

DATABASES

The Androgen Receptor Gene Mutations Database (ARDB): 2004 Update

Bruce Gottlieb,^{1,4*} Lenore K. Beitel,^{1,2} Jian Hui Wu,¹ and Mark Trifiro^{1,2,3}¹Lady Davis Institute for Medical Research, Sir Mortimer B. Davis-Jewish General Hospital, Montreal, Canada; ²Department of Medicine, McGill University, Montreal, Canada; ³Department of Human Genetics, McGill University, Montreal, Canada; ⁴Department of Biology, John Abbott College, Canada

Communicated by Alastair Brown

The current version of the androgen receptor (AR) gene mutations database is described. The total number of reported mutations has risen from 374 to 605, and the number of AR-interacting proteins described has increased from 23 to 70, both over the past 3 years. A 3D model of the AR ligand-binding domain (AR LBD) has been added to give a better understanding of gene structure–function relationships. In addition, silent mutations have now been reported in both androgen insensitivity syndrome (AIS) and prostate cancer (CaP) cases. The database also now incorporates information on the exon 1 CAG repeat expansion disease, spinobulbar muscular atrophy (SBMA), as well as CAG repeat length variations associated with risk for female breast, uterine endometrial, colorectal, and prostate cancer, as well as for male infertility. The possible implications of somatic mutations, as opposed to germline mutations, in the development of future locus-specific mutation databases (LSDBs) is discussed. The database is available on the Internet (www.mcgill.ca/androgendb/). *Hum Mutat* 23:527–533, 2004. © 2004 Wiley-Liss, Inc.

KEY WORDS: databases; androgen receptor gene; AR; androgen insensitivity syndrome; AIS; prostate cancer; CaP; Kennedy disease; spinobulbar muscular atrophy; SBMA

DATABASES:

AR – 313700, 300068 (AIS), 176807 (CaP), 313200 (SBMA); GDB: 120556; GenBank: NM_000044.2, AH002624. www.mcgill.ca/androgendb (ARDB)

INTRODUCTION

As we enter the 21st century, most genetic databases remain firmly entrenched in the 20th. Their function has been to act as passive repositories for listings of gene alterations that result in proteins whose putative altered structure presumably results in disease or nondisease phenotypes. While these databases have been relatively simple, they have been immensely useful in establishing many valuable links between specific mutations and disease. Meanwhile, the growth of the Internet has allowed this information to be readily accessible to almost all interested parties. In addition, it has allowed for online submission of data with the curator assuming responsibility for validating the data.

The effectiveness of these databases has been predicated on a number of assumptions. First, it is assumed that a particular mutation is indeed responsible for an altered phenotype. However, this clearly needs to be experimentally proven, by expressing the protein and then examining it. In most cases this is not done, based on the assumption that all mutations are likely to have a functional effect on the expressed protein. In addition, even when the gene is expressed, it is usually expressed, at least in the case of human genes, in a nonhuman cell line. Second, it is assumed that the mutation has been

inherited, i.e., is germline, and will therefore be an excellent predictor of the inheritance pattern of a particular condition. However, it is being increasingly observed, particularly in diseases such as cancer, that the mutations are often somatic rather than germline in origin. The third assumption is that there is a clear distinction between mutations and polymorphisms (gene alterations that occur in >1% of a population). However, the Human Genome Project has resulted in the discovery of hundreds of thousands of single nucleotide polymorphisms (SNPs) and other polymorphisms that, in many instances, have been found to be far from benign. Finally, it is often assumed that once a locus-specific gene mutation database (LSDB) has been created, it is routinely updated and maintained.

Received 7 September 2003; accepted revised manuscript 23 January 2004.

Grant sponsor: Canadian Institutes of Health Research.

*Correspondence to: Bruce Gottlieb, Ph.D., Lady Davis Institute for Medical Research, Sir Mortimer B. Davis - Jewish General Hospital, 3755 Cote Ste. Catherine Road, Montreal, Quebec, Canada H3T 2E1. E-mail: bruce.gottlieb@mcgill.ca

DOI 10.1002/humu.20044

Published online in Wiley InterScience (www.interscience.wiley.com).

However, in many cases this does not occur and the LSDBs gradually become obsolete [Stenson et al., 2003].

