Thymic involution and rising disease incidence with age

Citation for published version:
Palmer, S, Albergante, L, Blackburn, C & Newman, TJ 2018, 'Thymic involution and rising disease incidence with age', Proceedings of the National Academy of Sciences.
https://doi.org/10.1073/pnas.1714478115

Digital Object Identifier (DOI):
10.1073/pnas.1714478115

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proceedings of the National Academy of Sciences

Publisher Rights Statement:
This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Thymic involution and rising disease incidence with age

Sam Palmera,b,1, Luca Albergantec,a, Clare C. Blackburnd, and T. J. Newmanb,1,2

aSchool of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom; bSchool of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, 62200 Putrajaya, Malaysia; cInstitut Curie, Université de Recherche Paris Sciences et Lettres, Mines ParisTech, INSERM U900, F-75005 Paris, France; and dMedical Research Council Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, EH16 4UH Edinburgh, United Kingdom

Edited by Bruce S. McEwen, The Rockefeller University, New York, NY, and approved December 26, 2017 (received for review August 28, 2017)

For many cancer types, incidence rises rapidly with age as an apparent power law, supporting the idea that cancer is caused by a gradual accumulation of genetic mutations. Similarly, the incidence of many infectious diseases strongly increases with age. Here, combining data from immunology and epidemiology, we show that many of these dramatic age-related increases in incidence can be modeled based on immune system decline, rather than mutation accumulation. In humans, the thymus atrophies from infancy, resulting in an exponential decline in T cell production with a half-life of ~16 years, which we use as the basis for a minimal mathematical model of disease incidence. Our model outperforms the power law model with the same number of fitting parameters in describing cancer incidence data across a wide spectrum of different cancers, and provides excellent fits to infectious disease data. This framework provides mechanistic insight into cancer emergence, suggesting that age-related decline in T cell output is a major risk factor.

Significance

Understanding the risk factors of carcinogenesis is a major goal of biomedical research. Historically, the focus has been on the role of somatic mutations, and the reason for cancer typically occurring late in life is predominantly attributed to a gradual accumulation of such mutations. We challenge that view and propose that the decline of the immune system is the primary reason why cancer is an age-related disease. The immunological model featured here captures risk profiles for many cancer types and infectious diseases, suggesting that therapies reversing T cell exhaustion or restoring T cell proliferation will be promising avenues of treatment.

Author contributions: S.P. and T.J.N. designed research; S.P., L.A., C.C.B., and T.J.N. performed research; S.P., L.A., and T.J.N. contributed new reagents/analytic tools; S.P., L.A., and T.J.N. analyzed data; and S.P., L.A., C.C.B., and T.J.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: s.palmer@hw.ac.uk or tjnewman@solaravus.com.

2Present address: Solaravus, Cupar KY15 5AS, United Kingdom.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714478115/-/DCSupplemental.
(Methods). This gives a first-approximation prediction that the risk of immune escape, which we denote by $R$, rises exponentially with age at the same rate that T cell production declines. This defines a model for disease incidence with one fitting parameter, that being an overall prefactor (Table 1). If the random walk is biased (e.g., if the rates of cell division and cell death are not equal), a similar calculation produces a more general prediction for incidence with one additional parameter (Table 1). We will refer to these one- and two-parameter model predictions as, respectively, immune model I (IM-I) and immune model II (IM-II). The additional fitting parameter of IM-II can be interpreted as a “pivot age,” which marks a transition from very low to relatively much higher risk (Methods). We stress that $a$ is not a fitting parameter, but the empirically derived rate from thymus involution, given by $0.044 \text{y}^{-1}$, which we use for all of our analysis.

