A Spatial Relations Study of Virus Infected Cells and the Human Immune Response through the V-Proportionality Measurement

Leo Ehnborg

August 30, 2014

Abstract

Biotechnological tools have never been stronger than today and the data they provide is absolutely fascinating. As we get a clearer picture of the intricate workings of living systems effective mathematical and statistical tools become a necessity in order to reach a comprehensive understanding of said systems.

The purpose of this thesis is to statistically explore cutting edge biomedical data taken from virus infected human tissue samples in the hopes of finding interesting correlations amongst the different components in the samples. We will also show that spatial statistical methods can be used to draw valuable and significant conclusions about biological systems.

The method of choice for the statistical analysis in this thesis is the V-proportionality measurement. In theory it can distinguish positive, negative and lack of spatial correlation in datasets through clever use of the Voronoi diagram. The code used for the implementation of the V-proportionality measurement is both explained and provided within the confines of this paper.

Acknowledgements

I would like to thank my supervisor Lena Zetterqvist at Lunds University for her support and guidance. My special thanks to Caroline Sandén and Jonas Erjefält at Medetect, for providing me the opportunity to work with the spectacular data that resulted in this thesis. I would also like to mention Pontus Ingvarsson here, because true kindness is rare thing.

Contents

1 Introduction

The focus of this thesis will be on the exploration of a new interesting type of medical data. Its aspiration to explore the spatial correlations between the intricate parts of the human immune response and virus infected tissue. A quite unexplored statistical method is used, which in theory should be able to distinguish positive, negative and the lack of spatial correlation.

The data used in this thesis is provided by a biotechnology company by the name Medetect who among other things develop tools for the analysis of diseased tissue. One of these tools allow for the detection of virus infected cells alongside many parts of the immune response, in the same sample. This is achieved with the use of cutting edge technology and would have been impossible merely years ago.

Extracting all the necessary information for a medical diagnosis in a single tissue sample is not just convenient, but comes with additional benefits. Since all the information is provided within the same sample, the spatial interdependencies of infections and the immune response can be studied. The researchers at Medetect are confident that these tissue samples will one day serve as a valuable tool in practical medicine. Hopefully the contents of this thesis will help the researchers in their endeavours by presenting an effective method for the study of spatial correlation.

1.1 Overview of thesis

Section 2 introduces important concepts and definitions, then proceeds to describe and discuss the statistical method employed in the thesis. Section 3 uses simulated well understood data to validate the method. In section 4, analysis of real world data is conducted. The conclusions draw from the theses and suggestions for further work is presented in section 5.

2 Method

The following statistical method was first introduced in Óscar Martínez Mozos' intriguing paper on the spatial relations of neurons in the retina (2010). The method was chosen since paper shows promising results on data similar to ours.

This thesis will only be concerned with two-dimensional Euclidean spaces, hence the definitions have been altered specifically for this. The reason being that the method can be explained both clearer and more efficiently to the reader. However, this does mean that some of the statements made are not true in general, especially when discussing the Voronoi diagram.

2.1 The Voronoi diagram

First defined by Georgy Feodosevich Voronoy (1908), the Voronoi diagram is a quite natural construction. Given a set of points in some space, the Voronoi diagram simply splits the space into regions containing whatever is deemed closest to a point according to some metric. We will use the following definition based on the one provided in Óscar Martínez Mozos' paper.

Definition 1 *Let d(.,.) be the Euclidean distance function and* $S = \{s_1, s_2, ..., s_n\}$ *a finite set of points in the plane. Let p be all the points in the plane such that* $p \notin S$*. Now we introduce:*

$$V(s_i) = \{p \mid d(s_i, p) \leq d(s_j, p);\ p \notin S,\ \forall s_i \neq s_j\}$$

$V(S)$ *will hence be a collection convex polygons. This collection is called the Voronoi diagram.*

Each s_i is usually referred to as a *Voronoi site*. Our Voronoi diagram can be seen as a way to produce convex regions around each Voronoi site where the interior points are atleast as close to the site as any other site. These regions denoted as $V(s_i)$ are most commonly known as *Voronoi polygons*.

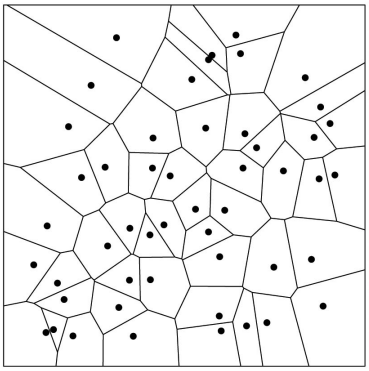

Figure 1: The Voronoi diagram applied to a realization of uniformly distributed pattern.

