A Random Approach to the Determination of Amino Acid Pairs in Von Hippel-Lindau Disease Tumor Suppressor (G7 Protein)

G. Wu, MD, PhD
S-M. Yan, MD, PhD

DreamSciTech Consulting Co. Ltd., Shenzhen, Guangdong Province, China

KEY WORDS: G7 protein, probability, randomness, Von Hippel-Lindau disease

ABSTRACT

Objective: We used a random approach to analyze amino acid pairs in human Von Hippel-Lindau disease tumor suppressor (G7 protein) in order to determine which amino acid pairs are more sensitive to 109 mutations from human G7 protein. The rationale for this study is based on our hypothesis and findings that a harmful mutation is more likely to occur at randomly unpredictable amino acid pairs and a harmless mutation is more likely to occur at randomly predictable amino acid pairs. We argue that the randomly predictable amino acid pairs should not be deliberately evolved, whereas the randomly unpredictable amino acid pairs should be deliberately evolved with connection of protein function.

Design: The amino acid sequence of the human G7 protein and its 109 mutations with point mutant were obtained from the Swiss-Protein data bank. There are a total of 212 amino acid pairs in human G7 protein—less than the 400 that are theoretically possible. We calculated the randomly predicted frequency and actual frequency in human G7 protein, the randomly predictable present amino acid pairs, randomly unpredictable present amino acid pairs, randomly predictable absent amino acid pairs, mutations in randomly predictable and unpredictable amino acid pairs, and the difference between actual and randomly predicted frequencies. In this way, we compared the frequency difference in the amino acid pairs affected by mutations.

Results: We found that 94.5% of 109 mutations occur at randomly unpredictable amino acid pairs, which account for 78.3% of amino acid pairs in human G7 protein.

Conclusion: The randomly unpredictable amino acid pairs are more sensitive to mutations in human G7 protein. The results also suggest that the human G7 protein has a natural tendency to mutate.

INTRODUCTION

Von Hippel-Lindau disease is a tumor syndrome in multiple organs. Retinal capillary hemangioma is the most frequent and often the earliest manifestation of Von Hippel-Lindau disease.[1] The original descriptions of retinal capillary hemangioma appeared more than 125 years ago. Since then a tremendous amount of new information has become available.[2] Typical extraocular lesions associated with Von Hippel-Lindau disease are central nervous system heman-

Table 1. Occurrences of Mutations with Respect to Randomly Predictable and Unpredictable Amino Acid Pairs in Human G7 Protein

G7 protein	Kinds		Pairs		Mutations		Ratio	
	Number	%	Number	%	Number	%	Mutations/Kinds	Mutations/Pairs
Predictable	38	27.54	46	21.70	6	5.50	6/38 ≠ 0.16	6/46 ≠ 0.13
Unpredictable	100	72.46	166	78.30	103	94.50	103/100 ≠ 1.03	103/166 ≠ 0.62
Total	138	100.00	212	100.00	109	100.00	109/138 ≠ 0.79	109/212 ≠ 0.51

gioma, renal cell carcinoma, pheochromocytoma, pancreatic islet cell tumors, endolymphatic sac tumor of the inner ear, and cysts and cystadenoma in the kidney, pancreas, epididymis, and broad ligament.[3] Current estimates of the prevalence of Von Hippel-Lindau disease range between two and three per 100,000 persons.[4-6] The incidence of Von Hippel-Lindau disease is approximately 1 in 40,000 live births.

In 1929 Moller was the first to suggest that Von Hippel-Lindau disease has an autosomal dominant pattern of inheritance.[7] About 60 years later, a genetic locus for the disease was mapped to the short arm of chromosome 3 by linkage studies in 9 Von Hippel-Lindau families.[8] The Von Hippel-Lindau tumor suppressor gene was identified in 1993;[9] subsequently, germline mutations in the Von Hippel-Lindau gene confirmed the molecular genetic basis of familial inheritance of the disease.[10] Many different intragenic Von Hippel-Lindau germline mutations have been detected and they are scattered over the Von Hippel-Lindau gene. Missense mutations (leading to an amino acid substitution in G7 protein) are found in 40% of the families with an identified Von Hippel-Lindau germline mutation,[11] and in 96% of the mutations in Von Hippel-Lindau disease with pheochromocytoma.[12]