**Infectious Disease Incidence.** For most infectious diseases, the increase in risk with age is believed to be due to changes in the immune system and therefore provides a good first test for our model. The assumption that the immunogenic cells arise with equal probability at any age amounts to assuming constant exposure across age groups. We found that six of the seven bacterial infections monitored by the Active Bacterial Core (ABC) surveillance program (Data Sources) fit IM-II well ($R^2 > 0.9$), with better fitting for those incidence curves underpinned by higher incidence and larger population sizes and hence associated with a smaller relative uncertainty [i.e., smaller confidence intervals (CIs)]. Turning to viral diseases, the incidence of West Nile virus (WNV) disease is particularly well fit by IM-II (and indeed IM-I). However, influenza A is not fit well, instead rising exponentially at a faster rate (Fig. 2). Prevalence of tuberculosis infection in Cambodia also fits the model well (Fig. S4). Indeed, even IM-I fits these infectious diseases very well, which confirms the importance of the thymic involution timescale. This provides confidence in applying our approach further.

**Cancer Incidence.** We next tested our model against cancer incidence curves, across 101 cancer types under the ICD03 WHO2008 classification (11). Fitting IM-II to the incidence curves, the median $R^2$ was found to be 0.956, with 57 cancer types fitting very well ($R^2 > 0.95$). Since IM-II has the same number of fitting parameters as the widely used PLM of cancer incidence, a direct comparison is possible. The PLM performs slightly worse overall ($R^2 > 0.95$ for 48 cancer types, median $R^2 = 0.947$; $R^2$ and associated fitting measures for each cancer type can be found in Dataset S1), with cancers whose incidence rises exponentially, such as chronic myeloid leukemia (CML) and brain cancer, fitting IM-I and IM-II better than the PLM. Many cancer types, including colon and gallbladder, fit both the PLM and IM-II very well (Fig. 3). There are no examples of PLM fitting well and notably better than IM-II. The ability of IM-II to capture the power law behavior seen in cancer incidence curves is an unexpected feature of the model and is discussed further in SI Theory (where we show that IM-II exhibits an apparent power law with power $c(e - 2) \approx 3.78$ in the age range of 33–82 y). We note that, of the top 10 best-fit cancers, the 9 carcinomas have pivot ages tightly clustered from 56.3 to 60.5 y (Dataset S1), suggesting a clinical significance of the mid- to late fifties as an age of particular importance for screening and intervention. In contrast, the PLM by definition is “scale-free” and thus has no associated age range of particular importance from a clinical perspective.

From the form of the equation of IM-II we can see that, up to a shift in age and an overall multiplicative factor, all incidence curves should follow the same function (Methods and Fig. 3D). This “universal scaling function” shows the range of behaviors possible within the model. Indeed, the quality of data collapse of incidence data onto the universal scaling function for IM-II is excellent, giving strong support to our model and highlighting those cancer types that fit the model particularly well. One such cancer, CML, is characterized by a single translocation event resulting in the formation of the Philadelphia chromosome (12). This is a good candidate for the type of initiating event featured in our model. Assuming this translocation event can happen at

![Image](https://www.pnas.org/cgi/doi/10.1073/pnas.1714478115)

*Fig. 1.* Declining T cell production leads to increasing disease incidence. Our model assumes that immunogenic cells arise with the same probability at any age, and, after a period of being targeted, the population may overwhelm the immune system by crossing an immune escape threshold. This threshold is assumed to be proportional to T cell production, which decreases with age. This provides a prediction for the possible forms of disease incidence curves.

### Table 1. Summary of the mathematical forms of the immune models and the PLM

| Model                               | Predicted risk profiles                       | Free parameters | Brief description                        |
|-------------------------------------|-----------------------------------------------|-----------------|------------------------------------------|
| Immune model I                      | $R = A e^{\text{ct}}$                         | $A$             | Risk doubles every 16 y                  |
| Immune model II                     | $R = A I (e^{\text{ert}} - 1)$                | $A, \text{r}$  | Risk profile shifts around a pivot age of r years |
| Power law (multistep) model         | $R = A t^r$                                   | $A, \gamma$    | Waiting time for $\gamma + 1$ rare events (6, 7) |

The risk of contracting a disease, $R$, as a function of age, $t$, can be modeled by immunological models or the PLM. The model parameter $a$ is given by $0.044 \text{y}^{-1}$, while the free parameters are obtained by fitting the models to each disease and are available in Dataset S1.
any age, on neglecting the IET one might expect that incidence would be approximately constant. Instead, incidence doubles every 16 y, mirroring the exponential decay of T cell production, consistent with our model.