2.2 The V-proportionality

Now that the Voronoi diagram is well defined we will extend the concept of it into a useful spatial statistical model, namely the *V-proportion measurement*.

In order to explain the method clearly one may imagine two sets of points Q and S on a plane. Using S as sites we construct the Voronoi diagram. For each of the polygons created, we will produce a δ smaller polygon proportional to its corresponding site. The area in between every Voronoi polygon and its smaller counterpart make up a set of bands on the plane. The entire construction of bands and polygons can be seen in Fig. 2.

We now investigate how many of the points in Q reside within our set of bands and compare it with the total amount of points in Q. This ratio is known as the *V-proportion*. More formally we will use the following definition, slightly more general than the one provided by Óscar Martínez Mozos.

Definition 2 *Let $Q = \{q_1, q_2, ..., q_m\}$ and $S = \{s_1, s_2, ..., s_n\}$ be two sets of points. Produce the Voronoi diagram of S and construct the set $E = \{e_1, e_2, ..., e_m\}$ where*

$$e_j = \arg\min_x \left(d(x, q_j) \;\middle|\; x \in \bigcup_{i=1}^{n} \partial V(s_i) \right)$$

and $\partial V(s_i)$ denotes the boundary of the Voronoi polygon. Introduce the set $Q_B \subseteq Q$ where

$$Q_B = \{q_j \in Q \mid d(q_j, e_j) \leq \delta \cdot d(s_i, e_j), \text{ where } q_j \in V(s_i)\}$$

Finally we define the V-proportion as the following ratio

$$\frac{\mid Q_B \mid}{\mid Q \mid}.$$

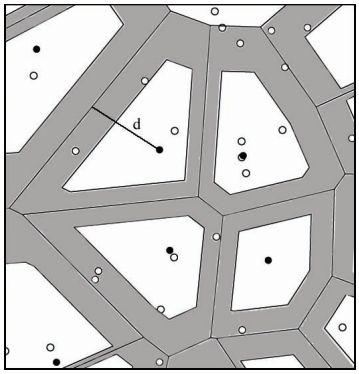

Figure 2: A visualisation of the V-proportionality construction. Sites are represented by filled circles and comparison points by empty ones. The grey area running along the edge of each Voronoi polygon shows the previously discussed bands. Here the created by taking 30% of the distance d between edge and site.

The usefulness of the V-proportion may not be abundantly clear but the thinking is quite straight forward. For simplicity imagine our sets Q and S from before and create the V-proportionality construction using S as sites. If the sets completely lack spatial correlation, one can expect the average point density to be almost the same inside and outside of the bands. If our two sets are positively

spatially correlated and the sites promote the occurrence of points in Q, the average point density will be significantly higher outside of the bands. Lastly, if the points in S actively inhibit the occurrence of points in Q the average point density will be significantly higher inside of the bands since the points of Q will in general cluster around the edges of the Voronoi polygons. Throughout this thesis spatial relations of this kind will be referred to as negative spatial correlation.

Even though the V-proportionality measurement has now been both defined and discussed, the way it should be used in order to draw conclusions has not. First and foremost, we need to introduce a Monte Carlo test procedure to serve as a reference point for lack of spatial correlation. Two completely spatially independent data sets are simulated in an area identical to the one under investigation. The sets are made to contain the exact same number of points as our sets of interest. The V-proportionality of the simulated data is then calculated. This process is repeated as many times as deemed necessary for the construction of a reliable confidence interval.

As can be seen from **Definition 2** our measurement is completely dependant on the the size of the bands, decided by the free parameter δ. The procedure of calculating the V-proportion and simulating confidence intervals will now be repeated for bands of different sizes. Finally the results are plotted and the spatial correlation of the two sets of interest can be analysed. If V-proportion values ever exceed that of confidence intervals it is interpreted as negative spatial correlation. Further, if the opposite is found positive spatial correlation can be assumed. More on this in section 3, where examples based on different kinds of simulated data can be found.

2.2.1 Handling of the edge effect

The data for which the V-proportionality is an effective analysis tool usually comes as a rectangular sampling window of a much larger pattern. This has an inherent problem that needs to be addressed, namely the edge effect. As the observant reader might have noticed in Fig. 1, many of the sites close to the boundary of the sampling window will result in open polygons when the Voronoi diagram is calculated. It is hard to construct meaningful bands for these open polygons as the necessary information is partly missing.

In order to get around this problem, the open polygons and comparison points residing within them are removed. This leads to more dependable results as it eliminates the chance that information not reflecting reality is introduced, but does decrease the total amount of data. This is the same course of action taken in Óscar Martínez Mozos' paper.

