Trends in Developmental Biology
Vol. 11, 2018

Review

Incomplete penetrance and variable expressivity of congenital disease: what's luck got to do with it?

Donald F. Newgreen[1,*], James M. Osborne[2], Benjamin J. Binder[3], Bevan L. Cheeseman[4] and Kerry A. Landman[2]

[1]The Murdoch Children's Research Institute, The Royal Children's Hospital, Parkville, Victoria 3052;
[2]School of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria 3010;
[3]School of Mathematical Sciences, University of Adelaide, SA 5005, Australia.
[4]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.

ABSTRACT

Genetic birth defects often show seemingly random variability of occurrence or of severity between individuals. These are termed incomplete penetrance and variable expressivity, respectively. Variability occurs even within monozygotic twins, where most sources of genetic variation are absent. This poses problems clinically and has been difficult to explain. Archetypal examples of unpredictable variability occur in the development of the enteric nervous system (ENS). In Hirschsprung disease the distal ENS fails to form, and in slow-transit constipation the distal ENS is present but is dysfunctional. We have developed mathematical cellular automata models of ENS formation and unpredictable but distally-amplified variabilities resembling (Hirscshprung) or allowing (slow-transit constipation) these conditions emerged spontaneously. These results are *in silico* examples of incomplete penetrance and variable expressivity. These differences in outcome at the level of the entire system between replications with identical initial conditions and identical rules of behaviour for ENS agents (cells) were driven by stochastic choice of local motile and proliferative behaviour of each of the many agents, iterated over many of cycles. Surprisingly, given the number of agents and iterations, the entire system did not 'balance out' to end identically. These examples show in detail how the inability to totally determine specific aspects of cells' morphogenetic behaviour can result in incomplete penetrance and variable expressivity in birth defects. Transposed into biological systems, these mechanisms therefore impose limits on disease predictability.

KEYWORDS: enteric nervous system, cellular automata models, Hirschsprung disease, slow-transit constipation, incomplete penetrance.

1. Introduction

1.1. Variability of congenital traits: a long-standing puzzle

Variability of congenital traits, especially of disease traits, is common yet often unpredictable to a degree described as stochastic. Even with gene alleles classically described as 'dominant', discordance of genotype and phenotype occurs. This genotype/ phenotype mismatch was named incomplete (or variable) penetrance if, despite the dominant allele being present, the phenotype was absent in some individuals but present in others. Where the phenotype occurred but varied in degree between individuals, this was termed variable expressivity [1]. (Here we will refer to both as incomplete penetrance). To add to this, discordancy of phenotype was observed between twins [2] and, in laterally paired structures in the one individual, between left and right [3]. Thus, although variability such as incomplete penetrance was named and discussed, its cause was mysterious. More practically, when

*Corresponding author: don.newgreen@mcri.edu.au

this unpredictability concerned a congenital disease, this clinical uncertainty posed difficult problems for treatment and prognosis, especially now for genetic counsellors [4].

Because of this clinical importance, virtually every genetic, epigenetic and environmental variable has been postulated as determining this variability [5]. The technical advances of molecular genetics and epigenetics are revolutionising health sciences. In the age of 'big data' and 'personalized' and 'precision' medicine', there is a sanguine hope that all or very nearly all determinants will become accountable even at the level of the individual patient, and this will resolve all or much of the current clinical uncertainty and unpredictability posed by the old problem of incomplete penetrance in congenital diseases.

1.1.1. Genetic and environmental differences contribute to incomplete penetrance

Differences in alleles of modifier genes contribute to differences in penetrance of identical mutant genes between different inbred mouse strains [6]. In a genetically diverse population such as humans, different alleles of modifier genes certainly contribute to variability because these alleles differ between individuals who share the same 'disease-causing' mutation. Exposure of individuals to different environmental conditions pre-natally is also an important contributor to differences in penetrance of congenital diseases [7]. Although often complex, these genetic and environmental effects are in principle determinative, clearly account for much phenotypic discordance and with sufficient information they could be predictive. However there are instances where it is much more difficult to invoke these as causes of variability between individuals.

1.1.2. Incomplete penetrance persists when genetic and environmental differences are minimised

In animal models heroic efforts have been made to achieve identity in genotype and in pre- and post-natal environments yet phenotypic variability stubbornly persists [8]. In humans, monozygotic (MZ or identical) twins share a genetic background and inhabit one uterus, but even when exhibiting birth defects, their phenotype is frequently discordant [9-11]. Of course differential microenvironmental effects cannot be entirely dismissed even in MZ twins [12], but these tend to result in gross differences such as size and weight variation rather than in presence, absence or degree of specific birth defects. Discordancies between neonatal MZ twins are a lens to focus on the missing elements of variability in phenotype and therefore on incomplete penetrance of developmental disease whenever it occurs [13].

1.1.3. Where are the missing determinants of developmental discordancy?

Discordance of phenotype might be based on changes involving genes without change of DNA sequence. Arguments can be made for variation between individuals, even between MZ twins, based on epigenetic discordancy and miRNA functions [12, 14-16]. In addition, expression levels of the same gene may show random fluctuations between cells. If in early embryogenesis these straddle a notional threshold level for attainment of some development milestone then intra-individual variability may result stochastically despite genetic identity [17]. However, these are unlikely to account for the frequency of trait-variations notably between genetically near-identical individuals such as MZ twins.

1.1.4. Are somatic mutations a source of developmental discordancy?

Another potential source of variation is post-zygotic somatic mutations, and these fulfil the qualities of originating stochastically and differing between MZ twins [18, 19]. Each somatic mutation occurs in a single cell and is copied in all clonally descendant cells, creating clonal mosaicism within tissues of one individual. Already several discordancies in MZ twins have been ascribed to somatic mutations [12]. Recently the importance has been realised of somatic mutations in functional brain diseases [20, 21].

Despite these examples, the general view has been that phenotypic effects of somatic mutations are not common. This lack of effect is ascribed to cells of each mutated clone being narrowly distributed and at a very low proportion in the entire cell population. Thus mutant cells, while individually having a phenotype, are below the threshold level to cause an overall disease [22]. By this reasoning, if somatic mutations are rarely phenotypically important in the one individual, they cannot commonly account for discordancy of birth defects between MZ twins.

The major exception to this phenotypic incapacity of somatic mutations is the rare occurrence when the mutation occurs at or before the blastocyst stage, when there are relatively few cells. This could permit large clonal contributions capable of exceeding a phenotypic threshold. In such cases the mosaicism would typically extend across derivatives of several germ layers [23]. Thus, early mutations affecting for example nervous tissue (ectoderm germ layer origin) could be identified by mutation analysis of readily obtainable cells such as lymphocytes (mesoderm germ layer origin). Multi-germ layer mosaicism does not suffice to explain the frequency of discordancy in MZ twins and most searches for differential somatic mutations in discordant MZ twins have been fruitless [24-27].

1.1.5. Could variation in cell behaviour contribute to developmental discordancy?

Discordance of phenotype could notionally be based on stochastic changes at a quite different level, involving variations in behaviours of normal cells rather than mutational and microenvironmental variations. The cellular and extracellular microenvironment of metazoan tissues in development is enormously and dynamically complex and the repertoire of cell behaviours -migration, proliferation, differentiation, apoptosis etc.- is diverse and reflects continually changing interactions with this microenvironment. For this reason, in time-lapse observations of cells *in situ*, it is impossible to predict exactly what each individual cell will do even minutes into the future [28]. However even if cells do not behave in exactly predictable ways, they do behave in broadly similar ways. Given that in tissues there are very large numbers of cells, the behaviour of each cell population is a time-average of the iterated behaviours of all the constituent cells. This would mean that the overall development of phenotype would be highly invariant and predictable from individual to individual. If this were not so, metazoan development would devolve into chaos. Viewed this way, this cell-scale detailed stochastic behavioural variation seems not to be a promising candidate for phenotypic discordancy between individuals such as MZ twins.

