microbial infection with microarrays include global analysis of microbial virulence determinants and host defense systems, development of mathematical approaches to model the interplay between host resistance and microbial infectivity and construction of databases to manage and analyze vast amounts of data derived from genotypic and phenotypic infectomics. (***See*** Microarray Bioinformatics.)

See also

DNA Chips and Microarrays
Microarrays and Expression Profiling in Cancer
Microarrays in Disease Diagnosis and Prognosis

References

Baldi P and Brunak S (eds.) (2001) *Bioinformatics: The Machine Learning Approach*, 2nd edn, pp. 299–321. Cambridge, MA: MIT Press.

Björkholm B, Lundin A, Sillen A, *et al*. (2001) Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. *Infection and Immunity* **69**(12): 7832–7838.

Cummings CA and Relman DA (2000) Using DNA microarrays to study host–microbe interactions. *Emerging Infectious Diseases* **6**(5): 513–525.

Fauci AS (2001) Infectious diseases: considerations for the twenty-first century. *Clinical Infectious Diseases* **32**(5): 675–685.

Grifantini R, Bartolini E, Muzzi A, *et al*. (2002) Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. *Nature Biotechnology* **20**(9): 914–921.

Huang SH, Triche T and Jong AY (2002) Infectomics: genomics and proteomics of microbial infections. *Functional and Integrative Genomics* **1**(6): 331–344.

McClane BA and Mietzner TA (eds.) (1999) *Microbial Pathogenesis: A Principles-oriented Approach*, pp. 1–5. Madison, CT: Fence Creek Publishing.

Shmulevich I, Dougherty ER, Kim S and Zhang W (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18**(2): 261–274.

Walter G, Bussow K, Lueking A and Glokler J (2002) High-throughput protein arrays: prospects for molecular diagnostics. *Trends in Molecular Medicine* **8**(6): 250–253.

Wang D, Liu S, Trummer BJ, Deng C and Wang A (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nature Biotechnology* **20**(3): 275–281.

Further Reading

Casadevall A and Pirofski LA (2000) Host–pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infection and Immunity* **68**(12): 6511–6518.

Dhiman N, Bonilla R, O'Kane DJ and Poland GA (2001) Gene expression microarrays: a twenty-first century tool for directed vaccine design. *Vaccine* **20**(1 and 2): 22–30.

Fukui S, Feizi T, Galustian C, Lawson AM and Chai W (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate–protein interactions. *Nature Biotechnology* **20**: 1–7.

Gronow S and Brade H (2001) Lipopolysaccharide biosynthesis: which steps do bacteria need to survive? *Journal of Endotoxin Research* **7**(1): 3–23.

Hooper LV and Gordon JI (2001) Commensal host–bacterial relationships in the gut. *Science* **292**: 1115–1118.

Joyce EA, Chan K, Salama NR and Falkow S (2002) Redefining bacterial populations: a post-genomic reformation. *Nature Reviews Genetics* **3**(6): 462–473.

Lucchini S, Thompson A and Hinton JC (2001) Microarrays for microbiologists. *Microbiology* **147**(6): 1403–1414.

Mitchell P (2002) A perspective on protein microarrays. *Nature Biotechnology* **20**(3): 225–229.

Rathod PK, Ganesan K, Hayward RE, Bozdech Z and DeRisi JL (2002) DNA microarrays for malaria. *Trends in Parasitology* **18**(1): 39–45.

Schweitzer B and Kingsmore SF (2002) Measuring proteins on microarrays. *Current Opinion in Biotechnology* **13**(1): 14–19.

Infertility: Genetic Disorders

Intermediate article

Peter H Vogt, *University of Heidelberg, Heidelberg, Germany*

Identification of genetic disorders causing human male or female infertility is of paramount importance in each infertility clinic that has to consider the possible risk of transfer of a specified genetic abnormality (i.e. a distinct chromosomal aberration and/or a specific gene mutation) to the offspring created by the applied artificial fertilization protocol. This is useful especially for those men and women where no obstructions of the germ line can be observed and who are therefore diagnosed in the clinic as having idiopathic infertility.

