Series

正しい知識が捏造を防ぐ

データを正確に解釈するための6つのポイント

No. 4

微妙なデータをどう表現するか

骨研究分野での実験データ解釈を例として

加藤茂明

はじめに

実験データの解釈は，研究をとりまとめるうえでの核心・肝心となる部分である．しかしながら，これまで，このシリーズ「正しい知識が捏造を防ぐ」で紹介されてきたように，この実験データの解釈に客観性を欠き恣意的に脚色されると，もはや正しいサイエンスからは遠のくことになる．実験データの解釈にも，実験系や分野によっておのおの流儀があり，それを知らないと投稿した論文が不当に評価されることにもなる．いちばんの問題は，これら“流儀”が大切な情報なのにもかかわらず，実際の論文には記載されておらず，また，実験書にも明確には書かれていないことが多いことである．これらは，いわゆる口伝の類といえよう．今回は，その分野における流儀やコンセンサスを正しい知識として習得しておかないと，実験データを間違って解釈する可能性のあることについて考えてみる．筆者が研究対象としている，骨組織でのさまざまな実験データ解釈の例をとりあげて，その解釈について議論してみたい．

骨組織と骨量

骨組織は，運動器として，また，体を支える硬組織として，あまねく認識されている．しかし，たとえば，ヒトの骨組織を生物進化の視点からながめてみると，血液環境が海水に近づくためのミネラル貯蔵庫ととらえることもできる．そのため，骨組織はたえず骨形成と骨吸収とをくり返しており，そのダイナミックな骨代謝回転により血中のミネラル濃度は一定に保たれているのである．これら骨代謝を担う細胞としては，骨吸収を行なう破骨細胞と，骨形成を担う骨芽細胞とが主役である．これらの骨細胞の数または機能に異常が生じると，さまざまな骨疾患が現われる（図1）．骨形成の亢進した大理石病では，骨量は増えるが骨吸収が行なわれないため，骨硬組織部分が劣化し突然の骨折をまねくことになる．一方，骨吸収が骨形成をうわまわると，骨量は徐々に減少し，いわゆる骨粗鬆症となる．骨粗鬆症の原因は多数あるが，一般的にわれわれの知る骨粗鬆症は，いわゆる閉経期後の骨粗鬆症である．これは，閉経により女性ホルモンの産生が失われ，骨吸収が異常に亢進するため骨量が減少するものである．したがって，この女性ホルモン欠乏性の骨粗鬆症は高齢の女性に顕在化する．このような骨粗鬆症が社会的に深刻かつ重要である理由は，骨折頻度の上昇にある．わが国をはじめとした先進諸国では，社会全体の著しい高齢化によって予想もしなかった新たな疾病が顕在化してきている．メタボリックシンドロームがそのよい例であるが，骨粗鬆症のリスクについて周知ではない．骨折は男女をとわず，高齢者の場合には寝たきりとなってしまうため，そののち，床に伏したまま死にいたる例も多い．そのため，本人の苦痛のみならず，医療経済を非常に圧迫しているという事実もあるのである．