The Von Hippel-Lindau protein (G7 protein) binds with other proteins to form a complex, which targets the hypoxia-inducible factors for degradation.[13] In the absence of G7 protein, the hypoxia-inducible factors are not degraded, with resultant excessive production of vascular endothelial growth factor. Inappropriate expression of vascular endothelial growth factor causes blood vessels to proliferate and form hemangioma.[14]

The Von Hippel-Lindau gene is a tumor suppressor gene, according to Knudson's "two-hit" hypothesis: inactivation of both copies of the Von Hippel-Lindau gene is required for a normal cell to develop into a tumor cell,[15] a model that is supported by recent clinical,[16] statistical,[17] and genetic[18,19] studies. However, there are no reports dealing with questions of why so many mutations occur in the G7 protein and which amino acids are more sensitive to mutations; that is, it is still difficult to draw a general rule of which amino acid subsequences are more sensitive to mutations and which amino acid subsequences are less sensitive to mutations. If such a general rule can be drawn, then we can get not only more insight into the relationship between G7 protein and Von Hippel-Lindau disease, but, more important, we can pay greater attention to these sensitive subsequences in order to prevent them from mutating. Moreover, we can even predict the possible subsequences sensitive to the currently unknown mutations.

This problem can be assessed from different approaches, including empiric (regression analysis), experimental (artificial and natural mutations), and computational (multiple sequence comparisons and alignments). Two explanations are currently proposed to explain why some amino acids mutate more frequently than others. The first is targeted mutagenesis, which defines the "hotspot" sites sensitive to endogenous and