All these potential variations do not seem to be strong candidates to account for the frequency and spectrum of discordant birth defects in general or in MZ twins. The high frequency of phenotypic variability between MZ twins with a birth defect suggests that there is are important and still missing pieces contributing to this variability.

2. The enteric nervous system as a model to study phenotypic variability

Here, we focus on one important developmental system, the neural crest (NC)-derived enteric nervous system (ENS) and its developmental abnormalities (collectively termed neurocristopathies) [29, 30]. Using as models one structural and one functional enteric neurocristopthy, both highly and unpredictably variable, we describe novel and feasible mechanisms that could produce the inherent trait variability in these diseases. Both mechanisms depend on credible cell behavioural events that are stochastic in the true sense and these therefore impose limits to predictability. We also emphasise the utility of biomathematical modelling for discovery of such non-intuitive and emergent processes.

2.1. NC derivatives are models to focus on variable congenital disease

The NC system provides examples of all morphogenetic and differentiation events in metazoan development [29]. The NC originates in the neural epithelium, the precursor of the CNS, and contributes a vast range of cells and tissues throughout the body of vertebrates and is involved in a wide range of neurocristopathies affecting many tissues. Neurocristopathies form a model for complex congenital diseases and often show incomplete penetrance [10, 31, 32]. Of all the NC-derived systems, the easiest to understand in broad developmental outline is the enteric nervous system (ENS).

2.2. ENS development and its variable birth defects

The ENS occurs throughout the gastro-intestinal tract, and is the largest and most complex division of the autonomic nervous system. The ENS consists of a vast number of small group of neural cells termed ganglia arranged in two layers in the wall of the intestine. There are at least 15 types of neurons and several types of attendant glia in the ganglia of the ENS [33]. The ENS controls gut muscular activity including peristalsis, as well as local blood supply and electrolyte and water balance across the mucosa.

In early development, the progenitors of the ENS migrate from the NC in the neural epithelium, with the largest number from a small hindbrain or vagal level [34]. Broadly speaking, most of these NC cells colonise the gut mesoderm in an oral-to-anal wave (Fig. 1A). The migration of these cells is in an almost 2D layer in the gut mesoderm, and involves high proliferation of cells [35] (Fig. 1B). The progeny of these migratory NC-derived cells then differentiate into all the neurons and glial cells of the ENS.

Of the many ENS birth defects [36] the easiest to describe and one of the most serious is Hirschsprung disease or colonic aganglionosis. In this disease the ganglia of the ENS are missing from a variable extent of the anal end of the colon. It can be grasped intuitively that the oral-to-anal colonisation by NC cells could, if not completed, cause Hirschsprung disease (Fig. 1A inset). This disease is established over the first 7 weeks of human gestation, the period of intestinal colonisation. The genetics of Hirschsprung disease is complex [37] and it serves as a paradigm for multigene, multifactorial developmental diseases. The most common identified cause of Hirschsprung disease in humans is haploinsufficiency due to a loss-of-function mutation in one copy of the gene *RET*, which encodes a growth factor receptor in ENS cells. In this case with identified mutations, penetrance is around 50-70% [38] with highly variable expressivity which is easily quantified as the length of colon lacking ganglia. In affected MZ twins, the disease is more often discordant than concordant [39, 40]. This disease can also occur as part of a syndrome, and in such MZ twins, the Hirschsprung phenotype may be discordant while other elements of the syndrome are concordant [41]. Hirschsprung disease is a very clear example of a serious genetically produced structural birth defect that displays incomplete penetrance and variable expressivity.

Other more subtle ENS motor neuropathologies occur [42]. In slow-transit constipation, in contrast to Hirschsprung disease, the ENS is present throughout the gastrointestinal tract but the peristaltic functions it controls are, as the name suggests, disturbed. This disturbance chiefly involves the distal bowel [43],

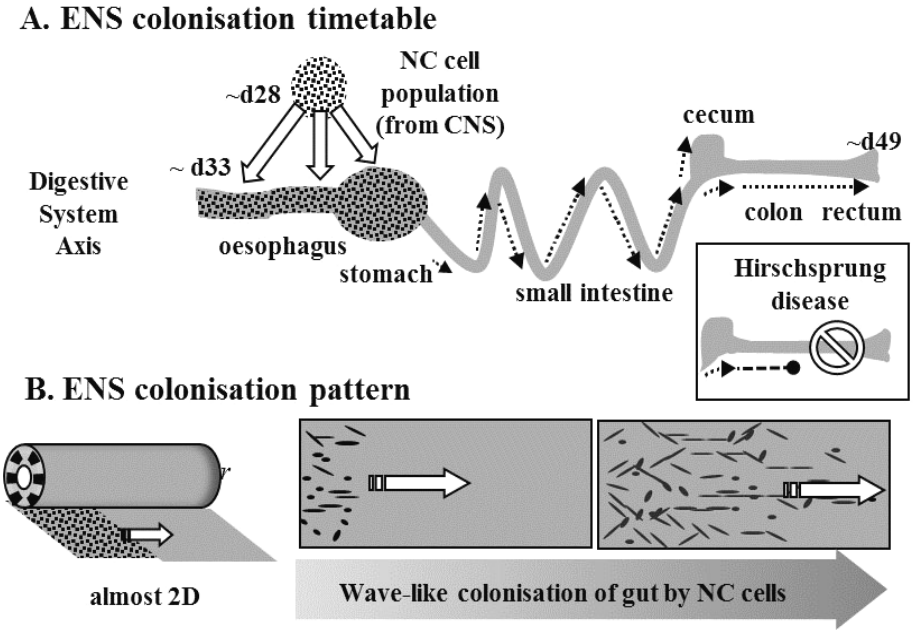

Fig. 1. Scheme of ENS formation. **A**. The ENS forms as NC cells (black dots) migrate from the vagal level of the CNS to the digestive tract (grey) which they colonise progressively in an oral-to-anal direction. In the human embryo this occurs from about post-fertilization day d28 to d49. In Hirschsprung disease the colonisation is not completed. **B**. The pattern of ENS colonisation is by occupation of an almost 2D layer of gut mesoderm in an oral-to-anal wave. (Adapted from ref. 63).

but the severity and extent of the functional disturbance varies widely between individuals. Causes of this disease are still being sought, with inconsistent data in paediatric and adult forms [44-46].

3. Mathematical models of ENS development

While developing mathematical models of development of the ENS [47], outcomes resembling or permitting birth defects emerged spontaneously under certain conditions and, despite identical inputs, these were stochastically variable. We will describe how these stochastic variations come about, and argue that these are not just mathematic abstractions, but are strong candidates for much of the 'missing variability' in phenotype seen in development and especially in developmental disease.

3.1. The mathematical model

For ease of visual comparison with real biology, we use agent-based cellular automata (CA) models [48] of ENS development [49-52]. In the following section we briefly describe the basic model in order to explain later how different outcomes emerge from different individual simulations of the identical model.

