Detecting Higher Order Genomic Variant Interactions with Spectral Analysis

David Uminsky
Mathematics and Statistics
University of San Francisco
San Francisco, CA 94117 USA

Mario Banuelos
Mathematics
California State University, Fresno
Fresno, CA 93740 USA

Lillian González-Albino
Computer Science
University of Puerto Rico, Río Piedras
San Juan, 00931, Puerto Rico

Rosa Garza
Mathematics and Statistics
California State University, Monterey Bay
Seaside, CA 93955 USA

Sylvia Akueze Nwakanma
Mathematics
Pomona College
Claremont, CA 91711 USA

***Abstract*—Genomic variations among a species consisting of one nucleotide change are known as single nucleotide polymorphisms (SNPs). Often these mutations result in a change in phenotype, but detecting higher order interaction of multiple SNPs remains a challenging problem. Common approaches to find groups of interacting SNPs associated with a phenotypic response, a problem under the umbrella of epistasis, often suffers from a combinatorial explosion and require Bonferroni or similar corrections. In this work, we develop and apply a novel Fourier transformation on the symmetric group to uncover higher order interactions of SNPs associated with a quantitative phenotypic response. We present results for simulated data and then apply our method to previously published data to detect, for the first time using a signal processing approach, new and statistically significant higher order SNP interaction phenotypes related to muscle mice genomic variants.**

***Index Terms*—Spectral analysis, Fourier transform, algebraic signal processing, epistasis, genomic variation**

I. Introduction

Although advances in DNA sequencing technologies have led to large sequencing studies [1]–[3], understanding how genomic variation and interaction influence phenotypes remains a challenging problem. This variation consists of a single basepair change at the population level, known as single nucleotide polymorphisms (SNPs), or rearrangements of larger regions, termed structural variants (SVs) [4]–[6]. For SNPs, these are often encoded measured with a binary response (0 or 1 copy of the variant) and either a quantitative or qualitative phenotypic response (e.g. presence of a disease or hemoglobin count) is recorded. For some genomic mutations, studies have demonstrated how certain variants may contribute to cancer susceptibility [7], [8]. At the same time, recent work has shown how recent variation leads to increased fitness [9], [10].

Methods to detect SNPs have been well studied and over the past several decades, thousands of SNPs have been associated to diseases and other complex traits [11], [12]. Statistical analysis typically looks for association between a phenotype and a SNP taken individually via single-locus tests, though geneticists admit this is an oversimplified approach to tackle the complexity of underlying biological mechanisms [13].

Since one SNP rarely describes a quantitative phenotype completely, we consider the interaction between SNPs, known as epistasis [11], [14], [15]. Figure 1 demonstrates this phenomenon when only considering two variants at a time. Unfortunately, effective epistasis detection gives rise to significant analytical and computational challenges. Two main challenges include: 1) the computational complexity of exhaustive approaches to epistasis grows exponentially with order of interactions and 2) many of the more traditional statistical methods increase type I error associated with too many hypothesis tests and require Bonferroni-like corrections to partially address this issue [13].

In this paper, we propose an algebraic signal processing method to detect epistasis by considering the Fourier transform over the associated irreducible representations of the symmetric group [16]. This method has more recently been used in machine learning problems where symmetry of the underlying data proves useful [17]–[19] as well in ranking sports and social choice applications [20]–[23]. By applying our method in genomic data, the resulting spectrum yields precise insight into the higher order interaction and their effect on response variables. Specifically the Fourier transform takes the measured phenotypic response to a set of candidate SNPs and orthogonally decomposes the response into pure higher order interactions. This work addresses the previous challenges by avoiding exhaustive searches and not limiting higher order interactions to SNPs that are individually statistically significant. In Section 2, we introduce our signal processing approach for identifying potential higher order interactions for a quantitative phenotype. For simplicity, we develop our approach for both homozygous (2 copies of the variant) and exclusive heterozygous (at least 1 copy of the variant). In Section 3, we present the results of our method on both simulated and real sequencing data from previously

This work was supported by NSF Grant DMS-1659138, NSA Grant H98230-18-1-0008, Sloan Grant G-2017-9876, and the Wicklow AI and Medicine Research Initiative at the USF Data Institute

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

published mice data and reveal new interactions that influence mice body weight and lean mass [24].