All of these assumptions have resulted in making most databases, in their present form, less than ideal for determining the significance of an alteration to a gene as the cause of a particular disease or condition. The androgen receptor gene mutation database (ARDB; www.mcgill.ca/androgendb) has confronted many of the same problems. In this article, we report on the developments that have occurred in the ARDB over the past few years, and the changes that we have introduced to make it both more relevant and more useful.

DATABASE INFORMATION

Loss-of-Function Mutations

Androgen insensitivity syndrome. The androgen receptor (AR; MIM# 313700) is a member of the superfamily of nuclear receptors that function as ligand-dependent transcription factors. Intracellular AR is essential for androgen action, whether of testosterone (T) or of its 5α -reduced derivative, 5α -dihydrotestosterone (DHT). Hence, the AR is essential for normal primary male sexual development before birth (masculinization), and for normal secondary male sexual development around puberty (virilization). AR dysfunctions in XY individuals result in androgen insensitivity syndromes (AIS; MIM# 300068).

The present version of the ARDB is based on GenBank reference sequence NM_000044.2, and contains over 500 entries of mutations causing androgen insensitivity syndrome (AIS), representing over 300 different AR gene (AR) mutations, from more than 600 patients with AIS. There has been a large increase in the number of reported AR mutations since the last published report on the database [Gottlieb et al., 1999]; the number of entries almost doubled, from 374 to 605. This increase has been partly attributed to the ease of sequencing the AR, but the increase might easily have been even larger; this is due to the fact many mutations are not reported in the literature unless they are unique. Further, even unique mutations may not be reported, as they may have been found as a result of routine blood tests that are administered solely to establish a diagnosis of AIS. Currently, it is still the policy of the curator not to accept submissions to the ARDB unless they have been accepted for publication, in order to ensure adequate quality control of the data.

There is still an unequal distribution of the mutations along the length of the AR, as shown in Figure 1. It has previously been suggested that these mutation-dense regions are hot spots that reflect the high density of mutable CpG sites in the region [Gottlieb et al., 1996]. It is also apparent that the types of mutations differ along the length of the AR. In particular, nearly all mutations in exon 1 (Fig. 1) cause complete AIS (CAIS), and nearly all are of the premature translation termination variety (Table 1), whether by direct mutation to a stop

codon, or indirectly, by frameshifts after small deletions or insertions. To date, only 54 mutations have been reported in exon 1 of the AR in patients suffering from some form of AIS, despite the fact that it encodes more than half of the AR protein [Gottlieb et al., 1999], and even fewer have been reported in splicing and untranslated regions of the AR gene (Table 1). In the C-terminal ligand binding domain (LBD), there is a striking preponderance of missense mutations, with a significantly greater number of CAIS than partial AIS (PAIS) cases (Table 1).

The appearance of the database has been modified, and it is now presented in a portable document format (pdf; Adobe Systems, www.adobe.com) for ease of use. We have expanded the database by providing information on proven AR-interacting proteins; and a rough map of the region of the AR with which they interact (Fig. 2) is also now available on the database website, together with a table that lists details of the properties of each of these proteins. In the past few years, there has been a tremendous expansion in the number of these proteins, the number rising from 23 to 70 in just 3 years.

Of particular interest is the appearance of 12 silent mutations (nine in AIS, three in prostate cancer) in regions of exons not close to splice sites, which raised the possibility that mutations can cause problems at the mRNA level [Gottlieb et al., 1999]. Further, an increasing number of mutation entries (128 out of 605) now contain data proving the pathogenicity of the putative mutation by reconstituting the mutation and seeing its effect on a reporter gene, which greatly improves the quality of the data.

Gain-of-Function Mutations

Prostate cancer. To date, 85 AR mutations have been found in prostate cancer (CaP) tissue (MIM# 176807), almost all being single-base substitutions due to somatic mutations, rather than germline mutations. These are now indicated in the database by being color-coded orange. As can be seen in Figure 1 and Table 1, the majority of somatic mutations occur in the LBD ($\approx 45\%$), with a substantial number occurring in exon 1 ($\approx 30\%$). Originally, it was thought that AR was not expressed in CaP tissues, but this does not appear to be the case [Edwards et al., 2003]. Considerable controversy has revolved around conflicting studies that only sometimes have found a significant number of AR mutations in CaP tissues [Culig et al., 2002]. It has been argued that AR mutations only appear during the latter stages of CaP and, in addition, some studies have indicated that antiandrogen treatments have resulted in AR mutations [Hyytinen et al., 2002]. This data is now incorporated into the database.