Article contents

Introduction

The prevalence of human infertility (male and female) in Western countries has been estimated to be between 10% and 15%. In 20% of the cases the causative agent was determined to be a purely male factor and in 38% of the cases a purely female factor (World Health Organization, 1987). In 27% of the cases both partners

presented with clinical abnormalities and in the remaining 15% of the cases no clear cause for the observed infertility could be diagnosed (i.e. idiopathic infertility). At the moment it is not possible to say how often gene mutations cause or contribute to the occurrence of infertility in humans. For most of the genes shown to be expressed in the germ-line mutations causing male or female infertility have still to be identified. One major reason for this lack of knowledge is the fact that the clinical analysis of infertility is often compromised by the lack of a detailed histological analysis of the patient's germ-line cells. Therefore, the exact interruption phase in the germ line and the functional state of the different germ-line cells are often not known. Consequently, the selection of specific patient subgroups, considered to be the most promising candidates for mutations in specific germ-line genes, is difficult and molecular mutation analysis is usually performed on the large patient group diagnosed as idiopathic. Genetic disorders of human male and female infertility first came into focus 34 years ago as a result of the cytogenetic characterization of microscopically visible chromosome aberrations in patients with reproductive failures.

Chromosome Aberrations and Male Infertility

Chromosome abnormalities are found in about 14% of azoospermic men (i.e. infertile patients with no spermatozoa in their semen fluid) and in 5% of oligozoospermic men (i.e. infertile patients with a severe reduction in the number of spermatozoa in their semen fluid) (Van Assche *et al*., 1996). Among azoospermic men, the karyotype 47,XXY associated with the Klinefelter syndrome is most frequent. Klinefelter patients have small testes (hypogonadism), indicating their inability to produce a normal number of spermatozoa (i.e. 20–100 million ml^{-1} ejaculate). However, the presence of a low number of spermatozoa in some of these patients indicates that the second X chromosome does not interrupt spermatogenesis but only reduces its quantitative outcome. This reduction of postmeiotic germ cells is probably the result of an interference of the premeiotic pairing process of the sex chromosomes by the second X chromosome competing with the Y chromosome. Further chromosome abnormalities associated with azoospermia are the 46,XX and 45,X karyotypes in phenotypic males and most aberrations of the euchromatic part of the long arm of the Y chromosome (Yq11). Autosome aneuploidies, like trisomies of chromosome 21 (Down syndrome), have individual effects on male fertility, and autosomal translocations (Robertsonian and reciprocal) were found more frequently in men with oligozoospermia (3%) than with azoospermia (1.6%) (Van Assche *et al*., 1996). In 60% of all infertile men with Robertsonian translocation chromosomes, the two acrocentric chromosomes, 13 and 14, were involved. Reciprocal translocations involving either the X or the Y chromosome always result in male infertility (Gabriel-Robez and Rumpler, 1996), whereas autosomal translocations are also found in fertile men. The variable expression of the chromosomally based infertility factors are most probably associated with variations of the meiotic cell division process. It can be assumed that chromosome mutations are generally counter-selected in a manner individually and specifically oriented by the meiotic segregation mechanisms. With the aid of fluorescence *in situ* hybridization (FISH) it is now possible to analyze chromosome abnormalities in sperms directly. A high incidence of sex chromosomal aneuploidies was observed especially in sperm samples with a low morphological quality (Pfeffer *et al*., 1999; Lewis-Jones et al., 2003). This is most important for the therapeutic perspective of this patient group since sex chromosome aneuploidies have a high risk to be transmitted to the offspring applying intra-cytoplasmatic sperm injection (ICSI) which is not selecting towards bad sperm morphology.

Chromosome Aberrations and Female Infertility

The 45,X karyotype is the most frequent chromosome aneuploidy associated with female infertility. A white fibrous streak gonad completely depleted of germ cells is usually observed in this patient group (primary amenorrhea). About 50% of females with gonadal dysgenesis are 45,X and 25% have a 45,X/46,XX karyotype (Simpson and Rajkovic, 1999). An increased atresia of the oocytes during fetal development and before puberty is the primary cause of infertility in these females (primary ovarian failure). In contrast to males, the full complement of germ cells (i.e. about 2 million oocytes) is already stored in females at birth. Multiple genes then control the rate of the apoptotic processes of the germ cell reduction to about 400 000 at puberty and to about 1000 at the age of 50 years (Morita and Tilly, 1999). The increased apoptotic germ cell death in females with a 45,X karyotype suggests that at least some genes important for the maintenance of folliculogenesis are located on the X chromosome (**Figure 1**).

