

序

『試験に出る英単語』という受験参考書をご存知だろうか。大学入試問題に頻出する英単語をよく出てくる順に単語帳にした本である。その本が出版される以前の参考書はアルファベット順で、abandonで「諦める」と皮肉られていたのを一新したインパクトのある本であった。本書はその考え方を活かさないかと考え、3千7百万件もの論文が収録されているPubMedのデータ全体のbibliome (bibliographyと-omeの造語で、ビブリオームと読む) から「遺伝子」と「文献」の関係性をヒト遺伝子に関して抽出し、各遺伝子ごとの論文言及数を数え上げた。そのうち、数の多かった遺伝子300に関して、トップ100遺伝子については1ページで、残りの200遺伝子は半ページで、その数の多いもの順にその概要を定型で解説するという「遺伝子の単語帳」となっている。論文に頻出するトップ300遺伝子を俯瞰することで、これまで気づかなかった有名ジーン (遺伝子) を発見し「知り合い」が増え、研究の深みが増すことだろう。

ヒトの遺伝子は、ヒトゲノム解読の前には10万以上あるだろうと考えられていた。2003年4月14日に解読完了が宣言された時点でのヒトの遺伝子数の推定値は3万2615個であった。しかしながら、2024年8月末時点ではタンパク質コード遺伝子の数だと約2万弱、タンパク質をコードしない遺伝子を加えても約4.6万ほどである (Ensemblの遺伝子アノテーションによるhttps://may2024.archive.ensembl.org/Homo_sapiens/Info/Annotation)。それらのうち遺伝子名がついている遺伝子は26,281であり、本書で紹介する300の遺伝子は論文に頻出する遺伝子の多くをカバーしていると言えるのではないだろうか。詳しい統計値に関しては、一緒に300遺伝子の選定を担当してくれた小野擁子氏が「論文に出る300遺伝子の選定について」にまとめているのでぜひご一読いただきたい。

ヒト遺伝子に限定してその数え上げを行ったため、自分の感覚と合わないと思われる方もいるだろう。そういう方はぜひ同様の解析をぜひお目当ての生物で行い、本書の結果と比べていただきたい。思わぬ発見があるかもしれない。bibliomeの活用を皆さんもぜひ。

2024年10月

酒まつり周辺の酒都西条にて
坊農秀雅