Examining which incidence curves fit poorly can give insight into the underlying diseases (Fig. S11). We fit 20 cancer types with IM-II and 34 cancer types with PLM). For example, breast and thyroid cancer both rise rapidly and then plateau from middle age onward, possibly due to the significant hormonal influences for these cancers. Many cancer types have a plateau or even a dip in incidence around age 80. This cannot be explained by either IM-II or the PLM, since both give strictly increasing incidence. For example, colorectal cancer.

Discussion
We have shown that there is a strong link between T cell production and incidence of both infectious diseases and cancer. Some disease incidence curves rise exponentially, inversely proportional to T cell production (Fig. 3 and Fig. S8), while some rise in a manner well-captured by our two-parameter model, IM-II. This simple model, comprising (i) a threshold proportional to T cell production and (ii) a biased random walk characterizing the population dynamics of the immunogenic cells, can explain, to a large extent, cancer and infectious disease incidence, including gender differences. Further research is needed on the precise form of the IET to understand how it interfaces with declining T cell production in different diseases and individuals. The immunological model provides a fresh perspective on carcinogenesis, strongly supporting the idea that cancer can be caused by a single event in one cell that subsequently manages to
beat the odds and evade the immune system through rare stochastic fluctuations in population dynamics. This is in stark contrast to the PLM, where the increase in risk with age arises from the waiting time for multiple independent events. We also predict that, for those animals that do not experience thymic involution, for example, some species of shark (16), cancer risk would not increase dramatically with age, and would thus be a relatively rare cause of death.

Mutations do indeed accumulate with age (17, 18), and although the premise of the PLM is logically and mathematically sound, this model predicts that several rare independent driver mutations are necessary for carcinogenesis. The fitted curves from the PLM and IM-II often overlap and can explain equally well many incidence curves. Further research is therefore necessary to estimate the number of driver mutations via other lines of inquiry. A recent paper (15) attempted to address this question from a new direction, by comparing groups with different independently determined values for $n$ in males than females, reflecting the gender bias in T cell production (Methods). A purely exponential incidence curve would correspond to a pivot age of negative infinity, and therefore, for the purposes of plotting, we set a minimum pivot age of $-50$ y. Models are fitted only for ages greater than 18 y. Error bars show 95% CIs for all diseases.
does not show a highly nonlinear relationship, which would be expected from the mutation accumulation hypothesis (20). Indeed, if we apply the PLM to this dataset (Dataset S2), we find that the number of driver mutations is just \( n \sim 0.91 \) (SI Theory) consistent with the speculations underpinning the immunological model.

While IM-II has only two emergent fitting parameters, the underlying random-walk model has three biological parameters, resulting in an undetermined system (Methods). To get estimates for these biological parameters, such as the size of the IET, additional assumptions are required. Given that the estimated total number of stem cell divisions provides a good predictor of cancer risk (19), the rate of stem cell divisions can be assumed to be proportional to the rate of cancer initiation attempts in the immunological model (SI Theory). From this, we can obtain values for the model parameters of IM-II. We found that the size of the IET is typically \( \sim 10^6 \), which would imply that a population growing beyond \( 10^6 \) cancer cells would overwhelm the immune system and result in immune escape (see Dataset S1 for values for each cancer). In mouse experiments, primary inoculations with \( >10^6 \) cancer cells rendered mice unable to control subsequent tumor inoculations (21), providing a degree of qualitative and quantitative support for our model assumptions. This effect is related to the phenomenon of “T cell exhaustion,” which was initially defined as the clonal deletion of antigen-specific T cells due to chronic stimulation (22), and is now understood to not only act via induction of T cell exhaustion but also changes in T cell phenotype and function (5). Therapies targeting T cell exhaustion have already been widely successful in cancer and infectious disease therapy in the form of immune checkpoint blockades such as PD-1 and CTLA-4 inhibitors (5). Our model provides a theoretical framework for such treatments and predicts that treatment efficacy could be enhanced if new naive T cell production were also increased. Additionally, evidence for a causative link between thymic activity and cancer risk has been found in mouse models, as thymectomized mice develop significantly more tumors (23,24) and thymus grafts on nude mice can induce cancer remission (25,26).