A considerable limiting factor in the value of this data is that, with a few exceptions, experiments to prove pathogenicity have not been done for CaP mutations, which considerably reduces the value of the data. In a recent functional analysis of disease-associated mutations of the AR, we have shown that most of

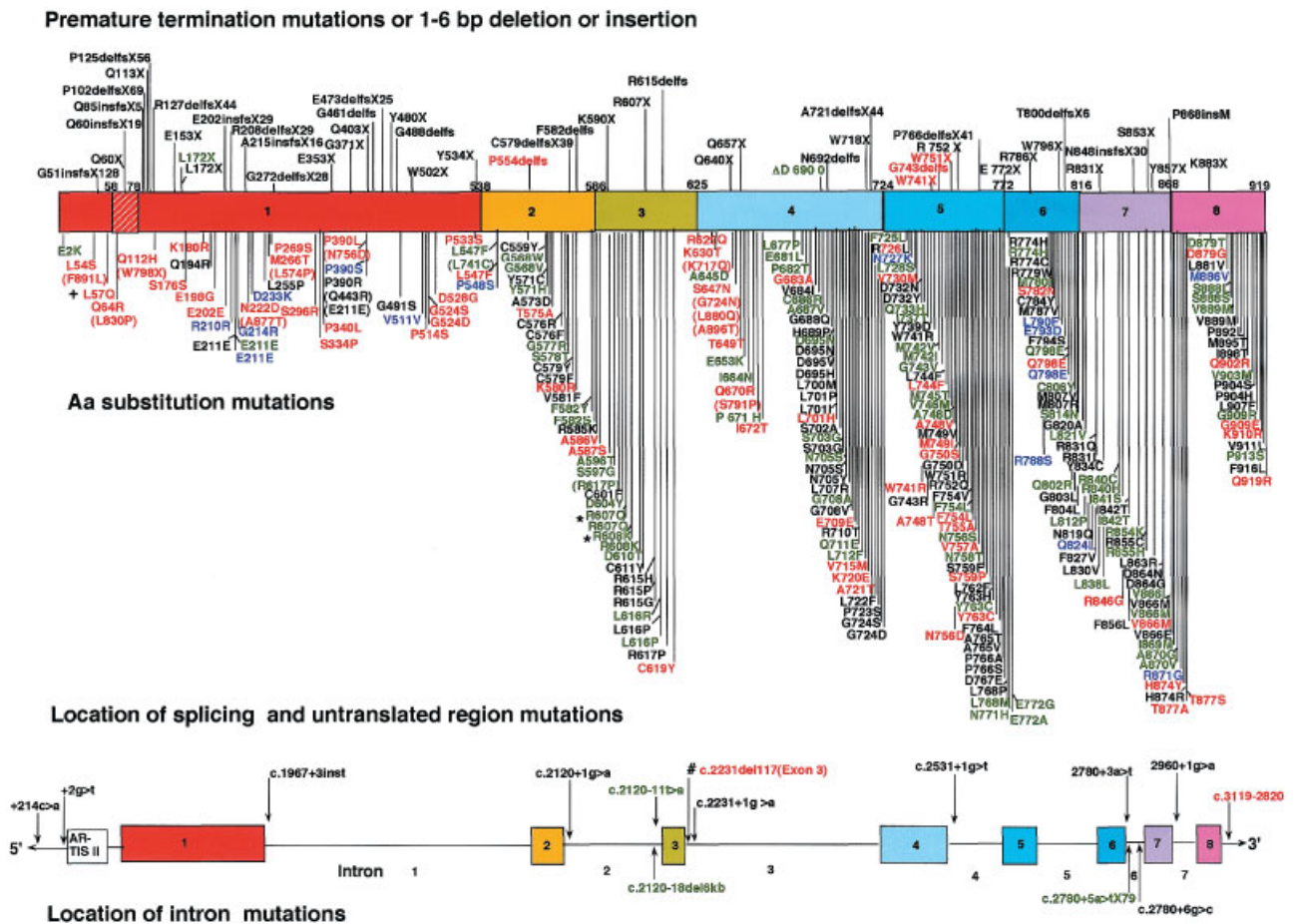


FIGURE 1. Structure of the AR indicating the location of all exon and intron mutations causing disease as of July 30, 2003. del, 1–6 bp deleted; ins, 1–6 bp inserted. X, a termination codon at the site of the mutation or at the frameshifted (fs) codon, which is identified by the number that follows. *, mutations found in male breast cancer tissue; #, mutations found in female breast cancer tissue; +, mutations found in laryngeal cancer tissue. Mutations in black type: CAIS-germline. Mutations in green: PAIS-germline. Mutations in blue: MAIS-germline. Mutations underlined cause both CAIS and PAIS. Mutations in red: CaP-somatic. When more than one mutation is present in a patient, additional mutations are in brackets. Exon mutations are protein-based, intron and untranslated regions are cDNA-based. GenBank reference sequence NM_000044.2, version AH002624.

the cancer-associated mutations reported in the database have a significantly lower degree of base conservation when compared to the few cases where the pathogenicity of the mutation has been proven [Mooney et al., 2003].

Kennedy disease. Kennedy disease, or spinobulbar muscular atrophy (SBMA; MIM# 313200), a spinobulbar motor neuronopathy associated with mild AIS (MAIS), is one of the classic trinucleotide repeat expansion diseases that cause inherited neurodegenerative disorders. It is caused by expansion of the glutamine-coding (CAG)_{8–35} CAA tract in exon 1 of the AR to a total number of at least 38 trinucleotide repeats [Pinsky et al., 2001]. The MAIS component of SBMA may reflect a loss of AR transcriptional regulatory activity by virtue of a pathologically-expanded polyglutamine (polyGln) tract. It should be noted that in SBMA, the AI phenotype is quite variable. Since subjects with CAIS, including those with complete AR deletions, do not develop SBMA, this knowledge mandated the logic that the polyCAG-expanded AR or the polyGln-expanded AR protein is somehow motor neuronotoxic by a gain, not a loss, of function.