2.3 Implementation

The code for the implementation of the V-proportionality was written in the language R (2014). The better part of the code relies on routines contained within the packages `deldir, sgeostat` and `spatstat`. The full code can be found in appendix A, but the general concept can be described as follows.

Step 1

The package `deldir` is used to create a Voronoi diagram of the data chosen

as sites.

Step 2
A help function that takes advantage of the fact that `deldir` provides the corner coordinates for the convex polygons created is used to produce bands of size δ.

Step 3
Each comparison point is examined to determine whether it resides within a band or not using the package `sgeostat`. The V-proportion is then calculated.

Step 4
Two independent randomly distributed sets of the same cardinality as the data sets are simulated. This is done in an area equal to the one of the data sets using the package `spatstat`.

Step 5
The help function is used to create bands.

Step 6
Each simulated comparison point is investigated to determine how many of them reside within the bands. The simulated V-proportion is then calculated.

Step 7
Steps 4,5,6 are repeated T times in order to calculated the mean and confidence intervals of the simulated datasets.

As explained in the previous section, this process needs to be repeated for different δ values and then plotted to provide us with the final product.

3 Validation of the method

Both the theory and its implementation has been discussed in detail. In order to validate the method, patterns with and without spatial correlation are simulated. The V-proportionality method is then used to analyse the sets in an effort to prove that the results obtained are in accordance with the theory. In each one of these analyses we simulate 1000 populations for comparison in order to establish reliable 95% confidence intervals.

3.1 Lack of spatial correlation

Completely independent random patterns can easily be simulated using routines built into R. This is done using the function `runifpoint()` contained within the package `spatstat`. The function simulates a given amount of independent uniformly distributed points in a two dimensional plane.

Two sets of this kind are simulated in a square area of 300^2 pixels. One containing 50 points to be used as sites, alongside one containing 300 points for comparison. The sets simulated are then analysed and the results can be seen in Fig. 3. As expected no significant deviation from the mean can be detected and presumptions of spatial correlation may be dismissed.

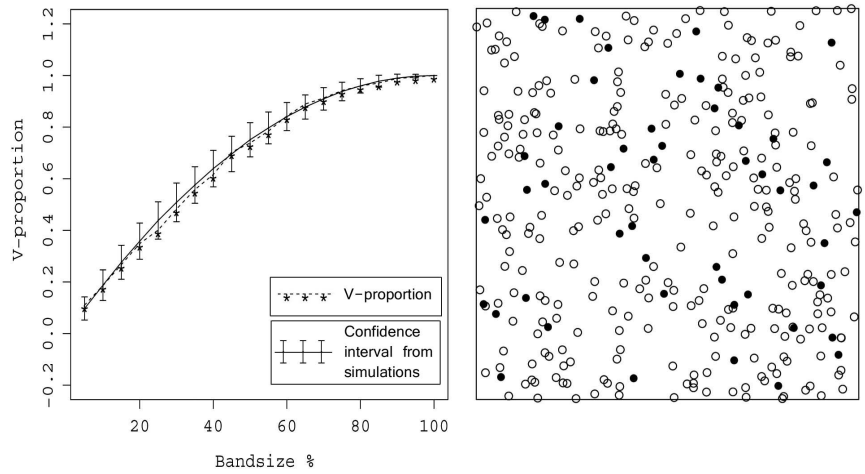

Figure 3: The two simulated random patterns alongside their V-proportionality measurement. Sites are represented by filled circles and the comparison set by empty ones.

3.2 Positive spatial correlation

Aggregated patterns can effectively be simulated using a *Poisson cluster process*, first introduced by Neyman and Scott (1958). This process was later described methodically by Peter J. Diggle (2013), in three postulates.

Postulate 1 Parent events form a Poisson process with intensity ρ

Postulate 2 Each parent produces a random number S of offspring, realized independently and identically for each parent according to a probability distribution $p_s : s = 0, 1, ...$

Postulate 3 The positions of the offspring relative to their parents are independently and identically distributed according to a bivariate pdf $h(\cdot)$.

In order to obtain positive spatial correlation, the offspring are created using a bivariate normal distribution relative to their parent. The simulation of these sets can be done in R using the function `pcp.sim()` contained within the package `splancs`.

In the same manner as before, the parent and offspring sets are simulated in a square area of 300^2 pixels. The results presented in Fig. 4, shows that the V-proportionality values are significantly lower than the random equivalent, indicating a strong positive spatial correlation between the two sets.

