



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Tait, BD

Title:

The importance of establishing genetic phase in clinical medicine

Date:

2022-02-01

Citation:

Tait, B. D. (2022). The importance of establishing genetic phase in clinical medicine. *International Journal of Immunogenetics*, 49 (1), pp.1-7. <https://doi.org/10.1111/iji.12567>.

Persistent Link:

<https://hdl.handle.net/11343/311213>

The importance of establishing genetic phase in clinical medicine

**Brian D Tait**<sup>1,2</sup>

1. Haplomic Technologies, Melbourne, Australia.  
(btait@haplomictechnologies.com)
2. Department of Medicine, University of Melbourne, Royal Melbourne Hospital, Australia

## **Abstract**

Haplotyping or determination of genetic phase has always played a pivotal role in MHC (HLA studies) both in helping to understand inheritance patterns in diseases such as type 1 diabetes (T1D) and in ensuring better matching in transplantation scenarios such as haematopoietic stem cell transplantation (HSCT), using donors genetically related to the patient. In recent years the need to establish genetic phase in a number of clinical scenarios has become apparent. These include:

1. Genetic phasing for hematopoietic stem cell transplants using unrelated donors, where the HLA haplotypes are not known but where haplotype matched recipients fare better clinically than allele matched, but haplotype mismatched patients.
2. The use of checkpoint inhibitors is one of the most innovative and exciting developments in cancer treatment in years. An example is the use of the monoclonal ipilimumab to block the CTLA-4 receptor which is known to contain polymorphic sites. Until the phase of these polymorphisms is known it will not be possible to determine how effectively this monoclonal will perform in individual patients.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1112/iji.12567](#).

This article is protected by copyright. All rights reserved.

3. The role of miRNA single strand molecules and their effect on gene expression. Thousands of non-coding genes have been identified, have been shown to be polymorphic, as have their target genes. Genetic phasing of polymorphism both in the miRNA source genes and their targets, is clearly a fertile area of research

4. In areas such as drug metabolism where the polymorphic family of CYP genes are responsible for the metabolism of the majority of prescription drugs, determining phase of SNPs is critical to understanding drug metabolism and efficacy.

5. In multigenic disease studies combinations of single nucleotide polymorphisms (SNPs) in participating genes require accurate phasing in order to fully appreciate their role in the disease process. In addition, the level of expression of genes (point 3) is also important in understanding disease processes at the functional level.

This review outlines the techniques that are currently available for approximating phase and discusses the clinical relevance of establishing genetic phase in areas of clinical medicine outlined in points 1-3.

## ***1. Introduction***

Protein is coded for by DNA and RNA nucleotides in cis phase, (haplotypic phase). When two or more non-synonymous nucleotide variations are present in a sequence therefore, unless the phase is established, it is not possible to assign protein sequences.

The word haplotype was introduced into the scientific lexicon by world renowned geneticist and HLA pioneer Ruggero Ceppellini in 1967 at the Third Histocompatibility Workshop in Turin, Italy, (Ceppellini et al, 1967), to describe alleles in the HLA system at the neighbouring HLA-A and B loci, which segregated together in family

studies. Ceppellini stated “a new term can be introduced without increasing confusion, it is suggested to substitute pheno-group with haplotype.” (Petersdorf,2017).

While the word haplotype still has an unambiguous meaning to those working primarily with the Major Histocompatibility Complex (MHC) the word has taken on a new meaning over the succeeding years, when describing polymorphisms of non-MHC genes. The original definition of haplotypes has been changed over the succeeding 60 years to the extent that two SNPs in one allele are now referred to as a haplotype (Davidson, 2000). In order to reduce any confusion that may arise over the use of the word “haplotype” In most instances I will refer to two linked polymorphisms as in genetic cis phase. Where two polymorphisms have been shown to be on homologous chromosomes, they are said to be in trans phase.

The phase of any genetic polymorphism whether they be two or more SNPs or two alleles is critical to understand, in order to fully appreciate the effect that genetic polymorphism can have on function. The fundamental reason for this is that proteins are coded for by gene sequences in cis phase. A genotype merely gives the sum of polymorphisms that exist in two inherited gene sequences, but tell us nothing about the functional protein sequences.

I will attempt in this review to describe methods for assessing phase, their advantages and drawbacks and then provide three examples in clinical medicine where I believe phase is critical in ascertaining true effects of therapy, but currently is not routinely performed.

## **2. Assessing genetic phase-current approaches**

### **2.1. Family studies**

Family studies to determine gene segregation within a nuclear family context is still the “gold standard”, having been since the early 1970s to demonstrate linkage of HLA A and B alleles (Ceppellini et al, 1974). Despite the unambiguous results obtained, in most cases family studies are generally not performed for non-MHC genes, particularly in a research setting, the main reasons being logistical and the cost involved.

### **2.2 Linkage Disequilibrium (LD)**

When two genes are closely linked, and occur together more often than is expected from their individual allele frequencies, the alleles are said to be in positive (LD). LD therefore can be used to predict genetic phase in the absence of family studies.