Shigeaki Kato
東京大学分子細胞生物学研究所 核内情報研究分野
E-mail：uskato@mail.ecc.u-tokyo.ac.jp

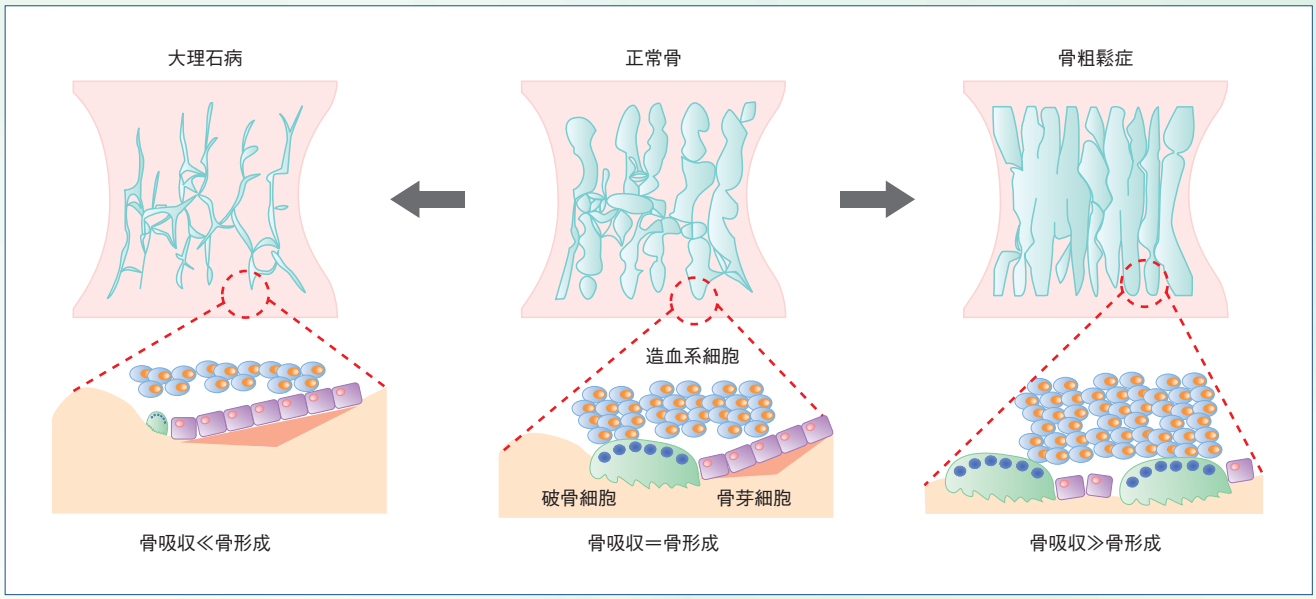

図1　骨形成と疾患

骨吸収と骨形成とのバランスが保たれていれば骨密度は正常であるが，骨吸収が骨形成を大幅にうわまわると骨粗鬆症となり，骨吸収が抑制され骨形成がこれを大きくうわまわると大理石病となる．

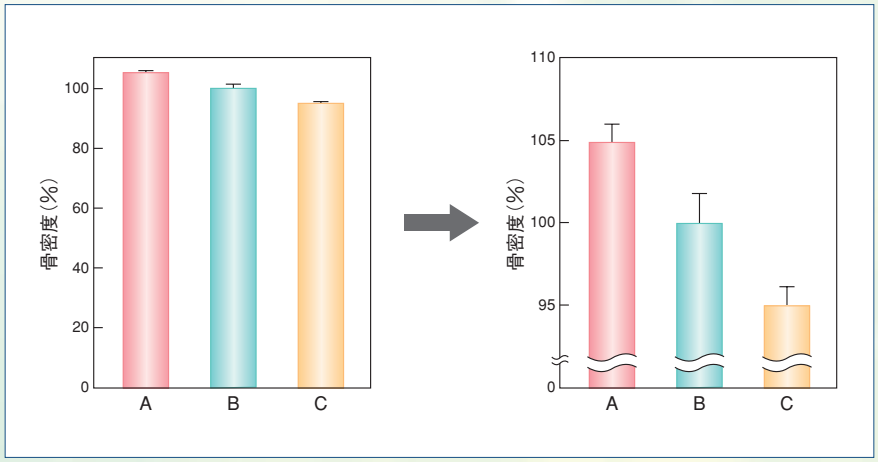

図2　骨密度データの表わし方

骨密度は，数%の差異で骨強度に重大な影響をあたえる．左に示す表わし方では，有意差があってもわかりにくい．そのため，骨代謝の研究分野では，右のように，下部を切り取ることで差異を強調したかたちで表わすのが普通である．これは，決して誇張したわけではない．

骨量をどのように評価するか？

現在，骨組織の強度を評価するものとしてもっとも汎用されているのは，骨量（ミネラルの蓄積量）である．それでは，骨量の減少と骨折の頻度とにはどれくらいの相関があるのだろうか．著しい骨粗鬆症でも，正常時よりわずか5%の骨量減少を示すだけである．しかしながら，1%の骨量減少によって骨折の頻度は10倍にも上昇することが疫学的に証明されている．一方，この5%の増減は，現代生物学，とくに分子生物学の世界では，はなはだ信用できない程度とされる数値の変動である．にもかかわらず，現在の骨量測定法で測定した数値と骨折頻度との相関に関しては，ヒトのみならず，マウスやラットにおいても，予想どおりの結果が確認されている．

