THE ROLE OF SUNLIGHT IN THE ETIOLOGY OF MELANOMA AND OTHER SKIN CANCERS

Much has been written on the possible role of sunlight in the etiology of malignant melanoma of the skin. Evidence in support of exposure to sunlight as a causal factor includes the increased susceptibility of fair-skinned individuals (1, 2) and of patients with XP (3); a higher rate of melanoma in populations living nearer the equator (4); variations in incidence in apparent response to changes in UV radiation level due to seasonal, climatic, and cosmic factors (5–7); and a significantly greater risk in association with increased total outdoor exposure (1, 2), sunbathing, and use of artificial UV lamps (8).

There are, however, other features of the epidemiology of melanoma that are inconsistent with a relationship with accumulative sun exposure. Unlike basal and squamous cell carcinomas, which occur predominantly on exposed sites in the elderly (9), melanoma arises most commonly on the trunk in men and lower limb in women and affects a younger group of patients (10). It has been argued that these dissimilarities exclude sunlight as a major cause of melanoma, because of the established association of other skin cancers with sun exposure (11).

The higher incidence of melanoma in urban populations than in rural populations (12) and in indoor rather than in outdoor workers (10) also appears to be contrary to a direct relationship with sunlight.

To account for these inconsistencies, it has been suggested that melanomas are mainly produced by intermittent exposure to sunlight and are less common when sun exposure is received more or less continuously, whereas other skin cancers occur as a function of the total accumulated dose of sunlight (10, 13–17). This hypothesis could account for the appearance of melanoma on less frequently exposed sites of the body (most of which are exposed to sun intermittently during recreation) and its higher incidence in those who work indoors, and therefore receive only intermittent sun exposure, than in those who work outdoors. Recently, Houghton and Viola (18) proposed that the role of recreational sun exposure in the pathogenesis of melanoma may be that of a cocarcinogen or promoter rather than a dose-dependent carcinogen.

ETIOLOGIC HETEROGENEITY OF MELANOMA

Before the biological basis for differences in epidemiology between melanoma and other skin cancers is explored, it should be recognized that melanoma itself may be heterogeneous with respect to its relationship to sunlight. Mishima (19) proposed that melanoma is not one but two diseases. Using electron microscopy, he found that melanoma cells of the HMF type had ultrastructural features similar to those of normal melanocytes, whereas the subcellular features of other types of melanomas were more like those of nevus cells. On this basis, he proposed that melanomas may develop by either of two pathways, one involving the transformation of melanocytes to HMF and the other giving melanoma as the end result of the development of junctional activity in pigmented nevi or simple lentigines which may be an early stage in nevus development.

These histopathologic differences appear to be reflected in epidemiologic and pathologic differences between HMF and other types of melanomas. HMF is reported to account for 23–36% of melanomas of the head and neck, compared with 2–4% of lesions occurring at other body sites (20–22). The age-specific incidence of preinvasive melanomas of the SSM and UCM types rises steeply in the early adult years to reach a peak in middle life and then falls, whereas the risk of preinvasive HMF increases gradually with increasing age (10). HMF is much more frequently associated with solar elastosis in the surrounding skin than other melanoma types even when consideration is confined to melanomas arising on...
the head and neck (23). Furthermore, there is evidence that the association of melanoma with indoor work does not apply to lesions arising on the face, head, and neck (and presumably, therefore, HMF), which like squamous and basal cell carcinomas occur at a higher rate in persons with outdoor occupations (24). It appears, therefore, that HMF has more features in common with nonmelanotic skin cancers than with other melanomas.

**SUNLIGHT: A PROMOTER OF MELANOMA ARISING IN PIGMENTED NEVI**

Nicholls (25) has argued that pigmented nevi (as well as freckles and lentigines) are not simple developmental abnormalities, but rather they are clones of mutated cells arising either spontaneously or in response to mutagenic radiation, chemicals, or viruses. It is possible that some, if not all, mutations leading to lentigo or nevus cells represent completion of an early stage in carcinogenesis. We propose, therefore, that some benign melanocytic lesions are clones of genetically altered cells capable of malignant proliferation under appropriate conditions (we shall refer to these genetically altered cells as *initiated nevus cells*). This theory is consistent with both clinical and histopathologic evidence that a substantial proportion of melanomas arise in nevi (26). Whether all acquired benign melanocytic lesions contain initiated nevus cells is moot and, in any case, is not necessary to our theory. The status of freckles, in particular, is uncertain because of their lack of melanocytic hyperplasia. Congenital nevi may contain initiated cells as evidenced by the increased risk of melanoma associated with them (26).