Our view supports the idea that as little as one single genomic aberration could be at the root of tumorigenesis. This event could be the emergence of a potent driver mutation, for example, a growth-inducing chromosomal translocation. Interestingly, it has been pointed out that a relatively small number of oncogenes have been confirmed across multiple biological experiments and all of these genes control cellular growth (27). Moreover, karyotypic analysis indicates that chromosomal rearrangements are encountered in most cancers in a way that is generally unique to that specific cancer under consideration (28). This led some to suggest that such changes are causative to cancer (29). Our analysis indicates that a single event (e.g., the emergence of a key mutation) could be enough to generate a malignancy that is able to evolve into a clinically manifest cancer if it escapes immune control. The immunological model also identifies a potential smoking gun in cancer risk in the form of the exponential decline of T cell production with age. Despite the decrease in T cell production from the thymus, overall T cell counts in the blood remain approximately constant due to increased peripheral clonal expansion (1). We therefore make the prediction that T cell efficacy is not increased by clonal expansion.

Our hypothesis and results add to the understanding of infectious disease and cancer incidence, suggesting in the latter case that immunosenescence, rather than gradual accumulation of mutations, serves as the predominant reason for an increase in cancer incidence with age for many cancers. For future therapies, including preventative therapies, strengthening the functionality of the aging immune system (30) appears to be more feasible than limiting genetic mutations, which raises hope for effective new treatments.

Methods

Immunological Model. Simple models can often be very powerful in explaining complex phenomena (31,32). With this in mind, we formulated a minimal model for disease incidence that does not attempt to explain the data exhaustively, but rather aims to be as simple as possible for the purposes of investigating the primary factors and rate-limiting steps.

During an immune response, immunogenic cells will be eliminated, while also increasing in number through division, such that the number of immunogenic cells follows a (biased) random walk. This stochastic birth-death process has been studied previously (10). The probability for reaching a population threshold \( K \) is given by the following:

\[
b^{r-1} \frac{d-b}{d^2-b^2}.
\]

[1]

where \( b \) and \( d \) are the birth (division) rates and death rates, respectively. The threshold \( K \) is interpreted as the largest number of immunogenic cells that can be effectively controlled by the immune system, and is thus the IET. Multiplying by the rate of initiating events \( r \), we arrive at the predicted risk profile:

\[
R = r b^{r-1} \frac{d-b}{d^2-b^2}.
\]

[2]

We assume that the only factor depending on age is \( K \). The decrease of the IET with age is supported by experiments in mice showing a decline in proliferative capacity of activated T cells with age (33,34). Specifically, we assume that the IET is proportional to the rate of export of naive T cells from the thymus. This would be the case if, for example, each T cell progenitor can only produce a finite number of daughter T cells and respond effectively to a finite maximum number of immunogenic cells, analogous to the Hayflick limit of replicative senescence (35). This gives \( K = K_0 e^{-\alpha r} \), leading to a predicted risk profile of the form \( R = A/(e^{\alpha r} - 1) \), where \( A = r(d-b)b/K_0 \log(db) \).