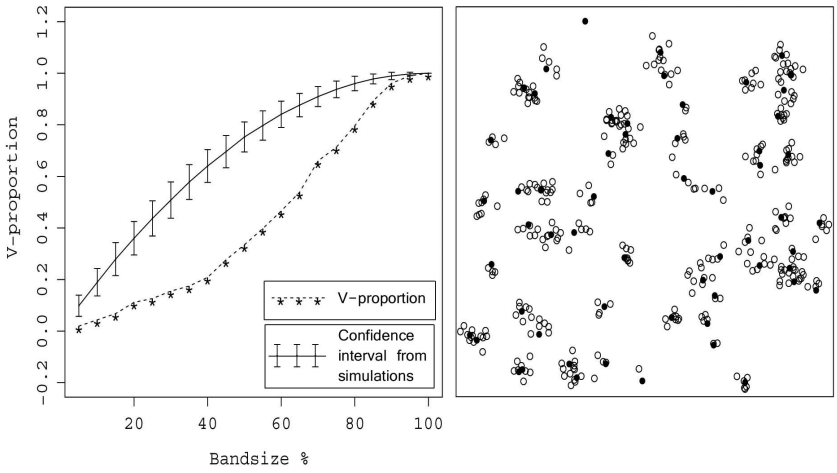

Figure 4: The two simulated positively spatially correlated patterns alongside their V-proportionality measurement. Sites are represented by filled circles and the comparison set by empty ones.

3.3 Negative spatial correlation

To simulate negatively correlated patterns we utilize a process described by Swedish statistician Bertil Matérn (1960). Matérn's method marks the birth time of points in a Poisson process, if a new event lies within distance δ of an older one it is removed. This creates a simple sequential inhibition process defined by Peter J. Diggle (2013, pp. 111) as:

Definition 3 *Consider a sequence of n events X_i in a finite region A. Then*

Matérn 1 *X_1 is uniformly distributed in A*

Matérn 2 *Given $\{X_j = x_j, j = 1, ..., i-1\}$, X_i is uniformly distributed on the intersection of A with $\{y : \|y - x_j\| \geq \delta, j = 1, ..., i-1\}$.*

A set X is simulated in a area of 300^2 pixels using the Matérn process with inhibition distance 25 pixels. Another set Z of uniformly distributed points is simulated into the same area as X. If a point in Z lies within the radius η of a point in X, it is removed. This procedure produces to two sets that both visually and V-proportionality wise exhibits a clear negative spatial correlation, as seen in Fig. 5.

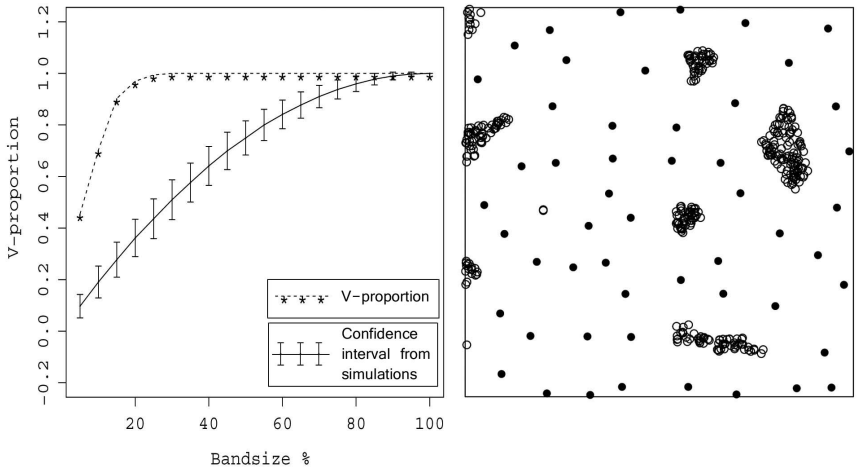

Figure 5: The two simulated negatively correlated patterns alongside their V-proportionality measurement. Sites are represented by filled circles and the comparison set by empty ones.

4 Spatial correlation analysis on real data

As briefly mentioned in the introduction, the data analysed in this thesis comes from diseased human tissue samples where different cell types have been mapped out. This is done by applying molecules of a certain color that bind to the receptors of a specific cell type. The sample is then analysed and the position of each coloured cell is recorded. The process of colouring is repeated for different cell types until the position each cell of interest has been recorded. Two different types of these tissue samples were analysed. The first taken from virus infected lungs where both the virus infected cells and the immune response is mapped out. The second from diseased tonsil tissue where only the immune response is mapped out.

Reading the chapter on non-specific(innate) host resistance in *Prescott, Harley and Klein's Microbiology* (2008) is suggested for the interested reader as it provides a comprehensive understanding of the human immune system. This is however just a suggestion and not a necessity for understanding the results in this thesis.

The statistical investigation of the data was preformed in two steps. First, a wide array of V-proportionality tests were carried out to examine the spatial relation between the different cell types in each sample. After the broad initial examination, the pairs of cells that showed some sort of interesting correlation were subjugated to a more thorough analysis. Confidence intervals of 95% were used in the following V-proportionality tests.