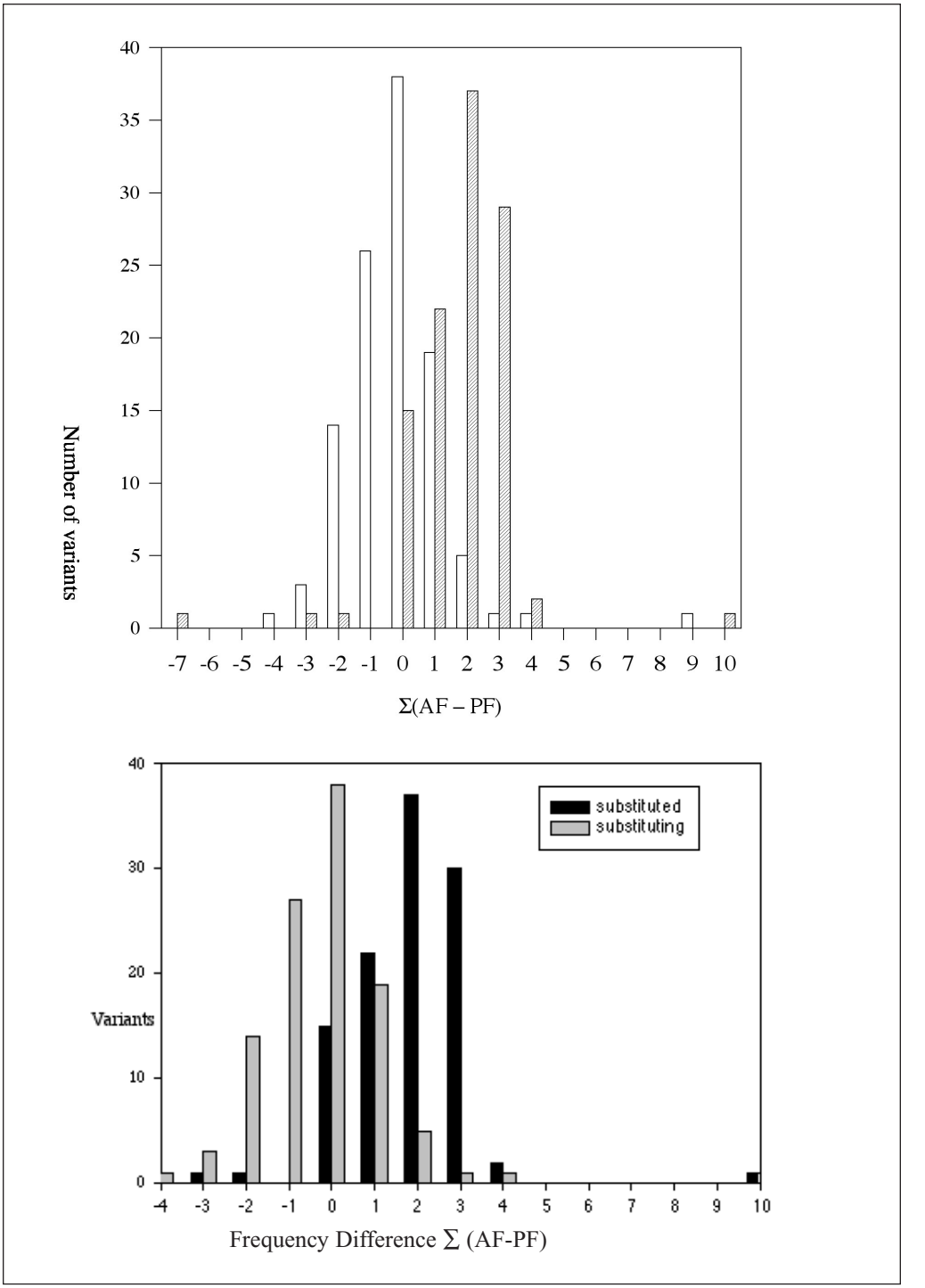

Figure 1. Frequency difference between substituted and substituting amino acid pairs induced by mutations.

Table 2. Classification of substituted amino acid pairs induced by mutations in human G7 protein.

	Pair I	Pair II	Mutations		Total
			Number	%	%
Predictable	AF = PF	AF = PF	6	5.50	5.50
Unpredictable	AF > PF	AF > PF	55	50.46	94.50
	AF > PF	AF = PF	34	31.19	
	AF > PF	AF < PF	12	11.01	
	AF < PF	AF = PF	0	0	
	AF < PF	AF < PF	2	1.83	

AF: actual frequency; PF: predicted frequency.

exogenous mutagens.[20-22] The second explanation is function selection, which posits that the disruption of protein functions may depend upon the position of the mutation in the protein.[23-25] Neither of these explanations answer why some amino acid sequences are sensitive to mutations and some are not. We believe that a probability approach can contribute to the understanding of this problem because in the past we used similar approaches to analyze the primary structure of different proteins to cast light on protein constructions and disease related to them.

In general, our approach can predict which amino acid subsequences are present or absent in a protein's primary structure. We argue that the randomly predictable present and absent subsequences should not be deliberately evolved, whereas the randomly unpredictable present and absent subsequences should be deliberately evolved. Accordingly, our first approach can classify the present amino acid subsequences as randomly predictable and unpredictable subsequences. We suggest that the randomly unpredictable amino acid subsequences are more related to protein function, and the mutations in these subsequences may lead to protein dysfunction. In a recent study,[26] we found that a mutation, which leads to the dysfunction of rat monoamine oxidase B, is located in a randomly unpredictable amino-acid pair. In contrast, another mutation, which does not affect rat monoamine oxidase B function, is located in randomly predictable amino acid pairs.[26]

In this study, we attempt to use a random approach to analyze amino acid pairs in human G7 protein, with its 109 mutations, in order to determine which amino acid pairs are more sensitive to the mutations.

MATERIALS AND METHODS

The amino acid sequence of the human G7 protein and its 109 mutations with point mutations were obtained from the Swiss-Protein data bank (access number P40337).[27] The detailed calculations and its rationales have already been published in a number of previous studies.[28] Briefly, the calculation procedure with examples is as follows.

Amino acid pairs in human G7 protein

The human G7 protein is composed of 213 amino acids. We count the first and second amino acids as an amino acid pair, the second and third as another amino acid pair, the third and fourth, and so on, until the 212th and 213th. Thus there is a total of 212 amino acid pairs. As there are 20 types of amino

Table 3. Classification of substituting amino acid pairs induced by mutations in human G7 protein.