Such a mathematical approach has been used successfully in the HLA system to calculate the probability of haplotypes within the US Bone Marrow Registry (Mori et al ,1997). At the time of the analysis there were in excess of 1.35 million HLA typed donors registered. By examining the A, B, DR phenotypes of 406,503 donors they observed 302,867 distinguishable HLA-A, -B, -DR phenotypes. Such large numbers permitted the assignment of haplotypes even at relatively low frequency. In a separate analysis of bone marrow donors Schmidt et al (2009) performed a similar estimation of haplotype frequency, using HLA-A, B, C and DR donors typed at the allele level. Using this information, the authors were able to predict the probability of a given recipient finding a matched donor in a registry of a given size.

The sequencing of the human genome in 2001 (Venter et al, 2001, Lander et al 2001) and the mapping of single nucleotide polymorphisms (SNPs) necessitated the need for defining haplotypes across the entire genome. As with the HLA system a mathematical approach to haplotyping was used to measure linkage disequilibrium and several methods have been developed. A full review of mathematical models of genomic haplotyping are summarised by Bonizzoni et al, 2003.

The genome consists of haplotype blocks, therefore a single nucleotide polymorphism can act as a marker for a disease gene as a result of LD, if it is located in the same haplotype block. One of the issues surrounding the use of tag SNPs is the variation observed in L.D between racial and ethnic groups. (Di Paolo 2005, Du et al 2007, Seyerle et al, 2014).

Despite this observation and despite other inherent risks (Hodge et al, 1999, Barton et al, 2001, Reich et al 2001) many papers appear in the literature relying on linkage disequilibrium to assign putative SNP haplotypes, or searching for SNP haplotype disease markers (Hohjoh et al, 1999, Mori et al, 2001, Yang et al, 2020, Hao et al, 2020.)

The use of LD is the most common one used when analysing two or more single nucleotide polymorphisms throughout the genome. In many cases LD provides a high level of accuracy in the absence of definitiveness (Hodge et al, 1999). Another issue to consider when applying LD to assign disease haplotypes is that at least as observed in the HLA system, the pattern of LD observed in disease can be different from that observed in controls (Tait et al,1995, Hanifi Moghaddam et al,1998) when selective pressure is applied to two neighbouring genetic regions, which presents a problem in relying on LD for definitive haplotyping in disease studies.

### **2.3. Long range sequencing**

Firstly, Next Generation Sequencing (NGS) and then long-range sequencing (LRS) have provided a platform, in combination with other unique technical and software approaches for defining haplotypes (Weitzel et al,2007, Peters et al, 2012, Hosomitchi et al, 2015, Mantere et al, 2019, Midha et al, 2019, Chintalaphani et al, 2021) which was not possible with conventional sequencing.

Although there have been reports of longer sequencing, the most common nucleotide length reported is of the order of 10-20 kb. In the future it is to be expected that the reliable reproducible range will increase. A technique called Hi-C has been added to the sequencing armamentarium which makes it possible to sequence entire chromosomes in haploid phase (Lieberman-Aiden et al, 2009). Used in conjunction with Falcon-phase up to 90% accuracy in phasing human haplotypes was achieved (Kronenberg et al, 2021). LRS and Hi-C require a deal of sophisticated software to achieve accurate phase.

### **2.4 Single chromosome sequencing**

Only accuracy in excess of 98% is satisfactory for phasing in clinical medicine. Conceptually, single chromosome sequencing would replace all the above techniques and provide unambiguous haploid information. Unfortunately to date this has not been achieved in a reproducible manner. In 2011 Christina Fan, Stephen Quake and co-workers from the Bioengineering Department, Stanford University US, (Fan et al, 2011) described a microfluidic devise which was able to trap single metaphase cells, lyse, protease digest the cytoplasm and collect single chromosomes which could be amplified by multiple strand displacement amplification (MSDA) and then sequenced. This was an exciting development which does not seem to have been

pursued, at least to the market stage, by individual research groups, or commercial companies. While single cell sequencing has been pursued, the determination of genetic phase or isolation of single chromosomes has not been the focus of further work for this group (Gawad et al, 2016).

### ***3.Examples in Clinical Medicine where knowledge of genetic phasing is important***

#### ***3.1 Haematopoietic Stem Cell Transplantation (Bone Marrow Transplantation)***

The term Bone Marrow Transplantation refers to the time when this procedure was restricted to the use of syngeneic or allogeneic adult bone marrow to treat patients with leukemias and other blood disorders whose bone marrow is depleted either as a result of a disease (aplastic anaemia) or treatment for conditions such as leukemias. The use of cord blood or stem cell mobilised peripheral blood has broadened the definition so that haematopoietic stem cell transplantation (HSCT) is now the preferred term.

The first procedure to find an HLA compatible donor for a patient is to search within the nuclear family. Considering each parent has two HLA haplotypes, the chance of a sibling inheriting the two HLA haplotypes of the patient is 1:4. Given the average size of families in western countries is comparatively small, compared with previous generations, only approximately 30% of patients have an HLA matched sibling.

HLA cis phase (in the HLA field this is commonly known as haplotypes) is established for the patient at the time of family studies. However, the World Marrow Donor Association (WMDA) has recently (July 2021) reported over 39 million donors, both adult and

cord, registered on their database (World Marrow Donor Association <https://wmda.info/>), the overwhelming majority of whom do not have HLA haplotypes established. This means when the database is searched the primary aim is to find donors who are matched at multiple HLA loci (usually A, B, C, DRB1, DQA1, DQB1 and sometimes DPB1). No information is provided on haplotype matching.