3.1.1. The substrate of colonisation

The ENS colonisation substrate, the gut mesoderm tube, can be approximated by rolling out the gut cylinder into a 2D sheet (Fig. 1B). In order to use a 2D CA model this sheet is gridded, each grid site representing a gut mesoderm cell unit; for simplicity we use a square grid (Fig. 2). We define

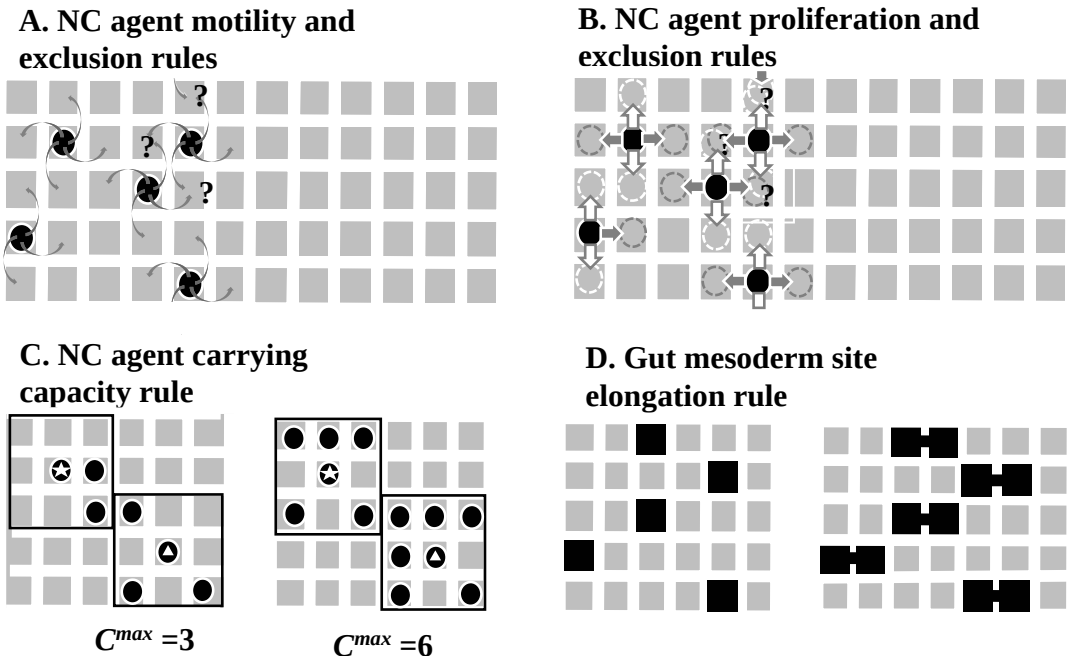

Fig. 2. CA model rules for colonisation by NC agents (black) of gut mesoderm sites (grey). **A**. At each motility cycle at probability P_m the NC agent moves one site east, west, north or south, chosen randomly. An attempt to co-occupy an occupied site (event marked by ?) results in that agent's motility cycle being aborted. **B.** For proliferation, at each cycle at probability P_p the NC agent can produce two agents placed east+west (dark grey) or north+south (white) of the original site. Attempts at co-placement (?) results in the division being aborted. **C**. A chosen NC agent may proliferate only if the number of other NC agents in its surrounding Moore neighbourhood (black square) is less than the local carrying capacity C^{max}. For low carrying capacity setting, say C^{max} = 3, proliferation is permitted (star) because C = 2, but only north+south (see Rule 2B) and disallowed (triangle) where C = 3. For higher carrying capacity setting, say C^{max} = 6, proliferation is permitted (star) where C=5, but only east+west (see Rule 2B) and disallowed (triangle) where C = 6. **D**. For gut mesoderm elongation, one site is chosen (black square) at random per row at probability P_g and duplicated east or west at random, so increasing the length of the gut one column per cycle.

the oral-anal gut length as the west-east axis and distribute NC cells (in the model termed NC agents) on grid sites at the west (i.e. oral) end of the sheet. Because the sheet represents a cylinder, the north and south margins have periodic boundary conditions whereby a NC agent moving off the south margin then appears on the north margin and *vice versa*.

3.1.2. Rules for colonisation

Based on observations of ENS development, we initially imposed a few simple rules for the motility and proliferation and for co-placement of the NC agents (Fig. 2A-C). We later imposed a rule for mesoderm site proliferation (Fig. 2D). Motility and proliferation cycles of each agent were enacted sequentially and updated.

i) In the gut of laboratory embryos, NC cell populations advance predictably and individual cells move about four cell diameters (taken as ~40 μm) per hour on average [53, 54]. To approximate this we dictated that each NC agent on a 2D mesodermal grid could move to its adjacent grid square (=one cell diameter, about 10 μm) at a probability P_m corresponding to one movement cycle per time base equivalent to ¼ hour. In time-lapse movies, isolated NC cells in the gut move unpredictably as a random walk [28]. To model this, NC agent movement direction is generated stochastically, north, south, east or west with equal probability for every movement cycle (Fig. 2A). Diagonal movements are not permitted for computational convenience.

ii) In animal models the ENS cells proliferate [35], and cell division occurs throughout the population [55], but a) the exact location of each division is unpredictable and b) the orientation of each cell division is unpredictable. In the CA model we dictate that a NC agent can divide into two daughter agents at an average proliferation rate P_p. We include different rates of division over a range corresponding to 10 to 100 times slower than that of the movement cycle i.e. average division cycle duration equivalents as short as 2.5 h up to as long as 25 h. To model the individual locational unpredictability, the precise NC agent to divide in each cycle is assigned stochastically. To model mitotic orientation unpredictability, the two daughter agents are placed stochastically either north+south or east+west of the original NC agent site (Fig. 2B).

iii) Notwithstanding these rules, we also established an exclusion rule to prevent co-placement. If any NC agent attempts to move to a grid site already occupied by another NC agent (indicated by '?' in Fig. 2A), then that movement cycle is aborted. Likewise, if there is an attempt to place a daughter agent in an occupied site, then that proliferation cycle is aborted (Fig. 2B). This rule imposes restrictions and hence biases the outcomes of the previous rules to favour movement into areas previously free of NC agents.

iv) In animal tissues the final ENS cell density comes to a constant value [53] which is dictated by the gut tissue. A situation like this usually arises from logistical competition for space, nutrient or, often, growth factors. We termed this value the carrying capacity (C^{max}), and to represent this we dictate that division of a chosen NC agent is suppressed if the number of NC agents in its local neighbourhood (termed the Moore neighbourhood; the 8 gut grid sites around the chosen NC agent) equals or exceeds a number representing the local C^{max}. The setting of C^{max} is altered to represent a range of agent densities (Fig. 2C).

3.2. Predictive successes of this model

NC agents were loaded at below carrying capacity C^{max} onto the west (oral) end of the mesodermal gridded sheet and the model was allowed to run. This produces a rise up to C^{max} without advance, then when C^{max} is attained, an eastward (anally) directed travelling wave of NC agent colonisation is generated (Fig. 3A). This closely matches the real situation where there is a 'loading-up time' in the most oral foregut before the wave of individually meandering NC cells advances anally. In addition, by marking different phalanxes of NC agents and their daughters, the model reveals that most of the colonisation is generated from the front agents, a process we called 'frontal expansion' (Fig. 3B) [56]. This contrasts with the previous intuitive notion based on NC cell density termed 'population pressure' [57, 58] which involved NC cell population shunting from behind. Cell labelling experiments in avian, mouse and fish models subsequently confirmed this mathematical prediction of 'frontal expansion'. These results and the prediction of many other emergent properties [56] bespeak of the superiority of mathematical modelling over biological intuition, and are consistent with our model having captured key elements of ENS formation.