The biochemical, histopathologic, and neurophysiologic features of SBMA are, unremarkably, those secondary to motor denervation. A number of possible causes for this gain of function are now listed on the database website (Table 2).

AR Protein Structure–Function Relationships

In an effort to better understand the structure–function relationship of how specific mutations in the LBD cause AIS, the 3D structure of the AR LBD has been added to the database. We have produced this model with reference to X-ray crystallography data [Matias et al., 2000]. This is particularly important in the case of mutations that alter amino acids, but clearly less so for mutations that cause premature terminations. The model revealed a structure that consisted of 12 α -helices (Fig. 3). To understand the possible affect of each specific gene alteration, we have labeled our model to show the exact position of all 12 α -helices (Fig. 3). Thus any mutation can be physically placed in the actual 3D protein structure. It can be seen that the putative crystal

TABLE 1. Nature and Distribution of Unique AR Mutations That Cause Disease

Loss of function disease	Type of mutation	N-terminal domain ^a	DNA-binding domain ^b	Hinge region ^c	Ligand-binding domain ^d	Splice site	Intron
CAIS	Single base substitution	13	17	1	100	7	1
	Complete gene deletion	3					
	Partial gene deletion	6	4		3		
	Deletion (1-6 bases)	8	3				
	Insertion	5				1	
	Duplication	2					
PAIS	Single base substitution	4	18	1	60	2	
	Multiple base substitution		1		1		
	Deletion (1-6 bases)				1	1	
MAIS	Single base substitution	7			10		
	Partial gene deletion				1?		
	Deletion (1-6 bases)				1		
Gain of function disease							UTR
Prostate cancer	Single base substitution	25	6	2	40		2
	Deletion (1-6 bases)	1			1		
	Insertion						1
Male breast cancer	Single base substitution		2				
Larynx cancer ^e	Deletion (30 bases)	1					
Female breast cancer ^e	Partial gene deletion					1	
Total in each region		76	51	4	218	13	3
Total all mutations	382						

^aaa 1-534; ^baa 559-624; ^caa between DBD and LBD 625-663; ^daa 664-919; ^eSomatic mutation.