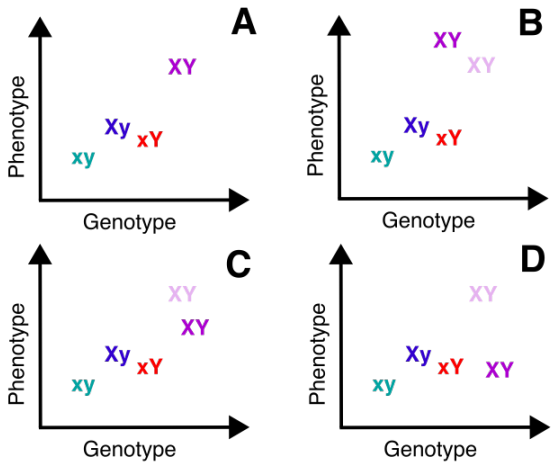

Fig. 1. **Epistasis Example.** Illustration of epistasis where mutations X and Y, separately, result in an increased observed phenotype. Each subfigure illustrates the observed quantitative phenotype when both X and Y are present. In A) effects are purely additive, and is represented by the transparent XY in subfigures B) effects $>$ additive, C) effects $<$ additive, and D) effects are less than each individual mutation.

II. Method

The Fourier transform [25] is a popular signal processing tool that takes in as input (usually noisy) data and projects the data onto a new basis that often is far more clarifying of the original form of the data. Classic examples of Fourier analysis include transforming raw audio signal (as a time series) into the frequency domain for easier analysis and classification. Large amplitudes associated to certain frequencies characterize the audio signal. In this case, the core assumption is that an *orthogonal* basis of $\sin(t)$ and $\cos(t)$ provide useful insight into the underlying signal. For our method, the orthogonal basis functions are representations of the abelian cyclic group. Therefore, the classic Fourier transform can be viewed as projections of the data onto the associated irreducible representations, which, in this abelian case, are all one dimensional homogeneous spaces.

Our objective is to develop a Fourier transform that detects the subsets (k) of interacting SNP mutations (n) that have a positive phenotypic response. In our data set of study [24], the response is on mice body weight. In general, we would like to understand the phenotypic response to first order, second order,..., up to i^{th} order interactions between mutations. The appropriate group structure to detect "subset" behavior is not the cyclic group but instead the symmetric group S_n, which is the group of permutations on a set of n elements. The foundational development of computing Fourier transforms with *nonabelian* group structure can be found in [16].

A. Vector Space Representation of the Mutation-Phenotype Data

We begin by labeling each SNP from 1 to n. We define X be the elements of the power set of n SNPs up to $\lfloor n/2 \rfloor$. To analyze samples where the subset of k SNPs are greater than $\lfloor n/2 \rfloor$, we separate, relabel and map back to a subset of $n-k$.

We now define the vector space of all functions, $f \in M$, from X to $\mathbb{R}$. We index the basis elements of this vector space in lexicographic order by size of subsets (i.e. the first $\binom{n}{1}$ elements of the vector f correspond to single subsets $\{1\},\{2\},\ldots,\{n\}$ the next $\binom{n}{2}$ elements of f correspond to pair subsets $\{1,2\}\{1,3\}\ldots\{n,n-1\}$). The vector space M will be essential to encode our data. Before introducing the analogous orthogonal decomposition of M via a Fourier transform, we look at an example that illustrates the data encoding.

Let A, B, C, D be four identified SNPs, BW body weight as the phenotypic response, then a data set of 10 samples can be encoded as the matrix S,

$$S = \begin{pmatrix} A & B & C & D & BW \\ 1 & 0 & 0 & 0 & 196 \\ 0 & 0 & 1 & 0 & 248 \\ 0 & 1 & 0 & 1 & 169 \\ 0 & 1 & 1 & 0 & 124 \\ 0 & 0 & 0 & 1 & 78 \\ 1 & 0 & 0 & 1 & 158 \\ 0 & 0 & 0 & 1 & 119 \\ 0 & 1 & 1 & 0 & 189 \\ 0 & 0 & 0 & 1 & 172 \\ 0 & 1 & 1 & 0 & 161 \end{pmatrix},$$

where an entry of 1 indicates presence of a SNP at that genomic locus. Thus, the individual in the first row has one SNP (A) and a corresponding body weight of 196. We now translate the rows of samples of S to our function space $f \in M$ in this example. The mapping for $S \to f$ is completed by the following:

1) If a specific mutation set is found once in S, then it is simply added into its location in the vector f lexicographically.
2) If more than 1 sample has the identical set of mutations, we average all phenotypic responses for that set of mutations.
3) If a specific subset of SNPs does not occur in the samples, we impute with the mean of all samples.