したがって，この分野の研究者は，骨密度における5%の有意差を非常に大きな差と感じる．そこで，この数値を棒グラフにする際には，図2に示すように，上部の差のある部分を強調する表わし方が一般的である．しかし，これは，ほかの分野の研究者にとっては，むしろ無理やり差異を強調しているようにうつるであろう．実際，筆者の研究室に所属する大学院生の学位審査の際，副査の教官から，実験データの信頼性について疑念をいだいたかのようなコ

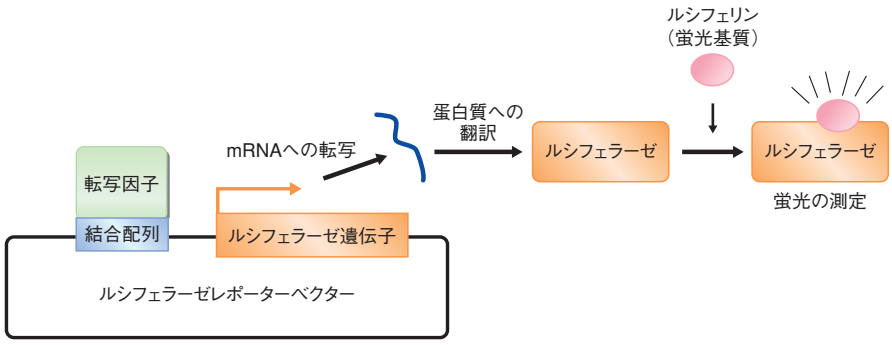

図3　ルシフェラーゼレポーターアッセイの原理

ルシフェラーゼレポーター遺伝子を用いた転写制御能の測定は，非常によくもちいられる測定法である．簡便で精度もよいが，その原理からわかるように，転写制御能を，転写，翻訳をへて酵素活性へと振り替えているため，大きな測定誤差をもたらすことがある．

メントをうけたことがある．

5％の差には意味があるのか？

一方，筆者の研究室では，転写制御に関連するさまざまな因子の転写制御能を，ルシフェラーゼの遺伝子をレポーターとして用いて，培養細胞の場合にはこれを一時的強制発現系により導入して測定することが多い．このアッセイは簡便かつ迅速で，また，非常に感度がよいため，転写制御の研究分野ではもっとも汎用されるアッセイ系のひとつである．しかしながら，逆に，感度がよすぎるため，このアッセイ系では1.5倍（50％）前後の差異では，通常，実際に差異があったとは考えない．このアッセイ系にくわしくないと思われる研究者や学生の発表あるいは論文に，苦しい実験データに統計的な有意差をつけることで差異を強調している場合をたびたびみかけるが，実験データの解釈としては賛成できない．5％の有意差があっても実際の差異としては疑わしいのだ．

その理由を考えるのに，まずこのアッセイ系の原理をみてみよう（図3）．このアッセイ系では，ルシフェラーゼの酵素活性を測っている．酵素すなわち蛋白質であり，転写活性をその最終産物である蛋白質によって測定していることになる．したがって，このアッセイ系では，転写により産生されたmRNAの安定性や，mRNAから蛋白質への翻訳の段階での影響を排除できない．そのため，mRNAの安定性や，もちいた細胞全体の翻訳過程に影響を及ぼすような因子については，正しく転写制御能を評価することができないのである．測定対象となる因子が細胞全般になんらかの影響をあたえることを考えると，差異として5％程度はつねに振れることになるだろう．

さらに，ルシフェラーゼ酵素活性の測定は，瞬間的な転写制御能を反映するものではなく，転写制御能の積分値的な蓄積を表わすものである．転写反応は，転写開始反応から，転写伸長反応，そして，転写終結反応へとつづく，連続しつつも独立した反応からなりたっている．そのため，差異が現われたとしても，このうちのどの反応段階に特異的に作用したのかは知ることができない．また，最近のエピゲノムの研究から，転写反応は染色体の構造やヒストンの修飾により大きく左右されることが明らかにされつつある．このため，これらエピゲノム制御因子も転写制御に重要な役割をはたしている．このように，転写反応に影響をあたえる因子は無数にあって，測定対象となる因子の転写制御能については，対照をきちんとおくことによって慎重に調べる必要がある．