Immunogenic cells are likely to have a higher division rate than normal cells, but since they are eliminated by the immune system, they will also have a higher death rate. Under the approximation that the division rate is equal to the death rate, Eq. 3 reduces to \( R = A/e^{\alpha r} \), while for ages greater than \( r \), the risk profile can be approximated as a pure exponential \( R = A e^{-\alpha r} \). In more biological terms, the pivot age represents the age when a cancer type transitions from very rare to relatively less rare. The median pivot age across all cancer types is \( \sim 0.041 \) for females. 1. For the universal scaling function separated by gender, we have used values of exponent \( \alpha \) higher in males than females. Since the available data on gender-separated TREC decline found in ref. 4 are very noisy (\( \alpha \) for male TREC is given by 0.08, with 0.05–0.11 95% CI, while \( \alpha \) for female TREC is given by 0.04, with 0.01–0.07 95% CI), we have arrived at values for \( \alpha \) in males and females based on disease data. The cancer type which fits IM-I data on gender-separated TREC decline found in ref. 4 are very noisy (\( \alpha \) for male TREC is given by 0.08, with 0.05–0.11 95% CI, while \( \alpha \) for female TREC is given by 0.04, with 0.01–0.07 95% CI), we have arrived at values for \( \alpha \) in males and females based on disease data. The cancer type which fits IM-I
The universal scaling function in Fig. 3 depicts the top 20 best-fitting cancers as measured by the Akaike information criterion (AIC). Other choices of measure give similar results (Fig. 510).

The immunological model above can be combined with the PLM to produce a model with three fitting parameters. To do so, we alter the assumption that potentially cancerous cells are produced at a constant rate, r, and assume instead that they arise from the gradual accumulation of driver mutations. Using the framework of the PLM (6, 7), the rate of attempts then takes the form \( r = r_0 \tau \), corresponding to the waiting time for \( r + 1 \) rare independent events. This PLM predicts risk profiles of the following form:

\[
R = \frac{A}{\alpha \tau} \tau^\gamma, \quad [5]
\]

where \( A = r_0d - d/bb \), \( B = K_0 \log(db) \).

Data Sources. Data sources for incidence rates are chosen based on largest possible sample sizes.

All cancer incidence data are obtained from the Surveillance, Epidemiology, and End Results Program (SEER) in the United States (11).

Bacterial incidence data are obtained from the ABC surveillance program run by the Centers for Disease Control and Prevention (CDC). This program studies seven key bacterial diseases in detail (https://www.cdc.gov/abc/reports-findings/surv-reports.html).

Incidence data for viral diseases are obtained from studies with the largest possible sample sizes. WWV disease incidence data are obtained from a y-survey conducted in the United States from 1998 to 2008 (available at https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5902a1.htm; accessed February 23, 2016).

Influenza A incidence data are obtained from a 22-year survey obtained in the United States (36).