4.1 Virus infected cells and myeloperoxidase in lung tissue

The primary analysis of the lung tissue data suggested the possibility of a significant correlation between myeloperoxidase(MPO) and virus infected cells. Myeloperoxidase is a peroxidase enzyme most commonly expressed in neutrophil

granulocytes. This enzyme uses hydrogen peroxide to produce hypochlorous acid, a cytotoxin that kills pathogens.

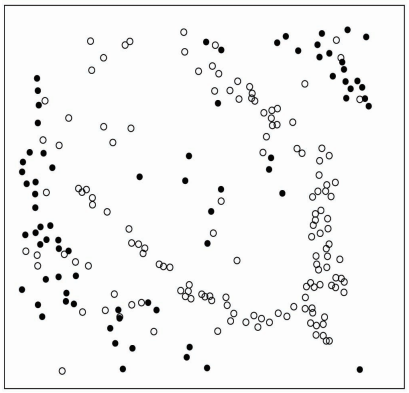

Figure 6: The pattern of MPO(filled circles) and Virus infected cells(empty circles) in human lung tissue.

The spatial analysis of these two cell types was preformed on 20 different band sizes each with a 1000 simulated populations to ensure the accuracy of the confidence bands. Both MPO and the virus infected cells were used as sites respectively, the results can be seen in Fig. 7.

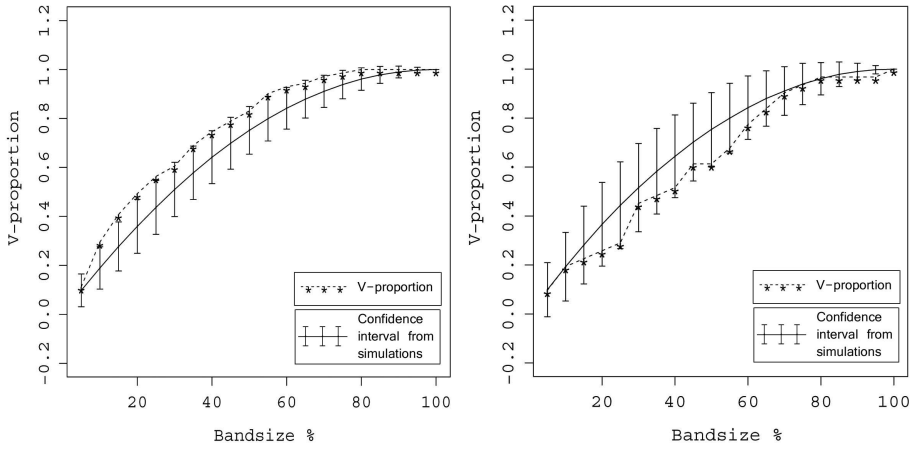

Figure 7: To the left MPO as sites and to the right viruses as sites.

When the MPO points are used as sites a significant negative spatial correlation can be detected in several band sizes. This result seems reasonable seeing as the function of MPO is to produce a virus killing cytotoxin.

No significant correlation is found when using the virus infected cells as sites, this is most likely explained by the handling of open polygons. Because of the way the samples have been taken, the wast majority of MPO points are located close to the boundary of the window, whereas the virus infected cells are located more towards the center. This leads to the removal of a large portion of the data when the virus infected cells are used as sites and far less can be concluded from the results.

4.2 MPO and CD68 in tonsil tissue

The data acquired from the mapping of the immune response in human tonsil tissue indicated that a further analysis of the correlation between MPO and Cluster of Differentiation 68(CD68) was warranted. CD68 is a glycoprotein expressed on different macrophages in the human immune system and is hence a good marker for the immunity response to pathogens.

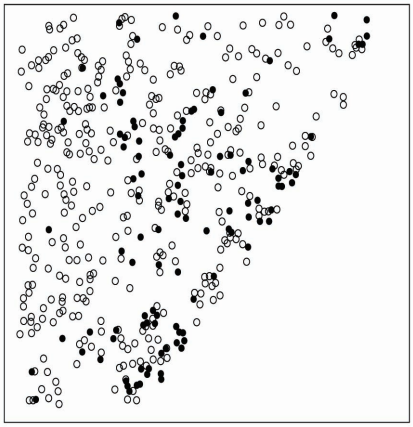

Figure 8: The pattern of MPO(filled circles) and CD68(empty circles) in human tonsil tissue.

The spatial analysis of these two cell types was preformed on 20 different band sizes each with a 1000 simulated populations to ensure the accuracy of the confidence bands. Both CD68 and MPO were used as sites respectively, the results can be seen in Fig. 9.