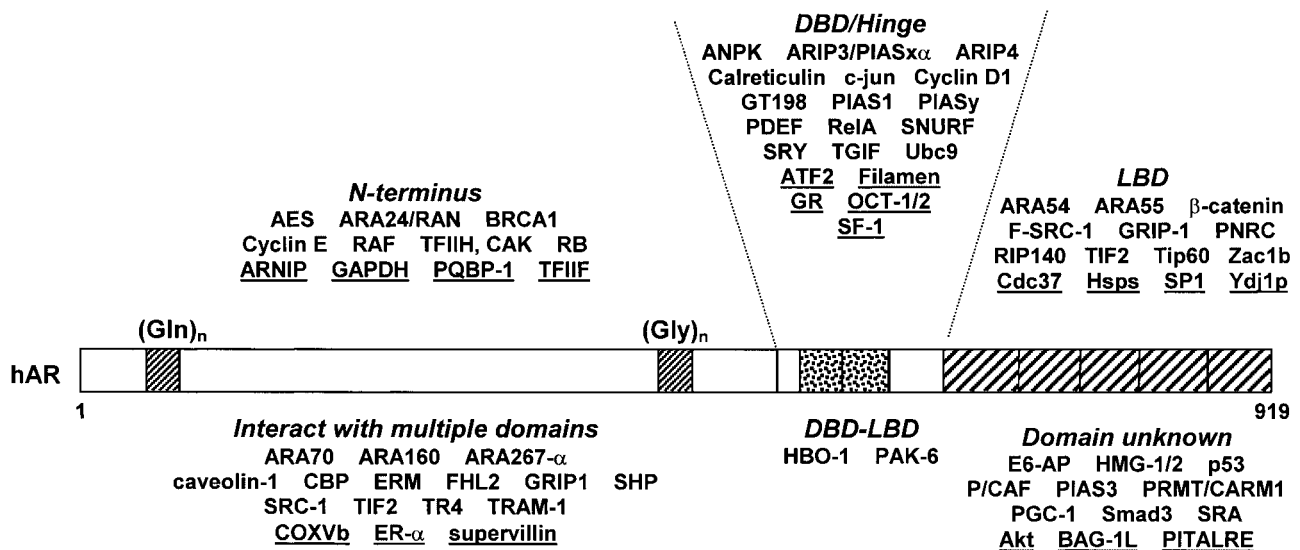


FIGURE 2. Coregulatory proteins that interact with different portions of the AR. A depiction of some of the coregulatory proteins that interact with different portions of the AR and positively affect transcription (coactivators) and negatively affect transcription (corepressors; underlined). The finely hatched rectangles in the left half of the AR represent the polyglutamine (Gln)_n and polyglycine tracts (Gly)_n in its N-terminal domain. The portion of the AR devoted to DNA-binding (DBD) is stippled; the portion devoted to ligand-binding (LBD) is coarsely hatched.

structure shows very few residues that are in close contact with the ligand-binding pocket when modeled with the synthetic ligand R1881 (Fig. 3). The availability of this model allows researchers to realize that the position of the mutant residue near the ligand-binding pocket is not necessary to produce severe ligand binding problems.

Therefore, to try to elucidate how such mutations could potentially affect the structure of the ligand-binding pocket, and so explain the lack of ligand binding in these mutants, we have recently used molecular dynamic modeling techniques over extended periods of

time (up to 4 nsec) to, in effect, create 4D structures of AR mutants [Wu et al., 2003]. In this study, a CAIS mutation some distance from the ligand-binding pocket produced a local structural distortion that also affected the ligand-binding pocket conformation. It is our intention to incorporate this type of information into the database in the near future. We believe that this will be particularly important where the pathogenicity of the mutation has not been proven.

Another aspect of the structure–function relationship is variable phenotypic expression, in which identical

TABLE 2. Diseases Associated With AR CAG Tract Length Variation

Direct association	CAG tract	Androgen sensitivity	Gain of function—possible causes	Symptoms	Reference
SBMA	≥ 38	Reduced	1. Misfolding 2. Truncation 3. Aggregation 4. Sequestration of AR protein/transcription factors 5. Proteasome inhibition 6. Mitochondrial dysfunction	1. Adult onset motor neuropathy of proximal muscles of hip and shoulder 2. Hypogonadism results in gynecomastia and testicular atrophy	Pinsky et al. [2001]
Indirect association	Relative length of tract ^a		Associated risk factors	Comments	Reference
Prostate cancer	Shorter	Increased	1. Ethnicity 2. Family history	Inconclusive studies—possible somatic alterations	Ferro et al. [2002]
Male infertility	Longer	Reduced	Ethnicity	Inconclusive studies—possible somatic alterations	Casella et al. [2003]
Female breast cancer	Longer	Reduced	BRCA1 mutation carriers		Elhaji et al. [2001]
Endometrial cancer	Longer	Reduced		Somatic alterations Selective growth advantage—somatic alterations	Sasaki et al. [2000]
Colon cancer	Shorter	Increased			Ferro et al. [2002]
Esophageal cancer	Shorter	Increased	Ethnicity	Inconclusive studies—except in African males	Dietzsch et al. [2003]

Relative length of CAG tract compared with control populations.