Pair I	Pair II	Mutations		Total
		Number	%	%
AF = 0, PF > 0	AF = 0, PF > 0	6†	5.50	77.99
AF = 0, PF > 0	AF = PF = 0	11†	10.09	
AF = 0, PF > 0	AF = PF > 0	11†	10.09	
AF = 0, PF > 0	AF < PF, AF ≠ 0	5†	4.59	
AF = 0, PF > 0	AF > PF	7†	6.42	
AF = PF = 0	AF = PF = 0	18	16.51	
AF = PF = 0	AF = PF > 0	7	6.42	
AF = PF = 0	AF < PF, AF ≠ 0	3†	2.75	
AF = PF = 0	AF > PF	17	15.60	
AF < PF, AF ≠ 0	AF < PF, AF ≠ 0	6†	5.50	22.01
AF < PF, AF ≠ 0	AF = PF > 0	3†	2.75	
AF < PF, AF ≠ 0	AF > PF	5†	4.59	
AF = PF > 0	AF = PF > 0	4	3.67	
AF > PF	AF > PF	2	1.83	
AF = PF > 0	AF > PF	4	3.67	

†Indicates the mutations that target one or both substituting amino acid pairs with their actual frequency smaller than predicted (52.28%).

acids, any amino acid pair can be composed of any of 20 types of amino acids; therefore there are 400 (20^2) kinds of amino acid pairs. Again, there are 212 amino acid pairs in human G7 protein—less than the 400 theoretically possible amino acid pairs. Clearly, some of the 400 potential amino acid pairs are absent from human G7 protein.

Randomly predicted frequency vs actual frequency

The predicted frequency of amino acid pairs is calculated according to the simple permutation principle.[29] For example, there are 20 arginines (R) and 30 glutamic acids (E) in human G7 protein. The predicted frequency of amino acid pair "ER" would be 3 (calculation: $30/213 \times 20/212 \times 212 = 2.817$). In actuality we find 3 "ER"s in human G7 protein, so the actual frequency of "ER" is 3. Hence we have 3 relationships between actual and predicted frequencies: the actual frequency can be smaller than, equal to, or larger than the predicted frequency, respectively.

Randomly predictable present amino acid pairs

As described in the last section, the frequency of a randomly present amino acid pair "ER" would be 3 and "ER" really appears 3 times in human G7 protein, so the presence of "ER" is randomly predictable.

Randomly unpredictable present amino acid pairs

The predicted frequency of a randomly present amino acid pair "EE" would be 4 (calculation: $30/213 \times 29/212 \times 212 = 4.085$), so that there would be 4 "EE"s in human G7 protein. But in actuality "EE" appears 9 times in human G7 protein, so the presence of "EE" is randomly unpredictable. This is also the case that the actual frequency of "EE" is larger than the predicted frequency of "EE." Another case is that the actual frequency is smaller than the predicted frequency. For example, there are 20 leucines (L) in human G7 protein, so the predicted frequency of "EL" is 3 (calculation: $30/213 \times 20/212 \times 212 = 2.817$), but the actual frequency is 2.

Randomly predictable absent amino acid pairs

There are 3 lysines (K) in human G7 protein. The expected frequency of randomly present "EK" would be 0 (calculation: $30/213 \times 3/212 \times 212 = 0.425$). Thus the amino acid pair "EK" would not appear in human G7 protein, which is true in the real situation. Thus the absence of "EK" is randomly predictable.

Randomly unpredictable absent amino acid pairs

There are 18 glycines (G) in human G7 protein. The predicted frequency of randomly present "EG" would be 3 (calculation: $18/213 \times 30/212 \times 212 = 2.535$), so that there 3 "EG"s would be expected in human G7 protein. However, there is no "EG" in human G7 protein; therefore, the absence of "EG" from human G7 protein is randomly unpredictable.

Mutations in randomly predictable and unpredictable amino acid pairs

Our rationale for determination of mutations in randomly predictable and unpredictable present amino acid pairs is based on the finding of our previous study,[26] which is described as follows: There are two mutations in rat monoamine oxidase B. The first mutation occurs at position 139, changing leucine (L) to histidine (H), and the amino acids at positions 138 and 140 are proline (P) and alanine (A). Thus this mutation leads to 4 amino acid pairs being changed: "PL" → "PH" and "LA" → "HA." As "PL" and "LA" are randomly predictable amino acid pairs according to our random analysis, consequently we would not expect the first mutation to lead to a substantial change in enzymic activity, which is true in the real situation. The second mutation occurs at position 199, changing "I" to "F" and leading to changes in amino acid pairs as "II" → "IF" and "IS" → "FS." As "IS" belongs to unpredictable amino acid pair according to our random analysis, we would expect the second mutation to lead to a substantial change in enzymic activity, and this expectation also is true in the real situation. In this manner we hope to determine whether a mutation occurs at randomly predictable or unpredictable amino acid pairs in human G7 protein in order to get more insight into the relationship between mutations and sensitivity of amino acid pairs.