Premature ovarian failure (POF) after puberty (i.e. secondary amenorrhea) is found in women with

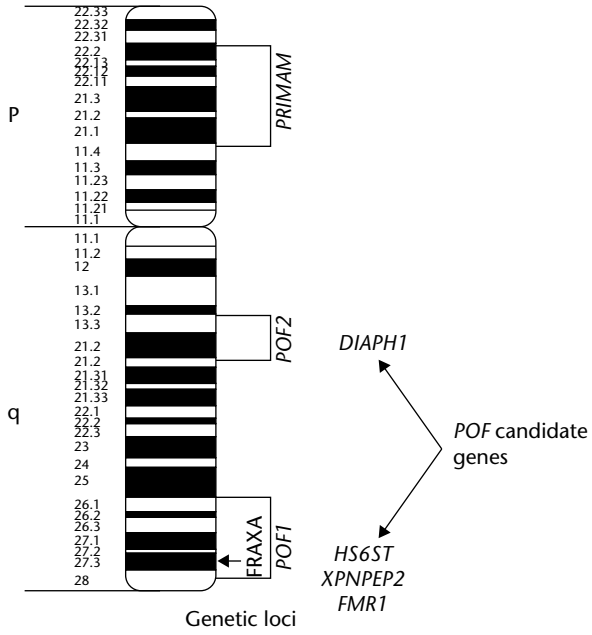

Figure 1 X-chromosome G banding pattern. Multiple genetic loci for the maintenance of folliculogenesis are located on the human X chromosome. Deletions in the primary amenorrhea (*PRIMAM*) locus in Xp21.1–22.2 cause primary ovarian failure (POF), before puberty. Deletions in *POF1* locus (Xq26.1–28) and *POF2* (Xq13.3–21.1) cause premature ovarian failure during the reproductive cycle of the female, that is, after puberty. Putative POF candidate genes are listed at the right. See text for further descriptions.

deletions of the long arm of the X chromosome (Xq; **Figure 1**). These so-called POF syndromes affect at least 1% of women and it has been postulated that gene mutations or chromosome abnormalities are the causative agents in $<31\%$ of cases (Vegetti *et al.*, 1998).

Naturally, the availability of human female germ cells for genetic analyses is limited and only possible in the context of the use of an *in vitro* fertilization protocol. Chromosome aneuploidies and structural aberrations were analyzed with high frequencies in oocytes unable to form pronuclei and in those with an abnormal morphology (Pellestor, 1991). It is now possible to analyze the karyotypes of oocytes also by chromosome analysis of the second polar body during genetic preimplantation diagnosis. About 80% of D and G chromosome trisomies (i.e. trisomies of chromosomes 13, 14, 15, and 21, 22) were found to be of maternal origin, resulting from first meiotic nondisjunctions in the maturating oocytes. Chromosome aneuploidies in the female germ cell increase with age, which is probably because of declining levels of meiotic recombination events which are different for each chromosome.

Molecular Genetic Disorders and Male Infertility

Genes functional in the male germ line are recognized by mutations causing male infertility. In humans, this can become a cumbersome task (see Introduction) unless the gene to be analyzed has a high mutation rate. This is the case for the cystic fibrosis (CF) transmembrane conductance regulator (*CFTR*) gene located on the long arm of chromosome 7 (7q31.2). Carriers of a *CFTR* mutation (i.e. heterozygotes) are found with 1 : 25 carrier rate in different Caucasian populations. Most males with CF also suffer from a congenital bilateral absence of the vas deferens (CBAVD) leading to obstructive azoospermia, that is, the transport of the mature spermatozoa through the testicular and epididymal structures is blocked. The *CFTR* gene functions as a low-voltage, cyclic-AMP-regulated chloride channel. Its genetic activity in epithelial cells of the male germ line is mainly in the columnar epithelium of the head of the epididymis. It has been shown that just a small reduction of functional *CFTR* transcripts can result in male infertility as a result of CBAVD without the occurrence of the somatic CF phenotype (Estivill, 1996). This is linked to the increased frequency (84%) of the 5T allele in the splice acceptor site of *CFTR* intron 8 (**Figure 2**) in CBAVD patients. The *CFTR* gene copies with the 5T variant only produce about 10% of functional *CFTR* mRNA. Epithelial cells in the male genital tract appear to be more sensitive to a quantitative CFTR protein deficiency than the same cell type in other organs. The pathology of obstructive azoospermia occurs with approximately 30% in the group of patients with azoospermia, and of those 25% suffer from CBAVD.