If intermittent sun exposure is an important risk factor for the development of melanomas of other than the HMF type, then the predominant action of sunlight received in this manner may be to promote development of invasive cancer from initiated nevus cells. Thus in agreement with Houghton and Viola (18), we suspect that UV radiation, in addition to its known potential as a cancer initiator (27), may act, in the case of melanoma, as a promoter similar to that described in the classical two-stage model of carcinogenesis (28). There is some experimental evidence to support this theory. Melanomas have been observed in response to UV irradiation of pigmented nevi induced in hairless mice by a single application of the initiating carcinogen DMBA (29). Melanomas were not seen in animals exposed to either DMBA or UV radiation alone.

A physiologic basis exists for a promotional effect of UV radiation on initiated nevus cells. In response to UVB (280–320 nm) radiation, melanocytes proliferate and increase their production of melanin (30, 31). Morpurgo et al. (32) have postulated that minor damage to DNA, resulting from a local effect of UV radiation, stimulates melanogenesis and melanocyte proliferation by induction of a repair system that includes increased synthesis of MSH receptors. They further suggested that binding of MSH to receptors on the cell membrane, through a chain of intracellular events, leads to increased synthesis of catechol oxidase and increased cell division. Thus it appears that the physiologic consequences of direct exposure to UV radiation include alterations in gene expression and cell growth, both considered to be general properties of cancer promoters (33). These effects may be mediated through systemic as well as local mechanisms (34). Raised levels of β-LPH, the human equivalent of MSH (35), have been observed in patients receiving PUVA therapy [i.e., psoralen plus UVA (320–400 nm) radiation] (32). The promoting effects of increased levels of a melanotropic hormone such as β-LPH, in response to UV radiation exposure, could possibly play a role in the development of melanomas at sites rarely exposed directly to sunlight.

Intermittent stimulation of the melanocyte system by UV radiation might be expected to have a more powerful promotional effect than continuous stimulation. In the latter situation, the occurrence of tanning would reduce the amount of UV radiation reaching the melanocytes and thus probably would give rise to a new steady state in which division of melanocytes would revert to or near its resting level. In the former situation, however, no prolonged tanning would occur and so each new exposure to the sun would lead again to melanocyte proliferation. There are analogous effects in some experimental carcinogen–cancer systems. For example, the feeding of raw soy flour, which contains a trypsin inhibitor, to rats produces pancreatic acinar hypertrophy and hyperplasia and increases the incidence of pancreatic cancer after carcinogen administration (36, 37). This effect is greater (weight for weight) if the soy flour is fed intermittently than if it is fed continuously (Morgan R: Personal communication).

The age-incidence curves of melanoma are consistent with a promotional effect of UV radiation on melanoma arising in nevi. The incidence of melanomas on the head and neck (where HMF is predominant) increases progressively with age, whereas the incidence of melanomas at other body sites (where HMF is rare) reaches a plateau or declines after the age of 40 or 50 years (10, 16). The latter pattern is consistent with the removal of a promotional stimulus in middle life corresponding, perhaps, to a decline in recreational exposure to the sun. It could, however, also be explained by reduction in exposure to an initiator some 20 or 30 years earlier. That SSM continues to be diagnosed, although at a reduced rate, in later years of life may reflect both variation in the induction period of melanoma and variation in the age at which recreational exposure to the sun declines. Among persons 80 years of age and older, SSM is uncommon and HMF clearly predominates (10).

**EFFECT OF SUNLIGHT ON MELANOMA ARISING IN HMF**

The similarities between the epidemiologic features of HMF and those of nonmelanotic skin cancers suggest that animal models in which UV radiation acts as an initiator of nonmelanoma skin cancer, without subsequent promotion by another agent, may be applicable to
HMF. Blum (38), in summarizing animal experiments of this type, put forward the notion that this form of UV radiation carcinogenesis is a continuous process that begins with the first exposure and is augmented by each subsequent dose. Thus it might be reasonable to suggest that HMF results from an accumulation of UV radiation-induced damage in the genome of melanocytes, which becomes too extensive for control by repressive mechanisms. The progressive increase in incidence of HMF with age (10) is consistent with this mechanism.