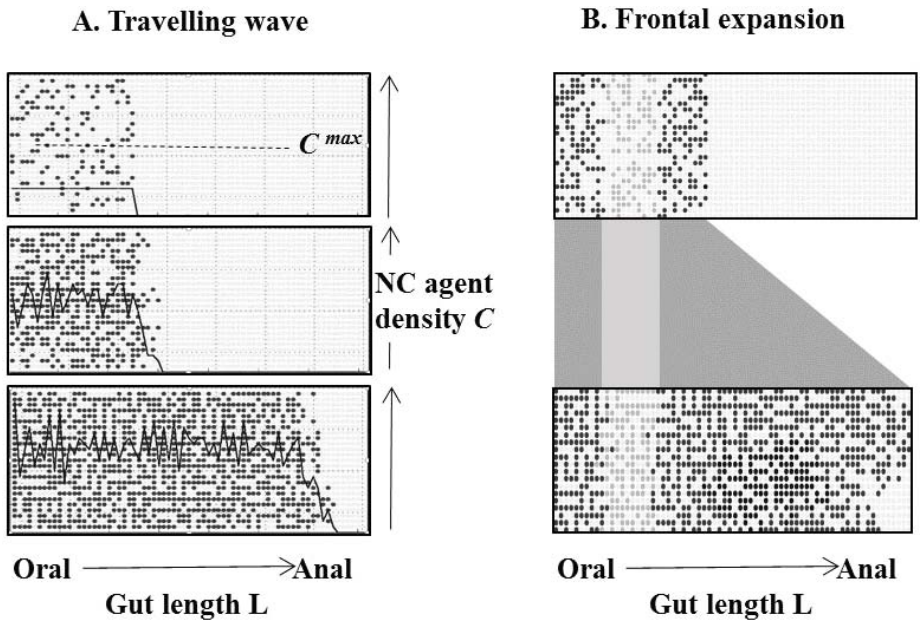

Fig. 3. Simulation of ENS colonisation by the rules described in Fig. 2. **A**. NC agents (dots) distributed at the oral end of a gut field proliferate to the density of the carrying capacity C^{max} then an oral-to-anal travelling wave is generated which colonises the entire gut field. **B**. Marking bands of NC agents reveals that most colonisation is due to proliferative advance of the front phalanx of agents, a process termed 'frontal expansion'. (Adapted from ref. 56).

4. Hirschsprung outcomes emerge stochastically in models of ENS development

4.1. Emergence of Hirschsprung outcome with incomplete penetrance

Many of the 'Hirschsprung genes' affect NC cell proliferation [58] and on mutation, proliferation would be reduced. However the initial CA model had a conspicuous failure: no reasonable decrease of proliferation probability P_p (or motility P_m) to resemble loss-of-function of *RET* produced a Hirschsprung-like result. ENS agent colonisation was always completed, just more or less rapidly. However the model failed to include the real occurrence of simultaneous elongation of the gut [57]. Based on measurements of gut growth [52], in the model we allowed the gut mesoderm grid sites to divide at probability P_g. Particular sites, one in each row, are chosen at random to produce two daughter-sites which are placed adjacently in the west-east (oral-anal) axis. This produces a lengthening of the gut domain (Fig. 2D). When this is included along with NC agent motility and proliferation the 'frontal expansion' travelling wave and other realistic features are not altered. Now, however, full colonisation by NC agents pertains when P_p is high and Hirschsprung-like failure of full colonisation emerges when proliferation probability P_p is low (Fig. 4A).

The real surprise emerges when running identical intermediate P_p-value simulations repeatedly; these produce some outcomes with full colonisation, and some Hirschsprung-like partial colonisation outcomes (Fig. 4B) [50]. This occurs probabilistically and sweeping the value of P_p in this intermediate range raises (or lowers) the probability of colonisation success by NC agents, that is, the degree of penetrance is altered (Fig. 4B). In those simulations where full colonisation is not obtained, the length of the uncolonised zone varies unpredictably between individual simulations but is greater on average at lower settings of P_p. Thus random variation in expressivity also emerges from the model [59].

In this mathematical system all variables are determined except for those that are explicitly allowed a stochastic choice at the level of individual agents. Thus, it is certain that the variable failure to complete colonisation at the east (anal) end -the Hirschsprung phenotype - is stochastic in the true sense.

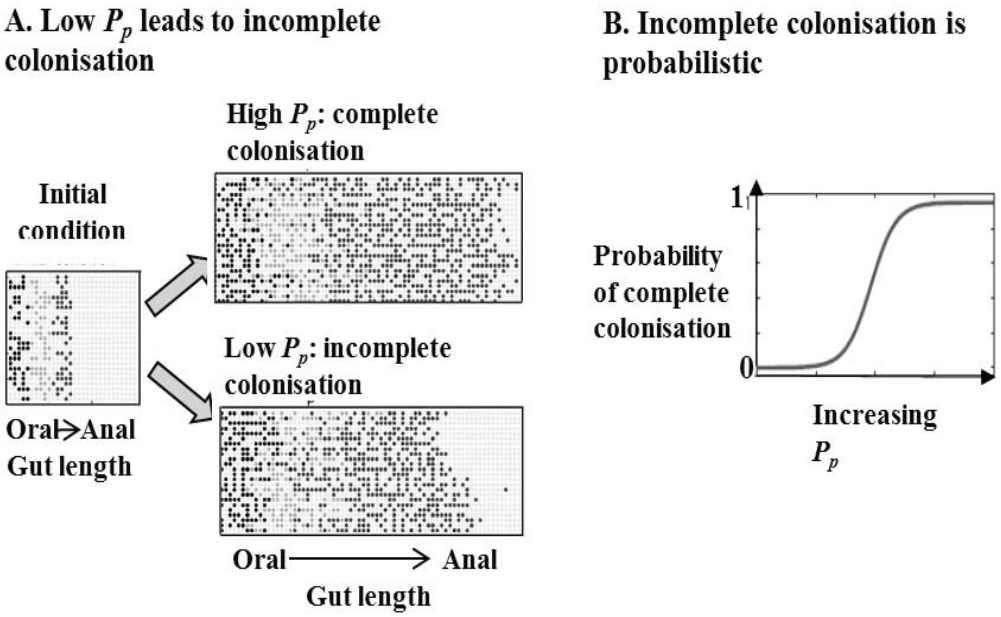

Fig. 4. Emergence of stochastically variable Hirschsprung-like incomplete colonisation when the gut mesoderm elongates (see Fig. 2D). **A**. NC agents with high proliferation probability P_p fully colonise the growing gut but fail to colonise the anal end at low P_p. **B**. At intermediate values of P_p colonisation completion is probabilistic. (Adapted from ref 50).

4.2. How do stochastic variations in cell behaviour result in a Hirschsprung outcome with incomplete penetrance?

In the CA model all NC agents are intrinsically identical and variations in behaviour are initiated at the level of individual agents, due solely to stochastic differences in their directional movement and the positioning of their daughter agents. These are enacted entirely locally and independently in space, and reset independently at each motility and division cycle. This reflects the unpredictable detailed behaviour of cells observed *in situ* with time-lapse imaging [28]. There is no possibility in this *milieu* of the future behaviours of individual cells being determined precisely. Intuitively, it might be thought that with large numbers of agents, and we modelled up to 150,000 in some simulations [35], these agent-to-agent, cycle-to-cycle variations would always even out at the level of the entire agent-population. This would parallel the broad predictability of cell populations in normal development.

Surprisingly, the detailed analyses permitted by these models show that outcomes at the system level do not entirely even out, and are always variable between simulations in which all conditions able to be determined are identical. However in most normal conditions these variations are invisible, the system having a large safety margin. In the developing ENS for example, experimentally reducing the number of ENS progenitor cells reveals a large 'reserve capacity' which allows full colonisation [35] and this masks any underlying variability. However this reserve capacity has limits, and when sufficiently eroded, for example by reducing agent proliferation P_p in the model, a border condition is reached where the stochastic variability of proliferation and motility at the level of single NC agents can make variable differences in outcome at the system level, in this case between the success of colonisation or Hirschsprung-like failure [59]. This means that this variable phenotype is only exhibited in patients with a pre-existing defect such as mutation of the *RET* gene. Consistent with this, in humans, loss-of-function mutation of one copy of the *RET* gene causes Hirschsprung disease with 50-70% penetrance [38]. As predicted by further reduction of P_p in the CA model, when both copies of *RET* are mutated, aganglionosis becomes 100% penetrant and expressivity (ie. length of intestine affected) is increased [60]. An identical

result occurs in mice with loss of both copies of *Ret* [61]. However mice seem to be quite different to humans when one copy of *Ret* is mutated: this never results in a murine Hirschsprung phenotype. When this is explored further using mice genetically engineered to drive *Ret* expression and Ret protein levels to about 30% of normal (i.e. significantly lower than that achieved by mutation of one copy of *Ret*) an incompletely penetrant Hirschsprung phenotype emerges [62]. The CA model can reproduce this difference by giving 'mice' and 'human' agents different P_p settings.