What is the evidence that patients who are transplanted with a haplotype matched unrelated donor have superior clinical outcome compared with patients who receive an HLA allele matched donor?

The Perth group led by Roger Dawkins and Frank Christiansen were at the forefront of defining and understanding the nature of HLA haplotypes. They demonstrated that HLA haplotypes are generated by combining segments of ancestral haplotypes (Christiansen et al 1991, Degli-Esposti 1992). They also demonstrated that recombination was not random across the HLA region of the MHC, and that haplotypes could be defined as blocks of DNA which are recombined to form new haplotypes. The regions were referred to as alpha, beta, gamma delta and epsilon blocks. The alpha block includes HLA-A, the beta HLA-B and HLA-C, the gamma the class 3 region which includes complement genes, the delta the class 2 DR and DQ genes and epsilon which includes the DP genes. Using microsatellite (Msat) markers specific for each segment, the group were able to show that HLA matched pairs were not necessarily haplotype matched.

The main criticism of the work, which can be made retrospectively, is that matching was performed largely by serology and not sequencing. However, in 2006 a follow up paper using PCR primers for the beta block showed that in the presence of sequence identity

at both HLA-B and C genes there was evidence of haplotype disparity (Kitcharoen et al, 2006).

Further evidence was provided in three seminal papers by Effie Petersdorf's group from the Fred Hutchison Centre in Seattle. The first (Malkki, 2005) described Msats specific for HLA haplotypes which had the obvious potential of providing haplotype matching in HSCT. The second (Guo, 2006) described a method for separating haplotypes using genomic DNA from individuals of known HLA genotype. Probes specific for the two HLA-B alleles were adhered to a glass surface and hybridised with the genomic DNA. After washing off excess DNA the two captured DNA fragments were placed in separate tubes and PCR primers were used to amplify the HLA-A and DRB1 genes.

Using this method of physical separation Petersdorf et al were able to study the clinical effect of HLA haplotype matching in 246 unrelated donor HSCT (Petersdorf et al 2007). Of the 246 recipients who were all HLA-A, B, C, DRB1, and DQB1 matched with their donors, 191 (78%) were haplotype matched with their donors while 55 (22%) were mismatched.

There was a strong association of HLA haplotype mismatching and increased levels of acute GVHD Grades 3 and 4. The effect was seen in the first few days, maximised at 20 days and then plateaued so that at 100 days the difference in incidence of severe acute GVHD was of the order of 35% (approx. 25% in the matched group versus approx. 60% in the mismatched group). In addition, a significant difference was seen in survival rates between the two groups.

These results demonstrated clearly the value of HLA haplotype matching in A, B, C DR, DQ donor/recipient allele matched pairs in HSCT. It is relevant to point out that HLA-C and DPB1 were not

considered in this study, however it is unlikely that the differences seen could be explained by incompatibilities at the DPB1 locus. Surprisingly, there has been no follow up by other groups to use this method as a way of determining haplotype matching.

### **3.2. Immunotherapy -check point inhibitors (CPI)**

One of the more exciting developments over the last few years has been the application of CPIs to inhibit the growth of cancer cells (Zhang et al, 2021). One of the more successful examples is the clinical use in metastatic melanoma of the monoclonal antibody Ipilimumab which targets the human CTLA- 4 (CD 152) receptor (Langer et al, 2007).

When antigen is presented via MHC molecules to the responding T cells, two co-stimulatory molecules expressed on the T cell can either amplify (CD28), or inhibit the T cell response (CTLA-4 called CD152). CTLA-4 binds to B7-1 (CD 80) and B7-2 (CD86) as part of the inhibitory mechanism pathway. Blockade of CTLA-4 therefore has the effect of amplifying the anti-tumour response, first described by James Allison (Leach et al, 1996) for which he and Tasuku Honjo received the Nobel Prize in 2018, and has subsequently been shown to be particularly effective in melanoma (Attia et al, 2005, Hodi et al, 2010), which led to FDA approval for clinical use in 2011, making it the first approval of a CPI for use in human cancer treatment.

With respect to the question of genetic phase it is noteworthy that CTLA-4 is polymorphic (Hodi et al, 2010). Several polymorphic sites have been identified within the non-coding and leader peptide coding region of the gene which appear to influence both expression and cancer risk (Wagner et al, 2021). These findings suggest that these disease associations are primarily due to differences in gene control and expression, but do not exclude the possibility that as yet

unidentified coding variants may interfere with binding of CTLA-4 to its ligand.

Elegant crystallography studies have revealed the molecular interaction between Ipilimumab and the CTLA-4 receptor (Ramagopal, 2017). Critical residues of CTLA-4 involved in binding, as shown by creating mutants, are S20, R35, R40, Q76, D88, K95, E97, Y104, L106, and I108 none of which have been shown to date to be polymorphic. However in other antibody /antigen systems non-conservative changes in amino acid sequence which are not part of the binding site can be responsible for altering the structure of the antibody binding site which in turn compromises antibody binding. The full impact of the described polymorphisms will not be fully appreciated until accurate SNP phasing is established and a range of ethnic groups are studied. The importance of SNP phasing in CTLA-4 has been demonstrated in a study of liver cancer in Han Chinese. Different SNP haplotypes were shown to be both protective and susceptibility markers (Yang et al 2019).