Frequent molecular mutations resulting in male infertility were also found in the azoospermia factor (*AZF*) locus on the long arm of the human Y chromosome (Yq11). However, these were not point mutations in a specific Y gene functional in the male germ line, but three different microdeletions with a molecular extension between 1 and 7 Mb of genomic Y DNA. They were designated as AZoospermia Factor regions, AZFa, AZFb and AZFc (Vogt *et al.*, 1996) because each deletion seemed to be associated with the occurrence of a specific testicular pathology. Men with deletion of AZFa were suffering from a complete absence of germ cells in their testicular tissue (Sertoli-cell-only syndrome). Men with deletion of AZFb displayed a meiotic arrest of their spermatogenesis and men with deletion of AZFc were still able to produce mature spermatozoa, although only in small numbers (i.e. severe oligozoospermia). Therefore, in rare cases, AZFc deletions were found to be hereditary from father to son.

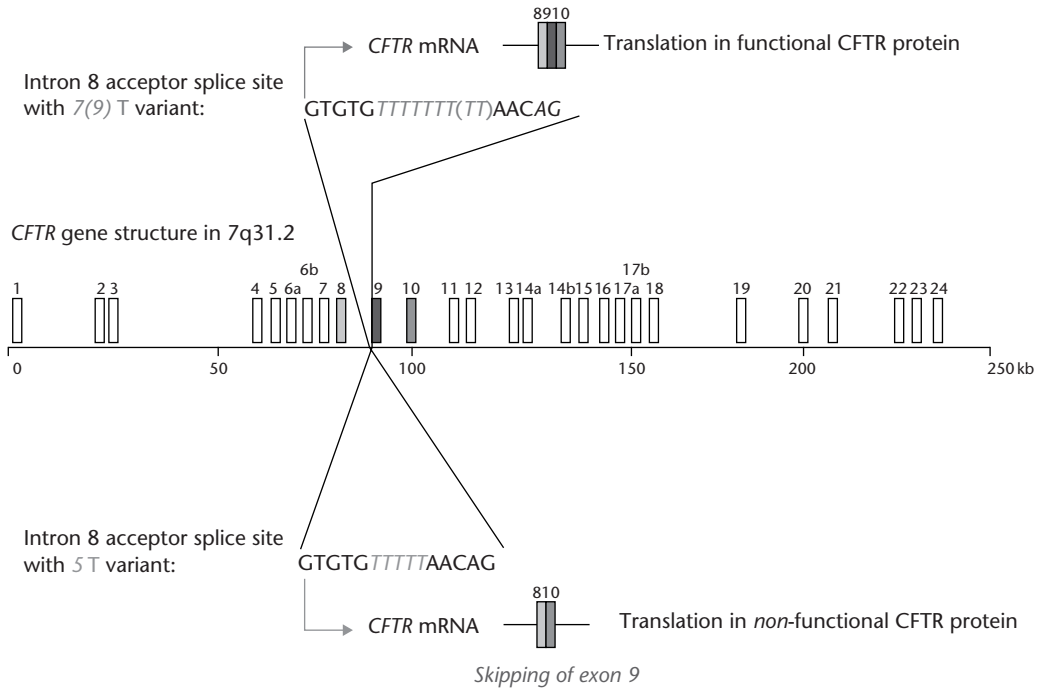

Figure 2 Increased frequency of 5T variant in the splice acceptor site of *CFTR* intron 8 in men with congenital bilateral absence of the vas deferens (CBAVD). Reduction of the T nucleotides to five causes most often skipping of *CFTR* exon 9, which results in the translation of nonfunctional CFTR protein (90%).