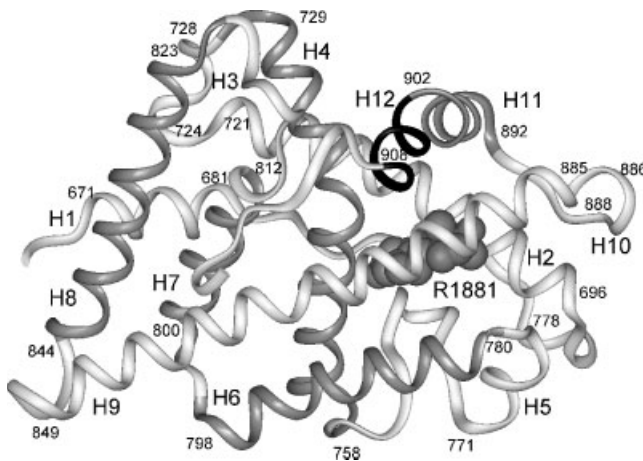


FIGURE 3. The AR ligand-binding domain as defined by the X-ray crystal structure (1e3g). Note that the α helices are numbered and the codon number at the start and finish of each helix is given. Spheres illustrate the bound ligand, R1881.

mutations produce different phenotypes. The mutations showing some degree of variable expressivity now numbers 28; these are indicated in the database by being shown in green. Further, it is highly probable that many more cases of variable expressivity exist but, due to the limited phenotype descriptions available, they are not always apparent. A number of possible causes of variable phenotype expression including somatic mosaicism have been discussed previously [Gottlieb et al., 2001].

AR CAG Tract Length Variation as a Risk Factor for Disease

The database now lists the length of the CAG (polyglutamine) and GGC (polyglycine) tracts that are

present in exon 1. At the present time, there are 58 entries in the database that show the lengths of these two tracts. The mean value of the CAG tract length is 22.34, which is significantly longer than in controls [Elhaji et al., 2001], though the number of database entries is too small to draw any specific conclusions. However, as more data becomes available, it will be interesting to see if increases in CAG repeat length, which reduce the efficacy of the AR [Mhatre et al., 1993], play any role in determining the AIS or CaP phenotype.

In the past few years, a number of studies have examined possible relationships between the length of the CAG repeat and the risk of getting certain diseases and conditions. These include: female breast cancer (MIM# 114480) [Elhaji et al., 2001]; male infertility [Casella et al., 2003]; prostate cancer, reviewed by Ferro et al. [2002]; uterine endometrial cancer (MIM# 608089) [Sasaki et al., 2000]; colorectal cancer [Ferro et al., 2002]; and esophageal cancer (MIM# 133239) [Dietzsch et al., 2003]. Table 2 has been added to the database that lists details of all of these diseases.

DISCUSSION

Germline Vs. Somatic Mutations

A most significant issue facing LSBD curators and designers in the future is likely to be the calculating the relative genetic significance of somatic mutations, as opposed to germline mutations, particularly in so called "late-onset" diseases. At the present time, the only databases that have found it necessary to distinguish between somatic and germline mutations are those of certain cancer-associated genes, such as adenomatous polyposis coli (APC) [Laurent-Puig et al., 1998]. The

significance of these observations is primarily limited to whether or not a particular mutation is inherited.

In discussing the significance of somatic mutations in LSDBs, one possible consideration is that their effect on phenotype expression will be different from germline mutations. As reported in the ARDB, the somatic mutations found in CaP tissues seem to elicit a gain of function, as opposed to a loss of function. This results in individuals that have a normal male phenotype, except that they suffer from CaP. Thus, somatic mutations produce a degree of genetic heterogeneity within an individual organism. In fact, in the case of CaP tumors, which (like most tumors) consist of heterogeneous tissue, there is often a wide range of cell types within each tumor, ranging from normal to advanced cancerous, possibly indicating a concomitant genetic heterogeneity within the tumors. Thus mutation databases will have to reflect such possible genetic variation, by much more closely identifying the tissue phenotype associated with a specific gene alteration, whether mutation or polymorphism. Fortunately, the arrival of techniques such as laser capture microdissection (LCM) have allowed us to examine genes in very specific cell and tissue types. Recently we have used LCM in a study into variation in AR CAG repeat length in prostate cancer tissue [Gottlieb et al., 2002], and it is our intention to eventually expand the database to include specific tissue phenotypes when listing patients with somatic mutations.