Difference between actual and randomly predicted frequencies

For the numerical analysis, we calculate the difference between actual frequency (AF) and predicted frequency (PF) of affected amino acid pairs: $\sum (AF - PF)$. For instance, a variant at position 93 substitutes "G" for "D," which results in 2 amino acid pairs, "DG" and "GE," changing to "DD" and "DE," because the amino acid is "D" at position 92 and "E" at position 94. The actual frequency and predicted frequency are 4 and 1 for "DG," 2 and 3 for "GE," 0 and 1 for "DD," and 1 and 2 for "DE," respectively. Thus, the difference between actual frequency and predicted frequency is 2 with regard to the substituted amino acid pairs $(4 - 1) + (2 - 3)$, and –2 with regard to the substituting amino acid pairs $(0 - 1) + (1 - 2)$. In this way, we can compare the frequency difference in the amino acid pairs affected by mutations.

RESULTS

General information on amino acid pairs in human G7 protein

Of the 400 theoretically possible amino acid pairs, 262 are absent from human G7 protein, including 207 randomly predictable and 55 randomly unpredictable pairs. Consequently, the 212 amino acid pairs found in human G7 protein include only 138 kinds of theoretically possible amino acid pairs (400 – 262 = 138; some amino acid pairs should appear more than once). Actually, of 212 amino acid pairs in human G7 protein, 94 kinds of theoretical amino acid pairs appear once, 25 kinds twice, 14 kinds three times, 3 kinds four times, 1 kind five times, and 1 kind nine times.

Of 138 kinds of theoretical amino acid pairs in human G7 protein, 38 kinds are randomly predictable and 100 kinds are randomly unpredictable. As mentioned earlier, some kinds of amino acid pairs appear more than once; thus, of 212 amino acid pairs in human G7 protein, 46 are randomly predictable and 166 are randomly unpredictable. We therefore can find how many mutations occur with respect to these present amino acid pairs in human G7 protein (Table 1).

Mutations of G7 protein in randomly predictable and unpredictable present amino acid pairs

As mentioned in "Materials and Methods," a protein with a point mutation leads two amino acid pairs to be substituted by another two, and their actual frequency can be smaller than, equal to, or larger than their predicted frequency. Tables 2 and 3 detail the situations related to substituted and substituting amino acid pairs, respectively, and the relationship between their actual and predicted frequencies.

Table 2 can be read as follows. The first column classifies the amino acid pairs into randomly predictable and unpredictable. The second and third columns show which type of amino acid pairs the mutation occurs in; for example, the first two cells in columns 2 and 3 indicate that the actual frequencies are equal to the predicated frequencies in amino acid pairs I and II. The fourth and five columns indicate how many mutations occur in amino acid pairs I and II; for example, 6 of 109 (5.5%) mutations occur in amino acid pairs whose actual frequencies are equal to predicted frequencies. The sixth column indicates the percentage of 109 mutations occurring at predictable and unpredictable amino acid pairs.

Tables 1 and 2 indicate that 94.5% of mutations occur at randomly unpredictable present amino acid pairs and 5.5% of mutations occur in randomly predictable amino acid pairs. These results mean that 100 kinds of randomly unpredictable present amino acid pairs account for 94.5% mutations in human G7 protein, whereas 38 kinds of randomly predictable present amino acid pairs account for only 5.5%. These results strongly support our rationale that the harmful mutations are more likely to occur at randomly unpredictable present amino acid pairs, which therefore are more sensitive to the mutations.