Intrachromosomal recombination events between specific human endogenous retroviral sequence blocks were found to be the major cause of AZFa deletions (Kamp *et al*., 2000) and, similarly, recombinations between large homologous sequence blocks composed of different repetitive sequence families are the major cause of complete AZFb and AZFc deletions and their molecular extension was estimated with 6.2 Mb (AZFb) and 4.5 Mb (AZFc), respectively (Kuroda-Kawaguchi *et al*., 2001; Repping *et al*. 2002). Each AZF region contains a series of genes expressed in testis tissue. Some of these genes are also expressed in other tissues and have a homolog on the X chromosome (in AZFa: Dead Box Y, DBY/DBX and Ubiquitin Specific Protease 9, USP9Y/USP9X; in AZFb: Essential-Initiation-Translation Factor 1AY, EIF1AY/EIF1AX and Selected Mouse cDNA Y, SMCY/SCMX). Some have multiple copies in the AZFc region (Deleted in Azoospermia, DAZ1-4, Basic protein Y2, BPY2.1-3) or are dispersed on the Y chromosome with nonfunctional pseudogenes (AZFb: RNA binding motif Y, RBMY, Testis-specific protein Y encoded, TSPY; AZFc: Chromo-domain Y, CDY, Protein phosphatase related Y, PRY). These multicopy structures of the AZF candidate genes complicate the molecular analyses of specific mutation events in some of these Y genes in patients with idiopathic infertility. A laborious screening protocol for USP9Y point mutations in 576 patients with an idiopathic low sperm count revealed one individual with a 4 nt deletion in the splice acceptor site of USP9Y exon 7 (Sun *et al*., 1999). However, although this mutation severely truncated the USP9Y protein (about 90%), the patient's testis tissue did not display the Sertoli-cell-only syndrome expected from an AZFa mutation (see above); all premeiotic germ cells and even spermatids were present in the patient's testis tubules. It has therefore been postulated that the different testicular AZF pathologies may be observed only after deletion of the complete gene content of each AZF region. However, this might not be true for AZFc deletions because DAZ1/DAZ2 gene doublet deletions in AZFc seem to cause oligozoospermia as well (Fernandes *et al*., 2002).

Large mitochondrial DNA deletions resulting in the formation of an incomplete electron transport chain are assumed to play an important role in the pathophysiology of sperms with a reduced motility (i.e. asthenozoospermia) (Kao *et al*., 1998), although a distinct correlation of a specific etiology to any of these DNA deletions could not yet be made. Asthenozoospermia contributes to male infertility in about 30% of all cases. Ruiz-Pesini *et al*. (2000) observed that the mitochondrial DNA haplotype T was significantly overrepresented in men with asthenozoospermia and that this haplotype was associated with point

mutations in some mitochondrial tRNAs, the pathology of which is not yet known. A significantly higher variability of the *N*-terminal polyglutamine tract of the mitochondrial DNA polymerase (POLG) was found in 5–10% of men with oligozoospermia (Rovio *et al*., 2001). This peptide is encoded by a CAG microsatellite repeat, which usually has a length of 10 repeats. This allele is absent in infertile men in all populations studied so far. It is therefore assumed that any impairment of the sperms' mitochondrial genetic activities controlling their energy metabolism will result in a reduced sperm motility.

It has been found that men with meiotic arrest who suffer from azoospermia or oligozoospermia have a decreased efficiency of their meiotic DNA repair mechanisms (Nudell *et al*., 2000). This suggests that defects in genes required in DNA repair could contribute to this germ-line pathology at least in some of them. The use of testicular spermatozoa from this patient group in artificial insemination protocols should therefore be accompanied by DNA quality controls.

Molecular Genetic Disorders and Female Infertility

A series of candidate genes, which cause ovarian failure if mutated, were isolated (Simpson and Rajkovic, 1999). Disruption of the human homolog of the *Drosophila* Diaphanous (*DIAPH2*) gene on the X chromosome was mapped to the POF2 region (**Figure 1**) and the heparan-sulfate-6-sulfotransferase (*HS6ST*) and Xaa-Pro aminopeptidase (*XPNPEP2*) genes were identified as strong candidate genes for expression of POF1. In 16% of pedigrees with familial POF, the expansion of the CCG trinucleotide in exon 1 of the *FMR1* gene in Xq28 (also associated with the fragile X syndrome) seems to be a causative agent (Conway *et al*., 1998). It is assumed that each POF locus on the X chromosome contains more than one gene functional important for the maintenance of the ovary function. Alternatively, it has been argued that any disruption of the meiotic sex chromosome pairing structure in Xq by autosomal translocations would naturally interfere with the meiotic maturation process of the oocyte and therefore cause a POF syndrome. Targeted gene mutation analyses in mice and mice transgenesis experiments (Greenhouse *et al*., 1998) indicated that a series of additional autosomal genes must also be important for female fertility. For the maintenance of ovarian follicles, a haploinsuffient forkhead transcription factor (FOXL2 on chromosome 3; 3q23) was identified, which also cause the autosomal dominant blepharophimosis/ptosis/epicanthus inversus syndrome (BPES).