5. Slow-transit constipation-enabling conditions emerge stochastically in models of ENS development

This is a two-step process, one step explained below involves stochastic generation of clonal dominance, and the second step is the universal stochasticism of occurrence of somatic mutations in all cells.

5.1. Emergence of distally-amplified stochastic clonal dominance

An advantage of the CA model is that every agent can be tracked, even if there are thousands. We followed NC agents individually to study their clonal evolution, expecting that clones would be reasonably similar in size. The counter-intuitive outcome is that most clones remain small but a few clones, which we termed 'superstars', become huge with disproportionately large contributions to the entire final population of NC agents (Fig. 5A). Although few in each starting population, 'superstars' occur in every simulation (Fig. 5B). Increasing the probability of proliferation P_p, which reduces the cycle time, increases the disproportionality. This was achieved by chance [63]; which agent would attain this clonal dominance is not predictable although there is a probabilistic bias favouring NC agents originally at or near the front. In addition, as colonisation proceeds, the more east (i.e. anal)

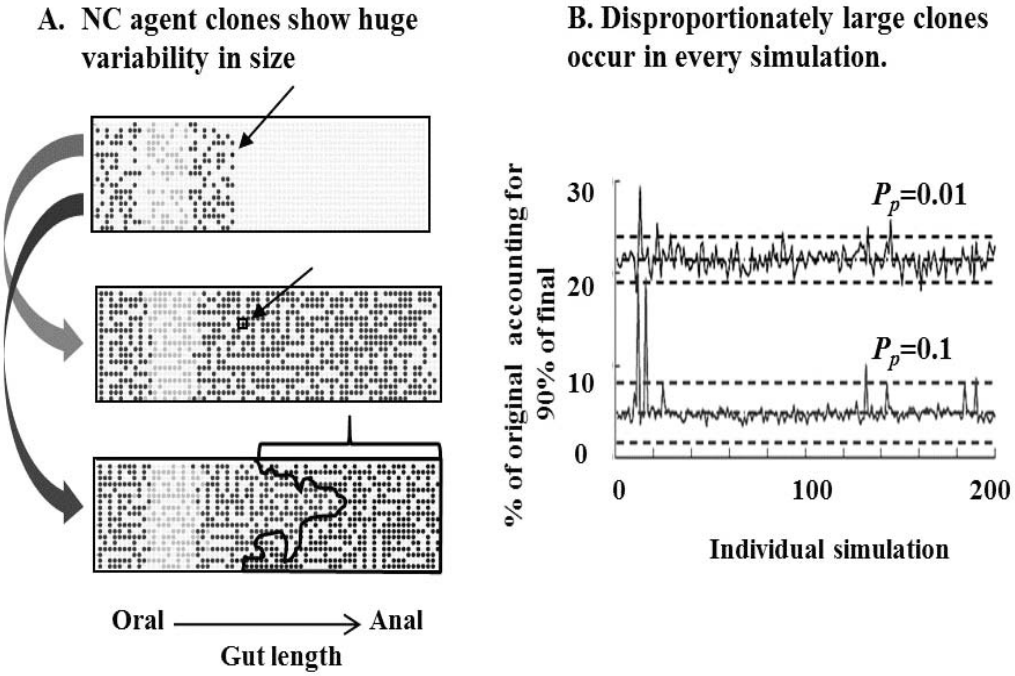

Fig. 5. Stochastic emergence of 'superstar' clones of NC agents leads to clonal dominance in CA models. **A**. Single NC agents tagged (arrow, top image) to follow clonal descendants in a colonising population mostly form small clones (arrow, middle image). A few clones become huge and dominate the colonisation wave (bottom image, bracketed zone). These simulations are in a non-growing gut domain. **B**. Plotting the % of the initiating NC agent population that produces 90% of the final population confirms that this disparity of clone size is found in every simulation (here 200 consecutive simulations) and reveals that the disparity is increased at higher proliferation probability P_p. (A adapted from ref 50, B from ref. 64).

the position the more extreme the reduction in clonal diversity [64].

Normal distributions of frequencies are often seen in biological populations, and hence this fat-tailed distribution of clone sizes was unexpected. Therefore confirmation was sought in a biological ENS system by tracking one avian NC cell with a genome-integrated fluorescent label which identified every daughter cell. This labelled cell was mixed with a much larger population (in this case 600 cells) of unlabelled but otherwise similar NC cells; we termed this 'cloning-in-a-crowd' (Fig. 6A). The clonal growth of the labelled NC cell was measured by counting the labelled cells after the avian gut in organ culture grafts had been colonised by the NC cells. The results of the CA model (Fig. 6B upper panel) and of the experiments (Fig. 6B lower panel) were similar in that most clones were small (less than 0.1% of the entire final NC cell population) but a few were huge (more than 10% of the final NC cell population) [35, 64].

Agreement between the biological result and the CA model do not prove that the biological 'superstars' are of stochastic origin, as they are in the model. For example, it is possible that just a few cells were pre-determined to form huge clones, similarly to cells in a carcinoma with activation of oncogenes determining clonal dominance. However, if the 'superstar' NC cells were pre-determined, reduction in the number of unlabelled NC cells (ie. 'cloning-in-a-smaller-crowd') would not alter the probability of superstardom occurring in the

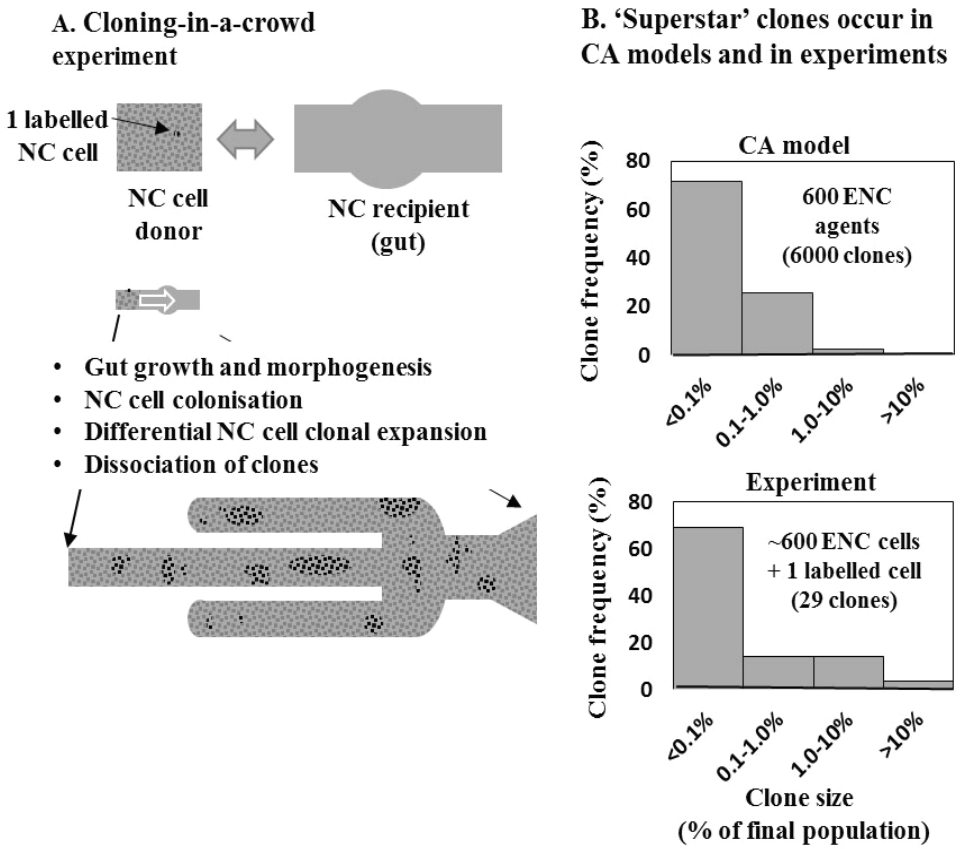

Fig. 6. NC cell clone size frequency in experiments is similar to that in CA models. **A**. A labelled avian NC cell (black dot) with many unlabelled NC cells (dark grey dots) is co-cultured (double-headed arrow) with uncolonised recipient avian intestine (grey). Later the entire intestine, after extensive growth (arrows), is colonised by NC cells. The number of NC cells derived from the single labelled cell is counted. **B**. CA models and experiments show similar frequency versus clone size distributions, with the ENS formed by numerous small clones and a few 'superstar' clones. (A adapted from ref. 64, B from ref. 35).