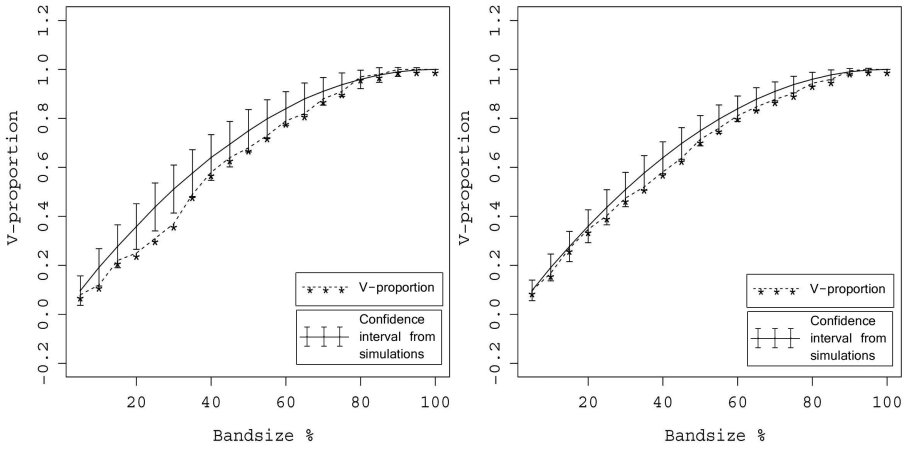

Figure 9: To the left CD68 as sites and to the right MPO as sites.

Using CD68 points as sites suggests a positive spatial correlation between the two cell types within the 95% confidence interval on smaller band sizes. That positive spatial correlation is found between these two cell types is not particularity surprising. They can be seen as two different weapons deployed by the human body in an attempt to combat an infection.

However, no significant spatial correlation can be detected when using MPO points as sites. Further analysis of the interdependencies of these cell types is suggested.

5 Conclusions

To summarize, this thesis has shown that the V-proportionality approach is a valid option for statistical analysis when dealing with biological data of this kind. Significant spatial correlation was found in multiple samples and basic hypotheses as to why was provided. While this was the aspiration of the thesis, a weakness of the V-proportionality approach was also found. As discussed earlier, the interaction between the way the samples have been taken and how the edge effect was handled led to a grate loss of data in certain cases.

5.1 Further work

The patterns of the virus infected cells and the immune response is only part of the picture. In many of these samples, areas of different tissue types are present and detectable. These serve as a backdrop for the patterns and may or may not affect the correlation between the populations. Devising some sort of method that incorporates this additional information could lead to a stronger analysis.

References

Diggle, P.J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, ISBN 9781466560239

Matérn, B. (1960). *Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations.* Meddelanden fran Statens Skogsforskningsinstitut, Vol. 49, No. 5, pp. 144.

Mozos, Ó.M., Bolea, J.A., Ferrández, J.M., Ahnelt, P.K. and Fernández, E. (2010). *V-Proportion: A method based on the Voronoi diagram to study spatial relations in neuronal mosaics of the retina.* Neurocomputing, Vol. 74, Issue 1-3, pp. 418-427.

Neyman, J. and Scott, E.L. (1958). *Statistical approach to problems of cosmology.* Journal of the Royal Statistical Society. Series B, Vol. 20, No. 1, pp. 1-43.

R Development Core Team. (2014). *R 3.1.0.* http://www.r-project.org/

Voronoy, G.F. (1908). *Nouvelles applications des paramètres continus à la théorie de formes quadratiques.* Journal für die reine und angewandte Mathematik, Vol. 134, pp. 198-287.

Willey, J.M., Sherwood L.M. and Woolverton, C.J. (2008). *Prescott, Harley, and Klein's Microbiology, seventh edition.* McGraw-Hill Higher Education, ISBN 9780071102315