Mutations in the androgen receptor (AR) gene leading to complete androgen insensitivity result in testicular feminization. These individuals are phenotypically female but genotypically male with a normal 46,XY karyotype. They lack not only a uterus, oviducts and the cervix, but also the vas deferens, epididymis and penis, the development of which is dependent on a normal AR binding of the hormones, testosterone and dihydrotestosterone. Infertility is therefore a secondary consequence in these females with a primary defect in their gonadal development. Depending on the severity of the androgen binding deficiency to the androgen receptor, only male infertility occurs and no sex reversal. Interestingly, the same AR mutation can cause severe and moderate androgen binding deficiencies (Gottlieb *et al*., 1999). Like other members of the steroid/nuclear receptor superfamily of transcription factors, the AR has four main functional domains, comprising the *N*-terminal transactivation domain, a DNA-binding domain, a hinge region and *C*-terminal ligand-binding domain. Analysis of the tertiary architecture of the AR protein has shown that AR mutations causing the androgen insensitivity syndromes are induced by even small structural alterations of the helical sandwich structure of the ligand-binding domain (Ong *et al*., 2002).

See also

Clinical Molecular Cytogenetics
Databases in Genetics Clinics
Gonadotropin Hormones: Disorders

References

Conway GS, Payne NN, Webb J, Murray A and Jacobs PA (1998) Fragile X premutation screening in women with premature ovarian failure. *Human Reproduction* **13**: 1184–1187.

Estivill X (1996) Complexity in a monogenic disease. *Nature Genetics* **12**: 348–350.

Fernandes S, Huellen K, Goncalves J, *et al*. (2002) High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. *Molecular Human Reproduction* **8**: 286–298.

Gabriel-Robez O and Rumpler Y (1996) The meiotic pairing behaviour in human spermatocytes carrier of chromosome anomalies and their repercussions on reproductive fitness. II. Robertsonian and reciprocal translocations. A European collaborative study. *Journal Annual Genetics* **39**: 17–25.

Gottlieb B, Pinskky L, Beitel LK and Trifiro M (1999) Androgen insensitivity. *American Journal of Medical Genetics* **89**: 210–217.

Greenhouse S, Rankin T and Dean J (1998) Genetic causes of female infertility: targeted mutagenesis in mice. *American Journal of Human Genetics* **62**: 1282–1287.

Kamp C, Hirschmann P, Voss H, Huellen K and Vogt PH (2000) Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Human Molecular Genetics* **9**: 2563–2572.

Kao SH, Chao HT and Wei YH (1998) Multiple deletions of mitochondrial DNA are associated with the decline of motility and fertility of human spermatozoa. *Molecular Human Reproduction* **4**: 657–666.

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, *et al*. (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genetics* **29**: 279–286.

Lewis-Jones I, Seshadri S, Douglas A and Howard P (2003) Sperm chromosomal abnormalities are linked to sperm morphologic deformities. *Fertility and Sterility* **79**: 212–215.

Morita Y and Tilly JL (1999) Oocyte apoptosis: like sand through an hourglass. *Developmental Biology* **213**: 1–17.

Nudell D, Castillo M, Turek P and Reijo-Pera R (2000) Increased frequency of mutations in DNA from infertile men with meiotic arrest. *Human Reproduction* **15**: 1289–1294.

Ong YC, Kolatkar PR and Yong EL (2002) Androgen receptor mutations causing human androgen insensitivity syndromes show a key role of residue M807 in helix 8–helix 10 interactions and in receptor ligand-binding domain stability. *Molecular Human Reproduction* **8**: 101–108.

Pellestor F (1991) Frequency and distribution of aneuploidy in human female gametes. *Human Genetics* **86**: 283–288.