### **3.3 Micro RNA molecules**

Non -coding DNA was originally thought to be “junk” DNA. However, since only approximately 2% of human DNA codes for proteins (Elgar and Vavouri, 2008), it is difficult to imagine that nature invests in this degree of redundancy, in the absence of a useful biological purpose.

The discovery of both long non-coding RNA and microRNA as essential components of the gene regulation network has been one of the most significant and exciting discoveries of the past few years, which has implications for a host of human diseases. In this review I will concentrate on the micro- RNA molecules, their known polymorphisms, the implications on gene expression, and the

importance of genetic phase on understanding the precise effects these molecules have on biological processes.

Micro RNA molecules (miRNA) are single stranded RNA molecules, of usually 22 nucleotides long which act at the post transcription stage by binding onto part of the 3' regulatory region of the gene, which inhibits the translation of the gene. The first report of miRNA was described in the bacteria *Caenorhabditis elegans* (Lau et al, 2001). Lau and colleagues described 55 miRNA molecules, the expression of which varied at different stages of development. In the same issue of the journal *Science*, Lagos-Quintana described miRNAs in both invertebrate and vertebrates (Lagos-Quintana M, 2001). Zeng et al demonstrated that miRNAs were produced *in vivo* in human cells and had the capacity to regulate the expression of genes (Zeng et al, 2002). According to the latest version of miRBase version 22 ([miRNA Database](#), 2021), there are now 1,881 loci which code for mature miRNA molecules. This therefore represents an enormously large and complex regulation of gene expression.

From the viewpoint of genetic phase, there is convincing data that despite the evolutionary conserved nature of these molecules there is documented polymorphisms in miRNA (Carbonell et al, 2012) There is now ample evidence in a range of different diseases including pre-eclampsia (Li et al, 2021), type 2 diabetes (Yan et al, 2021), lung fibrosis after lung transplantation (Wang et al, 2020), non-alcoholic fatty liver disease (Lopez-Sanchez et al, 2021), liver cancer (Ghafouri-Fard et al, 2021) and multiple sclerosis (Shareef et al, 2021) ,that polymorphisms in the relevant miRNA can confer susceptibility and determine the effectiveness of suppression of gene expression. In addition, polymorphisms in the miRNA binding domain of target gene regions play a role in diseases such as colorectal cancer where SNPs in the solute carrier transporter gene predict

clinical outcome in colorectal cancer (Bendova et al, 2020) and the KRT81 gene in breast cancer (Sha et al 2020). There are numerous other examples in this ever- expanding field of research (Vohra et al, 2020, Liu et al, 2021).

This effect on gene expression has created a new paradigm for explaining at a functional level, associations of disease with gene alleles. For example, the MHC contains a greater density of disease associations than any other part of the human genome (Clark et al, 2015). This group identified 89 miRNA transcripts (Clark et al, 2018) within the MHC, approximately 50% of which lie in linkage disequilibrium blocks that contain disease related SNPs. These data hint that at least some HLA and disease associations are functionally related to levels of expression. An example is a SNP in the upstream region of HLA-C (-35C/T) which is associated with a high level of HLA-C expression and HIV+ individuals with the T allele progress more slowly to AIDS (Thomas et al, 2009). The authors interpreted the results in terms of more effective viral peptide presentation to cytotoxic T cells. The C allele is a proxy for high HLA-C mRNA allele expression which in turn makes for more effective control of HIV infection. About 3.5 million years ago, a sequence exchange event occurred between HLA-C and HLA-B giving rise to a variant which escaped miRNA regulation, the marker being the SNP at position-35. This sequence exchange resulted in 7 of the 14 HLA-C lineages identified today (O’Uigin et al 2011). The important point to note from the viewpoint of phase is that the --35 SNP is in linkage disequilibrium with other polymorphisms which have been shown to directly affect the binding of miRNA 148a, such as the deletion at position 263 downstream of the stop codon of HLA-C.

Location of the target gene does not inform where a miRNA which influences expression is located. For, example, in an elegant series of

experiments ( Rooda et al, 2020) studying follicular development of ovarian cells examined two previously identified miRNA (Velthut-Meikas et al ,2013) and derived from the introns of two genes, namely FSHR (follicular stimulating hormone receptor), located on chromosome 2 and CYP19A1, (cytochrome P450 family 19 subfamily A number 1) located on chromosome 15. These two miRNA molecules, namely miRNA 548bc and miRNA 7973 respectively, influence the expression of at least seven genes. miRNA 548bc influences expression of the genes Neogenin, a member of the immunoglobulin superfamily NEO1, (chromosome 15), LIFR, Leukemia Inhibitory Factor Receptor (chromosome 5), PTEN Phosphatase Tensin Homologue (chromosome 10) and SP110 Nuclear Body Protein (chromosome 2). miRNA 7973 influences the expression of ADAM 15 which is a member of the metalloproteinase family ADAM and is located on chromosome 1, FMNL3 a formin like gene found on chromosome 12 and FXDN, Peroxidasin gene found on chromosome 2. Further complexity in interpretation is added by the fact these regulatory elements act in both a cis and trans configuration on their target genes (Elcheva et al, 2020).