When looking at the unpredictable pairs in Table 2, we find that the vast majority of these pairs are characterized by one or both substituted pairs whose actual frequency is larger than the predicted frequency (the first 3 rows in unpredictable pairs). Comparing each mutation, we find that the impact of mutations is to narrow the difference between actual and predicted frequencies by means of reducing the actual frequency. This means that the mutations lead to the construction of amino acid pairs to be randomly predictable. In other words, the mutations lead amino acid pairs to occur more easily. It is interesting to note that only 2 mutations occur in the amino acid pairs whose actual frequency is smaller than predicted frequency in both pairs. This suggests that it is difficult for mutations to narrow the difference between actual and predicted frequencies by means of increasing the actu-

al frequency; however, reducing actual frequency would lead to the construction of amino acid pairs against natural direction.

Table 3 can be read as follows. The first and second columns indicate the actual and predicted situations in amino acid pairs I and II, the third and fourth columns indicate the number of mutations that occurs at amino acid pairs I and II and their percents, the fifth column is the total of our classifications.

Table 3 shows that 77.99% of mutations result in one or both substituting amino acid pairs, which are absent in normal human G7 protein (AF = 0). Table 3 also tells us that 52.28% of mutations target one or both substituting amino acid pairs with their actual frequency smaller than predicted frequency (see table footnote). These phenomena indicate that the amino acid pairs in mutant G7 proteins are more randomly constructed.

Frequency difference of amino acid pairs affected by mutations

The difference between actual and predicted frequencies represents a measure of randomness of construction of amino acid pairs; that is, the smaller the difference, the more random the construction of amino acid pairs. In particular, (i) the larger the positive difference, the more randomly unpredictable amino acid pairs were present; and (ii) the larger the negative difference, the more randomly unpredictable amino acid pairs were absent.

Considering all 109 variants, the difference between actual and predicted frequencies is 1.83 ± 0.14 (mean ± SE, ranging from –3 to 10) for substituted amino acid pairs. This means that the variants occur in the amino acid pairs, which appear more than their predicted frequency. Meanwhile, the difference between actual and predicted frequencies is –0.29 ± 0.12 (mean ± SE, ranging from –4 to 4) for substituting amino acid pairs, which implies that the substituting amino acid pairs are randomly constructed in the mutant G7 proteins, as their actual and predicted frequencies are about the same. Striking statistical difference is found between the substituted and substituting amino acid pairs ($P < 0.0001$). Figure 1 shows the distribution of difference between actual and predicted frequencies.

DISCUSSION

In this study we use the random approach to analyze amino acid pairs in human G7 protein in order to determine which amino acid pairs are more sensitive to mutations. The results confirm our hypothesis that the randomly unpredictable amino acid pairs are more sensitive to mutations. This data-based theoretical analysis may provide a clue for preventing human G7 protein from mutations and cast light on the nature of G7 protein mutations.

Based on our previous studies,[28] we argue that functional amino acid pairs should be deliberately evolved, and thus the actual frequency should be different from the predicted frequency. As the predicted frequency is the highest chance for construction of amino acid pairs, it is important to find whether the mutations lead to the actual frequency approaching the predicted frequency. If so, we can understand that the protein has a natural trend toward mutations; if not, we can understand the protein does not have a natural trend to mutate. The present study demonstrates that the human G7 protein has a natural trend toward mutations.

With respect to randomly unpredictable absent and present amino acid pairs, we are interested in the difference between actual and predicted frequencies because the randomly predicable absent and present frequency represents the easiest, most naturally occurring event. The construction of amino acid pairs should be the least energy- and time-consuming. Thus the difference between actual and predicted frequencies should be engineered by the evolutionary process: the larger the difference, the larger the impact of the evolutionary process. The narrowing of the difference between actual and predicted frequencies has been shown in

this study; thus the mutations in fact are a degeneration process inducing tumors.