Pfeffer J, Pang M-G, Hoegerman SF, *et al*. (1999) Aneuploidy frequencies in semen fractions from ten oligoasthenoteratozoospermic patients donating sperm for intracytoplasmatic sperm injection. *Fertility and Sterility* **72**: 472–478.

Repping S, Skaletsky H, Lange J, *et al*. (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *American Journal of Human Genetics* **71**: 906–922

Rovio AT, Marchington DR, Donat S, *et al*. (2001) Mutations at the mitochondrial DNA polymerase (POLG) locus associated with male infertility. *Nature Genetics* **29**: 261–262.

Ruiz-Pesini E, Lapena A, Diez-Sanchez C, *et al*. (2000) Human mitochondrial DNA haplogroups associated with high or reduced spermatozoa motility. *American Journal of Human Genetics* **67**: 682–696.

Simpson JL and Rajkovic A (1999) Ovarian differentiation and gonadal failure. *American Journal of Medical Genetics* **89**: 186–200.

Sun C, Skaletsky H, Birren B, *et al*. (1999) An azoospermic man with a *de novo* point mutation in the Y-chromosome gene USP9Y. *Nature Genetics* **23**: 429–432.

Van Assche E, Bonduelle M, Tournaye H, *et al*. (1996) Cytogenetics of infertile men. *Human Reproduction* **11**: 1–24.

Vegetti W, Tibiletti MG, Testa G, *et al*. (1998) Inheritance in idiopathic premature ovarian failure: analysis of 71 cases. *Human Reproduction* **13**: 1796–1800.

Vogt PH, Edelmann A, Kirsch S, *et al*. (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics* **5**: 933–943.

World Health Organization (1987) Towards more objectivity in diagnosis and management of male infertility. *International Journal of Andrology* **7**: 1–35.

Further Reading

Cummins J, Jequier AM and Kann R (1994) Molecular biology of human male infertility: links with aging mitochondrial genetics, and oxidative stress? *Molecular Reproductive Development* **37**: 345–362.

Faraut T, Mermet M-A, Demongeot J and Cohen O (2000) Cooperation of selection and meiotic mechanisms in the production of imbalances in reciprocal translocations. *Cytogenetics and Cell Genetics* **88**: 15–21.

Finkelstein S, Mukamel E, Yavetz H, Paz G and Avivi L (1998) Increased rate of nondisjunction in sex cells derived from low-quality semen. *Human Genetics* **102**: 129–137.

Hassold T, Sherman S and Hunt P (2000) Counting cross-overs: characterizing meiotic recombination in mammals. *Human Molecular Genetics* **9**: 2409–2419.

Marozzi A, Manfredini E, Tibiletti MG, *et al*. (2000) Molecular definition of Xq common-deleted region in patients affected by premature ovarian failure. *Human Genetics* **107**: 304–311.

Matias PM, Donner P, Coelho R, *et al*. (2000) Structural evidence for ligand binding specificity in the binding domain of the human androgen receptor: implications for pathogenic gene mutations. *Journal of Biological Chemistry* **275**: 26 164–26 171.

Mumm S, Herrera L, Waeltz PW, *et al*. (2001) X/autosomal translocations in the Xq critical region associated with premature ovarian failure fall within and outside genes. *Genomics* **76**: 30–36

Nature supplement on fertility
http://www.nature.com/fertility

Pieters MH, Geraedts JP, Meyer H, *et al*. (1990) Human gametes and zygotes studied by nonradioactive *in situ* hybridization. *Cytogenetics and Cell Genetics* **53**: 15–19.

Rajpert-De Meyts E, Leffers H, Petersen JH, *et al*. (2002) CAG repeat length in androgen-receptor gene and reproductive variables of fertile and infertile men. *Lancet* **359**: 44–46.

Stenchever MA, Macintyre MN, Jarvis JA and Hempel JM (1969) Cytogenetic studies of 32 infertile couples. *Obstetrics and Gynecology* **33**: 380–382.

Vogt PH (1996) Human Y chromosome function in male germ cell development. *Advances in Developmental Biology* **4**: 191–257.

Vogt PH (1998) Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. *Molecular Human Reproduction* **4**: 739–744.

Vogt PH, Affara N, Davey P, *et al*. (1997) Report on the third international workshop on Y chromosome mapping 1997. *Cytogenetics and Cell Genetics* **79**: 1–20.