一方，さきに述べた骨量は，このような間接的な方法ではなく，骨組織そのものを測定している．そのため，僅少な差でも，測定対象の数を増やすときれいに差異のみられることが多い．

筆者の研究室では，転写活性能また骨量と，まったく異なる測定法によりさまざまな実験データを得てこれらの解釈を行なっているため，同じ研究室内でも，実験データの有意差の解釈について混乱をきたすこともある．しかし，ここで強調したいことは，測定原理の十二分な理解が間違った実験データの解釈をさける大きな知恵であることである．

骨組織のどこを測定すべきか？

さて，ふたたび骨組織の実験データ解釈に戻る．骨は，図4に示すように，体の部位によって形状およびなりたちが異なっている．そのため，どの部位の骨量を測定するかが

大きな問題となる．重大な骨折は大腿骨に集中するため，大腿骨の骨量測定がひとつのスタンダードにはなっている．ところが，大腿骨は内軟骨から骨芽細胞に置き換わった骨組織であるのに対し，頭蓋骨は軟骨をへずに骨芽細胞から直接に生じる．そのため，大腿骨の骨量のみを測定しても片手落ちであって，頭蓋骨を研究する場合には別の実験データが必要となる．

また，たとえば大腿骨でも，部位により形状となりたちが大きく異なる（図4）．骨端は成長軟骨層から置き換わった部位であり，その内部は海綿骨とよばれる柱状の骨により埋まっている．一方，骨中央は骨端に比べると海綿骨は疎で，骨の外側のいわゆる皮質骨とよばれる部分が骨量の大部分をしめている．したがって，ひとくちに骨量といっても，骨組織の構造の差異に起因する要因により，骨端を測定するのか骨中央を測定するのかで異なる測定結果が生じるのである（図5）．そのため，分野外の研究者がこれらを無視して骨量の増減を議論しても，専門家には正しい判断ができない．また，測定部位を変えることで恣意的な実験データを作成することも可能であり，残念ながら，実際にこの点を逆手にとった自己満足的な論文も散見される．かといって，このような実験データの解釈は非常にむずかしい判断であり，ひとえに実験データの信憑性の問題としてかたづけるわけにもいかないであろう．

何を主張したいのか？

このように，骨折との関係で骨量を考えるならば，骨折が頻繁にみられる部位の骨量や骨強度が重要となる．一方，骨疾患には，骨粗鬆症のように成人になってからの疾患も多くあるが，先天性の疾患も存在する．このような遺伝的な変異によりひき起こされる骨疾患では，さまざまな部位でさまざまな異常を示すことになる．さきにも述べたように，ひとくちに骨といっても，器官としての骨は発生生物学的には異なる形式によって組織化されている．頭蓋骨のみに病状が現われる疾患もあれば，鎖骨でのみ変異や骨折のみられる特殊な遺伝病も存在する．このような遺伝疾患を解析する場合には，当然ながら，胚発生までさかのぼり，そのなりたちのしくみと調節因子を調べる必要がある．

骨組織はまさに硬組織であるため，ほかの器官とは異なり，細胞を分離して培養系に持ち込むことは不可能に近いことが多い．そのため，最近では，ノックアウトマウスを用いて機能を解析することが常道となってきている．ノッ

