Title
A mathematical description of non-self for biallelic genetic systems in pregnancy, transfusion, and transplantation.

Author
Klaus Rieneck

Corresponding author:
Klaus Rieneck
Department of Clinical Immunology
Laboratory of Blood Genetics
Rigshospitalet
Copenhagen University Hospital
Blegdamsvej 9, DK-2100, Copenhagen, Denmark
Tel: +45 35 45 35 45
Email: kr@rh.dk

Klaus Rieneck ORCHID 0000-0002-7251-5754

Keywords: Immunology, assay design, pregnancy, blood transfusion, transplantation, biallelic systems
Abbreviations: cfDNA: cell-free DNA, GB: gigabase, HLA: Human Leucocyte Antigen, HSCT: Human Stem Cell Transplantation; Rh: Rhesus, STR: Short Tandem Repeat
MSC code 92D99
The author has no relevant financial or non-financial competing interests to disclose.

Abstract
A central issue in immunology is the immunological reaction against non-self. The prerequisite for a specific immunological reaction is the exposure to the immune system of a non-self-antigen.
A simple mathematical description of the fraction of a non-self-allele in biallelic systems at the population level in transfusion/transplantation medicine and in pregnancy is presented.
Furthermore, when designing assays, the mathematical descriptions may help estimate the number of genetic markers necessary to obtain a predetermined probability level in detecting non-self-alleles of a given frequency. For instance, the described equations can be helpful, in the design of assays where the non-self-allele can be detected by analysis of cfDNA in plasma from pregnant women to estimate fetal fraction or cfDNA and to monitor changes in cfDNA in plasma of transplantation patients.
Besides the equations that describe all non-self-situations in pregnancy and transfusion/transplantation, a novel way of estimating immunogenicity related to allele frequency is proposed.

Introduction
A non-self-situation arises when an individual is exposed to an antigen that this individual does not possess. Such a situation would arise in pregnancy where the mother can be exposed to an antigen that the fetus has inherited from its father or by blood transfusion or organ transplantation from an allogeneic donor. The term non-self is here used to encompass cfDNA with a primary sequence that is not found in the pregnant woman nor recipient of transfusion or organ donation. Simple mathematics was used to describe, in biallelic systems, the fraction of all situations with the presence of a non-self-allele irrespective of allele

frequency. The non-self-scenarios in pregnancy and blood donation/transplantation were addressed and the equations presented define the theoretically maximal fraction of situations where an immune response may arise and define all situations where the non-self-allele can be potentially detected by various assays based on primary DNA sequence. The equations presented can be helpful in assay development in biallelic systems and provide general insight. Most blood group antigens are biallelic as is also much other genetic variation.
Methods for determining fetal fraction have been published, e.g., (Ni et al., 2019).

Materials and methods
Only biallelic systems with alleles p and q were considered.
Basic mathematics and theoretical deliberations led to the development of three equations describing all non-self-situations in any biallelic system in pregnancy and transfusion/transplantation and based on the equations, simple equations were derived for calculating the number of biallelic markers needed to be combined into an assay to reach a given probability of detecting non-self in pregnancy and transfusion/transplantation including the scenario where both donor and recipient are homozygous albeit with different alleles.
Microsoft Excel was used to evaluate the equations in simulations *in silico* by creating 4000 samples in Hardy-Weinberg equilibrium but with varying allele frequency p of 0.1, 0.2, 0.3, 0.4, and 0.5 respectively. The 4000 samples with the same allele frequency were duplicated and both identical 4000 samples were randomized separately using the SLUMP function in Excel. A 9-digit number between 0 and 1 was generated and there were no duplicate numbers. The genotypes were coupled to the random number and sorted by the number thus generating a column of randomly sorted genotypes. For the prenatal testing, a column with only a single allele was randomized as the paternal contribution to the fetus. The two columns of the independently randomized 4000 samples – with the same allele frequency - were combined into one set of 4000 outcomes and the number of times, a non-self-situation was created, was counted. This was repeated for a total of 3 times for each allele frequency. The number of non-self-outcomes counted was compared with the number of non-self-outcomes predicted as calculated by equations (3), (9) and (12). The confidence intervals were calculated in GraphPad Prism 10.1.2.
The calculations have the Hardy-Weinberg principle as the basis. Frequency-wise in a biallelic system, the frequency of the genotypes in a population is $(p+q)^2=p^2+q^2+2pq=1$, when there is Hardy-Weinberg equilibrium (Hardy, 1908; Weinberg, 1908).

Results
In situations of non-self in pregnancy, the fetus has inherited an allele from its father that the mother does not have. We here extend the non-self term to cover genetic sequence variants that the mother does not have. Irrespective, cfDNA from both the fetus and the mother is present in maternal plasma and non-self cfDNA sequences can be used as identification tags for the presence of fetally derived cfDNA in pregnant women. Such considerations would also apply to transplantation situations albeit the donor has contributed two alleles.
In the pregnancy situation, the paternal allele inherited by the fetus may or may not be different from the alleles of the mother. In case the fetus inherits a p allele from the father, p_f (the $_f$ suffix solely denotes that the allele is inherited from the father, it is– indiscernible from other p alleles -) (Fig 1).

p^2p_f	pqp_f
pqp_f	q^2p_f

Figure 1

And in case the fetus inherits a q allele (q_f) from the father (Fig 2).

p^2q_f	pqq_f
pqq_f	q^2q_f

Figure 2

Figure 1 and 2. The maternal non-self-situations in pregnancy visualized.

The two situations, p^2q and q^2p are not the same, but both represent a non-self-situation that can provide the necessary. Still, by no means the sufficient basis for immunization and both situations can be informative as to the presence of fetal cfDNA in maternal plasma, i.e., fetal cfDNA that is qualitatively different from the maternal allotype as to the primary DNA sequence.
This means that for a given non-self-situation both the p^2q_f and the q^2p_f outcomes are relevant.
Mathematically, given that p+q=1, the first situation translates to

$$S_p=p^2q =-p^3+ p^2 \quad (1)$$

This formula gives the fraction of pregnancies for a given biallelic system where one allele has the frequency p where non-self is present, and the genetic precondition is fulfilled for a possible immunization event (Fig 3).

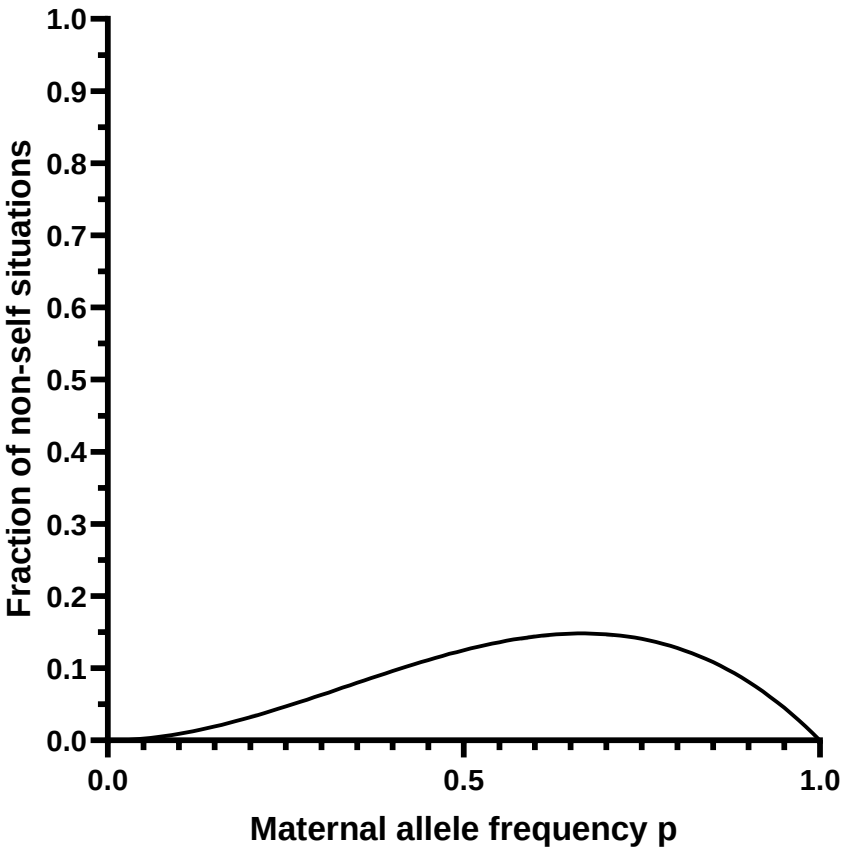

Fig 3. The fraction of pregnancies with non-self of a single allele in a biallelic system, p^2q ($-p^3+p^2$) (full line).