The discovery of the complexity of gene expression and the role that non coding DNA plays in this process, is arguably one of the most exciting and significant developments of this century. Control of gene expression via miRNA and other molecules will take years to fully understand. Elucidation of genetic phase in both polymorphic miRNA and their gene targets will play a vital role, and until genetic phase is definitively determined, the full story surrounding gene expression will not be fully understood.

## References

Attia, P., Phan, G.Q., Maker, A.V., Robinson, M.R., Quezado, M.M., Yang, J.C., Sherry, R.M., Topalian, S.L., Kammula, U.S., Royal, R.E., Restifo, N.P., Haworth, L.R., Levy, C., Mavroukakis, S.A., Nichol, G., Yellin, M.J. and Rosenberg, S.A. (2005). Autoimmunity correlates with tumor regression in patients with metastatic melanoma treated with anti-cytotoxic T-lymphocyte antigen-4. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, [online] 23(25), pp.6043–6053. Available at: <https://pubmed.ncbi.nlm.nih.gov/16087944/>.

Barton, A., Chapman, P., Myerscough, A., Pinel, T., Davies, N., Worthington, J. and John, S. (2001). The Single-Nucleotide Polymorphism Lottery: How Useful are a Few Common SNPs in Identifying Disease-Associated Alleles? *Genetic Epidemiology*, 21(S1), pp.S384–S389.

Bendova, P., Pardini, B., Susova, S., Rosendorf, J., Levy, M., Skrobanek, P., Buchler, T., Kral, J., Liska, V., Vodickova, L., Landi, S., Soucek, P., Naccarati, A., Vodicka, P. and Vymetalkova, V. (2020). Genetic variations in microRNA-binding sites of solute carrier transporter genes as predictors of clinical outcome in colorectal cancer. *Carcinogenesis*, 42(3), pp.378–394.

Bonizzoni, P., Della Vedova, G., Dondi, R. and Li, J. (2003). The Haplotyping problem: An overview of computational models and solutions. *Journal of Computer Science and Technology*, 18(6), pp.675–688.

Carbonell, J., Alloza, E., Arce, P., Borrego, S., Santoyo, J., Ruiz-Ferrer, M., Medina, I., Jiménez-Almazán, J., Méndez-Vidal, C., González-del Pozo, M., Vela, A., Bhattacharya, S.S., Antiñolo, G. and Dopazo, J.

(2012). A map of human microRNA variation uncovers unexpectedly high levels of variability. *Genome Medicine*, 4:pp 62-78.

<https://doi.org/10.1186/gm363>

Ceppellini, R., Curtoni, ES., Mattiuz, PL., Miggiano, C., Scudeller, G., Serra, A. Genetics of leukocyte antigens: a family study of segregation and linkage. In: Curtoni, ES.Mattiuz, PL., Tosi, RM., editors. *Histocompatibility Testing*. Copenhagen; Munksgaard: 1967. p. 149-187.

Ceppellini, R., Van Rood, J.J. (1974). The HL-A system. Genetics and Molecular Biology. *Seminars in Hematology*, 11 (3), pp.233-251.

Chintalaphani, S.R., Pineda, S.S., Deveson, I.W. and Kumar, K.R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathologica Communications*, 9,98. <https://doi.org/10.1186/s40478-021-01201-x>

Christiansen FT, Witt CS, Dawkins RL. (1991) Questions in marrow matching: the implications of ancestral haplotypes for routine practice *Bone Marrow Transplant* 8 pp 83-86.

Clark, P. M., Kunkel, M. & Monos, D. S. (2015). The dichotomy between disease phenotype databases and the implications for understanding complex diseases involving the major histocompatibility complex. *International journal of immunogenetics* 42, pp 413–422.

Clark, P.M., Chitnis, N., Shieh, M., Kamoun, M., Johnson, F.B. and Monos, D. (2018). Novel and Haplotype Specific MicroRNAs Encoded by the Major Histocompatibility Complex. *Scientific Reports*, 8(1). pp 1-10. DOI:10.1038/s41598-018-19427-6

Davidson, S. (2000). Research suggests importance of haplotypes over SNPs. *Nature Biotechnology*, 18(11), pp.1134–1135.

Degli-Esposti, MA, Leaver AL, Christiansen FT, Witt CS, Abraham LJ, Dawkins RL. (1992) Ancestral haplotypes conserved population MHC haplotypes *Hum. Immunol.* 34 (40 pp242-252.

Di Paola, J., Jugessur, A., Goldman, T., Reiland, J., Tallman N, D., Sayago, C. and Murray, J.C. (2005). Platelet glycoprotein I balpha and integrin alpha2beta1 polymorphisms: gene frequencies and linkage disequilibrium in a population diversity panel. *Journal of Thrombosis and Haemostasis*, 3(7), pp.1511–1521.

Du, J., Shi, Y., Zhang, A., Wang, L., Xuan, J., He, G., Xu, L., Xu, M., Shu, A., Yu, L., Li, X., Meng, J., Feng, G., Xing, Q. and He, L. (2007). Screening for SNPs and haplotypes in the CYP3A7 gene in Chinese populations. *Pharmacogenomics*, 8(6), pp.559–566.