one labelled cell. However if the 'superstars' were not pre-determined this would increase the probability that any one labelled cell would develop as a 'superstar'. We performed this experiment by reducing the unlabelled NC cells from around 600 to about 40 and found a strong increase in the probability that any one labelled cell was a 'superstar' [35]. The conclusion from this is that the 'superstar' potential is not pre-determined and is therefore likely to be stochastic as predicted by the model.

5.2. How do stochastic variations in cell behaviour result in clonal dominance?

'Superstar' clones occur when an advantageous proliferative outcome of an agent in an early cycle is achieved by any means and this biases the probability of an advantageous outcome for this agent's descendants in the next cycle. Thus, a system of spiralling cumulative advantage is established. This is known as the Matthew Principle [65]. This is widely operative in biology, as it is in social and economic networks. In the ENS system modelled here, the seeding advantage is stochastic but biased; agents at or near the front have increased likelihood of a favourable immediate outcome because they will be less likely to be impeded in proliferative opportunities by their neighbours, since they have no competing neighbours on one side. This unequal increase in agent-number improves the chance that the next advantageous outcome will occur in an agent of the same clone, leading to escalating clonal amplification. The cloning-in-a-crowd and cloning-in-a-smaller crowd experiments strongly support the occurrence of this mechanism in the ENS in real intestinal tissues.

The emergence in CA models of a relatively few numerically dominant 'superstar' clones in the context of a constant total number of agents (dictated by C^{max}) has the inevitable outcome of reducing clonal diversity. Whether this results in reduced clonal diversity in the human ENS as it does in animal models is a key unanswered question, but the high degree of conservation between species in ENS development [66, 67] suggests that this is likely.

5.3. Clonal dominance may allow somatic mutations to cause ENS dysfunction

Reduced clonal diversity of itself would have no phenotypic effect if all the NC-derived cells of all the clones were and remained identical. However, cells never remain identical due to somatic mutation, itself a stochastic process. Brain cells and fibroblast cells each acquire many somatic mutations although estimates of the mutational burden vary [68, 69]; it is certain that ENS cells also accumulate numerous somatic mutations. As mentioned previously, the argument against the importance of such mutations in generating a phenotype is that there are low levels of mosaicism unless the mutations occur at or before the brief blastocyst stage [22]. However the mode of development of the ENS may make it even more susceptible than the CNS [20, 21] to somatic mutations attaining phenotypic levels. First the likelihood of occurrence of such a mutation is increased because the period in which the mutation might occur is extended throughout the period of ENS generation rather than being restricted to the short blastocyst period. Second the ENS, a very large cell population, is produced mostly from a very small vagal NC population [35] and especially in 'superstar' clones, requires many cell division cycles [70]. As in cancer generation [71] this increases the chance of occurrence of a mutation, which mostly occurs during each cycle of DNA synthesis. Third the mode of ENS colonisation described by CA models actively drives clonal dominance and hence increases the potential for achieving high level mutational mosaicism. Fourth, the mode of ENS formation described here maximises the reduction of clonal diversity locally in the distal segment of the intestine. Thus even if the average level of the mutation in the ENS cell population over the entire intestine is low, it will be elevated in the distal segment of the colon. ENS failure even in only a small segment of bowel can totally disturb overall function because of the linear nature of the intestinal tract.

The genes and the types of somatic mutations acting in the context of clonal dominance are important. Several studies have investigated somatic mutations in Hirschsprung disease [72-74] but these face difficulties in interpretation because the cell populations sampled were not restricted to ENS cells. The CA model logic predicts that a gene and mutation causing a non-structural enteric neuropathy such as slow-transit constipation cannot be of the type that causes familial Hirschsprung disease (e.g. loss-of- function mutation in *RET)*. This is because

such mutations impair ENS cell proliferation or motility during colonisation. In such cases the model predicts that the clonally affected ENS cells would be out-competed by their non-affected neighbours, thereby preventing clonal dominance and resulting in no phenotype. In contrast a gain-of-function mutation in a 'Hirschsprung gene' could confer a deterministic clonal advantage; such mutations in *RET* occur in Multiple Endocrine Neoplasia-2B and can cause ENS hyperganglionation. This type of mutation can occur somatically [75]. However, from the model, a mutation that is functionally neutral at the colonisation phase can have phenotypic consequences if it affects later nervous function, such as altering neurochemical synthesis, connectivity, synapse formation, excitability etc in post-migratory differentiated ENS cells. This is in accord with the identification of enteric nerve fibre neurotransmitter abnormalities in biopsies of colon from children with slow-transit constipation [44].

The 'superstar' phenomenon therefore underpins a means to achieve a mosaicism threshold sufficient to affect ENS function. Moreover, the model predicts that this phenotype would be manifest only or mainly in the distal gut because there, clonal diversity is predicted to be least. This fits the locational profile of ENS functional disturbances such as slow-transit constipation, and the stochastic nature of somatic mutations is consistent with the variability of such effects.

6. Is there evidence for these mechanisms in patients?

Purely stochastic effects are notoriously difficult to demonstrate since by definition they have no material cause by which to be identified. In the case of Hirschsprung disease a conclusion of causation by stochasticity of cell behaviour, as in the CA model, may be reached because it is 'the last conclusion standing' after eliminating other potential deterministic causes. This is consistent with the level of Hirschsprung disease discordancy in MZ twins [39, 40] being comparable to its level of incomplete penetrance in non-identical populations [38], despite major sources of consistent genetic variability being eliminated in the twins.

For non-structural ENS defects like slow-transit constipation, direct evidence is currently lacking in humans but the underlying clonal dominance has been shown recently in experimental animal systems [64] and is likely in humans as ENS developmental mechanisms are highly conserved. Given the universality of somatic mutations, these must also occur in the cells of the human ENS. Direct assessment could be made by single cell whole genome sequencing of ENS cells. The identification of family trees of mutations in ENS cells would provide a bar code to delve into clonal lineages as well as to identify potential function-impairing somatic mutations. If these somatic mutations can occur relatively late, well after the blastocyst stage, such investigations would need to targeting the cell lineages involved, rather than using cells from a convenient but inappropriate source such as blood cells [76]. While currently technically challenging as well as expensive, this has been done in *post-mortem* CNS specimens where the frequency of somatic mutations, their clonal inheritance and functional relevance has been confirmed [68, 76]. Given that, unlike CNS cells, ENS cells are routinely harvested by biopsy in cases of suspected ENS dysfunction, single cell whole genome sequencing could be performed for the non-Hirschsprung enteric neuropathology patients. With the continual reduction of cost of single cell whole genome sequencing along with increases in accuracy, this could soon become a reality. This could then not only confirm this mechanism but also identify patient-specific, ENS-specific somatic mutations that could guide targeted treatments.