Appendix A

```
#Provide data as deldir tile.lists with rw=c()
   attribute corresponding to size of sampling window
#N=Number of bandsizes, Generations=Number of
   simulation generations,
#Mode="remove" or "keep" open polygons, Thesis=type of
    plot "yes"/"no"
Vprop<-function(tilelist,compset,N,Generations,mode="
   keep",thesis="no"){
  realV<-list();z=1 #PART 1, EMPIRICAL V-
     PROPORTIONALITY
  if(mode=="remove"){
    z=2
    modifiedcompset=list()
    bandset<-hull(tilelist,0, z)
    for(i in 1:length(compset)){
      xorg<-compset[[i]][[2]][[1]];yorg<-compset[[i
         ]][[2]][[2]]
      for(j in 1:length(bandset)){
        xycoords=xy.coords(bandset[[j]][,1],bandset[[j
           ]][,2])
        if((in.chull(xorg,yorg,xycoords$x,xycoords$y))
           ==TRUE){
          xytemp<-c(xorg,yorg)
          modifiedcompset=append(modifiedcompset,list(
             xytemp))
        }
      }
    }
    if(length(modifiedcompset)==0){
      plot(1,1)
      title(main="No points left after removing open
         polygons")
      return()
    }
    for(k in 1:N){
      Pb=0;bandset<-hull(tilelist,(100/N)*k, z)
      for(i in 1:length(modifiedcompset)){
        xorg<-modifiedcompset[[i]][1];yorg<-
           modifiedcompset[[i]][2]
        for(j in 1:length(bandset)){
          xycoords=xy.coords(bandset[[j]][,1],bandset
             [[j]][,2])
          if((in.chull(xorg,yorg,xycoords$x,xycoords$y
             ))==TRUE){
            Pb=Pb+1
            break
          }
```

```
      }
    }
    realV[k]=(1-Pb/length(modifiedcompset))
  }
  print("Open polygons removed")
}
if(mode!="remove"){
  for(k in 1:N){
    Pb=0;bandset<-hull(tilelist,(100/N)*k, z)
    for(i in 1:length(compset)){
      xorg<-compset[[i]][[2]][[1]];yorg<-compset[[i
          ]][[2]][[2]]
      for(j in 1:length(bandset)){
        xycoords=xy.coords(bandset[[j]][,1],bandset
            [[j]][,2])
        if((in.chull(xorg,yorg,xycoords$x,xycoords$y
            ))==TRUE){
          Pb=Pb+1
          break
        }
      }
    }
    realV[k]=(1-Pb/length(compset))
  }
  print("Open polygons kept")
}
print("Empirical Vproportion klar")
simuV<-matrix(0, nrow=Generations,ncol=N);w=0 #PART
   2, MONTECARLO PROCEDURE
for(l in 1:Generations){
  Tile=runifpoint(length(tilelist),win=owin(c(0,max=
      attr(tilelist,"rw")[2]),c(0,max=attr(tilelist,"
      rw")[4])))
  tiletess<-deldir(Tile$x,Tile$y,rw=c(attr(tilelist
      ,"rw")))
  Stileset<-tile.list(tiletess)
  if(z==1){
    Comp=runifpoint(length(compset),win=owin(c(0,max
        =attr(tilelist,"rw")[2]),c(0,max=attr(
        tilelist,"rw")[4])))
  }
  if(z==2){
    Snollbandset<-hull(Stileset,0,z);polylist=list()
    for(j in 1:length(Snollbandset)){
      tempmatrix=matrix(0,nrow=length(Snollbandset[[
          j]][,1]),ncol=2)
      tempmatrix[,1]=Snollbandset[[j]][,1];
          tempmatrix[,2]=Snollbandset[[j]][,2]
      tempmatrix=rbind(tempmatrix,tempmatrix[1,])
      polylist[[j]]=tempmatrix
```

```
    }
    Comp=runifpoint(length(modifiedcompset),win=owin
       (poly=polylist))
  }
  for(k in 1:N){
    if(z==1){
      SPb=0
      Sbandset<-hull(Stileset,(100/N)*k, z)
      for(i in 1:Comp[[2]]){
        for(j in 1:length(Sbandset)){
          xycoords=xy.coords(Sbandset[[j]][,1],
             Sbandset[[j]][,2])
          if((in.chull(Comp$x[i],Comp$y[i],
             xycoords$x,xycoords$y))==TRUE){
            SPb=SPb+1
            break
          }
        }
      }
      simuV[l,k]=(1-(SPb/Comp[[2]]))
    }
    if(z==2){
      SPb=0
      Sbandset<-hull(Stileset,(100/N)*k, z)
      for(i in 1:Comp[[2]]){
        for(j in 1:length(Sbandset)){
          xycoords=xy.coords(Sbandset[[j]][,1],
             Sbandset[[j]][,2])
          if((in.chull(Comp$x[i],Comp$y[i],
             xycoords$x,xycoords$y))==TRUE){
            SPb=SPb+1
            break
          }
        }
      }
      simuV[l,k]=(1-(SPb/Comp[[2]]))
    }
  }
  w=w+1
  print(c((w/Generations)*100,"%Simulering klar"))
}
Smean=list();Svar=list();Conf=matrix(0,nrow=2,ncol=N
   ) #PART 3, CALCULATION OF CONFIDENCE INTERVALS
for(i in 1:N){
  Smean[i]<-mean(simuV[,i])
}
for(i in 1:N){
  sum=0
  for(j in 1:Generations){
    sum=sum+((Smean[[i]]-simuV[[j,i]])^2)
```

```
    }
    Svar[i]<-(1/(Generations-1))*sum
  }
  for(i in 1:N){
    Conf[1,i]=norm.ci(t0=Smean[[i]],var.t0=Svar[[i]])
        [2]
    Conf[2,i]=norm.ci(t0=Smean[[i]],var.t0=Svar[[i]])
        [3]
  }
  if(thesis!="yes"){
    plot(seq(100/N,100,100/N), realV,type="o",col="red
        ",ylim=c(-0.2,1.2),xlab="Bandsize %",ylab="V-
        proportion")
    lines(seq(100/N,100,100/N), Smean,type="o",col="
        blue",xlab="Bandsize %",ylab="V-proportion")
    errbar(seq(100/N,100,100/N),Smean,Conf[1,],Conf
        [2,],add="TRUE")
    title(main = bquote("Sites:" ~ .(length(bandset))
        ~ "Points:" ~ .(Comp[[2]])))
  }
  if(thesis=="yes"){
    plot(seq(100/N,100,100/N), realV,type="o",lty=2,
        pch="*",ylim=c(-0.2,1.2),xlab="Bandsize %",ylab
        ="V-proportion")
    lines(seq(100/N,100,100/N), Smean,type="o",lty=1,
        pch=".",xlab="Bandsize %",ylab="V-proportion")
    errbar(seq(100/N,100,100/N),Smean,pch=".",Conf
        [1,],Conf[2,],add="TRUE")
  }
  if(z==2 & thesis!="yes"){
    title(sub="OPEN POLYGONS REMOVED")
  }
}


#Help function that calculates "inverted"
   proportionality bands.
#Open polygons are removed if mode=2
hull<-function(tilelist, delta, mode){
  rekt<-list();delta=delta/100
  for(i in 1:length(tilelist)){
    oriX<-tilelist[[i]][[2]][[1]];oriY<-tilelist[[i
        ]][[2]][[2]]
    temp1<-matrix(0,nrow=length(tilelist[[i]][[3]]),
        ncol=2)
    for(j in 1:length(tilelist[[i]][[3]])){
      diffX=(abs(oriX-tilelist[[i]][[3]][[j]]))*delta
      diffY=(abs(oriY-tilelist[[i]][[4]][[j]]))*delta
      if(tilelist[[i]][[3]][[j]]<oriX){
        temp1[j,1]<-(tilelist[[i]][[3]][[j]]+diffX)
```

```
      }
      if(tilelist[[i]][[3]][[j]]>oriX){
        temp1[j,1]<-(tilelist[[i]][[3]][[j]]-diffX)
      }
      if(tilelist[[i]][[4]][[j]]<oriY){
        temp1[j,2]<-(tilelist[[i]][[4]][[j]]+diffY)
      }
      if(tilelist[[i]][[4]][[j]]>oriY){
        temp1[j,2]<-(tilelist[[i]][[4]][[j]]-diffY)
      }
    }
    rekt[[i]]=temp1
  }
  if(mode==2){
    rekt2=list()
    for(k in 1:length(tilelist)){
      if(any(tilelist[[k]]$bp)==FALSE){
        rekt2<-append(rekt2,rekt[k])
      }
    }
    return(rekt2)
  }
  return(rekt)
}
```