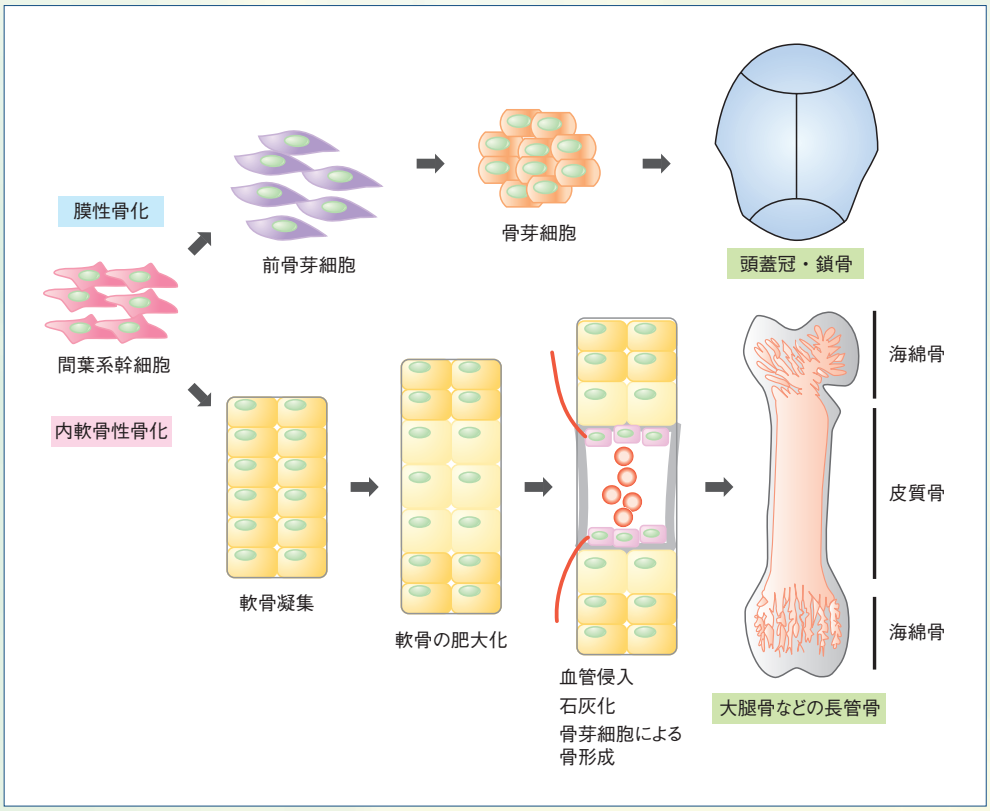

図4　骨化の様式には2つの形態がある

間葉系幹細胞が直接に骨芽細胞に分化し石灰化を導く膜性骨化により，頭蓋骨を形成する頭蓋冠および鎖骨が形成される（上段）．一方，長管骨などほかの骨組織は，間葉系幹細胞が凝集し軟骨分化をとげたのち，血管侵入のあと軟骨が骨組織に置き換わることで骨組織を形成する内軟骨性骨化によっている（下段）．

クアウトマウス法も，古典的な全身性の遺伝子破壊法から，細胞種に特異的な遺伝子破壊による時期特異的あるいは部位特異的なコンディショナルノックアウト法に移行しつつある．骨組織での研究も，徐々にではあるがコンディショナルノックアウトマウスを用いた解析に移行しつつあるが（図5），さきにも述べたように，骨組織はつねに骨代謝をくり返しており，刻々と骨を破壊しまた形成しているため，調べようとする骨の部位のみならず周辺の骨組織にも間接的な影響をあたえている可能性が考えられる．一方，造血系の視点からみれば，造血は骨髄内で行なわれるため（図1），骨組織は器だと考えられる．また，最近の研究の進歩から，造血の環境因子として骨細胞の重要性が明らかになりつつあり，双方の系から調節因子が分泌され，その機能的なクロストークが考えられている．したがって，造血系のなんらかの異常は，骨代謝系に対し2次的な影響をあたえる可能性が考えられる．その逆も真であろう．しかしながら，木を見て森を見ずの議論よろしく，多くの研究者は自分の専門的な興味でしか生物現象をとらえない例が多い．

このように，ひとくちに骨量やその測定部位を考えても，骨のよう複雑な器官では，何を主張したいかによって測定項目も異なり，また，測定すべき動物個体の時期も大きく異なることになる．また，何を主張するかによって，まったく同じ実験データに対しても議論する視点が異なってくるのである．