That the immediate precursor of HMF is the melanocyte and not the nevus cell is supported by the cyto­logenic findings of Mishima (19) described above and also by the observation that HMF is rarely associated with histologic evidence of a preexisting benign nevus. In a series of carefully reviewed cases from Western Australia (39), only 1 of 40 HMF was associated with a pigmented nevus as opposed to 28 of 156 melanomas of the SSM and UCM types (Heenan PJ: Unpublished results). A similar conclusion may be inferred from a study in Johannesburg, in which no histologic evidence of an associated nevus was found in 20 melanomas of the head and neck (where HMF is most common) compared with 12 melanomas with associated nevi in 44 melanomas of the trunk and limbs (40).

EVIDENCE OF PROMOTION OF JUNCTIONAL ACTIVITY IN PIGMENTED NEVI BY SUNLIGHT

It has been shown in several populations that malignant melanomas are diagnosed more frequently in summer than in winter (5, 41, 42). One possible explanation of this finding is that sunlight promotes a late stage in the development of invasive melanoma from precursor cells (5, 42). We have shown recently also that benign nevi are more likely to be excised in summer than in winter and that the proportion with a junctional component is higher in summer than in winter (43). In midsummer the proportion reached a peak of 50%, whereas in midwinter only 34% of excised nevi had a junctional component. In addition, there was evidence of greater degrees of inflammation and regression in junctional and compound nevi excised in summer than in winter. These features are all consistent with a promotional effect of sunlight on junctional nevi with an associated immune response possibly to new antigens that had begun to appear as a result of derepression of abnormal parts of the genome.

LACK OF EVIDENCE FOR CYCLIC VARIATION IN THE DIAGNOSIS OF MELANOMA ARISING IN HMF

Data on the numbers of melanomas diagnosed per month in Western Australia in 1980 and 1981 are shown in text-figure 1A; the melanomas are subdivided by histogenetic type (Holman CD, Heenan PJ, Armstrong BK: Unpublished data). When assessed by Edwards' test (44), statistically significant cyclic trends with summer-time peaks were observed in the SSM (χ² = 21.5; P = .00002) and UCM (χ² = 14.6; P = .0007) types. For tumors of the HMF type (χ² = 4.5; P = .098) and nodular type (χ² = 1.6; P = .458), however, no cyclic trend was confirmed statistically. On inspection of the data there was some evidence of seasonality in HMF, but this could be due to a behavioral tendency to present with skin lesions in summer, as has been postulated for intradermal nevi (43). Text-figure 1B shows variation in the ratio of cases of SSM to cases of HMF also in relation to month of biopsy. It is apparent that SSM in Western Australia is comparatively most common in the midsummer months of December and January, whereas the lowest numbers in comparison with cases of HMF are reached in the late autumn months of April and May.

These results suggest that overall seasonal variation in the incidence of malignant melanoma is mainly a reflection of a cyclic trend in the SSM and UCM histogenetic types. HMF appears to be comparatively unresponsive to seasonal factors, an observation consistent with the view that it is mainly due to accumulated lifetime dose of UV radiation. NM also showed little seasonal variation, perhaps because it usually represents a fairly advanced disease in which invasion has probably occurred at a considerably earlier time (45).

GENETIC FACTORS

Resting skin color and capacity of the skin to react to sunlight with a protective tan are probably the most important genetic determinants of risk of melanoma (1, 2, 8). The former affects the extent to which UV radiation can both initiate and promote melanoma; the latter may determine particularly the degree of promotion produced by any given regime of intermittent UV radiation exposure. There are, in addition, at least two inherited syndromes, XP and DNS, that affect risk of melanoma.

Patients with XP have a recessively inherited defect in their ability to repair UV radiation-damaged DNA and a greatly increased risk of nonmelanotic skin cancer and melanoma (3). Their repair deficiency would presumably increase both the extent of accumulation of DNA damage in melanocytes and the frequency with which initiated nevus cells developed. Therefore, it is hard to predict whether the distribution of histogenetic types of melanoma in XP patients would differ from that usually observed and what the direction of any such difference might be.