Elcheva, I.A. and Spiegelman, V.S. (2020). The Role of cis- and trans-Acting RNA Regulatory Elements in Leukemia. *Cancers*, 12(12), pp.3854-3884.

Elgar, G. and Vavouri, T. (2008). Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends in Genetics*, 24(7), pp.344–352.

Fan, H.C., Wang, J., Potanina, A. and Quake, S.R. (2011). Whole-genome molecular haplotyping of single cells. *Nature Biotechnology*, 29(1), pp.51–57.

Gawad, C., Koh, W. and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, [online] 17(3), pp.175–188. Available at: <https://www.nature.com/articles/nrg.2015.16>.

Ghafouri-Fard, S., Honarmand Tamizkar, K., Hussien, B.M. and Taheri, M. (2021). MicroRNA signature in liver cancer. *Pathology - Research and Practice*, 219:153369.

<https://doi.org/10.1016/j.prp.2021.153369> PP 1-20.

Guo, Z., Hood, L., Malkki, M. and Petersdorf, E.W. (2006). Long-range multilocus haplotype phasing of the MHC. *Proceedings of the National Academy of Sciences*, 103(18), pp.6964–6969.

Hanifi Moghaddam, P., de Knijf, P., Roep, B.O., Van der Auwera, B., Naipal, A., Gorus, F., Schuit, F. and Giphart, M.J. (1998). Genetic structure of IDDM1: two separate regions in the major histocompatibility complex contribute to susceptibility or protection. Belgian Diabetes Registry. *Diabetes*, 47(2), pp.263–269.

Hao, X., Wang, A., Li, C., Shao, L., Li, Y. and Yang, P. (2020). Genetic association of BIN1 and GAB2 in Alzheimer's disease: A meta-analysis and systematic review. *Geriatrics & Gerontology International*, 21(2), pp.185–191.

Hodge, S.E., Boehnke, M. and Spence, M.A. (1999). Loss of information due to ambiguous haplotyping of SNPs. *Nature Genetics*, 21(4), pp.360–361.

Hodi, F.S., O'Day, S.J. Mc Dermott, D.F. ,Weber, R.W, Sosman, J.A., J Haanen, J.B., Gonzalez, R. , Robert, C.,Schadendorf, D., Hassel, J.C., Akerley, W., van den Eertwegh, A.J.M., Lutzky, J., Lorigan, P., Vaubel, J.M., Linette, G.P., Hogg, D., Ottensmeier, C.H., Lebbé, C., Peschel, C., Quirt, I., Clark, J.I., Wolchok, J.D., Weber, J.S., Tian, J., Yellin, M.J., Nichol, G.M., Hoos, A., and Urba, W.J.(2010) Improved Survival with Ipilimumab in Patients with Metastatic. Melanoma. *New England Journal of Medicine*, 363(8) pp.711-723.

Hohjoh, H., Nakayama, T., Ohashi, J., Miyagawa, T., Tanaka, H., Akaza, T., Honda, Y., Juji, T. and Tokunaga, K. (1999). Significant association of a single nucleotide polymorphism in the tumor necrosis factor-alpha (TNF- $\alpha$ ) gene promoter with human narcolepsy. *Tissue Antigens*, 54(2), pp.138–145.

Hosomichi, K., Shiina, T., Tajima, A. and Inoue, I. (2015). The impact of next-generation sequencing technologies on HLA research. *Journal of Human Genetics*, 60(11), pp.665–673.

Kitcharoen, K., Witt, C.S., Romphruk, A.V., Christiansen, F.T. and Leelayuwat, C. (2006). MICA, MICB, and MHC Beta Block Matching in Bone Marrow Transplantation: Relevance to Transplantation Outcome. *Human Immunology*, 67(3), pp.238–246.

Kronenberg, Z.N., Rhie, A., Koren, S., Concepcion, G.T., Peluso, P., Munson, K.M., Porubsky, D., Kuhn, K., Mueller, K.A., Low, W.Y., Hiendleder, S., Fedrigo, O., Liachko, I., Hall, R.J., Phillippy, A.M., Eichler, E.E., Williams, J.L., Smith, T.P.L., Jarvis, E.D. and Sullivan, S.T. and Kingan, S.B. (2021). Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C. *Nature Communications*, 12:1935 <https://doi.org/10.1038/s41467-020-20536-y> . PP1-10

Lagos-Quintana, M. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. *Science*, 294(5543), pp.853–858.

Lander, E.S. et al International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, 409 pp 860-921.

Langer, L.F., Clay, T.M. and Morse, M.A. (2007). Update on anti-CTLA-4 antibodies in clinical trials. *Expert Opinion on Biological Therapy*, 7(8), pp.1245–1256.

Lau, N.C., Lim, L.P., Weinstein, E. G. and Bartel, D. P. (2001). An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*. *Science*, 294(5543), pp.858–862.

Leach, D.R., Krummel, M.F. and Allison, J.P. (1996). Enhancement of Antitumor Immunity by CTLA-4 Blockade. *Science*. 271(5256), pp.1734–1736.

<https://science.sciencemag.org/content/271/5256/1734>.