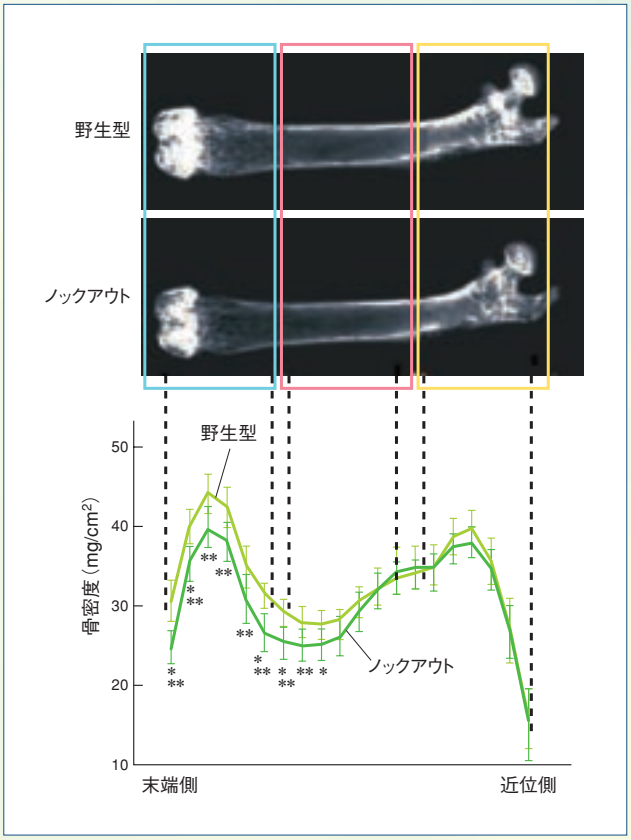

図5　部位別にみた大腿骨の骨量の違い

成長した骨組織は皮質骨と海綿骨からなりたつ．大腿骨の例を考えると，海綿骨と皮質骨のいずれの骨量に大きな変化があったかによって，測定部位により測定結果が異なることになる．エストロゲン受容体αを破骨細胞でのみノックアウトしたコンディショナルノックアウトマウスでは，海綿骨のみで骨量が大きく減少するため，骨量の減少が顕著な部位は海綿骨の多い末端側となる．そのため，この実験では末端側の実験データを用いて差異を主張することになる．

かけ離れた値の考え方

生物学の研究，とくに，マウスなど動物個体を用いた研究の場合，閉鎖系動物（クローズドコロニー）や近交系動物を用いたり，同胞間での比較を行なったり，など，個体差をなくすための工夫がされている．しかしながら，動物個体を用いた実験では，対照群・実験群の別にかかわらず，群内の平均値からかけ離れた値を示す個体の存在する場合がある．このような“はずれ値”をどのように扱うかは，非常にナーバスな問題である．

臨床研究の場合，たとえば，骨粗鬆症に対する薬物の治療効果を判定する大規模試験が試みられているが，そこではまず，一定の基準で骨粗鬆症を診断し母集団を形成する．さらに，母集団から対照群を絞り込む作業（除外の基準は，疾病既往や検査異常値の有無など）が必ず行なわれている．大規模試験では，母集団およそ18,000人に対し対照群が約8000人となるような絞り込みの行なわれている場合がしばしばみうけられる．これらの作業により，対照群はある程度の幅をもって均一化が図られているのである．しかしながら，筆者の行なっているような動物実験においては，個体差が大きいことだけでなく，犠牲死後のサンプリングまで骨密度などの値の正確な評価は困難であることから，対照群を絞り込むことのできないのが大きな問題となる．また，研究者の手によって行なわれる実験であるので，サンプルの取り違いや実験データの写し間違いなど，人為的なミスが皆無とはいえないのも事実である．

その結果，あまりにもかけ離れた値に遭遇することがあるが，はたして，これらの値を除外してもよいのだろうか？もちろん，自分の都合のいいように（研究仮説にあうように）実験データを除外するのは言語道断であるが，答えはイエスである．ただし，それにはある一定の条件を満たす必