By integrating equation (1) and calculating the area under the graph in Fig 3, maximally 1/12 of all pregnancies for all values of p have this genetic constellation that is a precondition for maternal immunization for a given single biallelic antigen system. The maximal pregnancy risk allele frequency is found when p is close to $\frac{2}{3}$ and the risk is low both when p is either very low or very high. This means that blood group systems where immunization is observed should be most prevalent for blood groups with an allele frequency around this maximum, all other things being equal.

But p^2q only describes half the possible situations. The formula for the converse situation:

$$S_q = q^2p = (1-p)^2p = p^3 - 2p^2 + p \qquad (2)$$

is depicted with dots in Fig 4. With this distribution, the maximal risk situations are when p is close to $\frac{1}{3}$.

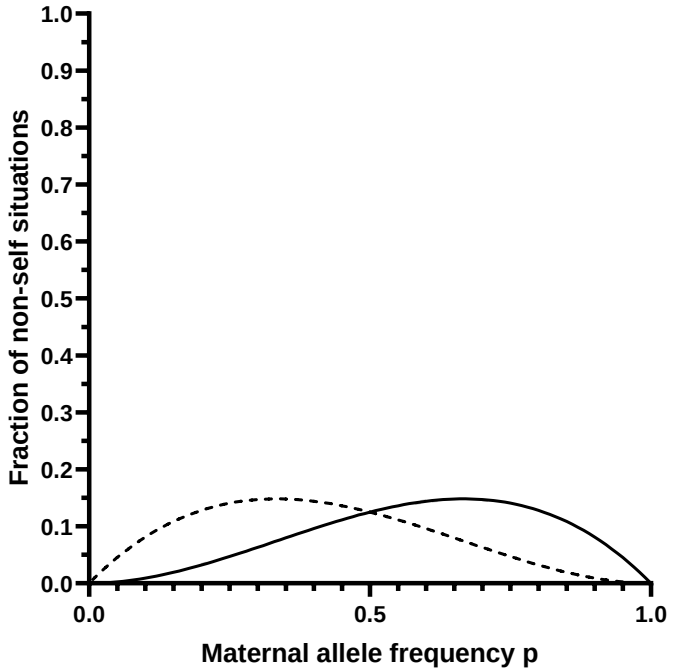

Fig 4. The fraction of non-self in pregnancy in relation to allele frequency for both alleles in a single biallelic system with the mother being either homozygous p; p^2q (full line) or homozygous q; q^2p (dotted line).

The sum of the two distributions is decisive for the total, theoretical immunization risk in a population for this biallelic locus and for all possible non-self-situations in a population. Determining which distribution, a given allele belongs to in a biallelic system is not possible.

In the case of an assay for determining either allele, both equation (1) and (2) must be considered.
To obtain a high probability for detection of fetal cfDNA in maternal plasma, it is necessary to use several biallelic variant markers to produce an assay to ascertain if one (or more) non-maternal alleles have been inherited by the fetus. Such an assay will have the added advantage that no prior knowledge of maternal or paternal alleles is needed. An assay addressing this can be useful as a control assay for the presence of fetal cfDNA in maternal plasma when making genetic predictions based on findings of fetal cfDNA in maternal plasma.
The cumulated information from several marker alleles will help establish the presence or non-detectable presence, that is, presumed absence or very low level of fetal cfDNA and this information will minimize the risk of a false negative result in relation to other tests based on detection of specific fetal cfDNA. The principle is illustrated in Fig 5 with an example of

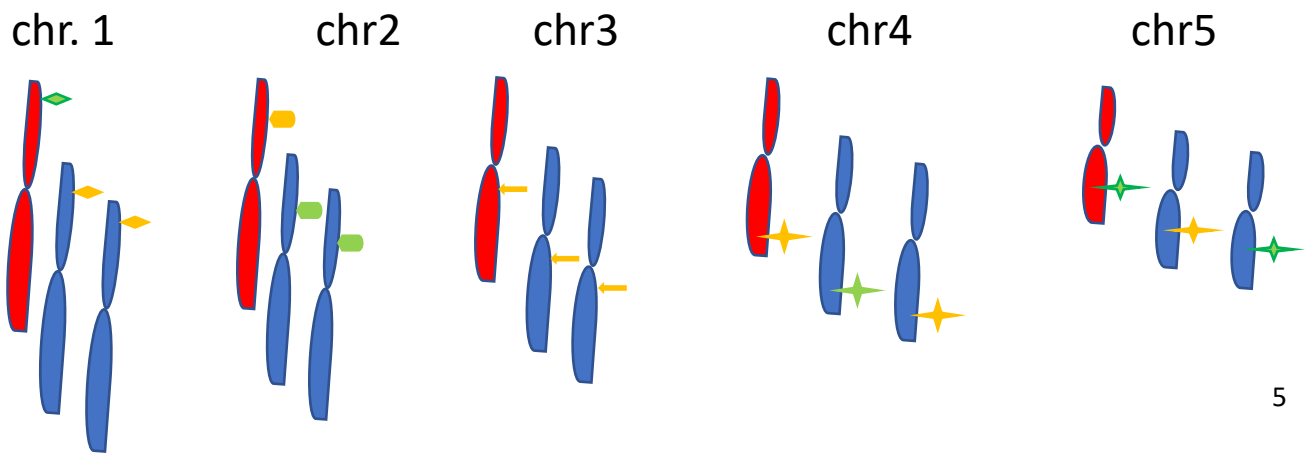

Fig 5. Marker informativity and non-informativity are exemplified.

five different primer sets targeting five different markers but only the markers on chromosomes 1 and 2 are informative of the presence of fetal cfDNA. So, for one marker either allele may be informative in these situations with both the (p^2q) and the (q^2p) situations being informative of the presence of fetal cfDNA in maternal plasma.
In both situations: the p^2q_f and the q^2p_f outcome (1/4 of all outcomes from Punnett squares, grey squares) are relevant (Fig 1 and 2) to detect the presence of non-self cfDNA from the fetus. Both situations are theoretically informative (S_I), and all other situations are non-informative of the presence of fetal cfDNA in maternal plasma, and without risk of immunization. The equation for the presence of non-self in the pregnant woman will be:

$$S_I = (p^2q) + (q^2p) = -p^2 + p \quad (3)$$

This is shown in Fig 6 for allele frequencies of $0 \leq p \leq 1$.