Li, T., Chen, Y., Lai, Y., He, G. and He, G. (2021). The relationship between polymorphisms of microRNA and preeclampsia. *Medicine*, 100(14), p.e25477. DOI 10.17605/OSF.IO/MJY2X.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S. and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, [online] 326(5950), pp.289–93. Available at:

<https://www.ncbi.nlm.nih.gov/pubmed/19815776>.

Liu, C.-J., Fu, X., Xia, M., Zhang, Q., Gu, Z. and Guo, A.-Y. (2021). miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucleic Acids Research*, 49 (Database Issue), pp.D1276–D1281. doi: 10.1093/nar/gkaa783.

López-Sánchez, G.N., Dóminguez-Pérez, M., Uribe, M., Chávez-Tapia, N.C. and Nuño-Lámbarri, N. (2021). Non-alcoholic fatty liver disease and microRNAs expression, how it affects the development and progression of the disease. *Annals of Hepatology*, 21:100212. pp 1-7.

[Http://doi.org/10.10106/j.aohep.2020.04.02](http://doi.org/10.10106/j.aohep.2020.04.02)

Malkki, M., Single, R., Carrington, M., Thomson, G. and Petersdorf, E. (2005). MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: implications for unrelated donor hematopoietic transplantation and disease association studies. *Tissue Antigens*, 66(2), pp.114–124.

Mantere, T., Kersten, S. and Hoischen, A. (2019). Long-Read Sequencing Emerging in Medical Genetics. *Frontiers in Genetics*, 10:426. doi: 10.3389/fgene.2019.00426

Midha, M.K., Wu, M. and Chiu, K.-P. (2019). Long-read sequencing in deciphering human genetics to a greater depth. *Human Genetics*, 138(11-12), pp.1201–1215.

miRNA database [miRBase](#).

Mori, M., Beatty, P.G., Graves, M., Boucher, K.M. and Milford, E.L. (1997). HLA Gene and Haplotype Frequencies in the North American population. *Transplantation*, 64(7), pp.1017–1027.

Mori, H., Okazawa, H., Iwamoto, K., Maeda, E., Hashiramoto, M. and Kasuga, M. (2001). A polymorphism in the 5' untranslated region and a Met 229 - Leu variant in exon 5 of the human UCP1 gene are associated with susceptibility to Type II diabetes mellitus. *Diabetologia*, 44(3), pp.373–376.

O'hUigin, C., Kulkarni, S., Xu, Y., Deng, Z., Kidd, J., Kidd, K., Gao, X. and Carrington, M. (2011). The Molecular Origin and Consequences of Escape from miRNA Regulation by HLA-C Alleles. *The American Journal of Human Genetics*, 89(3), pp.424–431.

Peters, B.A., Kermani, B.G., Sparks, A.B., Alferov, O., Hong, P., Alexeev, A., Jiang, Y., Dahl, F., Tang, Y.T., Haas, J., Robasky, K., Zaranek, A.W., Lee, J.-H., Ball, M.P., Peterson, J.E., Perazich, H., Yeung, G., Liu, J., Chen, L. and Kennemer, M.I. (2012). Accurate

whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature*, 487(7406), pp.190–195.

Petersdorf, E.W., Malkki, M., Gooley, T.A., Martin, P.J. and Guo, Z. (2007). MHC Haplotype Matching for Unrelated Hematopoietic Cell Transplantation. *PLoS Medicine*, 4(1), e8.pp59-68.

Petersdorf, E.W. (2017). In celebration of Ruggero Ceppellini: HLA in transplantation. *HLA*, 89(2), pp.71–76.

Ramagopal, U.A., Liu, W., Garrett-Thomson, S.C., Bonanno, J.B., Yan, Q., Srinivasan, M., Wong, S.C., Bell, A., Mankikar, S., Rangan, V.S., Deshpande, S., Korman, A.J. and Almo, S.C. (2017). Structural basis for cancer immunotherapy by the first-in-class checkpoint inhibitor ipilimumab. *Proceedings of the National Academy of Sciences*, 114(21), pp.E4223–E4232.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R., & Lander, E.S. (2001). Linkage disequilibrium in the human genome. *Nature* 411 pp 199-204.

Rooda, I., Hensen, K., Kaselt, B., Kasvandik, S., Pook, M., Kurg, A., Salumets, A. and Velthut-Meikas, A. (2020). Target prediction and validation of microRNAs expressed from FSHR and aromatase genes in human ovarian granulosa cells. *Scientific Reports*, 10(1). P 2300. <https://doi.org/10.1038/S41598-020-59186-X>

Schmidt, A.H., Baier, D., Solloch, U.V., Stahr, A., Cereb, N., Wassmuth, R., Ehninger, G. and Rutt, C. (2009). Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Human Immunology*, 70(11), pp.895–902.

Seyerle, A.A., Young, A.M., Jeff, J.M., Melton, P.E., Jorgensen, N.W., Lin, Y., Carty, C.L., Deelman, E., Heckbert, S.R., Hindorff, L.A., Jackson, R.D., Martin, L.W., Okin, P.M., Perez, M.V., Psaty, B.M., Soliman, E.Z., Whitsel, E.A., North, K.L., Laston, S., Kooperberg, C. and Avery, C.L. (2014). Evidence of Heterogeneity by Race/Ethnicity in Genetic Determinants of QT Interval. *Epidemiology* 25(6) pp 790-798.

