capable of identifying one or more genetic variants of the *KITLG* gene that is associated with susceptibility of a human subject to testicular germ cell tumors (TGCT) wherein the genetic variant of the *KITLG* gene is a nucleotide sequence containing a single nucleotide polymorphism (SNP) selected from the group consisting of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, and a combination thereof;
and
identifying the occurrence of, or an increased susceptibility to, testicular germ cell cancer when one or more said variants are identified in said sample.
3. A method identifying a risk for testicular germ cell cancer comprising:
contacting a biological sample from a human subject with a reagent that is capable of identifying one or more genetic variants of the *SPRY4* gene that is associated with susceptibility of a human subject to testicular germ cell tumors (TGCT) wherein the genetic variant of the *SPRY4* gene is a nucleotide sequence containing a single nucleotide

polymorphism (SNP) selected from the group consisting of rs4324715, rs6897876, rs12521013 and a combination thereof; and

identifying the occurrence of, or an increased susceptibility to, testicular germ cell cancer when one or more said variants are identified in said sample.

4. The method according to any of claims 1 to 3, wherein said reagent is capable of forming a physical association with said SNP in a biological sample containing said one or combination of said SNPs.

5. The method according to claim 4, wherein said reagent is a nucleic acid sequence capable of hybridizing with one allele of a nucleotide sequence containing a selected SNP.

6. The method according to claim 5 wherein said reagent is a genomic probe and said physical association is the hybridization of said probe to the cDNA or mRNA of a sequence containing said SNP or SNPs.

7. The method according to claim 4 wherein said reagent is a PCR primer probe set and said physical association is the hybridization of said probe/primer to the mRNA of a sequence containing said SNP or SNPs.

8. The method according to claim 4, wherein said reagent is associated with a directly-detectable, or indirectly-detectable, label.

9. The method according to claim 4, wherein said reagent is immobilized on a substrate.

10. The method according to claim 4 wherein said substrate is computer chip or computer-readable chamber.

11. The method according to claim 4 wherein said substrate is a microfluidics card.
12. The method according to any of claims 1-7, comprising a microarray of two or more said reagents capable of identifying the presence of two or more SNPs in a biological sample.
13. The method according to claim 12, wherein said two or more SNPs are selected from the group of marker sequences consisting of: rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, rs4324715, rs6897876, rs12521013 and a combination thereof
14. The method according to claim 13, wherein said two or more SNPs are selected from the marker sequences consisting of: rs3782179, rs4324715, rs4474514, and rs6897876.
15. The method according to any of claims 1 to 14 wherein said contacting comprises forming a physical association between the reagent and the variant in said sample.
16. The method according to claim 15, further comprising transforming the detectable signals generated from the reagent in association with a SNP present in the biological sample into numerical or graphical data.
17. The method according to claim 16, wherein said transforming is performed by a suitably-programmed machine or instruments that can detect the detectable signals generated from the reagents associated with the SNPs present in the biological sample and transform same into numerical or graphical data useful in performing the identification.

18. The method according to claim 17, wherein identifying the occurrence of, or an increased susceptibility to, testicular germ cell cancer comprises coupling the identification of the selected SNPs with the presentation of clinical symptoms in a subject.

19. The method according to any one of claims 1 to 14, wherein identifying the occurrence of, or an increased susceptibility to, testicular germ cell cancer comprises coupling the identification of the selected SNPs with familial history of testicular cancer and/or a personal history of undescended testes.

20. The method according to any one of claims 1 to 14, wherein identifying the occurrence of, or an increased susceptibility to, testicular germ cell cancer provides a quantitative assessment of the likelihood or risk of TGCT occurrence in a subject that has not yet developed clinical symptoms of TGCT.

21. A composition for identifying a risk for testicular germ cell cancer comprising a reagent that is capable of identifying a genetic variant of the *KITLG* gene that is associated with susceptibility of a human subject to testicular germ cell tumors (TGCT), wherein the genetic variant of the *KITLG* gene is a nucleotide sequence containing a single nucleotide polymorphism (SNP) selected from the group consisting of rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, and a combination thereof.