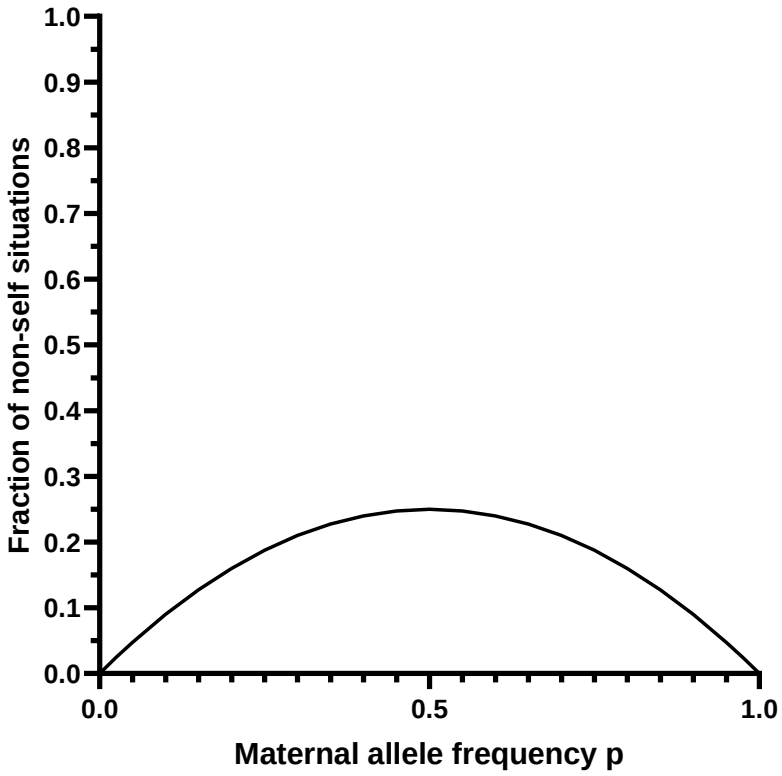

Fig 6. The fraction of all combined non-self-outcomes for one allele system in pregnancy, shown in relation to allele frequency p.

Equation (3) has a maximum at p=0.5. Thus, alleles with a frequency p=0.5 in a biallelic system are optimal for the detection of the presence of fetal cfDNA, as this allele frequency is most informative. By integrating equation (3) and calculating the area under the graph in Fig 6, maximally 1/6 of all situations can be informative in a single biallelic antigen system or result in immunization by contributing a non-self-antigen when both alleles from a biallelic system are taken into consideration. By using alleles with the most

informative frequencies in the narrow interval between 0.4 to 0.6, potentially 0.0493 can be accessed, that is 0.0493/0.1667 ~30% of all theoretically possible information in this setting.
By testing 20 alleles with p=0.5, on average about 5 alleles (0.25x20≈5) will be expected to be non-self and thus informative for a given blood sample; and at least one allele must be informative for an assay to be of use.
A non-informative situation, S_N can be calculated by

$$S_N=1-(p-p^2) =p^2-p+1 \quad (4)$$

For several alleles with differing allele frequencies p_1, p_2, p_3 ...p_i an assay, $S_{I(1-i)}$ using these allelic markers will be informative for at least one allelic marker, using equation (4) when.

$$S_{I(1-i)}=1-(p_1^2-p_1+1)(p_2^2-p_2+1)(p_3^2-p_3+1)\ldots\ldots (p_i^2-p_i+1) \quad (5)$$

If all allele frequencies are identical p, then equation (5) can be generalized and the fraction of informative situations testing n alleles can be calculated by

$$S_{I(n)}=1-(p^2-p+1)^n \quad (6)$$

Equation (6) can be rewritten to calculate the number of different allelic markers n, with the same allele frequency p, needed to obtain a desired level of informativity $S_{I(n)}$:

$$n = \frac{\ln (1-S_{I(n)})}{\ln (p^2-p+1)} \quad (7)$$

For instance, if information on the presence of fetal cfDNA is wanted in 99% of situations of testing a maternal plasma sample using alleles with a frequency of 0.5, it can be calculated that at least 16 biallelic markers would be needed in an assay as estimated from equation (7).
This is useful information when designing an assay for the detection of fetal cfDNA.

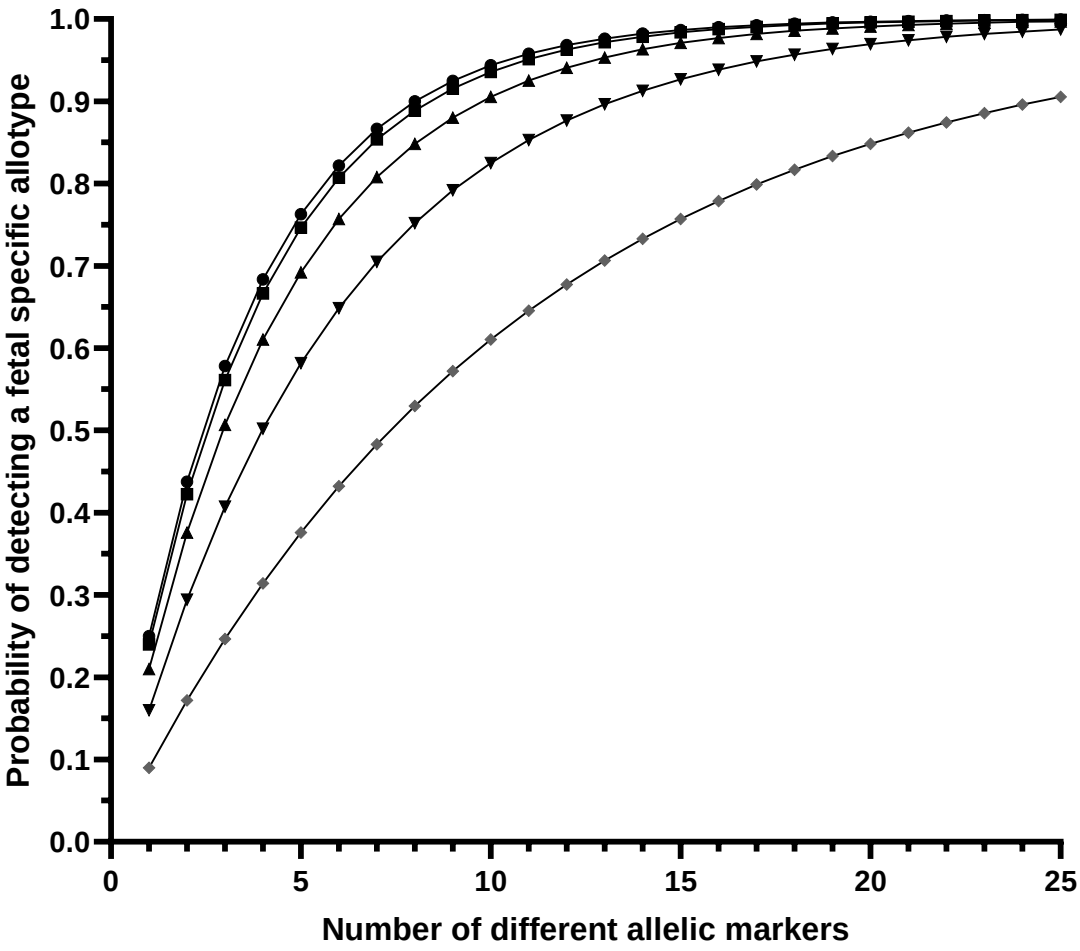

Fig 7. The cumulative effect of choosing multiple biallelic markers for detection of the presence of non-self in pregnancy i.e., fetal cfDNA in relation to allele frequencies of 0.5 (●), 0.4 (■), 0.3 (▲), 0.2 (▼), and 0.1 (◆), respectively.

In Fig 7 the application of equation (7) shows the effect of the number of markers with different allele frequencies in relation to the probability of detecting a fetal specific allotype. Alleles with allele frequencies down to about 0.3 are highly informative and can be included in an assay.
It should be added that if p is substituted with (1-q) in equation (3) $-p^2+p$, as $-(1-q)^2+(1-q)$, the result is $-q^2+q$.