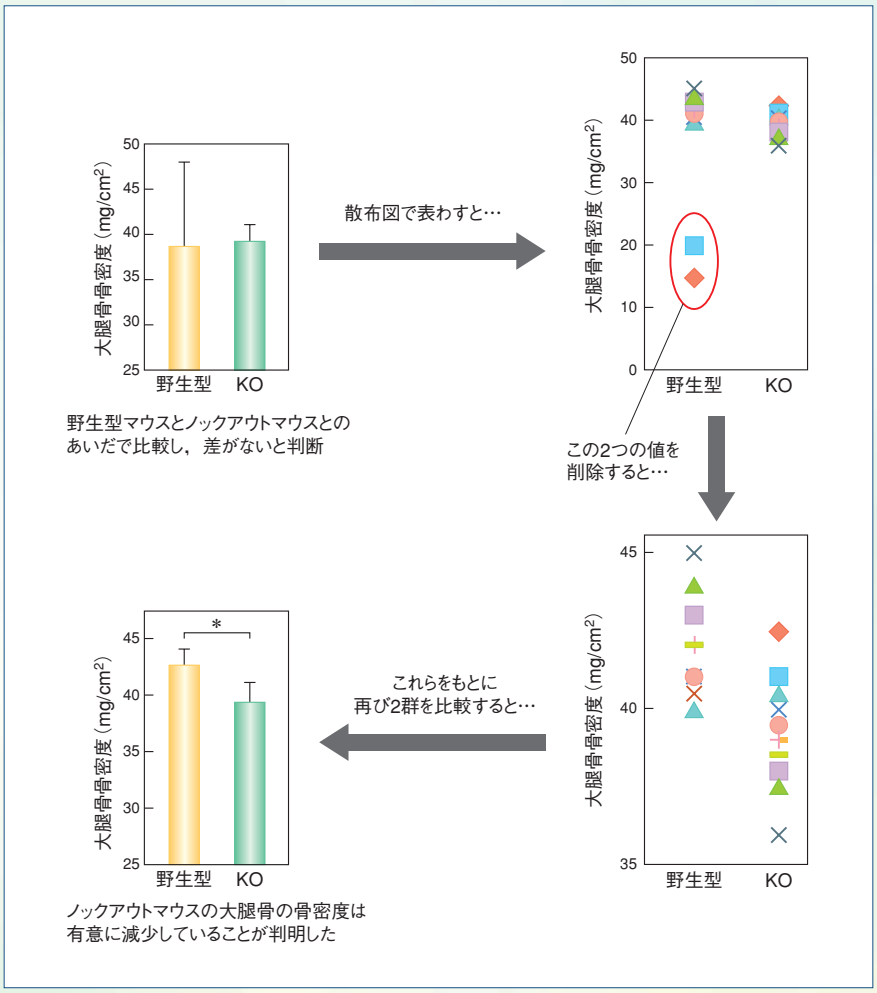

図6　かけ離れた値の考え方

たとえば，大腿骨の骨密度を野生型マウスとノックアウトマウス（KO）とで比較し，左上の棒グラフのような結果を得たため，大腿骨の密度には差がない（*t* 検定で $p = 0.75$）と判断したが，各サンプルを詳細に検討するため右上の散布図を作成したところ，野生型マウス群に2つの飛び離れた値を認めた（これらは，サンプリングの際に骨の欠損を生じるという人為的なミスであった）．この2つの値を削除して比較すると，左下の棒グラフのように，ノックアウトマウス群の大腿骨の骨密度は野生型マウス群よりも有意に減少していることが明らかとなった（*t* 検定で $p = 0.0002$）．

要のあることはいうまでもない．詳細な方法は統計学の教科書にゆだねるが，たとえば，平均値から標準偏差の2倍または3倍をこえる値を除外する，統計学的な検定（スミルノフ-グラブス棄却検定やディクソン検定など）を行なう，などが考えられる（図6）．しかし，標準偏差が小さい場合に除外してはならない重要な実験データを除外してしまったり，サンプル数が大きくないと正規分布をとっているかどうかを正確に評価できなかったりするなどの問題もある．さらには，かけ離れた値には新たな発見がひそんでいることも否定できず，実験データの取り扱いには慎重な考慮が必要である．

これらの問題をクリアするために考えられる手段として，個体差（かけ離れた値）を考慮して，たとえ生まれる確率が低い遺伝子型をもつトランスジェニックマウスであろうとも，十分なサンプル数をもって実験プロトコールをたてることを勧める．また，近年，めざましい進歩をとげている *in vivo* イメージングのよりいっそうの発展により，さまざまな表現型の評価が縦断的に行なえるようになれば，より正確な動物実験の評価が可能となり，対照群の均一化も図ることができるだろう．

性差をどう考えるか？

動物を用いた実験系のもうひとつの悩ましい点は性差である．明確には理由が記されていないのにもかかわらず，実験系によっては一方の性でのみ実験が行なわれている例が多々ある．しかしながら，ヒトを考えてもわかるように，健康診断で目にするような血中パラメーターや血球細胞数にしても，厳然とした性差が存在する．また，尿酸やコレステロール類の代謝のみならず，基礎代謝量にも性差の存在することはマウスでも確かめられている．骨組織にも性差が存在することはヒト男女の体格をみても自明であり，骨盤の形成パターンにも大きな差異が存在する．性決定には，それにかかわる多数の遺伝病が知られており，また，