DNS is a dominantly inherited but also sporadic syndrome in which multiple dysplastic nevi occur, continue to appear well into adult life, and progress to melanoma (46). Similarly to Elder et al. (46), we suspect that the dysplastic nevus is an intermediate stage between the common nevus and SSM-type melanoma. Dysplastic nevus may not be confined to the recognized syndrome; thus all clones of initiated nevus cells that eventually progress to melanoma may pass, even if only briefly, through a phase that would be described as dysplastic if the cells were examined at that time. The increased frequency with which this process occurs in DNS may be due to an increased sensitivity of melanocytes to UV radiation (47).
TEXT-Figure 1.—A) Distribution of melanomas of the HMF, SSM, UCM, and NM histogenetic types by month of biopsy in Western Australia, 1980–81. B) Variation in the ratio of cases of SSM to cases of HMF by month of biopsy in Western Australia, 1980–81.

facilitating both nevus initiation and promotion, or some other change increasing the probability of promotion of nevi to melanoma, or both.

A THEORY OF THE PATHOGENESIS OF HUMAN CUTANEOUS MALIGNANT MELANOMA

In text-figure 2 we propose in schematic form a theory of the pathogenesis of human malignant melanoma. Much of the theory, which retains Mishima’s original concept of the duality of melanoma (19), is derived from arguments developed in the preceding sections. We wish to clarify, however, some additional aspects of the theory. It is possible that some groups of initiated nevus cells, especially in small lentigines, are too small to be recognized as visible lesions. This would explain the reported occurrence of non-HMF melanomas in areas of skin that previously appeared normal on macroscopic examination. Intermittent promotion of initiated nevus cells may well occur not only in response to UV radiation exposure, but also as a result of other factors thought to be associated with melanoma, such as sex hormones (48) and trauma (49), although the evidence for such effects is inconclusive at present. A depressed immunologic response, which has been demonstrated in skin showing evidence of actinic damage (50), also might be a contributory factor, perhaps allowing in situ HMF to progress to dermal invasion.

As indicated in text-figure 2, we suggest that melanoma of the NM histogenetic type represents a common end stage of the other 3 types, the in situ components of HMF, SSM, or UCM having been obliterated by proliferating malignant cells. This view, which has been stated somewhat indirectly by Ackerman (51), is supported by the epidemiologic and associated histologic features of
NM, which are consistent with NM having originated in HMF or SSM (45). It has been commonplace to refer to NM as melanoma arising "de novo" because, by definition, it has no adjacent in situ component. The histologic appearance of NM at the time of excision, however, provides an indication of its histogenetic origin at only an instant in time; it is impossible on this basis, therefore, to determine whether or not it began de novo or in an in situ HMF or SSM. Thus although we cannot exclude the possibility that some NM may evolve through a pathway separate from those pathways proposed for HMF and SSM, we know of no evidence for it and have not included it in text-figure 2. UCM has been grouped with SSM because in situ UCM may be an early stage of in situ SSM (52). Atypical melanocytic hyperplasia is often seen in the epidermis adjacent to in situ SSM (52). Atypical melanocytic hyperplasia here refers to a form of in situ UCM (52) and not, as is sometimes the case, all level-1 melanomas.

We recognize that many aspects of this theory are speculative, but we believe that this theory provides a useful explanation for some of the puzzling features of malignant melanoma. It also provides some new hypotheses that can be tested by further research, for example: 1) HMF melanoma shows no special relationship to intermittent sun exposure, whereas other types of melanoma do; 2) increased intermittent (recreational) exposure to sunlight in fair-skinned populations has caused an increase in the proportion of non-HMF melanomas [greater proportional increases with time in melanomas of the trunk and lower limb compared with melanomas of the head and neck have already been reported by Magnus (16)]; 3) persistent tanning has a greater protective effect against non-HMF than HMF melanoma; 4) a plateau and/or decline in non-HMF melanoma incidence after the age of 40 years corresponds to changes in recreational exposure to the sun around that age; 5) NM forms a larger proportion of melanomas in regions where patients present later than in regions where early diagnosis is common.

REFERENCES
(2) Gellin GA, Kopf AW, Garfinkel L. Malignant melanoma. A


(35) BROWN JD, DOE RP. Pituitary pigmented hormones. Relationship of melanocyte-stimulating hormone to lipotropic hormone. JAMA 1978; 240:1273-1278.