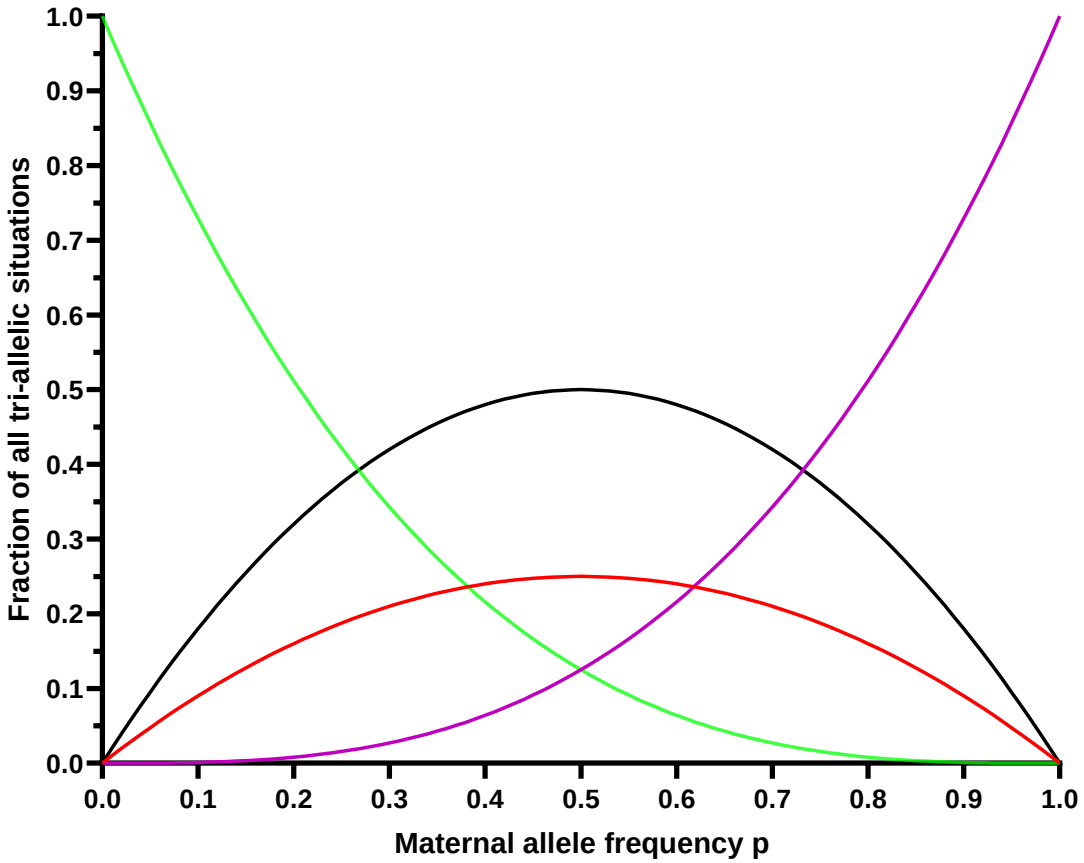

Fig 8. All theoretically possible outcomes of allele combinations in pregnancy for a single biallelic marker of any frequency: p^3 defined by (p^3) (lilac graph), q^3 defined by ($-p^3+3p^2-3p+1$) (green), and the situation with non-self defined by ($-p^2+p$) (red), mother heterozygous and fetus with either allele ($-2p^2+2p$) (black). All equations added give 1. ($-2p^2+2p-p^2+p-p^3+3p^2-3p+1+ p^3 = 1$)

An overview of all possible outcomes in pregnancy in relation to allele composition when considering the two maternal alleles and the allele contributed by the father is shown in Fig 8. As the mother will invariably contribute one allele to the fetus (except in situations as the recipient of an egg donation, which situation will be the same as described for the transfusion/transplantation situation) only three alleles are considered. All situations described by equation (3), $-p^2+p$ give rise to non-self and consecutively risk of immunization as well as being informative in prenatal assays that detect fetal-specific cfDNA sequences. No other situation gives rise to a non-self-situation for the mother.
If for instance, the allele frequency of p is 0.5, then 25% of all pregnancies will be at risk of immunization, in 12,5% of all pregnancies, the mother is homozygous for p and the fetus has received a p from its father, and in 12,5% of all pregnancies the mother will be homozygous for q and the father has passed on a q allele to the fetus. At p=0.5 then in 50% of all pregnancies, the mother is heterozygous, and in these situations, there is no risk of immunization by the p or q allele.

In the recipient of transfusion/transplantation, non-self-alleles are introduced in the following 4 situations:

1) Recipient*Donor p^2*pq2
2) Recipient*Donor p^2*q^2
3) Recipient*Donor q^2*pq2
4) Recipient*Donor q^2*p^2

This is shown in Table 1, cells marked with (A) indicating all the non-self-outcomes.
All non-self-outcomes can be written as:

$$s=p^2pq2+p^2q^2+ q^2pq2+ q^2p^2=p^2(p((1-p)^2))+p^2((1-p)^2)+((1-p)^2)p((1-p)^2)+((1-p)^2)p^2 \quad (8)$$

Equation (8) can be simplified as:

$$s=-2p^4+4p^3-4p^2+2p \quad (9)$$

Where p is allele frequency.
And for calculation of n markers with an allele frequency of p, cfr. above:

$$n = \frac{\ln(1-S_{I(n)})}{\ln(1-2p^4-4p^3+4p^2-2p)} \quad (10)$$

Setting $S_{I(n)}$ at 0.99 and p=0.5 gives n≈5.
This equation can be useful for calculating the number of primer sets needed in design for the detection of non-self in this situation where both a heterozygous and homozygous contribution to non-self can be used for assay purposes.

donor\recipient	p^2	q^2	pq	pq
p^2	p^2p^2	q^2p^2 (A,B)	pqp^2	pqp^2
q^2	p^2q^2 (A,B)	q^2q^2	pqq^2	pqq^2
pq	p^2pq (A)	q^2pq (A)	pqpq	pqpq
pq	p^2pq (A)	q^2pq (A)	pqpq	pqpq

Table 1. Overview of outcomes including non-self-outcomes. A and B are non-self-situations for transfusion/transplant recipients.

In Table 1. The cells marked (A) define the non-self-situations that must be considered for assay design when all non-self-situations in transfusion or transplantation must be considered. These situations are described as $-2p^4+4p^3-4p^2+2p$ (equation (9)).
The cells marked (B) define the non-self-situations that must be considered when only the two homozygous situations are relevant e.g., in some assays detecting rejection of a transplanted organ. These situations are described as $2p^4-4p^3+2p^2$ (equation (12)).
Setting p=0.5, maximally 37.5% of recipients of blood transfusion or a recipient of a donor organ will have a non-self-allele for a given biallelic system, equation (9) and Fig 11. An overview of all the fractions in transfusion and organ donation is shown in Fig 9.