Populärvetenskaplig sammanfattning

Möjligheten att utnyttja framsteg från nästintill alla vetenskapliga discipliner gör att forskning inom medicin och bioteknik går framåt i rasande fart. Med de nyaste teknikerna kan vi idag studera de mest komplexa sambanden i den mänskliga kroppen, något som läkare och forskare i tidigare generationer bara kunnat drömma om.

Med dessa nya tekniker produceras kopiösa mängder data som måste struktureras och analyseras, det är här matematisk statistik kommer in. Den tjänar som vetenskapens objektiva ögon och ser det som är för stort, för komplicerat eller för subtilt för våra egna att upptäcka.

Det är just detta denna uppsats strävar efter att uppnå. Att undersöka data från nyutvecklade medicinska forskningsmetoder med hjälp av matematik, för att hitta samband på ett objektivt sätt. I uppsatsen studeras bland annat sambanden mellan infektion och immunförsvar i sjuka mänskliga lungor samt sambanden mellan de olika delarna i själva immunförsvaret.

För att uppnå relevanta resultat behöver vi först övertyga oss själva att sättet vi arbetar med datan är effektivt och faktiskt fungerar. Därför tillägnas en stor del av detta arbete den matematiska metoden som används för vår analys. Metoden kallas för V-proportionalitetsmåttet dess generella tillvägagångssätt bör vara förståligt även för den utan en bakgrund inom matematik.