REFERENCES

1. Maher ER, Yates JR, Harries R., et al. Clinical features and natural history of von Hippel–Lindau disease. *Q J Med.* 1990;77:1151–1163.
2. Singh AD, Shields CL, Shields JA. Von Hippel–Lindau disease. *Surv. Ophthalmol.* 2001;46:117–142.
3. Hes FJ, van der Luijt RB, Lips CJM. Clinical management of Von Hippel–Lindau (VHL) disease. *Neth J Med.* 2001;59:225–234.
4. Maher ER, Iselius L, Yates JR, et al. Von Hippel–Lindau disease: a genetic study. *J Med Genet.* 1991;28:443–447.
5. Neumann HP, Wiestler OD. Clustering of features of von Hippel–Lindau syndrome: evidence for a complex genetic locus. *Lancet.* 1991;337:1052–1054.
6. Neumann HPH. Pheochromocytomas multiple endocrine neoplasia type 2 and von Hippel–Lindau disease. *N Eng.J Med*. 1994:330:1091–1092.
7. Moller HU. Familial angiomatosis retinae et cerebelli. *Acta Ophthalmol (Copenh).* 1929;7:244–260.
8. Seizinger BR, Rouleau GA, Ozelius LJ, et al. Von Hippel–Lindau disease maps to the region of chromosome 3 associated with renal cell carcinoma. *Nature.* 1988;332:268–269.
9. Latif F, Tory K, Gnarra J, et al. Identification of the von Hippel–Lindau disease tumor suppressor gene. *Science.* 1993;260:1317–1320.
10. Hes FJ, Lips CJM, van der Luijt RB. Molecular genetic aspects of Von Hippel–Lindau (VHL) disease and criteria for DNA analysis in subjects at risk. *Neth J Med.* 2001;59:235–243.
11. Stolle C, Glenn G, Zbar B, et al. Improved detection of germline mutations in the von Hippel–Lindau disease tumor suppressor gene. *Hum Mutat.* 1998;12:417–423.
12. Chen F, Kishida T, Yao M, et al. Germline mutations in the von Hippel–Lindau disease tumor suppressor gene: correlation with phenotype. *Hum Mutat.* 1995;5:66–75.
13. Maxwell PH, Wiesener MS, Chang GW. The tumor suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature.* 1999;399:271–275.
14. Kaelin WG Jr. Cancer. Many vessels, faulty gene. *Nature.* 1999;399:203–204.
15. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA.* 1971;68:820–823.
16. Chang JH, Spraul CW, Lynn ML, et al. The two-stage mutation model in retinal hemangioblastoma. *Ophthalmic Genet*. 1998;19:123–130.
17. Maher ER, Yates JR, Ferguson-Smith MA. Statistical analysis of the two stage mutation model in von Hippel–Lindau disease, and in sporadic cerebellar haemangioblastoma and renal cell carcinoma. *J Med Genet.* 1990;27:311–314.
18. Zber B, Brauch H, Talmadge C, Linehan M. Loss of alleles of loci on the short arm of chromosome 3 in renal cell carcinoma. *Nature.* 1987;327:721–724.
19. Kreusel KM, Bornfeld N, Lommatzsch A, et al. Ruthenium-106 brachytherapy for peripheral retinal capillary hemangioma. *Ophthalmology.* 1998;105:1386–1392.
20. Rideout WM, Coetzee GA, Olumi AF, Jones PA. 5-Methylcytosine as an endogenous mutagen in human LL receptor and p53 genes. *Science.* 1990;249:1288–1290.
21. Montesano R., Hainaut P, Wild CP. Hepatocellular carcinoma: from gene to public health. *J Natl Cancer Inst*. 1997;89:1844–1851.
22. Hainaut P, Pfeifer GP. Patterns of p53 GÆT transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis.* 2001;22:367–374.
23. Ory K, Legros Y, Auguin C, Soussi T. Analysis of the most representative tumour-derived p53 mutants reveals that changes in protein conformation are not correlated with loss of transactivation or inhibition of cell proliferation. *EMBO J*. 1994;13:3496–3504.
24. Forrester K, Lupold SE, Ott VL, Chay CH, Band V, Wang XW, et al. Effects of p53 mutants on wild-type p53-mediated transactivation are cell type dependent. *Oncogene* 1995;10, 2103–2111.
25. Aas T, Borresen AL, Geisler S, Smith-Sorensen B, Johnsen H, Varhaug, JE, et al. Specific p53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nat Med.* 1996;2:811–814.
26. Wu G, Yan SM. Prediction of presence and absence of two- and three-amino-acid sequence of human monoamine oxidase B from its amino acid composition according to the random mechanism. *Biomol. Eng*. 2001;18:23–27.
27. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 2000;28:45–48.
28. Wu G, Yan SM. Randomness in the primary structure of protein: methods and implications. *Mol Biol Today.* 2002;3:31–37.
29. Feller W. *An Introduction to Probability Theory and Its Applications.* Vol 1. 3rd ed. New York: John Wiley and Sons; 1968.