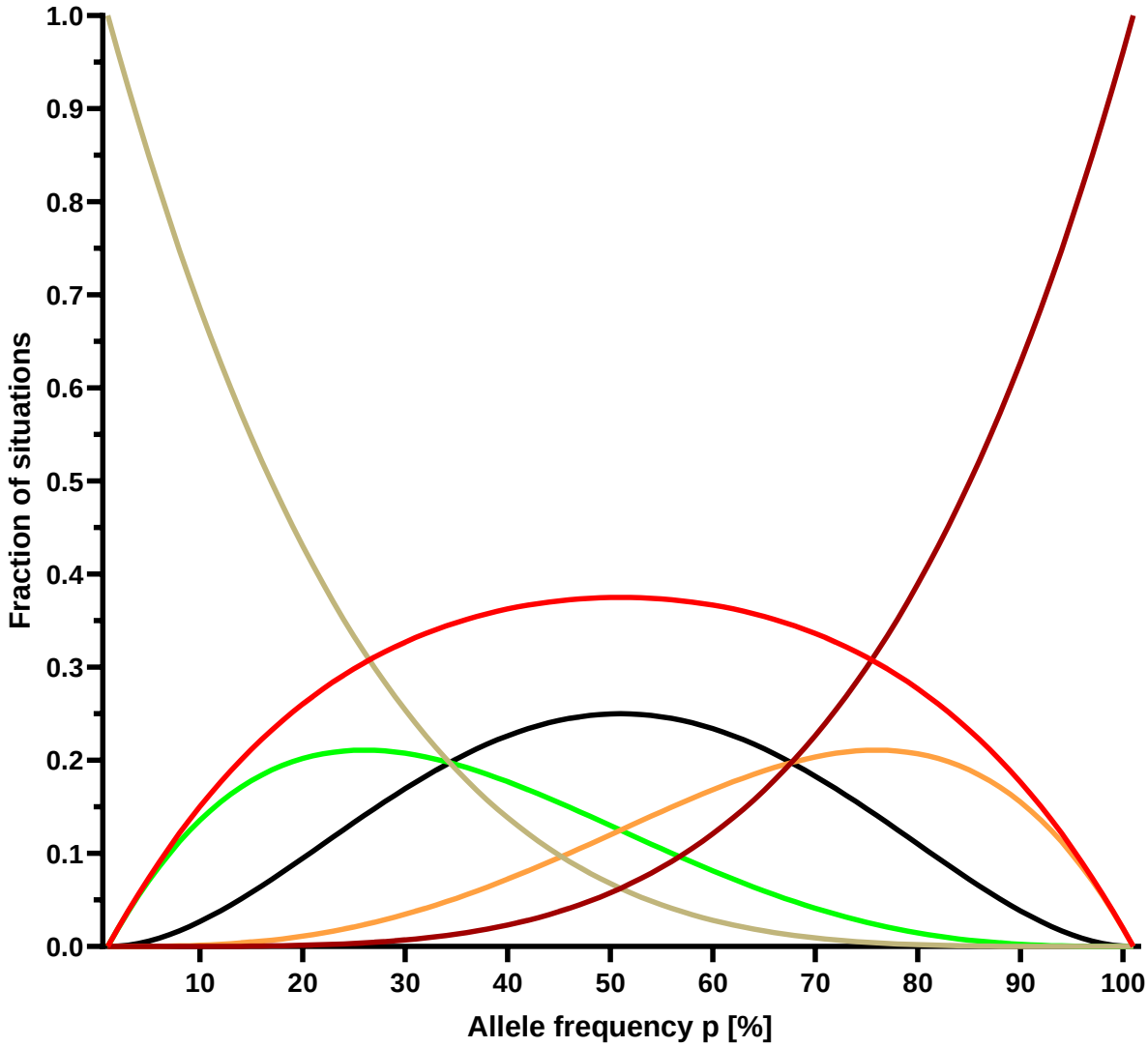

Fig 9. All theoretical situations from transfusion/transplantation in a biallelic system. Situations with non-self are defined by ($-2p^4+4p^3-4p^2+2p$) (red) and are the combined risk of the two homozygous situations p^2q^2 and q^2p^2 and the four situations of non-self for homozygous recipient and heterozygous donor p^2pq+p^2pq and q^2pq+q^2pq (see Table 1). The two latter situations occur with the same fraction as the green and orange graphs respectively. The situations without non-self: with a heterozygous recipient and two different homozygous donors: pqp^2+pqp^2 are defined by ($-2p^4+2p^3$) (orange) and pqq^2+pqq^2 are defined by ($-2p^4+6p^3-6p^2+2p$) (green). The situations where the recipient and donor have the same homozygous alleles are defined by p^4 (p^4) (brown), and for q^4 by ($p^4-4p^3+6p^2-4p+1$) (grey). The 4 situations where both recipient and donor are heterozygous are defined by ($4p^4-8p^3+4p^2$) (black). All the above equations added give 1. ($-2p^4+4p^3-4p^2+2p+ p^4+ p^4-4p^3+6p^2-4p+1-2p^4+2p^3-2p^4+6p^3-6p^2+2p+4p^4-8p^3+4p^2=1$)

From Fig 9 it seems that non-self-situations (the red graph) arise at a fairly constant level for values of p between 20% to 80% and make up a fraction of about 0.25-0.35 in this interval of p. None of the other genotype combinations pose any risk as to immunization of the recipient. The two equations for single alleles in pregnancy have maximal fractions exactly where the three graphs intersect at p=1/3 and 2/3 respectively.

In other situations, where it is desirable to detect admixed DNA from different individuals such as for chimerism measurements in HSCT or some cases of organ transplantation it can be desirable to investigate only the double homozygous situation; the mathematics is slightly different. Digital PCR technology may be advantageous in these situations.

An optimally informative situation S_I for digital PCR to monitor cfDNA in cases of chimerism for instance after transplantation would be when the donor is homozygous for a given marker and the recipient is homozygous for the alternative allele or vice versa (cells marked (B) in table 1):

$$S_I=p^2q^2+q^2p^2 \quad (11)$$

Given that p+q=1, this can be simplified to

$$S_I=2p^4-4p^3+2p^2 \quad (12)$$

And the situation F_N that is non-informative

$$S_N=1-(2p^4-4p^3+2p^2)=1-2p^4+4p^3-2p^2$$

For more markers with varying allele frequencies p_1, p_2, p_3...p_i, where at least one situation is informative.

$$S_{I(1\text{-}i)}=1-(1-2p_1^4+4p_1^3-2p_1^2)(1-2p_2^4+4p_2^3-2p_2^2)(1-2p_3^4+4p_3^3-2p_3^2).....(1-2p_i^4+4p_i^3-2p_i^2) \quad (13)$$

When p is the same for n different allelic markers, the formula can be simplified to

$$S_{I(n)}=1-(1-2p^4+4p^3-2p^2)^n \quad (14)$$

Rearranged:

$$n=\frac{\ln(1-S_{I(n)})}{\ln(1-2p^4+4p^3-2p^2)} \quad (15)$$

Setting $S_{I(n)}$ at 0.99 and p=0.5 gives n≈34, i.e., 34 primer sets with biallelic markers with an allele frequency of p=0.5 are needed to have, with 99% probability, at least one marker that can be used to discern between recipient and donor cells with a marker that is homozygous in the recipient as well as homozygous in the donor material albeit for the alternative allele. If p=0.4 then 38 allelic markers would be needed to achieve the same $S_{I(n)}$.