7. General role of stochastic variation of cell behaviour in reduced penetrance

We have explained how unpredictable variations in two major classes of developmental enteric neuropathologies could come about. The variable Hirschsprung as modelled is completely stochastic, being displayed at the cusp of haploinsufficiency for a pre-existing Hirschsprung gene mutation. The non-Hirschsprung ENS dysfunction is also stochastic but functional outcome requires one stochastic event, a somatic non-Hirschsprung ENS mutation, added to another stochastic event, the 'superstar' clonal expansion which reduces clonal diversity. The functional impairment due to the specific somatic mutation is then permitted to become functionally deterministic if it occurs in the cell lineage of the ENS 'superstar'. We argue that these mathematically derived models provide

explanations for much of the variation in MZ twins with discordant ENS birth defects and therefore for at least some of the unexplained variability in ENS birth defects in more diverse genetic backgrounds.

The characteristics of the development of the ENS may make it prone to the stochastic variabilities which we have described mathematically here. However these characteristics are merely exaggerated in the ENS rather than being unique. The mathematical model is in essence universal, and hence variations of a similar stochastic nature could apply to some degree to other aspects of development. These roles of non-determinative functions may therefore contribute to the variability of other congenital pathologies, and particularly to other neurocristopathies. This therefore places a limit on the predictability of birth defects for individuals in the same way that limits to cancer predictability have been proposed [77]. It is important to confirm the existence and extent of true stochastic variation because in such cases searching for predictive determinants would be futile.

ACKNOWLEDGEMENTS

The authors thank past and current collaborators for discussions including Matthew Simpson, Emily Hackett-Jones, Anthony Fernando, Lincon Stamp, Heather Young and Dongcheng Zhang. Work was carried out with grants from the NHMRC and ARC. The MCRI acknowledges the Victorian Government's Operational Infrastructure Support Program.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to disclose.

REFERENCES

1. Allen, G. 1952, Am. J. Hum. Genet., 4(3), 155.
2. Schinz, H. R. 1945, Arch. Juliu Klaus-Stift. Vererb.-Forsch., 20, 13.
3. Danforth, C. H. 1924, Genetics, 9(3), 99.
4. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. and Kehrer-Sawatzki, H. 2013, Hum. Genet., 132(10), 1077.
5. Burga, A. and Lehner, B. 2012, FEBS J., 279(20), 3765.
6. Walters, L. C., Cantrell, V. A., Weller, K. P., Mosher, J. T. and Southard-Smith, E. M. 2010, Hum. Mol. Genet., 19(22), 53.
7. Heuckeroth, R. O. and Schafer, K. H. 2016, Dev. Biol., 417(2), 188.
8. Gartner, K. 1990, Lab. Anim., 24(1), 71.
9. Goodship, J., Cross, I., Scambler, P. and Burn, J. 1995, J. Med. Genet., 32(9), 746.
10. Keyte, A. and Hutson, M. R. 2012, Differentiation, 84(1), 25.
11. Manning, N. and Archer, N. 2006, Prenat. Diagn., 26(11), 1062.
12. Czyz, W., Morahan, J. M., Ebers, G. C. and Ramagopalan, S. V. 2012, BMC Med., 10, 93.
13. Zwijnenburg, P. J., Meijers-Heijboer, H. and Boomsma, D. I. 2010, Am. J. Med. Genet. B Neuropsychiatr. Genet., 153B(6), 1134.
14. Ollikainen, M., Smith, K. R., Joo, E. J., Ng, H. K., Andronikos, R., Novakovic, B., Abdul Aziz, N. K., Carlin, J. B., Morley, R., Saffery, R. and Craig, J. M. 2010, Hum. Mol. Genet., 19(21), 4176.
15. Ahluwalia, J. K., Hariharan, M., Bargaje, R., Pillai, B. and Brahmachari, V. 2009, Bioessays, 31(9), 981.
16. Lyu, G., Zhang, C., Ling, T., Liu, R., Zong, L., Guan, Y., Huang, X., Sun, L., Zhang, L., Li, C., Nie, Y. and Tao, W. 2018, BMC Genomics, 19(1), 428.
17. Raj, A., Rifkin, S. A., Andersen, E. and van Oudenaarden, A. 2010, Nature, 463(7283), 913.
18. Breckpot, J., Thienpont, B., Gewillig, M., Allegaert, K., Vermeesch, J. R. and Devriendt, K. 2011, Molecular Syndromology, 2(2), 81.
19. Li, R., Montpetit, A., Rousseau, M., Wu, S. Y., Greenwood, C. M., Spector, T. D., Pollak, M., Polychronakos, C. and Richards, J. B. 2014, J. Med. Genet., 51(1), 28.
20. Nishioka, M., Bundo, M., Iwamoto, K. and Kato, T. 2018, Mol. Psychiatry, doi:10.1038/s41380-018-0129-y.
21. Rohrback, S., Siddoway, B., Liu, C. S. and Chun, J. 2018, Dev. Neurobiol., 78(11), 1026.
22. Erickson, R. P. 2010, Mutat. Res., 705(2), 96.
23. Ju, Y. S., Martincorena, I., Gerstung, M., Petljak, M., Alexandrov, L. B., Rahbari, R., Wedge, D. C., Davies, H. R., Ramakrishna, M., Fullam, A., Martin, S., Alder, C., Patel, N., Gamble, S., O'meara, S., Giri, D. D., Sauer, T., Pinder, S. E., Purdie, C. A., Borg, A.,

Stunnenberg, H., van De Vijver, M., Tan, B. K., Caldas, C., Tutt, A., Ueno, N. T., Van 'T Veer, L. J., Martens, J. W., Sotiriou, C., Knappskog, S., Span, P. N., Lakhani, S. R., Eyfjord, J. E., Borresen-Dale, A. L., Richardson, A., Thompson, A. M., Viari, A., Hurles, M. E., Nik-Zainal, S., Campbell, P. J. and Stratton, M. R. 2017, Nature, 543(7647), 714.