In examining the relationship between somatic mutations and disease, it is not insignificant that CaP is considered a late onset disease. This might be expected, as studies have shown that the rate of somatic mutations may increase with age [Jackson and Loeb, 1998], possibly due to genomic instability. This suggests that by incorporating data into the ARDB about changes in AR CAG repeat length, which is known to be quite unstable over time [Zhang et al., 1994], it might be possible to provide important insights into how the disease progresses. Further, age considerations could also be significant in other late-onset diseases, including other cancers. Indeed, in a recent study, we have examined the possible association between CAG repeat length changes and the onset and progression of female breast cancer [Elhaji et al., 2001].

Recently, polymorphisms in DNA repair genes have been identified as a risk factor for a number of cancers [Goode et al., 2002]. What is particularly interesting for the ARDB, is that recent studies of DNA mismatch repair enzymes (MMR) in CaP tissues showed a decrease in the activity of these enzymes, with low expression of some of the MMR proteins [Yeh et al., 2001; Chen et al., 2001]. Incorporating information into the ARDB on MMR activity in CaP tissues that have AR gene alterations, could possibly lead to a further understanding into the genetic events that may lead to the initiation and progression of CaP. Clearly, the more information of this type that can be accumulated in LSDBs, the more likely it will be that clearer associations can be identified between the genetic events that lead to disease phenotypes such as cancers. This is particularly true in

the age of genomics, in which mass screenings entail looking for many different genes associated with a particular condition. Perhaps it makes more sense to look for associations that might exist between genes that have already been identified with a particular condition, rather than looking at all possible random associations. Thus, in the case of genes having LSDBs, the genetic and phenotypic information contained within these databases would make such genes natural association partners for any putative genes that have been found by genomic screening to be associated with a similar disease phenotype.

CONCLUSION

At the present time, most LSDBs have been created with the intention of listing gene mutations and their phenotypic expression, in order to correlate specific phenotypes with specific genotypes. The emergence of somatic mutations as genetically significant events indicates that it may be necessary to adapt LSDBs to incorporate additional data on somatic mutations, genetic heterogeneity, and variable expressivity. It is becoming clear that the role of LSDB curators is becoming more, rather than less important, particularly as it relates to using LSDBs as tools to help in understanding specific structure–function relationships. In a previous discussion on the role of LSDB curators [Gottlieb et al., 1999], we stated that it was time for LSDB curators to seize the initiative and start to determine the nature of the data that needs to be part of their LSDBs. If curators set the criteria, it seems reasonable to believe that researchers will ultimately follow them. It has often been said that in any database, the quality of the data that you get out of it is only as good as the quality of the data that you put into it. Up until now, the quality control most curators have undertaken has been aimed at ensuring the accuracy of the data they have entered. However, perhaps it is now time for curators to take a more proactive role, particularly in light of the likelihood that, in the future, mutation databases are going to be playing a much more significant role in both biological and medical research.

ACKNOWLEDGMENT

This work is funded by grants to B.G. from the Canadian Institutes for Health Research.

REFERENCES

- Casella R, Madura MR, Misfud A, Lipshultz LI, Yong EL, Lamb DJ. 2003. Androgen receptor gene polyglutamine length is associated with testicular histology in infertile patients. *J Urol* 169:224–227.
- Chen Y, Wang J, Fraig MM, Metcalf J, Turner WR, Bissada NK, Watson DK, Schweinfest CW. 2001. Defects of DNA mismatch repair in human prostate cancer. *Cancer Res* 61:4112–4121.
- Culig Z, Klocker H, Bartsch G, Hobisch A. 2002. Androgen receptors in prostate cancer. *Endocr Relat Cancer* 9:155–170.