The dimension of of M is 2^{n-1}, so most subsets of SNPs are not sampled in practice. We apply step 3) to account for infrequent mutations that may appear to have a large positive effect on body weight.

Thus, f has final form with imputed mean $(175.\overline{3})$ as:

$$f = \begin{pmatrix} 196 \\ 175.\bar{3} \\ 248 \\ (78+119+172)/3 \\ 175.\bar{3} \\ 175.\bar{3} \\ 158 \\ 124+189+161)/3 \\ 169 \\ 175.\bar{3} \end{pmatrix} = \begin{pmatrix} 196 \\ 175.\bar{3} \\ 248 \\ 123 \\ 175.\bar{3} \\ 175.\bar{3} \\ 158 \\ 158 \\ 169 \\ 175.\bar{3} \end{pmatrix}$$

f is then partitioned into smaller data vectors $f^{(3,1)}$ (the first four entries of f) and $f^{(2,2)}$ (the next six entries of f). Each smaller "k-partitioned" data vector now contains the body weight for all mice with exactly 1 and 2 mutations respectively.

B. Orthogonal Decomposition of the Data, the Fourier transform step

In order to study the i^{th} order effects of k mutations on the response variable, we have to further decompose the data vectors, $f \in M$ into linearly independent, irreducible and interpretable vectors. We first note that $M = M^{(n,0)} \bigoplus M^{(n-1,1)} \bigoplus \cdots \bigoplus M^{(n-k,k)}$ decomposes naturally into orthogonal homogeneous spaces with respect to S_n, where $M^{(n-k,k)}$ is the $\binom{n}{k}$ dimensional space containing the data from samples with exactly k mutations [16]. We denote $f^{(n-k,k)}$ as the part of the data vector f that lives in the associated homogeneous space $M^{(n-k,k)}$.

The last step in this methodology is to project $f^{(n-k,k)} \in M^{(n-k,k)}$ onto orthogonal, interpretable subspaces so that we can observe the phenotypic response of the pure i^{th} order mutations for $0 \leq i \leq k$. Observing that $M^{(n-k,k)}$ becomes a $\mathbb{C}S_n$ module [26] if we define the action of S_n on the basis of $M^{(n-k,k)}$ in the usual way, [16], [26]. $M^{(n-k,k)}$ now admits an irreducible orthogonal decomposition of the data space in the form:

$$M^{(n-k,k)} = M_0^{(n-k,k)} \bigoplus M_1^{(n-k,k)} \bigoplus \cdots \bigoplus M_k^{(n-k,k)}, \tag{1}$$

where $M_i^{(n-k,k)}$ is an irreducible, orthogonal subspace corresponding to precisely the pure i^{th} order interaction effects. Taking $f^{(n-k,k)}$ and projecting them on to (1) is the Fourier transform of our data against the symmetric group.

There are a number of techniques developed to compute these transforms [27]–[31]. We follow [31] which exploits the fact that the i^{th} eigenspace of the adjacency matrix associated to the Johnson Graph $J(n,k)$ correspond precisely to the irreducible orthogonal subspace $M_i^{(n-k,k)}$. We then compute these eigenspaces from $J(n,k)$ and project $f^{(n-k,k)}$ onto (1) directly.

For a data set with $n = 15$ candidate SNPs, the full orthogonal decomposition is:

$$\begin{aligned}
f^{(14,1)} &= f_0^{(14,1)} + f_1^{(14,1)} \\
f^{(13,2)} &= f_0^{(13,2)} + f_1^{(13,2)} + f_2^{(13,2)} \\
f^{(12,3)} &= f_0^{(12,3)} + f_1^{(12,3)} + f_2^{(12,3)} + f_3^{(12,3)} \\
f^{(11,4)} &= f_0^{(11,4)} + f_1^{(11,4)} + \cdots + f_4^{(11,4)} \\
f^{(10,5)} &= f_0^{(10,5)} + f_1^{(10,5)} + \cdots + f_5^{(10,5)} \\
f^{(9,6)} &= f_0^{(9,6)} + f_1^{(9,6)} + \cdots + f_6^{(9,6)}. \\
f^{(8,7)} &= f_0^{(8,7)} + f_1^{(8,7)} + \cdots + f_7^{(8,7)}.
\end{aligned}$$

For each $f^{(15-k,k)}$, $f_i^{(15-k,k)}$ is the vector whose entries correspond to the pure i^{th} order effect on the phenotypic response that the samples with exactly k SNP mutations.