With p=0.5 a check of equation (15) for 34 markers gives

$$S_{I(n)}=(1-(2/16))^{34}=0.0107$$

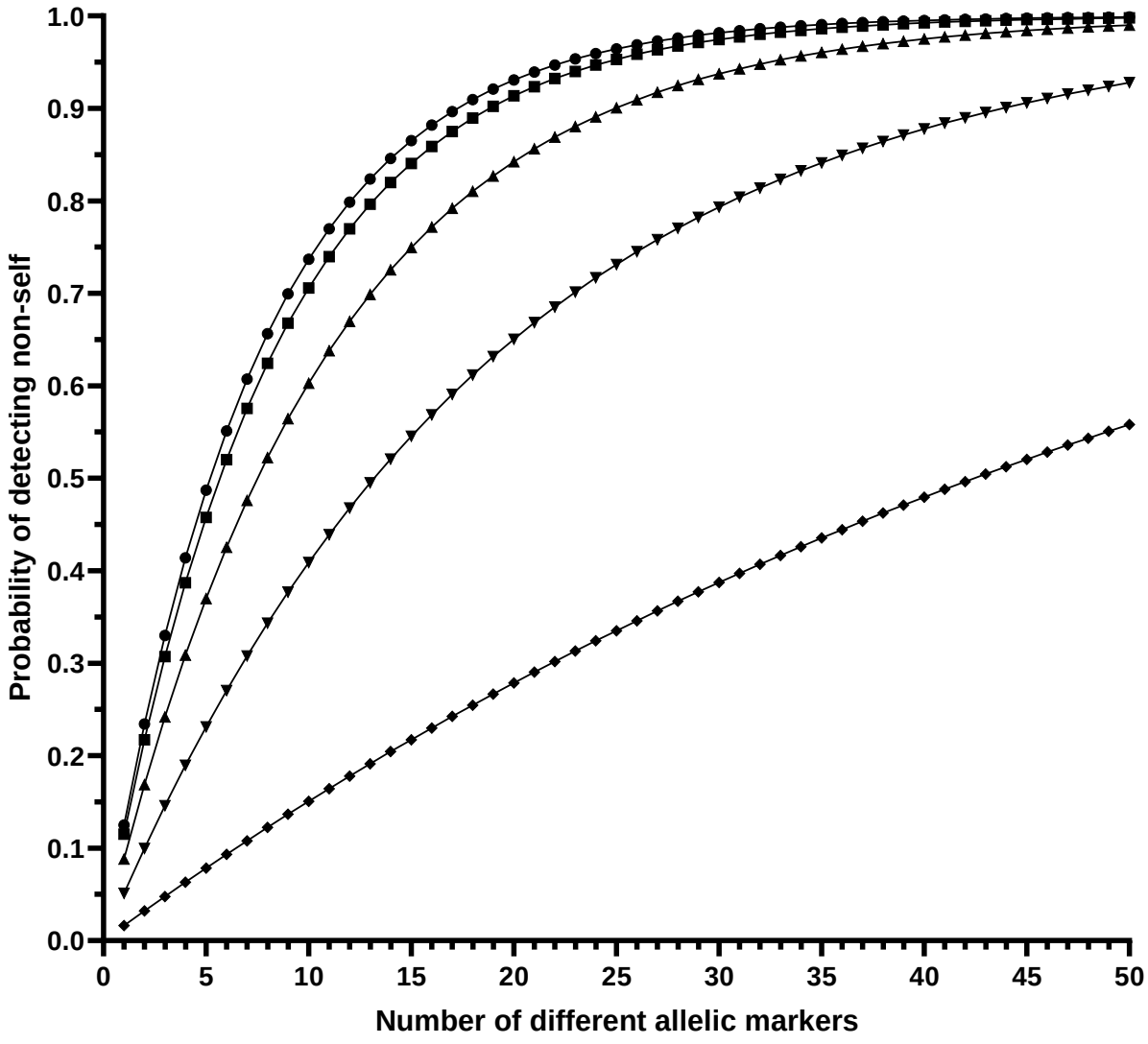

Fig 10. The probability of detecting non-self in relation to allele frequencies and the cumulative number of markers used in an assay of the double homozygous situation. Allele frequencies of 0.5 (●), 0.4 (■), 0.3 (▲), 0.2 (▼), and 0.1 (◆), respectively are shown.

In Fig 10 the application of equation (14) shows the number of markers with different allele frequencies needed to obtain a given level of probability of detecting non-self. Alleles with allele frequencies down to about 0.4 are highly informative and can be included in an assay.

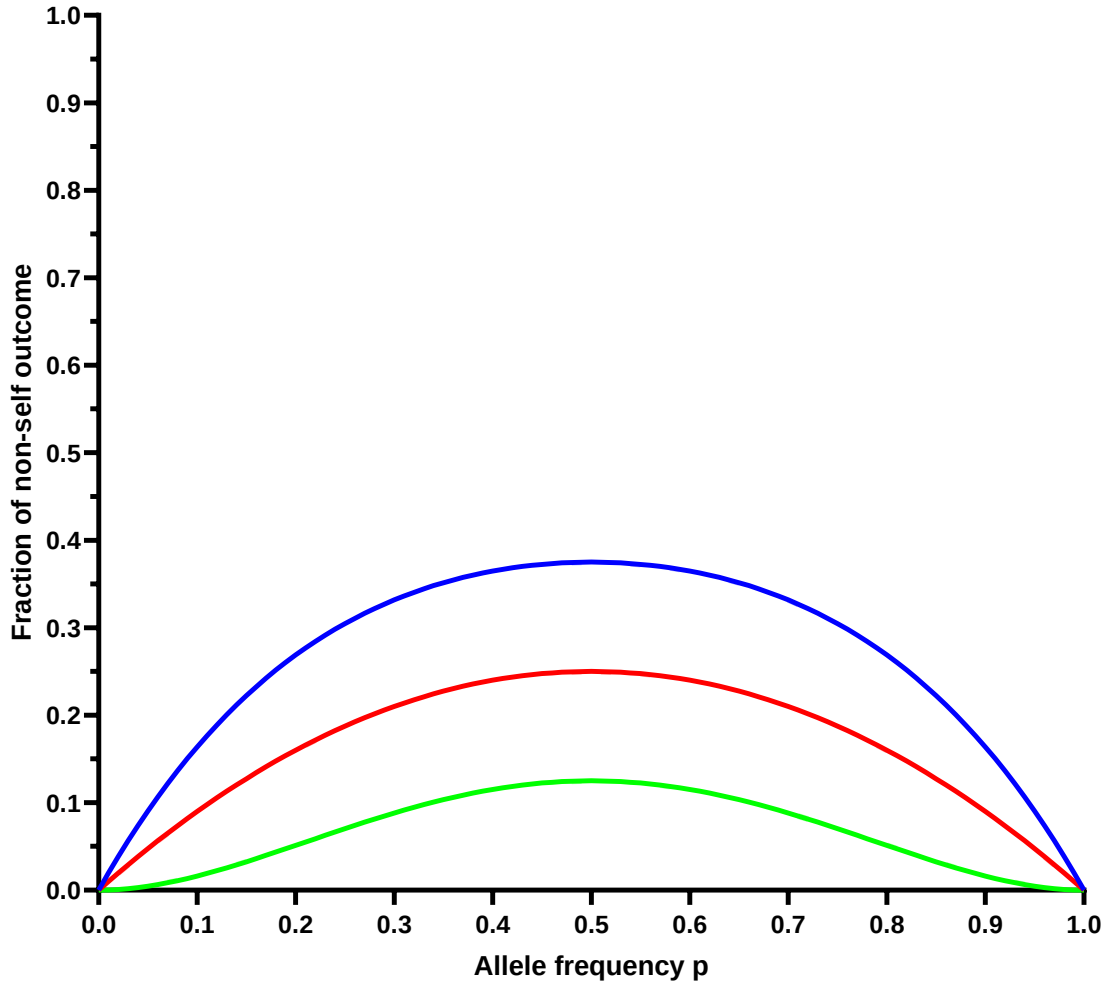

Fig 11. The fraction of non-self-outcomes predicted by the three equations as a function of allele frequency. The double homozygous non-self-situations defined by ($2p^4-4p^3+2p^2$) (green), the non-self-situations relevant for pregnancy defined by ($-p^2+p$) (red), and all non-self-situations in transfusion and transplantation are defined by ($-2p^4+4p^3-4p^2+2p$) (blue).

The graphs in Fig 11 depict the 3 different scenarios described by the three different equations for the contribution of non-self. At p=0.5, the graph describing the transfusion/transplantation situation has a maximum of 6/16, the graph describing pregnancy has a maximum of 4/16, and the graph describing the double homozygous situation has a maximum of 2/16, corresponding to the number of squares in table 1 with the genotypes used for the deduction of the equations.