- Dietzsch E, Laubscher R, Parker MI. 2003. Esophageal cancer risk in relation to GGC and CAG trinucleotide repeat lengths in the androgen receptor gene. *Int J Cancer* 107:38–45.
- Edwards J, Krishma NS, Grigor KM, Bartlett, JMS. 2003. Androgen receptor gene amplification and protein expression in hormone refractory prostate cancer. *Br J Cancer* 89:552–556.
- Elhaji YA, Gottlieb B, Lumbroso R, Beitel LK, Lumbroso R, Foulkes WD, Pinsky L, Trifiro MA. 2001. The polymorphic CAG repeat of the androgen receptor gene: a potential role in breast cancer in woman over 40. *Breast Cancer Res Treat* 70:109–116.
- Ferro P, Catalano MG, Dell'Eva R, Fortunati N, Pfeffer U. 2002. The androgen receptor CAG repeat: a modifier of carcinogenesis? *Mol Cell Endocrinol* 193:109–120.
- Goode EL, Ulrich CM, Potter JD. 2002. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev* 11:1513–1530.
- Gottlieb B, Trifiro M, Lumbroso R, Vasiliou DM, Pinsky L. 1996. The androgen receptor gene mutations database. *Nucleic Acids Res* 24:151–154.
- Gottlieb B, Beitel LK, Lumbroso R, Pinsky L, Trifiro M. 1999. Update of the androgen receptor gene mutations database. *Hum Mutat* 14:103–114.
- Gottlieb B, Beitel LK, Trifiro M. 2001. Somatic mosaicism and variable expressivity. *Trends Genet* 17:79–82.
- Gottlieb B, Lumbroso R, Gillespie J, Beitel LK, Trifiro M. 2002. A determination of the relationship between the length of the CAG tract within the androgen receptor gene and prostate cancer. In: Program of the 84th Meeting of the U.S. Endocrine Society, San Francisco. Abstract P2-222.
- Hyytinen E-J, Haapla K, Thompson J, Lappalainen I, Roiha M, Rantala I, Helin HJ, Janne OA, Vihinen M, Plavimo JJ, Koivisto PA. 2002. Pattern of somatic androgen receptor gene mutations in patients with hormone-refractory prostate cancer. *Lab Invest* 82:1591–1592.
- Jackson AL, Loeb LA. 1998. The mutation rate and cancer. *Genetics* 148:1483–1490.
- Laurent-Puig P, Beroud C, Soussi T. 1998. APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res* 26:269–270.
- Matias PM, Donner P, Coelho R, Thomax M, Peixoto C, Macedo S, Otto N, Joschko S, Scholtz P, Wegg A. 2000. Structural evidence for ligand specificity in the binding domain of the human androgen receptor. *J Biol Chem* 275:26264–26171.
- Mhatre A, Trifiro MA, Kaufman M, Kazemi EP, Figlewicz D, Rouleau G, Pinsky L. 1993. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nat Genet* 5:184–188.
- Mooney SD, Klein TE, Altman RB, Trifiro MA, Gottlieb B. 2003. A functional analysis of disease associated mutations in the androgen receptor gene. *Nucleic Acids Res* 31:e42.
- Pinsky L, Beitel LK, Trifiro MA. 2001. Spinobulbar muscular atrophy. In: Scriver CR, Beaudet AL, Valle D, Sly WS, editors. *Metabolic and molecular basis of inherited disease*. 8th ed. New York: McGraw-Hill. p 4147–4157.
- Sasaki M, Dahiya R, Fujimoto S, Ishikawa M, Oshimura M. 2000. The expansion of the CAG repeat in exon 1 of the human androgen receptor gene is associated with uterine endometrial carcinoma. *Mol Carcinog* 27:237–244.
- Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeyasinghe S, Krawczak M, Cooper DN. 2003. Human gene mutation database (HGMD): 2003 update. *Hum Mutat* 21:577–581.
- Wu J, Gottlieb B, Batist G, Sulea T, Purisima EO, Beitel LK, Trifiro M. 2003. Bridging structural biology and genetics by computational methods: an investigation into how the R774C mutation in the AR gene can result in complete androgen insensitivity syndrome. *Hum Mutat* 22:465–475.
- Yeh C-C, Lee C, Dahija R. 2001. DNA mismatch repair enzyme activity and gene expression in prostate cancer. *Biochem Biophys Res Commun* 285:409–413.
- Zhang L, Leeflang EP, Yu J, Arnheim N. 1994. Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the androgen receptor gene. *Nat Genet* 7:531–535.