22. A composition for identifying a risk for testicular germ cell cancer comprising a reagent that is capable of identifying a genetic variant of the *SPRY4* gene that is associated with susceptibility of a human subject to testicular germ cell tumors (TGCT), wherein the genetic variant of the *SPRY4* gene is a nucleotide sequence containing a single nucleotide polymorphism (SNP) selected from the group consisting of rs4324715, rs6897876, rs12521013 and a combination thereof.

23. The composition according to claim 21 or 22, wherein said reagent is capable of forming a physical association with said SNP in a biological sample containing said one or combination of said SNPs.

24. The composition according to claim 23, wherein said reagent is a nucleic acid sequence capable of hybridizing with one allele of a nucleotide sequence containing a selected SNP.

25. The composition according to claim 24 wherein said reagent is a genomic probe and said physical association is the hybridization of said probe to the cDNA or mRNA of a sequence containing said SNP or SNPs.

26. The composition according to claim 23 wherein said reagent is a PCR primer probe set and said physical association is the hybridization of said probe/primer to the mRNA of a sequence containing said SNP or SNPs.

27. The composition according to claim 23, wherein said reagent is associated with a directly-detectable, or indirectly-detectable, label.

28. The composition according to claim 23, wherein said reagent is immobilized on a substrate.

29. The composition according to claim 23 wherein said substrate is computer chip or computer-readable chamber.

30. The composition according to claim 23 wherein said substrate is a microfluidics card.

31. The composition according to any of claims 21-26, comprising a microarray of two or more said reagents capable of identifying the presence of two or more SNPs in a biological sample.

32. The composition according to claim 31, wherein said two or more SNPs are selected from the group of marker sequences consisting of: rs995030, rs1352947, rs1472899, rs3782179, rs3782181, rs4474514, rs11104952, rs4324715, rs6897876, rs12521013 and a combination thereof

33. The composition according to claim 32, wherein said two or more SNPs are selected from the marker sequences consisting of: rs3782179, rs4324715, rs4474514, and rs6897876.

34. A kit for identifying a risk for testicular germ cell cancer comprising one or more reagents that are capable of identifying a SNP or combination of SNPs associated with susceptibility of a human subject to testicular germ cell tumor (TGCT) in a biological sample.

35. The kit according to claim 34 that comprises a reagent or multiple reagents of any of claims 1-13.

36. The kit according to claim 35 comprising a positive control or a negative control.

37. The method according to claim 1, comprising contacting the biological sample with a reagent of any of claims 21-34.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/36606

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68, C07H 21/00, C12P 19/34 (2010.01)

USPC - 435/6, 536/23.1, 536/24.3

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(8) - C12Q 1/68, C07H 21/00, C12P 19/34 (2010.01)

USPC - 435/6, 536/23.1, 536/24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

IPC(8) - C12Q 1/68, C07H 21/00, C12P 19/34 (2010.01) - see keyword below

USPC - 435/6, 536/23.1, 536/24.3 - see keyword below

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubWEST(USPT,PGPB,EPAB,JPAB); Medline, Google: testicular germ cell cancer, KITLG, c-Kit ligand, single nucleotide polymorphism, SNP, variant, risk, susceptibility, rs995030, hybridizing, allele, probe, primer, human, cDNA, mRNA, PCR, primer, cDNA, mRNA, genomic, label, substrate, chip, immobilize, link, substrate, computer-readable chamber, micr

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HEANEY et al., Loss of the Transmembrane but not the Soluble Kit Ligand Isoform Increases Testicular Germ Cell Tumor Susceptibility in Mice. Cancer Res. 1 July 2008, Vol. 68, No 13, pp. 5193-5197. Abstract; pg 5194, col 2, para 2 and para 3; pg 5195, col 2, middle para, Fig 1B, and Table 2	1-2, 4-11, 21, 23-30; 34
Y	GALAN et al. Association of genetic markers within the KIT and KITLG genes with human male infertility. Human Reproduction, December 2006, Vol.21, No 12, pp 3185-3192. Abstract; and pg 3186, col 1, para 4, and col 2, para 4 and para 5	1-2, 4-11, 21, 23-30
Y	US 2006/0147955 A1 (ALLAWI et al.) 06 July 2006 (06.07.2006), para [0004], [0012], [0014], [0029], [0035], [0050], [0062], [0072], [0092], [0147], [0158], [0201], [0252], [0254], [0255], [0265], [0268], and [0414]	6-11, 25-30, 34
A	SUYAMA, DNA chips - Integrated Chemical Circuits for DNA Diagnosis and DNA computers. 2007, Institute of Physics, Graduate School of Arts and Sciences, The University of Tokyo, pp 1-6	10, 29
P, A	KANETSKY et al., Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat Genet. July 2009, Vol 41, No 7, pp 811-5. Epub 2009 May 31. Entire document	1-2, 4-11, 21, 23-30, 34