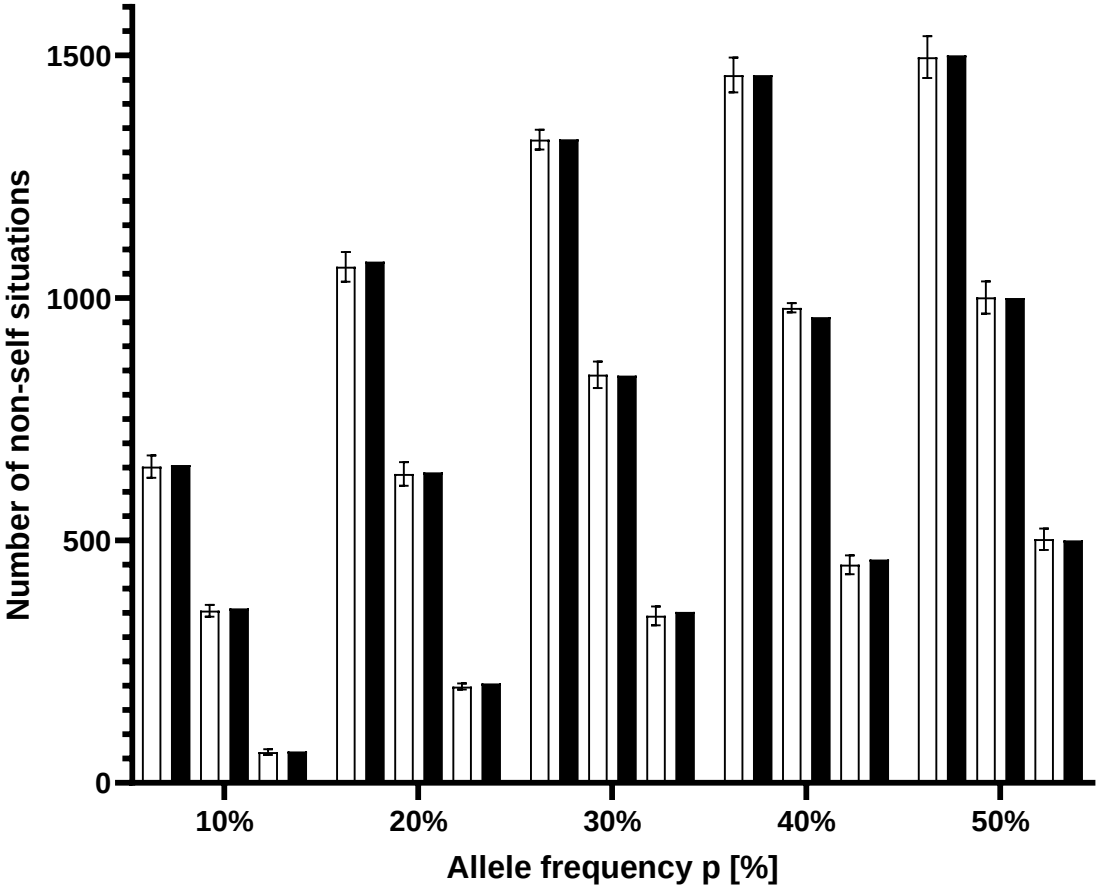

Fig 12. *In silico* simulation. Non-self-situations (transplantation, pregnancy, and double homozygous scenarios) were counted after *in silico* simulation of 4000 constructed alleles/genotypes in Hardy-Weinberg equilibrium with 5 different allele frequencies (0.1, 0.2, 0.3, 0.4, and 0.5), each with four replicates. The counted non-self-situations (white columns) were compared to the predicted situations (black columns) by the equations (9), (3), and (12). The 95% confidence interval is shown for the counted situations. For each allele frequency, the two first columns are from the transfusion/transplantation scenario, the next two columns are from the pregnancy scenario and the last two columns are from the double homozygote scenario.

The result of the simulation of the combined 2x4000 constructed and randomized genotypes in Hardy-Weinberg equilibrium showed good agreement with the results predicted by the equations (Fig 12) and Table 2. The number of non-self-situations that were counted from the simulation of the transfusion/organ recipient situation, the pregnancy situation, and the double homozygous scenario, were compared to the predicted numbers from the equations (9), (3), and (12). There was no significant difference in Fisher's exact test at $p < 0.05$.

Allele frequency	found_tx	predicted_tx	found_pre	predicted_pre	found_hom	predicted_hom
10%	659	655.2	357	360	64	64.8
	644		357		68	
	669		362		64	
	637		344		59	
20%	1068	1075.2	629	640	200	204.8
	1089		660		202	
	1056		628		193	
	1044		632		200	
30%	1324	1327.2	840	840	351	352.8
	1313		865		328	
	1324		837		356	
	1344		824		343	
40%	1493	1459.2	980	960	458	460.8
	1454		985		462	
	1448		971		442	
	1444		982		437	
50%	1512	1500	998	1000	519	500
	1510		980		492	
	1508		996		509	
	1456		1030		490	

Table 2. The numbers from the simulations and predictions with an allele frequency of p that underlies Fig 12 are shown.

The biggest differences were found at 40% allele frequency for the prenatal simulation with a mean of 980 and a predicted number of 960.
The 95% confidence intervals were calculated based on four generated replicates of random genotype combinations and all except one predicted value (p=40% for pregnancy) were within the 95% confidence intervals.
The simulation was also done once with 400 samples with consistent results (data not shown).

The three polynomial equations (9), (3), and (12) can be characterized for all theoretical outcomes valid for $0 \leq p \leq 1$ and the fraction ≥ 0 and ≤ 1:

$ap^4-bp^3+cp^2$	$a=2b=c,$	$0 \leq a \leq 16,$	1:2:1
$-ap^4+bp^3-cp^2+dp$	$a=2b=2c=d,$	$0 \leq a \leq 1/0.0625$	1:2:2:1
$-ap^2+bp$	$a=b,$	$0 \leq a \leq 4,$	1:1

However, only equations (9), (3), and (12), are valid in the stated biological context.