24. Kimani, J. W., Yoshiura, K., Shi, M., Jugessur, A., Moretti-Ferreira, D., Christensen, K. and Murray, J. C. 2009, Twin. Res. Hum. Genet., 12(5), 462.
25. Baudisch, F., Draaken, M., Bartels, E., Schmiedeke, E., Bagci, S., Bartmann, P., Nothen, M. M., Ludwig, M. and Reutter, H. 2013, Twin. Res. Hum. Genet., 16(4), 802.
26. Zhang, R., Thiele, H., Bartmann, P., Hilger, A. C., Berg, C., Herberg, U., Klingmuller, D., Nurnberg, P., Ludwig, M. and Reutter, H. 2016, Twin. Res. Hum. Genet., 19(1), 60.
27. Lyu, N., Guan, L. L., Ma, H., Wang, X. J., Wu, B. M., Shang, F. H., Wang, D., Wen, H. and Yu, X. 2016, Chin. Med. J. (Engl.), 129(6), 690.
28. Young, H. M., Bergner, A. J., Simpson, M. J., Mckeown, S. J., Hao, M. M., Anderson, C. R. and Enomoto, H. 2014, BMC Biol., 12(1), 23.
29. Zhang, D., Ighaniyan, S., Stathopoulos, L., Rollo, B., Landman, K., Hutson, J. and Newgreen, D. 2014, Birth Defects Res. C Embryo Today, 102(3), 275.
30. Watt, K. E. N. and Trainor, P. A. 2014, in Neural Crest Cells, Evolution, Development and Disease, P. Trainor, Editor. Elsevier, Burlington, 362.
31. Papangeli, I. and Scambler, P. 2013, Wiley Interdiscip. Rev. Dev. Biol., 2(3), 393.
32. Etchevers, H. C., Amiel, J. and Lyonnet, S. 2006, Adv. Exp. Med. Biol., 589, 213.
33. Furness, J. B., Callaghan, B. P., Rivera, L. R. and Cho, H. J. 2014, Adv. Exp. Med. Biol., 817, 39.
34. Simkin, J. E., Zhang, D., Stamp, L. A. and Newgreen, D. F. 2019, Dev. Biol., 446(1), 22-33.
35. Zhang, D., Osborne, J. M., Abu-Bonsrah, K. D., Cheeseman, B. L., Landman, K. A., Jurkowicz, B. and Newgreen, D. F. 2018, Dev. Biol., 444, S287-296.
36. Rao, M. and Gershon, M. D. 2018, Nat. Rev. Neurosci., 19(9), 552-565.
37. Amiel, J., Sproat-Emison, E., Garcia-Barcelo, M., Lantieri, F., Burzynski, G., Borrego, S., Pelet, A., Arnold, S., Miao, X., Griseri, P., Brooks, A. S., Antinolo, G., De Pontual, L., Clement-Ziza, M., Munnich, A., Kashuk, C., West, K., Wong, K. K., Lyonnet, S., Chakravarti, A., Tam, P. K., Ceccherini, I., Hofstra, R. M., Fernandez, R. and Hirschsprung Disease Consortium. 2008, J. Med. Genet., 45(1), 1.
38. Attie, T., Pelet, A. Edery, P., Eng, C., Mulligan, L. M., Amiel, J., Boutrand, L., Beldjord, C., Nihoul-Fékété, C., Munnich, A., Ponder, B. A. J. and Lyonnet, S. 1995, Hum. Mol. Genet., 4(8), 1381.
39. Henderson, D., Zimmer, J., Nakamura, H. and Puri, P. 2017, Pediatr. Surg. Int., 33(8), 855.
40. Jung, P. M. 1995, J. Pediatr. Surg., 30(5), 646.
41. Sarioglu, A., Senocak, M. E. and Hicsonmez, A. 2000, Eur. J. Pediatr. Surg., 10(3), 204.
42. Goldstein, A. M., Thapar, N., Karunaratne, T. B. and De Giorgio, R. 2016, Dev. Biol., 417(2), 217.
43. Frattini, J. C. and Nogueras, J. J. 2008, Clinics in Colon and Rectal Surgery, 21(2), 146.
44. King, S. K., Sutcliffe, J. R., Ong, S. Y., Lee, M., Koh, T. L., Wong, S. Q., Farmer, P. J., Peck, C. J., Stanton, M. P., Keck, J., Cook, D. J., Chow, C. W., Hutson, J. M. and Southwell, B. R. 2010, Neurogastroenterol Motil., 22(8), 883-92, e234.
45. Di Nardo, G., Blandizzi, C., Volta, U., Colucci, R., Stanghellini, V., Barbara, G., Del Tacca, M., Tonini, M., Corinaldesi, R. and De Giorgio, R. 2008, Aliment Pharmacol. Ther., 28(1), 25.
46. Knowles, C. H. and Martin, J. E. 2000, Neurogastroenterol. Motil., 12(2), 181.
47. Landman, K. A., Simpson, M. J. and Newgreen, D. F. 2007, Dev. Growth Differ., 49(4), 277.
48. Soheilypour, M. and Mofrad, M. R. K. 2018, Bioessays, 40(7), e1800020.
49. Simpson, M. J., Merrifield, A., Landman, K. A. and Hughes, B. D. 2007, Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys., 76(2 Pt 1), 021918.
50. Landman, K. A., Binder, B. J. and Newgreen, D. F. 2012, in 10th International Conference on Cellular Automata for Research and Industry, ed. G.C.S.a.S. Bandini, Berlin Heidelberg Springer-Verlag

51. Binder, B. J. and Landman, K. A. 2009, J. Theor. Biol., 259(3), 541.
52. Binder, B. J., Landman, K. A., Simpson, M. J., Mariani, M. and Newgreen, D. F. 2008, Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys, 78(3 Pt 1), 031912.
53. Allan, I. J. and Newgreen, D. F. 1980, Am. J. Anat., 157(2), 137.
54. Young, H. M., Bergner, A. J., Anderson, R. B., Enomoto, H., Milbrandt, J., Newgreen, D. F. and Whitington, P. M. 2004, Dev. Biol., 270(2), 455.
55. Young, H. M., Turner, K. N. and Bergner, A. J. 2005, Cell Tissue Res., 320(1), 1.
56. Newgreen, D. F., Dufour, S., Howard, M. J. and Landman, K. A. 2013, Dev. Biol., 382(1), 305.
57. Newgreen, D. F., Southwell, B., Hartley, L. and Allan, I. J. 1996, Acta. Anat. (Basel), 157(2), 105.
58. Hearn, C. J., Murphy, M. and Newgreen, D. 1998, Dev. Biol., 197(1), 93.
59. Binder, B. J., Landman, K. A., Newgreen, D. F. and Ross, J. V. 2015, J. Theor. Biol., 380, 309.
60. Ishii, K., Doi, T., Inoue, K., Okawada, M., Lane, G. J., Yamataka, A. and Akazawa, C. 2013, Pediatr. Surg. Int., 29(2), 157.
61. Schuchardt, A., D'Agati, V., Larsson-Blomberg, L., Costantini, F. and Pachnis, V. 1994, Nature, 367(6461), 380.
62. Uesaka, T., Nagashimada, M., Yonemura, S. and Enomoto, H. 2008, J. Clin. Invest., 118(5), 1890.
63. Newgreen, D. F., Zhang, D., Cheeseman, B. L., Binder, B. J. and Landman, K. A. 2017, Cells Tissues Organs, 203(2), 105.
64. Cheeseman, B. L., Zhang, D., Binder, B. J., Newgreen, D. F. and Landman, K. A. 2014, J. R. Soc. Interface, 11(93), 20130815.
65. Merton, R. K. 1968, Science, 159, 56.
66. Heanue, T. A., Shepherd, I. T. and Burns, A. J. 2016, Dev. Biol., 417(2), 129.
67. Bondurand, N. and Southard-Smith, E. M. 2016, Dev. Biol., 417(2), 139.
68. Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., Lee, S., Chittenden, T. W., D'gama, A. M., Cai, X., Luquette, L. J., Lee, E., Park, P. J. and Walsh, C. A. 2015, Science, 350(6256), 94.
69. Milholland, B., Dong, X., Zhang, L., Hao, X. X., Suh, Y. and Vijg, J. 2017, Nature Communications, 8, 15183.
70. Cheeseman, B. L., Newgreen, D. F. and Landman, K. A. 2014, J. Theor. Biol., 363, 344.
71. Tomasetti, C. and Vogelstein, B. 2015, Science, 347(6217), 78.
72. Jiang, Q., Liu, F., Miao, C., Li, Q., Zhang, Z., Xiao, P., Su, L., Yu, K., Chen, X., Zhang, F., Chakravarti, A. and Li, L. 2018, Genet. Med., 20(7), 770.
73. Moore, S. W. and Zaahl, M. G. 2014, J. Pediatr. Surg., 49(2), 258.
74. Zhang, Z., Jiang, Q., Li, Q., Cheng, W., Qiao, G., Xiao, P., Gan, L., Su, L., Miao, C. and Li, L. 2015, Int. J. Clin. Exp. Pathol., 8(5), 5528.
75. Elisei, R., Cosci, B., Romei, C., Bottici, V., Renzini, G., Molinaro, E., Agate, L., Vivaldi, A., Faviana, P., Basolo, F., Miccoli, P., Berti, P., Pacini, F. and Pinchera, A. 2008, J. Clin. Endocrinol. Metab., 93(3), 682.
76. Rohrback, S., Siddoway, B., Liu, C. S. and Chun, J. 2018, Dev. Neurobiol., 78, 1026.
77. Lipinski, K. A., Barber, L. J., Davies, M. N., Ashenden, M., Sottoriva, A. and Gerlinger, M. 2016, Trends Cancer, 2(1), 49.