Web Links

AR mutations database
http//www.mcgill.ca/androgendb

Cystic Fibrosis Mutation Data Base
http://www.genet.sickkids.on. ca/cftr/

Downs syndrome. Mendelian Inheritance of man Online (OMIM) website: reference number: 190685
http://www3.ncbi.nlm.nih.gov/omim/
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id = 605597

Human Y AZF (Azoospermia Factor) candidate spermatogenesis genes
http://humangenetik.uni-hd.de/ger/humgen/Vogt/AZFcandK.html

Cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7) (*CFTR*); LocusID: 1080. LocusLink:
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 1080

Diaphanous homolog 2 (Drosophila) (DIAPH2). LocusID: 1730. LocusLink:
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 1730

Forkhead transcription factor (FOXL2 on chromosome 3; 3q23) LocusID: 668. LocusLink:
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?I = 668

Heparan sulfate 6-O-sulfotransferase 1 (HS6ST1); LocusID: 9394. LocusLink:
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 9394

Premature ovarian failure, X-linked (POF); LocusID: 5421. LocusLink:
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 5421

X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-bound (*XPNPEP2*). LocusID: 7512. LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 7512
Azoospermia factor 1(*AZF1*); MIM number: 415000. OMIM: http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?415000
Blepharophimosis, ptosis, and epicanthus inversus (BPES); MIM number: 110100. OMIM: http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?110100
Cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7) (*CFTR*); MIM number: 602421. OMIM: http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?602421
Diaphanous homolog 2 (Drosophila) (DIAPH2). MIM number: 300108. OMIM: http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?300108
Heparan sulfate 6-O-sulfotransferase 1 (HS6ST1); MIM number: 604846. OMIM: http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?604846
Infertile male syndrome. MIM number: 308370. OMIM: http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?308370
Premature ovarian failure, X-linked (POF); MIM number: 311360.http OMIM: http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?311360
X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-bound (*XPNPEP2*). MIM number: 300145. OMIM: http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?300145

Information Theories in Molecular Biology and Genomics

Andrzej K Konopka, *BioLingua Research Inc.– Center for Advanced Sequence Analysis, Gaithersburg, Maryland, USA*

Advanced article

Information theory, statistical theory of signal transmission, communication theory and Shannon theory are synonymous names for the mathematical theory first published by Claude Shannon in the 1940s. The concept of entropy in this theory can be viewed as a measure of 'randomness' of sequences of symbols. Shannon entropy and its variants have been widely used in molecular biology and bioinformatics as statistical tools of choice for sequence and structure analyses.

Article contents

Introduction

'Information' is a many-sided metaphorical concept. It loosely signifies pieces of knowledge that are exchanged in acts of communication but there is no universally accepted, single definition (Bar-Hillel and Carnap, 1953; MacKay, 1969) of this term. What is known as information theory, that is the statistical theory of (physical) signal transmission (Shannon theory), originated in the 1920s in the telephone industry (Hartley, 1928; Shannon, 1948) but it later became a more general theory of printed texts (Shannon, 1951) and their cryptanalysis (Shannon, 1949). This latter, text-theoretical, variant of Shannon's work gained noticeable prominence in applications to nucleic acids and protein sequence analysis over the years (Gatlin, 1972; Campbell, 1982; Lipman and Wilbur, 1983; Konopka, 1984, 1985; Kuppers, 1990; Garnier *et al*., 1996). (A great many of these applications focused on deviation of a given probability distribution from a discrete uniform distribution and therefore the apparatus of Shannon theory was just one of many possible choices for calculating an equivalent of variance of a distribution.) Other theories of information, particularly the semantic ones (Bar-Hillel and Carnap, 1953; MacKay, 1969), were potentially interesting for biologists (Pattee, 1969) but their focus on formal logic was too abstract to be useful in biological applications.

The foundations of Shannon theory are carefully designed to avoid a meaningful definition (Tarski, 1944) of 'information' altogether. In a semantic and a pragmatic sense a meaningful definition needs to be formally (logically) correct and materially adequate to the defined thing, process or phenomenon. Shannon's definition is adequate to the mathematical formalism (pure syntax) that has nothing to do with any of the commonsense (i.e. semantic or pragmatic) meanings of the word 'information'. Instead it focuses on a mathematical convention to measure how much a given statistical distribution differs from the so-called