Note that the pure 0^{th} order effect corresponds to the average effect that the k-groupings of mutations have on the quantitative response. We also note that, as expected with a Fourier transform, we can reconstruct the original signal $f^{(15-k,k)}$ as the sum of all the orthogonal projections.

To analyze $f_i^{(n,k)}$ for insight into *which* set of mutations had significant effects, we apply Mallow's method [16], which creates a lower dimensional, interpretable Mallow vector, $\tilde{f}_i^{(n,k)}$. Each entry of the Mallow vector, $\tilde{f}_i^{(n,k)}$, contains the inner products of $f_i^{(n,k)}$ with indicator functions of all the i^{th} subsets, in lexicographic order. Mallow's method allows for simple interpretation of $f_i^{(n,k)}$ as will be seen in the results section. A large positive value in a Mallow vector means that the corresponding coalition has a relatively high positive effect on the phenotype while a large negative value implies a relatively large negative effect. Values closer to 0 could mean that those specific coalitions have little effect on the response.

III. Results

We implemented our method in Python and analyze the method's detection of genotype coalitions affecting observable phenotype on both simulated and real mice data from a previous study [24]. For both data sets, we compare predicted coalitions to known SNPs affecting the phenotypic response. In the case for the Karst mice data, we present results detecting new groups of SNPs that lead to a statistically significant increase in phenotype.

A. Simulated Data

To test our method, we simulated 1000 realizations of 400 individuals with 15 SNPs. Each individual had exactly 6 variants with equal probability of each variant, and a quantitative phenotype response drawn from $\mathcal{N}(\mu = 0, \sigma = 0.25)$. Out of these variants, we chose the coalitions $\{(B,D),(C,E,H)\}$ to lead to an increased response of $\mathcal{N}(\mu = 2, \sigma = 0.25)$ and $\mathcal{N}(\mu = 1, \sigma = 0.25)$, respectively. The presence of the A variant resulted in a decrease in the response by $\mathcal{N}(\mu = -2, \sigma = 0.25)$.

When we apply our method to first through third-order interactions, we consider up to $\binom{15}{3} = 455$ interactions of groups of SNPs. After applying Mallow's method, we report

the top highest three and lowest three frequencies, corresponding to groupings of 1, 2, and 3 SNPs. Figures 2 and 3 show a sample single realization of the second order and third order effects in the simulated data and we correctly identify the highest frequency spike to be BD and CEH, respectively. Furthermore, we observe similar results for A with a negative spike in the spectrum. As we expect, averaging over the 1000 realizations, only the groupings BD and CEH are the only groups that result in statistically significant peaks contributing to an increase in the phenotype.

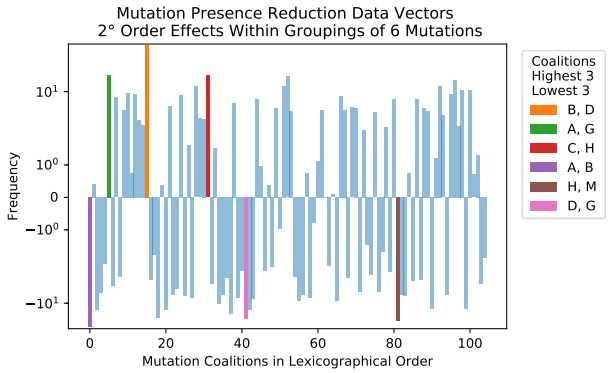

Fig. 2. **Second order spectrum.** Illustration of Mallow vector $\tilde{f}_2^{(9,6)}$ in one simulated data set. We report the highest and lowest three groupings of pairs of SNPs and correctly identify the pair BD as the strongest positively affecting phenotype.