Sha, Z., Lai, R., Zhang, X., Zhao, Y., Wu, J., Geng, C. and Guo, Z. (2020). A Polymorphism at the microRNA Binding Site in the 3' Untranslated Region of KRT81 Is Associated with Breast Cancer. *DNA and Cell Biology*.39(10) pp1886-1894.

Shareef, S., Ebrahim, S.O., Reisi, S. (2021). Contribution of hsa-miR-146a and hsa-miR-223 gene variations in patients with multiple sclerosis reveals association of rs2910164 and rs1044165 with risk of multiple sclerosis susceptibility. *Journal of Investigative Medicine; London* Vol. 69 (5), pp 1015-1021.

Tait, B.D., Drummond, B.P., Varney, M.D. and Harrison, L.C. (1995). HLA-DRB1\*0401 is associated with susceptibility to insulin dependent diabetes mellitus independently of the DQB1 locus. *European Journal of Immunogenetics*, 22(4), pp.289–297.

Thomas, R., Apps, R., Qi, Y., Gao, X., Male, V., O'hUigin, C., O'Connor, G., Ge, D., Fellay, J., Martin, J.N., Margolick, J., Goedert, J.J., Buchbinder, S., Kirk, G.D., Martin, M.P., Telenti, A., Deeks, S.G., Walker, B.D., Goldstein, D. and McVicar, D.W., Moffet A., and Carrington M. (2009). HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nature Genetics*, 41(12), pp.1290–1294.

Velthut-Meikas, A., Simm, J., Tuuri, T., Tapanainen, J.S., Metsis, M. and Salumets, A. (2013). Research Resource: Small RNA-seq of

Human Granulosa Cells Reveals miRNAs in FSHR and Aromatase Genes. *Molecular Endocrinology*, 27(7), pp.1128–1141.

Venter JC, Adams MD, Myers EW et al (2001). The sequence of the human genome. *Science*.16;291(5507), pp 1304-51.

Vohra, M., Sharma, A.R., Prabhu B, N. and Rai, P.S. (2020). SNPs in Sites for DNA Methylation, Transcription Factor Binding, and miRNA Targets Leading to Allele-Specific Gene Expression and Contributing to Complex Disease Risk: A Systematic Review. *Public Health Genomics*, 23(5-6), pp.155–170.

Wagner, M., Jasek, M. and Karabon, L. (2021). Immune Checkpoint Molecules—Inherited Variations as Markers for Cancer Risk.

*Frontiers in Immunology*, 11:

<https://doi.org/10.3389/fimmu.2020.606721>

Wang, C., Zhang, X., Ling, Q., Zheng, S. and Xu, X. (2020). A model integrating donor gene polymorphisms predicts fibrosis after liver transplantation. *Aging*, 13(1), pp.1264–1275.

Weitzel, J.N., Lagos, V.I., Herzog, J.S., Judkins, T., Hendrickson, B., Ho, J.S., Ricker, C.N., Lowstuter, K.J., Blazer, K.R., Tomlinson, G. and Scholl, T. (2007). Evidence for Common Ancestral Origin of a Recurring BRCA1 Genomic Rearrangement Identified in High-Risk Hispanic Families. *Cancer Epidemiology Biomarkers & Prevention*, 16(8), pp.1615–1620.

World Marrow Donor Association (WMDA) <https://wmda.info/>

Yan, Y., Xiao, H., Zhang, J., Wang, S., Dong, J. and Wu, L. (2021). Pri-miR-144 rs9279 is associated with type 2 diabetes and regulation of stress response. *Journal of Cellular Physiology*, 236(1), pp.561–569.

Yang, J., Liu, J., Chen, Y., Tang, W., Liu, C., Sun, Y. and Chen, J. (2019). Association of CTLA-4 tagging polymorphisms and haplotypes with hepatocellular carcinoma risk. *Medicine*, 98(29), e16266. pp1-6.

Yang, G., Collins, J.M., Rafiee, R., Singh, S., Langae, T., McDonough, C.W., Holliday, L.S., Wang, D., Lamba, J.K., Kim, Y.S., Pelliccioni, G.A., Vaszilko, M., Kosa, J.P., Balla, B., Lakatos, P.A., Katz, J., Moreb, J. and Gong, Y. (2020). SIRT1 Gene SNP rs932658 Is Associated With Medication-Related Osteonecrosis of the Jaw. *Journal of Bone and Mineral Research*, 36(2), pp.347–356.

Zeng, Y., Wagner, E.J. and Cullen, B.R. (2002). Both Natural and Designed Micro RNAs Can Inhibit the Expression of Cognate mRNAs When Expressed in Human Cells. *Molecular Cell*, 9(6), pp.1327–1333.

Zhang, H., Dai, Z., Wu, W., Wang, Z., Zhang, N., Zhang, L., Zeng, W.-J., Liu, Z. and Cheng, Q. (2021). Regulatory mechanisms of immune checkpoints PD-L1 and CTLA-4 in cancer. *Journal of Experimental & Clinical Cancer Research* 40:184 pp 1-22.