1. van den Dool C, de Boer RJ (2006) The effects of age, thymectomy, and HIV infection on alpha and beta TCR excision circles in naive T cells. J Immunol 177:4391-4401.
2. Murray JM, et al. (2003) Naive T cells are maintained by thymic output in early ages but by proliferation without phenotypic change after age twenty. Nat Immunol 8:407-495.
3. Douek DC, et al. (1998) Changes in thymic function with age and during the treatment of HIV infection. Nature 396:690-695.
4. Linhart A, et al. (2015) Simultaneous quantification of T-cell receptor excision circles (TRECs) and K-deleting recombinase excision circles (KRECs) by real-time PCR. J Virol Exp 2014:52:184.
5. Jiang Y, Li Y, Zhu B (2015) T-cell exhaustion in the tumor microenvironment. Cell Death Differ 22:173-193.
6. Nordling CO (1953) A new theory on cancer-inducing mechanism. Br J Cancer 7:68-72.
7. Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 6:1-12.
8. Bray F, Møller B (2006) Predicting the future burden of cancer. Nat Rev Cancer 6:63-74.
9. Weng NP (2006) Aging of the immune system: How much can the adaptive immune system adapt? Immunol Cell Biol 84:11-24.
10. Cinner LS, Newman TJ (2014) Quantifying meta-analytic inefficiency: Rare genotypes versus phenotypes. Theor Popul Biol 91:46003.
11. Surveillance, Epidemiology, and End Results (SEER) Program (2017) SEER*Stat Database: Incidence–SEER 18 Regs Research Data (2000–2014). International Agency for Research on Cancer (IARC). Available at https://seer.cancer.gov. Accessed November 2, 2015.
12. Drucker BJ, et al. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 344:1013-1037.
13. Dorak MT, Karpuzoglu E (2012) Gender differences in cancer susceptibility: An in-depth review. Front Genet 3:268.
14. Podolsky DJ, Lobanov AV, Kryukov GV, Gladyshev VN (2016) Analysis of cancer genomes reveals basic features of human aging and its role in cancer development. Nat Comm 7:21257.
15. Tomasetti C, Marzioni L, Novak MA, Parmigiani G, Vogelstein B (2015) Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci USA 112:118-123.
16. Stanely DP, Aw D, Manley NR, Palmer DB (2009) An evolutionary perspective on the mechanisms of immunosenescence. Trends Immunol 30:374-381.
17. Bilozhil F, et al. (2016) Tissue-specific mutation accumulation in human adult stem cells during life. Nature 538:260-264.
18. Miholland B, Aulton A, Suh Y, Vijg J (2015) Age-related somatic mutations in the cancer genome. Oncotarget 6:24627-24635.
19. Tomasetti C, Vogelstein B (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 347:78-81.
20. Tomasetti C, Vogelstein B (2015) Misalignments on the theory that variation in cancer risk among tissues can be explained by the number of divisions of normal stem cells. arXiv:1501.05035.
21. McBride WH, Howie SE (1986) Induction of tolerance to a murine fibrosarcoma in two different strains of mice by exhaustions of antiviral cytotoxic effector T cells. Nature 326:758-761.
22. Moskophidis D, Lechner F, Pircher H, Zinkernagel RM (1993) Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. Nature 362:578-761.

Tuberculosis prevalence in Cambodia is obtained from ref. 37. Stem cell counts and division rate estimates are taken from ref. 19.

Statistical Methods. For incidence of infectious diseases and cancers, CIs are calculated assuming a \( \gamma \) distribution. All fitting of incidence curves is performed on log-transformed values.

To calculate the overall ratio of male TREGs to female TREGs, we computed the relative ratio of the mean and then used a bootstrapping approach to calculate the SD of that measurement.

To show that cancer risk rises more steeply for males compared with females, we fit pure exponentials to the incidence curves and record the exponents as Female alpha and Male alpha in Dataset S1. To calculate the value of \( p \) for the statement that risk rises more steeply for WNV in males compared with females, we used the ANCOVA method.

All of the code for our analysis is available online at https://github.com/Alibluca/ImmuneModelSEER.

ACKNOWLEDGMENTS. We thank Toni Aebischer, Md. Al Mamun, Doreen Cantrell, Mel Greaves, Sarah Hovie, Philipp Kruger, Danbio Liu, Luke McNally, Jacques Miller, Rob Newton, and Rose Zamoy ska for useful discussions and comments on the manuscript. This work was supported by Scottish Universities Life Sciences Alliance and NIH through Physical Sciences in Oncology Centers Grant U54 CA143682 (to S.P., T.J.N., and L.A.), the Medical Research Council (C.C.B.), the European Union Seventh Framework Program (FP7/2007–2013) collaborative project ThymStem under Grant Agreement 602587 (to C.C.B.), and the Instituts Thématiques Multi-Organismes Cancer within the framework of the Plan Cancer 2014–2019 and convention Biologie des Systèmes BIO2014 (COMET project, to L.A.).