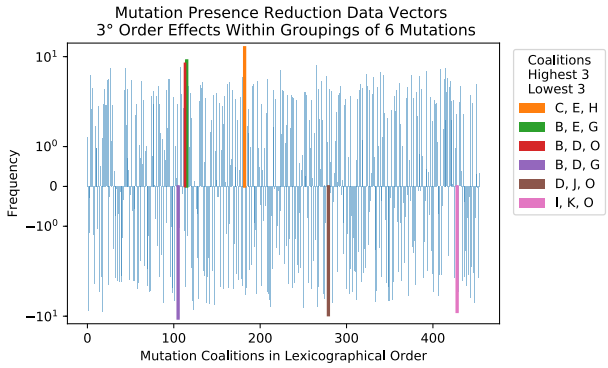

Fig. 3. **Third order spectrum.** Illustration of Mallow vector $\tilde{f}_3^{(9,6)}$ in one simulated data set. We report highest and lowest three groupings of 3 SNPs and correctly identify the CEH as positively affecting phenotype.

B. Karst Mice Data

We next apply our method to previously published mice data [24]. In this study, Karst et al. provided a candidate set of SNPs related to muscle mice in the intercrossed G_3 population. Siblings from the reciprocal cross generations interbred F_1 and F_2 populations were randomly mated to produce the G_3 population of mice [32]. These SNPs (in Chromosome 1), our labels, and their corresponding markers are summarized in Table I. For our analysis, we focus on all the candidate SNPs in Chromosome 1. Since mice can have 0, 1, or 2 copies of a SNP, we apply our method – which encodes the presence of a genomic variant – in two ways: 1) the *exclusive heterozygous* approach (i.e., if there is at least one SNP present) and 2) the *homozygous* (i.e., there are two copies of the variation present). For our analysis, we separated the data by sex and used both approaches for each data set. We refer to interacting SNPs as coalitions.

Applying the Fourier transform to exclusive heterozygous data set, our method detected several higher order coalitions that contributed to higher body weight and lean mass in female mice. For example, the coalition of SNPs ABL consistently positively contributed to increased lean mass. However, we note that having only at least one copy of the SNP does not yield statistically different means in phenotypes. Karst identified both A and B (on Chromosome 1) to have an individually statistically significant positive contribution to body weight and lean mass. For the heterozygous approach, we note that a simple two-sided difference of means t-test does not result in the same statistical significance as in [24], but our method still recovered these SNP interactions.

When we applied the Fourier transform to homozygous data set, several higher order coalitions were detected that significantly contributed to higher body weight and lean mass in male and female mice. We summarize these results in Table II with a two-sided difference of means t-test. In Figure 4, we plot both the male and female body weight and highlight the mice with the interacting variants detected by our model. For body weight, coalitions ABM for males and $ABDE$ for females were particularly significant with $p = 0.0017$ and $p = 0.0303$ respectively. As such, we independently find the SNPs previously correlated with these increased phenotypes, but also uncover that although A and B both contribute to higher lean mass and body weight, the additional presence of SNPs (e.g., E, N, M) lead to more statistically significant results. Thus, our model is able to detect the effects of the individual mutations as stated in [24], as well as discover previously undetected higher order mutation interactions contributing to these phenotypes.

TABLE I
CANDIDATE SNPS (AND ALPHABETICAL LABELS) IN CHROMOSOME 1[a]

Label	SNP	Label	SNP	Label	SNP
A	rs31194300	F	rs31684041	K	rs3672697
B	rs4222269	G	rs31234127	L	rs31424068
C	rs4222320	H	rs31791013	M	rs32257630
D	rs31991963	I	rs32520046	N	rs31474366
E	rs31886089	J	rs4222579	O	rs4222922

[a]Subset of data from [24].

IV. CONCLUSION

We propose a novel method to detect higher order interactions between single nucleotide polymorphisms (SNPs) associated with a phenotypic response from a candidate set of variants. Our method orthogonally decomposes mutation-phenotype data to interpretable i^{th} order interactions of coalitions of SNPs. We present results on both simulated and real data, verifying previous results, and our method uncovers new statistically significant groupings of SNPs related to phenotypic responses. In future studies, we intend to incorporate