社会的なインパクトも大きいため，学問的にも深く調べられている．また，発生生物学的な視点や，X染色体不活性化の分子機構の分子遺伝学的な解析には大きな興味がよせられている．一方，性差については，学問的にはあまり大きな興味ははらわれてこなかった．分子遺伝学的にみても，Y染色体には蛋白質をコードする遺伝子が極端に少ないため，長いあいだ，Y染色体の機能にはあまり注意がはらわれてこなかったように思う．しかしながら，最近のsmall RNAの同定やその機能解析から，予想もしなかった複雑な機能が浮かびあがろうとしている．Y染色体上には，このようなsmall RNAにくわえ，ノンコーディングRNAをコードする多数の遺伝子の存在することが明らかにされつつある（Y染色体上には多くのくり返し配列があるため，ヒトのゲノムでさえ解読はかなり不完全である）．

骨組織や骨形成における性差は，長いあいだ，性ホルモンの作用だと考えられてきた．しかしながら，筆者らが男性ホルモン受容体の遺伝子をノックアウトしたマウスを作製してみると，外見はほぼメス型になるにもかかわらず，骨長は依然としてメス型にはならなかった（松本ら：未発表データ）．ヒト，マウスをとわず，身長はオスのほうが高い．その要因は身長を規定する骨の成長に起因する．そのため，男性ホルモンがまったく作用しない男性ホルモン受容体ノックアウトマウスにおいて，依然としてオスの骨長がメス型よりも長い（実際には，メスとオスの中間型）のは，Y染色体にあるなんらかの遺伝子の機能によるものと推察された．実際，Y染色体の一部の欠失により低身長をきたす家系が見いだされているが，その原因遺伝子はいまだ同定されていない．すなわち，性差は従来，性ホルモンの作用の違いにのみ起因するものと考えられてきたが，性染色体，とくに，Y染色体にコードされる遺伝子の機能の重要性が明らかになりつつあるといってよいだろう．したがって，ある因子の機能を調べてこれを議論する場合に，同じ部位の骨量を測定していても，性差は無視できない要因なのである．

おわりに：微妙なデータを読む秘訣

微妙な実験データを正しく評価する秘訣は，決してむずかしいものではない．まず第1には，測定の原理と生物現象の意味を正しくそして統合的に理解することである．この点は，ある程度まで成書に依存することができるだろう．しかしながら，成書から得られるものだけでは足りない，第2の秘訣も必要となる．それは，人とのコミュニケーションの能力である．科学者が多くの学会に参加するのは，自分の実験データを公表したり，他人の実験データを聴きにいったりするだけのためではない．講演の合間のちょっとした会話に，決定的な情報を得ることが多いのである．すなわち，研究の動向だけではなく，技術や成果の真偽についての個人的な意見を聞くことができたりもするのである．主要な雑誌に公表された研究成果は大きな影響を及ぼすが，ときおり，間違った解釈や実験アプローチにより恣意的にねじまげられ再現性が得られなかったりすることもある．このような情報をいち早く知ることは，研究を立案するうえで計り知れない価値がある．同様に，実際にベンチで実験している研究者にとっては，学会に行き同様の実験を行なっている研究者と議論することはきわめて大切である．プロトコールに書くことがむずかしい実験のコツや秘訣などについて，技法の談議をすることができる．このことは，何も異なる研究グループのあいだだけでの話ではなく，同じ研究室や研究グループ内でも同じことである．日常会話の多いグループは，ふだんから技術論についての議論もさかんである．筆者の所属する大学で起こった不幸な論文捏造事件の報告書を読むと，研究グループ内での議論に乏しく，実験データに対する批判や評価が当事者どうし以外にはなされなかったようであり，これは筆者にとって驚きであった．ふだんから会話を心がけるだけで，実験データ解釈の鍛練になるのである．よい生命科学者は，一般に，人とのコミュニケーション能力が高い．

加藤茂明

略歴： 1988年 東京大学大学院農学研究科博士課程 修了，同年 東京農業大学農学部 助手，1989年 同 講師，1992年 同 助教授，1996年 東京大学分子細胞生物学研究所 助教授を経て，1998年より同 教授，2004年より科学技術振興機構ERATO加藤核内複合体プロジェクト 研究総括 兼務．