 Further documents are listed in the continuation of Box C.

* 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 application or patent 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)

"O" 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

"&" document member of the same patent family

Date of the actual completion of the international search

08 September 2010 (08.09.2010)

Date of mailing of the international search report

15 NOV 2010

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/36606

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.: 12-20, 31-33, and 35-37
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Group I-: Claims 1-2, 4-11, 21, 23-30, and 34, drawn to a method for identifying a risk for testicular germ cell cancer comprising: contacting a biological sample from a human subject with a reagent that is capable of identifying one or more genetic variants of the KITLG gene, and a composition for identifying a risk for testicular germ cell cancer comprising a reagent that is capable of identifying a genetic variant of the KITLG gene that is associated with susceptibility of a human subject to testicular germ cell tumors (TGCT). The first invention encompasses a single nucleotide polymorphism rs995030. Should an additional fee(s) be paid, Applicant is invited to elect an additional SNP(s) to be searched.

Group II-: Claim 1, 3-11, 22-30, and 34, drawn to a method for identifying a risk for testicular germ cell cancer comprising: contacting a biological sample from a human subject with a reagent that is capable of identifying one or more genetic variants of the SPRY4 gene, and a composition for identifying a risk for testicular germ cell cancer comprising a reagent that is capable of identifying a genetic variant of the SPRY4 gene that is associated with susceptibility of a human subject to testicular germ cell tumors (TGCT). The first invention encompasses a single nucleotide polymorphism rs4324715. Should an additional fee(s) be paid, Applicant is invited to elect an additional SNP(s) to be searched.

-----please see extra sheet-----

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-2, 4-11, 21, 23-30, and 34, limited to rs995030

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/36606

Continuation of:

Box No. III Observations where unity of invention is lacking

The inventions listed as Groups I-II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions of Group I+ do not include the inventive concept of the SPRY4 gene, that is required by Group II+. The inventions of Group II+ do not include the inventive concept of the KITLG gene, that is required by Group I+.

The inventions of Groups I+ and II+ share the technical feature of one or more genetic variants that are associated with susceptibility of a human subject to testicular germ cell cancer. However, this shared technical feature does not represent a contribution over prior art. Specifically, the article entitled "Loss of the Transmembrane but not the Soluble Kit Ligand Isoform Increases Testicular Germ Cell Tumor Susceptibility in Mice" by Heaney et al. (hereinafter 'Heaney'; Cancer Res. 2008, Vol. 68(13), p. 5193-5197) teaches that "Several genetic variants act as modifiers of testicular germ cell tumor (TGCT) susceptibility in the 129/Sv mouse model of human pediatric TGCTs. One such modifier, the Steel locus, encodes the transmembrane-bound and soluble ligand of the kit receptor. Some (Sl and SlJ) but not all (Sld) mutations of the Steel locus increase TGCT incidence in heterozygous mutant mice" (Abstract), wherein 'ligand of the kit' is encoded by the KITLG gene (Specification: pg 2, ln 23 - 'the KITLG (c-KIT ligand) gene'). Heaney further discloses SPN of the KITLG gene (Fig 1, B). Furthermore, SNPs of KITLG have been studied long before said invention was made. Specifically, the article entitled "Association of genetic markers within the KIT and KITLG genes with human male infertility" by Galan et al. (Human Reproduction Vol.21, No.12 pp. 3185?3192, 2006) teaches the SNP allele of rs995030 in humans (Abstract).

Groups I+-II+ therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.