TABLE II
HOMOZYGOUS COALITIONS FOR BODY WEIGHT AND LEAN MASS

Coalition	p-value		Sex
	Body Weight	*Lean Mass*	
AB	0.0315	0.0277	Male
AC	0.0305	0.0354	Male
ABM	**0.0017**	**0.0041**	Male
ABN	**0.0011**	**0.0011**	Male
BDE	0.0439	0.1651	Female
$ABDE$	**0.0303**	0.1711	Female
$ACDE$	0.0505	0.2367	Female

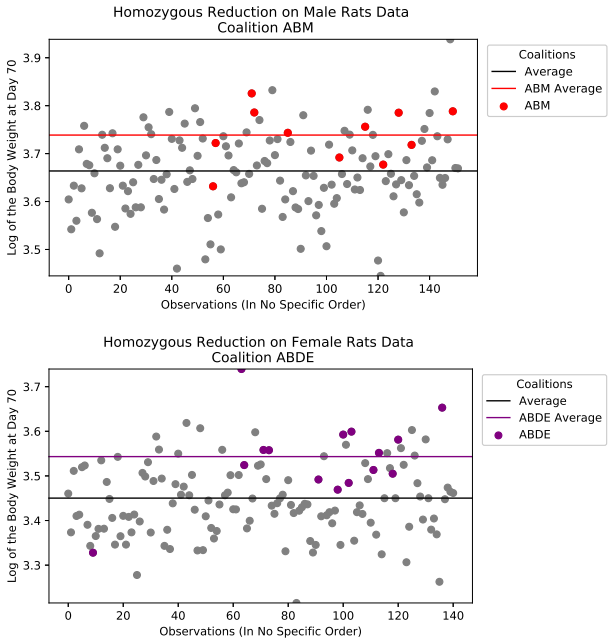

Fig. 4. G_3 mice population with log(body weight) plotted. *Top.* Male mice with coalition ABM are highlighted in red. *Bottom.* Female mice with coalition $ABDE$ are plotted in purple. In both, these groupings of SNPs lead to a statistically significant difference from the population's body weight.

additional comparisons with existing methods and include larger data sets from a variety of organisms and phenotypes.

REFERENCES

[1] 1000 Genomes Project Consortium et al., "An integrated map of genetic variation from 1,092 human genomes," *Nature*, vol. 491, no. 7422, pp. 56–65, 2012.

[2] J.-Y. Li, J. Wang, and R. S. Zeigler, "The 3,000 rice genomes project: new opportunities and challenges for future rice research," *GigaScience*, vol. 3, no. 1, pp. 1–3, 2014.

[3] Z. R Chalmers, C. F Connelly, D. Fabrizio, L. Gay, S. M Ali, R. Ennis, A. Schrock, B. Campbell, A. Shlien, J. Chmielecki, et al., "Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden," *Genome medicine*, vol. 9, no. 1, pp. 34, 2017.

[4] ENCODE Project Consortium et al., "Identification and analysis of functional elements in 1% of the human genome by the encode pilot project," *Nature*, vol. 447, no. 7146, pp. 799, 2007.

[5] C. Alkan, B. P. Coe, and E. E. Eichler, "Genome structural variation discovery and genotyping," *Nature Reviews Genetics*, vol. 12, no. 5, pp. 363, 2011.

[6] J. R. MacDonald, R. Ziman, R. K. C. Yuen, L. Feuk, and S. W. Scherer, "The database of genomic variants: a curated collection of structural variation in the human genome," *Nucleic acids research*, vol. 42, no. D1, pp. D986–D992, 2013.

[7] J. Weischenfeldt, O. Symmons, F. Spitz, and J. O. Korbel, "Phenotypic impact of genomic structural variation: insights from and for human disease," *Nature Reviews Genetics*, vol. 14, no. 2, pp. 125, 2013.

[8] I. Martincorena and P. J. Campbell, "Somatic mutation in cancer and normal cells," *Science*, vol. 349, no. 6255, pp. 1483–1489, 2015.

[9] C. Jeong, D. B. Witonsky, B. Basnyat, M. Neupane, C. M. Beall, G. Childs, S. R. Craig, J. Novembre, and A. Di Rienzo, "Detecting past and ongoing natural selection among ethnically tibetan women at high altitude in nepal," *PLOS Genetics*, vol. 14, no. 9, pp. 1–30, 09 2018.