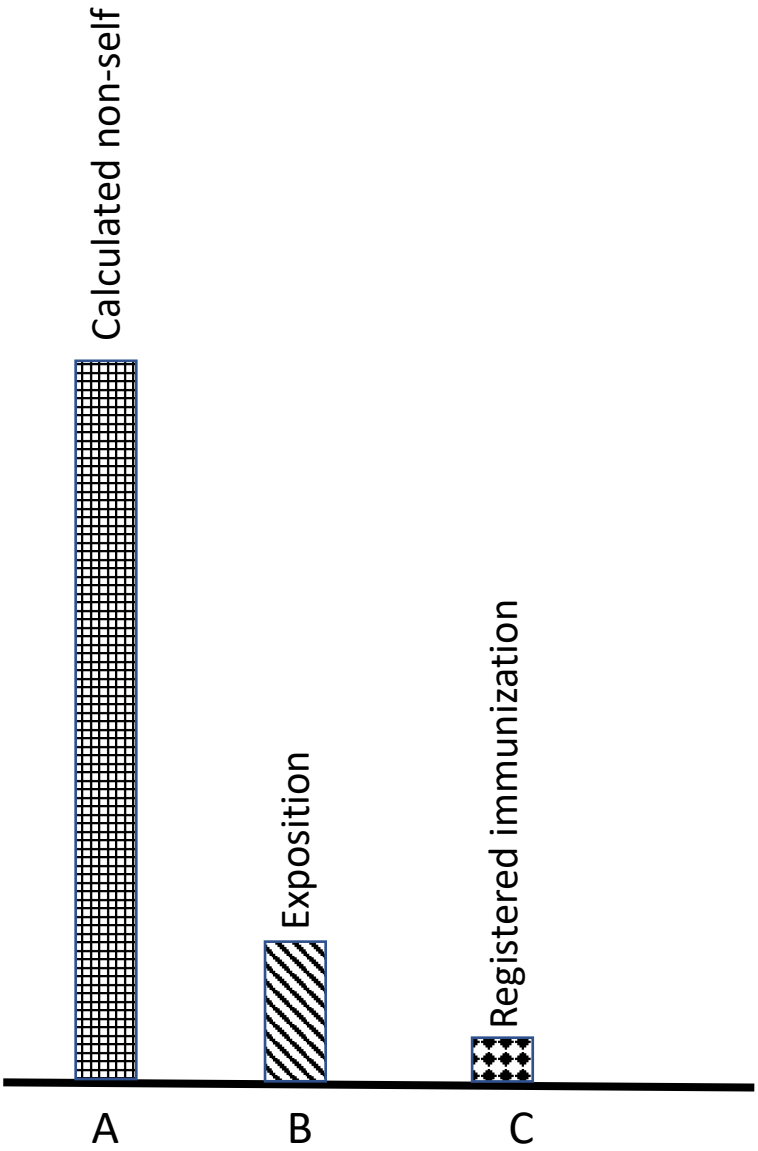

Fig 13. Proposal of an immunogenicity index related to maximum non-self calculation (not drawn to scale).

For a quantitative estimation of immunogenicity to enable easy, relative comparability among all biallelic serotypes, we propose calculating an immunogenicity index, $I=\frac{BC}{A}$ (Fig 13). The calculation is simple, and the exposition in the forms of the number of blood transfusions or transplantations or pregnancies should be relatively well documented, it would be more cumbersome to detect and register all immunizations as a consequence of exposition. Each allele can be considered separately.
An index for pregnancy (equation (3) $-p^2+p$) and transfusion/transplantation (equation (9) $-2p^4+4p^3-4p^2+2p$) should probably be calculated separately.

Discussion

The premise for the mathematical description of the number of markers needed is that alleles have undergone random assortment in accordance with Mendel's third law which says that alleles are independently assorted and thus that traits encoded by the alleles segregate independently of each other during gamete formation. The physiologic process of meiosis with independent segregation of non-disequilibrium alleles thus underlies the mathematical descriptions. In general, the same assumptions that apply to Hardy-Weinberg calculations would apply to the mathematical description.

All the non-self-scenarios in biallelic systems were described for pregnancy, transfusion/transplantation, and the special situation of a homozygous recipient and a donor homozygous for the alternative allele. In these situations, easy calculation of the number of markers needed for obtaining a desired information level in assays can be obtained regarding the presence of non-self-genetic variants from equations (7), (10), and (15).

Non-self-situations are also the prerequisite for an alloimmune response to occur, although importantly several other factors are needed.

Assays for determining chimerism in transplantation without prior knowledge of the genotypes of the involved individuals have been developed (Clausen et al., 2023). For instance, one group has chosen 24 indel markers using both homozygous and heterozygous informative marker genotypes (Pettersson et al., 2021).

The equations can be used to assess the number of markers to be used in prenatal control assay for the presence of fetal cfDNA to minimize the risk of false negative results. Also, it is important to note that the graph describing the pregnancy situation has a form that indicates that alleles with allele frequencies far from the optimal 0.5 are very informative as to non-self (Fig 6). This is also indicated in Fig 7 where alleles with a frequency as low as 0.3 appear to be reasonably informative (Lee et al., 2017).

By comparing Fig 7 and Fig 10, clearly, more markers are needed in the double homozygous situation to obtain the same probability of detecting non-self as compared to the pregnant situation.

In the case of the double homozygous situation, a large number of assays with individual markers must be designed to ensure a useful test with a high rate of useful outcomes. However, the markers should be assorted independently and therefore be spaced sufficiently. With a distance between loci of 40 GB (LaRue et al., 2014) about 75 markers can be designed from the human genome for a single assay.

Different equations describe the fraction of non-self in pregnancy on one hand and transfusion and transplantation on the other hand, which makes biological sense as in pregnancy, the father always passes only one of his two alleles to the fetus. The two scenarios of transfusion and transplantation are analogous in respect to the description of non-self.

In forensics multiallelic STR systems are routinely used and estimations of the number of biallelic SNPs needed to replace STRs have been done (Amorim and Pereira, 2005; Gill, 2001; Lee et al., 2017).

We also suggest a novel way of estimating immunogenicity that may better enable comparison across different biallelic systems. The basic idea is to relate immunization to the maximally theoretically possible immunization for any biallelic system and thus enable a different way of comparing immunogenicity among different allele systems. Perhaps the calculation of the immunogenicity index may give an alternative uniform way of comparing the immunogenicity of different biallelic systems. Once a reliable immunogenicity index has been calculated, it could be used to estimate the completeness of registration of immunization frequency in other populations of similar genetic backgrounds. This could be helpful in the registration of transfusion complications that often involve an antibody response. Also, in case the immunogenicity index does not vary among different populations, it may give hints as to the biological origins of immunogenicity for a given allele.

The suggested immunogenicity index calculation should be evaluated experimentally to gauge the relevance of this approach and the results compared to published data.
The mathematical descriptions were tested *in silico* to ascertain that the mathematical predictions were accurate. There was no significant deviation (at $p < 0.05$) by Fisher's exact test from the counted versus the expected numbers as calculated by equations (9), (3), and (12), see Fig 12. Thus, this *in silico* test does not invalidate the predictive accuracy of the equations, however, a more rigorous *in silico* validation would need both a much larger sample size and many more replicates. In three situations the predicted numbers fell just outside the 95% confidence interval, with a total of 4 replicates. With so many calculations and few replicates, this is not surprising.

In conclusion, a mathematical description is reported of biallelic systems of non-self-allele fractions in 3 different scenarios: pregnancy, transfusion/transplantation including the scenario with a homozygous donor and a recipient homozygous for the alternative allele. Also given, are derived equations to calculate the number of marker systems needed to reach a given probability of detecting non-self. These equations can be useful in quantitative estimations for the design of tests for identification purposes e.g., fetal fraction or chimerism determination and other purposes.

References

Amorim, A. and Pereira, L., 2005. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Sci Int. 150, 17-21.

Clausen, F.B., Jorgensen, K.M.C.L., Wardil, L.W., Nielsen, L.K. and Krog, G.R., 2023. Droplet digital PCR-based testing for donor-derived cell-free DNA in transplanted patients as noninvasive marker of allograft health: Methodological aspects. PLoS One. 18, e0282332.

Gill, P., 2001. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med. 114, 204-10.

Hardy, G.H., 1908. Mendelian proportions in a mixed population. Science. 28 49–50.

LaRue, B.L., Lagace, R., Chang, C.W., Holt, A., Hennessy, L., Ge, J., King, J.L., Chakraborty, R. and Budowle, B., 2014. Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selection for a global INDEL panel for human identification. Leg Med (Tokyo). 16, 26-32.

Lee, H.J., Lee, J.W., Jeong, S.J. and Park, M., 2017. How many single nucleotide polymorphisms (SNPs) are needed to replace short tandem repeats (STRs) in forensic applications? Int J Legal Med. 131, 1203-1210.

Ni, M., Peng, X.L. and Jiang, P., 2019. Bioinformatics Pipeline for Accurate Quantification of Fetal DNA Fraction in Maternal Plasma. Methods Mol Biol. 1909, 177-180.

Pettersson, L., Vezzi, F., Vonlanthen, S., Alwegren, K., Hedrum, A. and Hauzenberger, D., 2021. Development and performance of a next generation sequencing (NGS) assay for monitoring of mixed chimerism. Clin Chim Acta. 512, 40-48.

Weinberg, W., 1908. Über den Nachweis der Vererbung beim Menschen. Jahresh. Ver. Vaterl. Naturkd. 64 369–382.