[10] G. A. Gnecchi-Ruscone, P. Abondio, S. De Fanti, S. Sarno, M. G. Sherpa, P. T. Sherpa, G. Marinelli, L. Natali, M. Di Marcello, D. Peluzzi, et al., "Evidence of polygenic adaptation to high altitude from tibetan and sherpa genomes," *Genome biology and evolution*, vol. 10, no. 11, pp. 2919–2930, 2018.

[11] Clement Niel, Christine Sinoquet, Christian Dina, and Ghislain Rocheleau, "A survey about methods dedicated to epistasis detection," *Frontiers in Genetics*, vol. 6, pp. 285, 2015.

[12] M Mielczarek and J Szyda, "Review of alignment and snp calling algorithms for next-generation sequencing data," *Journal of applied genetics*, vol. 57, no. 1, pp. 71–79, 2016.

[13] Heather J Cordell, "Detecting gene–gene interactions that underlie human diseases," *Nature Reviews Genetics*, vol. 10, no. 6, pp. 392, 2009.

[14] Patrick C Phillips, "Epistasis?the essential role of gene interactions in the structure and evolution of genetic systems," *Nature Reviews Genetics*, vol. 9, no. 11, pp. 855, 2008.

[15] Trudy FC Mackay, "Epistasis and quantitative traits: using model organisms to study gene–gene interactions," *Nature Reviews Genetics*, vol. 15, no. 1, pp. 22, 2014.

[16] Persi Diaconis, "Group representations in probability and statistics," *Lecture Notes-Monograph Series*, vol. 11, pp. pp. i–vi+1–192, 1988.

[17] R. Kakarala, "A signal processing approach to fourier analysis of ranking data: The importance of phase," *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1518–1527, April 2011.

[18] Risi Kondor, Andrew Howard, and Tony Jebar, "Multi-object tracking with representations of the symmetric group," vol. 2, pp. 211–218, 2007.

[19] Risi Kondor and Walter Dempsey, "Multiresolution analysis on the symmetric group," in *Advances in Neural Information Processing Systems*, 2012, pp. 1637–1645.

[20] K. P. Paudel, M. Pandit, and M. A. Dunn, "Using spectral analysis and multinomial logit regression to explain households? choice patterns," *Empirical Economics*, vol. 44, no. 2, pp. 739–760, 2013.

[21] G. Jurman, S. Merler, A. Barla, S. Paoli, A. Galea, and C. Furlanello, "Algebraic stability indicators for ranked lists in molecular profiling," *Bioinformatics*, vol. 24, no. 2, pp. 258–264, 2008.

[22] Brian L. Lawson, Michael E. Orrison, and David T. Uminsky, "Spectral analysis of the supreme court," *Mathematics Magazine*, vol. 79, no. 5, pp. pp. 340–346, 2006.

[23] Stephen Devlin and David Uminsky, "Evaluating nba lineups through spectral analysis," Preprint, 2019.

[24] S. Kärst, R. Cheng, A. O. Schmitt, H. Yang, F. P. M. De Villena, A. A. Palmer, and G. A. Brockmann, "Genetic determinants for intramuscular fat content and water-holding capacity in mice selected for high muscle mass," *Mammalian genome*, vol. 22, no. 9-10, pp. 530, 2011.

[25] Michael James Lighthill, *An introduction to Fourier analysis and generalised functions*, Cambridge University Press, 1958.

[26] David Steven Dummit and Richard M. Foote, *Abstract algebra*, John Wiley & sons, Hoboken, NJ, 2004.

[27] Risi Kondor, "Snob: a c++ toolkit for fast fourier transforms on the symmetric group," 2006.

[28] Risi Kondor, "Non-commutative harmonic analysis in multi-object tracking," in *Bayesian time series models*. 2011, pp. 277–294, Cambridge University Press.

[29] D. Keith Maslen, M. Orrison, and D. N. Rockmore, "Computing isotypic projections with the lanczos iteration," vol. 25, pp. 784–803, 01 2003.

[30] M. Iwasaki, "Spectral analysis of multivariate binary data," *Journal of the Japan Statistical Society, Japanese Issue*, vol. 22, no. 1, pp. 45–65, 1992.

[31] Michael E Orrison, *An eigenspace approach to decomposing representations of finite groups*, Ph.D. thesis, Dartmouth College, 2001.

[32] A Darvasi and M Soller, "Advanced intercross lines, an experimental population for fine genetic mapping.," *Genetics*, vol. 141, no. 3, pp. 1199–1207, 1995.